Tampereen yliopisto

Людмила Омесь

КОРПУСНОЕ ИССЛЕДОВАНИЕ ЯЗЫКОВЫХ ОСОБЕННОСТЕЙ ИНФОРМАТИВНЫХ И ОПЕРАТИВНЫХ ТЕКСТОВ О COVID-19 НА РУССКОМ И ФИНСКОМ ЯЗЫКАХ, А ТАКЖЕ В ПЕРЕВОДАХ НА ЭТИ ЯЗЫКИ

Факультет информационных технологий и коммуникационных наук

Магистерская работа

Май 2022

АННОТАЦИЯ

Омесь, Людмила: Корпусное исследование языковых особенностей информативных и оперативных текстов о COVID-19 на русском и финском языках, а также в переводах на эти языки.
Магистерская работа
Университет Тампере
Магистерская программа "Многоязычная коммуникация и переводоведение (перевод русского языка)"
Май 2022

В данной работе исследуются языковые особенности информативных и оперативных текстов о COVID-19 на русском и финском языках, а также в переводах на эти языки. Тексты с информацией о мерах предосторожности заражения коронавирусом, о действиях граждан при заражении, о способах лечения, о возможности получении вакцины, а также о положении в стране и мире собраны с сайтов ВОЗ, министерств и центров здравоохранений, правительственных порталов, медицинских и социальных учреждений и т. д.

В работе было проведено корпусное исследование с применением анализа ключевых слов и частотных списков. Для исследования было создано четыре сопоставительных корпуса: 1) оригинальные тексты на русском языке, 2) переводные тексты на русский язык, 3) оригинальные тексты на финском языке и 4) переводные тексты на финский язык. Анализ корпусов осуществлен с помощью программы Sketch Engine (утилиты Wordlist, Word Sketch, Thesaurus, N-grams, Keywords и Concordance).

В теоретической части рассматриваются аспекты исследования и лингвистический статус подъязыка в языковой системе с точки зрения терминологии. Далее описываются особенности информативных и оперативных текстов и особенности языка переводных текстов. В данной работе освещаются этапы развития корпусной лингвистики, а также рассматривается классификация корпусов текстов и использование их в лингвистических исследованиях.

В практической части описываются результаты корпусного исследования, которые базируются на анализе ключевых слов и частотных списков. Отдельно рассматривается использование терминов *covid* и *коронавирус* во всех текстах. Для слова *коронавирус* с помощью утилиты Thesaurus получены облака слов из каждого корпуса. Данные, полученные из корпусов, продемонстрированы как в письменном виде, так и в виде таблиц и рисунков. Сделанные в результате анализа выводы подкрепляются 86 примерами предложений из текстов.

Результаты исследования показали, что тексты одного дискурса, одного жанра и, предназначенные выполнять одну и ту же функцию, могут отличаться друг от друга не только на лексическом, но и на морфологическом и синтаксическом уровнях. Выяснилось, что в текстах на русском языке медицинская тематика выражена ярче, чем в текстах на финском языке. Однако, тексты подъязыка о коронавирусе данного исследования связывают многие факторы. Интересное наблюдение было сделано при анализе оригинальных текстов на финском языке – присутствие темы труда. Эта тема может быть интересной для дальнейших исследований.

Ключевые слова: COVID-19, подъязык, корпусное исследование, ключевые слова, частотные слова.

Оригинальность исследования проверена с помощью программы Turnitin OriginalityCheck.

TIIVISTELMÄ

Ludmila Omes: COVID-19-diskursin kielellisten piirteiden korpustutkimus venäjän- ja suomenkielisissä informatiivisissa ja operatiivisissa teksteissä sekä käännöksissä näille kielille.
Pro gradu -tutkielma
Tampereen yliopisto
Monikielisen viestinnän ja käännöstieteen maisteriohjelma, venäjän opintosuunta
Toukokuu 2022

Tässä tutkielmassa tutkitaan COVID-19-diskurssin informatiivisten ja operatiivisten tekstien kielellisiä piirteitä autenttisissa venäjän- ja suomenkielisissä sekä näille kielille käännetyissä teksteissä. Tekstit, jotka sisälsivät tietoa kansalaisten toimista ja varotoimista koronavirustartunnan sattuessa, hoitomenetelmistä, rokotteen saatavuudesta sekä koronatilanteesta kansallisesti ja maailmanlaajuisesti, on kerätty hallituksen portaaleista, WHO:n, ministeriöiden, terveyskeskusten, terveys- ja sosiaalialan laitoksien jne. verkkosivuilta.

Opinnäytetyössä on tehty korpustutkimus avainsana-analyysin ja frekvenssiluetteloiden avulla. Tätä tutkimusta varten olen luonut Sketch Engine -ohjelmassa neljä vertailevaa korpusta: 1) alkuperäiset tekstit venäjäksi, 2) käännetyt tekstit venäjäksi, 3) alkuperäiset tekstit suomeksi ja 4) käännetyt tekstit suomeksi. Korpuksien analysoinnissa käytin Sketch Engineä ja sen apuohjelmia Wordlist, Word Sketsh, Thesaurus, N-grams, Keywords ja Concordance.

Teoreettisessa osassa käsitellään erikoiskielen tutkimusnäkökohtia ja sen asemaa kielijärjestelmässä terminologian näkökulmasta. Tutkimusaineisto sisältää informatiivisia ja operatiivisia tekstejä alkuperäisillä ja käännetyillä kielillä, joten kuvataan myös näiden tekstien erityispiirteitä. Tarkastelen tutkielmassani muun muassa korpusten ja korpuslingvistiikan kehitysvaiheita, korpusten luokittelua nykyaikaisessa kielitieteessä sekä niiden käyttöä kielentutkimuksessa.

Analyysissa kuvataan korpustutkimuksen tuloksia, jotka perustuvat avainsanojen ja frekvenssiluetteloiden analysointiin. Esimerkiksi, termien *covid* ja *koronavirus* käyttöä kaikissa teksteissä tarkastellaan erikseen, sanalle *koronavirus* saadaan sanapilvet kustakin korpuksesta käyttämällä Thesaurus-toimintoa. Tutkimuksen tulokset esitetään sekä sanallisesti että taulukoiden ja kuvioiden muodossa. Ensin kerron analyysin tuloksista kunkin korpuksen osalta erikseen, sitten vertaan tietoja keskenään ja lopuksi kirjoitan tuloksista yhteenvedon. Analyysin perusteella tehtyjä päätelmiä tuetaan 86 esimerkkilauseella teksteistä.

Tutkimuksen tulokset osittivat, että samaan diskurssiin ja genreen kuuluvat tekstit sekä alun perin samaan tehtävään tarkoitetut tekstit voivat poiketa toisistaan paitsi leksikaalisesti myös morfologisesti ja syntaktisesti. Analyysissa totesin, että venäjänkielisissä teksteissä käytetään enemmän lääketieteellisiä termejä ja termiä *koronavirus* käytetään näissä teksteissä oikeammin lääketieteen näkökulmasta. Tämän tutkimuksen tekstejä yhdistää kuitenkin erikoiskieli. Mielenkiintoisena havaintona on työideologian läsnäolo supisuomenkielisissä teksteissä. Tämä aihe voisi olla kiinnostavana jatkotutkimuskohteena.

Avainsanat: COVID-19, erikoiskieli, korpustutkimus, avainsanat, frekvenssisanat.

Tämän julkaisun alkuperäisyys on tarkastettu Turnitin OriginalityCheck –ohjelmalla.

ОГЛАВЛЕНИЕ

1. Введение 1

1.1. Актуальность исследования 1

1.2. Предмет и методы исследования 3

1.3. Цели и задачи исследования 4

1.4. Разделы работы 5

2. Лингвистическое исследование подъязыка с помощью корпусов 6

2.1. Подъязык 6

2.1.1. Что такое подъязык? 6

2.1.2. Место подъязыка в языковой системе 11

2.1.3. Особенности подъязыка информативных и оперативных текстов 13

2.1.4. Подъязык переводных текстов 16

2.1.5. Переводческие универсалии в корпусном переводоведении 18

2.2. Лингвистическое исследование с помощью сопоставимых корпусов 23

2.2.1. Появление корпуса и развитие корпусной лингвистики 23

2.2.2. Понятие корпус и классификация корпусов 27

2.2.3. Применение сопоставительных корпусов в лингвистических исследованиях 31

2.2.4. Система Sketch Engine для работы с корпусами 33

3. Корпусное исследование языковых особенностей текстов о COVID-19 38

3.1. Материал и методы исследования 38

3.1.1. Отбор материала исследования 38

3.1.2. Описание корпусов 40

3.1.3. Инструменты и методы исследования 42

3.2. Работа с сопоставительными корпусами в программе Sketch Engine 43

3.2.1. Частотные списки корпусов РУСО и РУСП 43

3.2.2. Сравнение N-грамм корпусов РУСО и РУСП 48

3.2.3. Ключевые слова корпусов РУСО и РУСП 49

3.2.4. Слова *covid* и *коронавирус* в корпусах РУСО и РУСП 51

3.2.5. Частотные списки корпусов ФИНО и ФИНП 54

3.2.6. Сравнение N-грамм ФИНО и ФИНП 58

3.2.7. Ключевые слова корпусов ФИНО и ФИНП 60

3.2.8. Слова *covid* и *коронавирус* в корпусах ФИНО и ФИНП 62

3.3. Общие и отличительные черты текстов исследования 64

3.3.1. Частотность слов 64

3.3.2. N-граммы 66

3.3.3. Сопоставление списков ключевых слов 67

3.3.4. Особенности использования слов *covid-19* и *коронавирус* в корпусах данного исследования 69

3.3.5. Особенности переводных текстов в корпусах РУСП и ФИНП 73

4. Заключение 77

4.1. О результатах исследования 77

4.2. Вопрос и гипотеза исследования 78

4.3. Всё ли нам удалось? 79

4.4. Перспективы и польза исследования 79

Список литературы 81

Приложение 85

Список веб-сайтов, с которых найдены тексты для исследования 85

1. ВВЕДЕНИЕ

Данная магистерская работа основывается на корпусном исследовании собранных нами текстов о коронавирусе на русском и финском языках, а также переводов на эти языки. В данной главе излагается важность и актуальность выбранной темы, кратко рассказывается о материале, предмете и методах исследования, освещаются цели и задачи работы. Также отмечается теоретическая обоснованность темы и описывается структура работы.

1.1. Актуальность исследования

С появлением новых технологий меняется характер чтения и характер передачи информации: зачастую читателю необходимо быстро адаптироваться в изменяющемся контексте и уметь адекватно воспринимать информацию. Ежедневно в мире происходят различные катастрофы, чрезвычайные ситуации, природные катаклизмы. Как правило, все эти явления сказываются на повседневной жизни людей одной или нескольких стран и отражаются на экономике и политике. Информация о различных событиях распространяется молниеносно по всему миру на разных языках посредством информационно-телекоммуникационных сетей, в том числе через Интернет. От восприятия и оценки подаваемой информации во многом зависит поведение человека и общества в целом.

В данной работе рассматриваются языковые особенности и коммуникативные факторы, влияющие на эффективность восприятия и понимания информации оригинальных[1] и переводных текстов.

То, что началось в 2019 году и продолжается по сей день затронуло всё человечество. Первое официальное сообщение о вспышке в китайском Ухане коронавируса SARS-CoV-2, вызывающего атипичную пневмонию, появилось в мировых СМИ 31 декабря 2019 года. С этого момента эпидемия, вызванная коронавирусом, начала мгновенно распространяться по всему миру, превращаясь в пандемию. Сейчас на карте мира

[1] Под термином *оригинальный текст* здесь и далее подразумевается непереведённый, созданный на исходном языке текст.

невозможно найти страны, которую не коснулась бы данная проблема. Во время пандемии некоторым городам, даже целым странам, пришлось прибегнуть к полной изоляции и к объявлению общенационального карантина, во время которого была приостановлена работа предприятий, учреждений, транспорта, существенно ограничена свобода передвижения людей в пределах и за пределами страны. Практически прекратилось международное авиасообщение, отменились конференции, концерты и спортивные мероприятия. Вся общественная жизнь была вынуждена перейти на онлайн.

Во всём мире проводится активная работа по борьбе с коронавирусом – это и памятки, и объявления в общественных местах, и челленджи с призывами, и даже штрафы за невыполнение предписаний.

По данным информационного агентства Интерфакс от 11.05.2022 во всём мире от коронавируса умерло 6 255 511 человек, количество зараженных – 518 739 086 (Интерфакс 2022). Только с появлением вакцины у человечества появилась надежда на улучшение ситуации, ежегодно вакцины спасают миллионы жизней. По данным Всемирной организации здравоохранения (ВОЗ) уже в феврале 2021 года во всём мире применялось по меньшей мере семь различных вакцин. Клинические исследования продолжаются, и более 200 вакцин-кандидатов находятся на этапе разработки (ВОЗ 2022).

Так как пандемия продолжается и по сей день, правительства всех стран бросают все силы и средства на мероприятия по борьбе с распространением коронавируса. Одно из таких действий – это своевременное оповещение населения о чрезвычайной ситуации и предоставление информации о правилах поведения, а также доступность этой информации.

Например, информация о мерах предосторожности заражения коронавирусом, о действиях граждан при заражении, о способах лечения или о возможности получении вакцины является общедоступной. Официальную информацию на разных языках можно получить на сайтах ВОЗ, министерств и центров здравоохранения, правительственных порталов, медицинских и социальных учреждений и т. д.

В данном корпусном исследовании были использованы тексты правил, рекомендаций, оповещений и сводок новостей о соблюдении мер предосторожности во время пандемии нового коронавируса, COVID-19. Например, «Инструкции по защите от коронавируса и лечению болезни», «Инструкции для прибывающих в Финляндию путешественников и

работников», которые опубликованы на сайте Национального института здравоохранения и социального благосостояния (THL) на финском и русском языках; информация о возможных путях заражения коронавирусом, о профилактике лечения коронавирусной инфекции, о вакцинации против COVID-19, которая доступная на сайте официальной информации о коронавирусе в России, «Стоп коронавирус» и. т. д.

Важность и актуальность данной работы заключается в том, что результатами исследования могут воспользоваться составители текстов и переводчики по данной тематике.

1.2. Предмет и методы исследования

Для данного исследования были использованы тексты определённой тематики и определённого жанра, что подразумевает наличие определённой терминологии и стиля языка, связывающих данные тексты. Появление новых реалий ведет к появлению новых наименований и специальных слов в языке. В данном случае мы имеем дело с подъязыком, который объединяет тексты исследования о COVID-19.

Ещё три года назад мы не поняли бы значение и смысл некоторых фраз, которые используются сегодня в СМИ, общественных местах и в разговорной речи. Например, знаки и таблички при входе в общественные заведения с надписями: «Без маски вход воспрещён!», «Используйте маску», «Внимание! Соблюдайте дистанцию 1,5 м!», «Используйте санитайзер» и др. сегодня мы воспринимаем абсолютно спокойно.

С появлением коронавируса появились новые тексты подъязыка о COVID-19. Как правило, новое явление вызывает появление неологизмов в языке. Так и с появлением нового коронавируса возникли слова *ковид, ковидный*, а также словосочетания *ковидные выплаты*, *ковидные каникулы*, *ковидный диагноз* и т. д. (Громенко и др. 2021, 56–59). В разговорной речи как в русском, так и в финском языке часто используется слово *корона/korona*, образованное от *коронавирус/koronavirus*. Так со словом *корона* появились выражения: *сдать тест на корону*, *симптомы короны*, *заразиться короной*, *поймать/подхватить корону* и др. Вполне возможно, что через некоторое время для понимания этих фраз читателю придётся знакомиться с реалиями нашего времени.

В данном исследовании не обошлось без использования компьютерных технологий, благодаря которым появилась возможность создания и использования «больших по объёму совокупностей языкового материала, которые стали называть корпусами»

(Копотев & Мустайоки 2008, 8). Результаты данной работы основываются на сделанном нами корпусном исследовании.

Корпусная лингвистика – это направление лингвистики, инструментом которой являются корпусы (Майорова 2017, 42). Исследование особенностей лексического состава корпусов текстов было произведено нами с помощью анализа ключевых слов.

По нашему мнению, важность и актуальность этой темы заключается ещё и в том, что представители различных сфер деятельности, включая лингвистов, активно участвуют в борьбе с коронавирусом. Участие лингвистов реализуется в исследовании текстов данной тематики как средства передачи коммуникации, в изучении языковых средств, влияющих на оперативность информации, лингвистами также создан корпус текстов о COVID-19 на английском языке.

1.3. Цели и задачи исследования

Цель данной работы – исследование языковых единиц на лексическом, морфологическом и синтаксическом уровне, анализ и сравнение функций текстов с помощью сопоставимых корпусов. Мы попытаемся исследовать, какую функцию тексты выполняют более эффективно, информативную или оперативную. Будут также рассмотрены языковые факторы, влияющие на эффективность восприятия и понимания информации.

Можно предположить, что оригинальные тексты могут отличаться от переводных, также тексты на русском и финском языках будут отличаться, т. к. в них существуют различия на лексическом и грамматическом уровне.

Для реализации поставленной цели необходимо было решить следующие задачи:

- Создать корпусы текстов определённого стиля и одной тематики с оригинальными и переводными текстами на русском и финском языках.
- Провести корпусное исследование для выявления необходимого материала.

После выполнения данной прикладной задачи мы приступили к лингвистическому исследованию и попытались решить следующие теоретические проблемы:

- Выделить ключевые языковые особенности в корпусе текстов на русском языке и в корпусе переводов на русский язык. Сравнить эти особенности.

- Выделить ключевые языковые особенности в корпусе текстов на финском языке и в корпусе переводов на финский язык. Сравнить эти особенности.
- Выяснить, отличается ли язык дискурса о COVID-19 в русском и финском языках. Если различия существуют, чём они заключаются?

1.4. Разделы работы

Магистерская работа состоит из четырёх разделов: введение, теоретическая часть, практическая часть и заключение.

В теоретической части исследования подробно рассматриваются понятия *функциональный стиль, подъязык и ЯСЦ*. Для этого были проанализированы работы В. В. Виноградова (1955), М. Н. Кожиной (2002), И. В. Кудашева (2007), Т. А. Воронцовой (2009), А. В. Раздуева (2010), З. С. Хасановой (2013) и др. Так как данная работа основана на корпусном исследовании, мы посчитали необходимым познакомить читателя с основными терминами корпусной лингвистики. Для описания корпуса, корпусной лингвистики, истории и классификации корпусов были использованы учебные пособия В. П. Захарова и С. Ю. Богдановой (2020), а также А. А. Шунейко (2020), научные труды и статьи М. Н. Михайлова (2003), С. А. Шарова (2003), А. Д. Майоровой (2017), М. А. Куниловской (2015; 2017; 2018) и др.

В практической части подробно описывается материал исследования, корпусы и инструменты для работы. Сначала освещаются результаты исследования для каждого корпуса отдельно, затем полученные данные сравниваются и предоставляются предположения, умозаключения и выводы.

В заключении демонстрируются основные результаты работы и ответы на вопросы исследования, рассказывается об удачах и неудачах и освещаются возможные перспективы дальнейшего исследования. К работе прилагается также список используемой литературы и реферат работы на финском языке. Приложение содержит список сайтов, с которых собраны тексты для исследования.

2. ЛИНГВИСТИЧЕСКОЕ ИССЛЕДОВАНИЕ ПОДЪЯЗЫКА С ПОМОЩЬЮ КОРПУСОВ

Как отмечалось ранее, предметом исследования данной работы являются языковые особенности дискурса о COVID-19, т. е. подъязык текстов для информирования населения о коронавирусе.

В данной главе мы познакомимся с термином *подъязык* и попытаемся выяснить, какое место он занимает в языковой системе. Для изучения каких-либо языковых явлений в определённых текстах необходимо знать, какую функцию тексты предназначены выполнять. Тексты данного исследования относятся к информативному и публицистическому подстилю публицистического стиля, поэтому мы кратко осветим классификацию функциональных стилей и также рассмотрим особенности приёмов передачи информации в оперативных и информативных текстах данных подстилей как в оригинальном языке, так и в переводе.

Исследовательская работа была выполнена с помощью сопоставительных корпусов, поэтому в данной главе рассматриваются также термины *корпус* и *корпусная лингвистика*, приводятся примеры корпусных исследований и описываются возможности программы для работы с корпусами, Sketch Engine.

2.1. Подъязык

Один из важных терминов данного исследования – это подъязык, используемый для специального общения. Анализ литературы показал, что в современной лингвистике нет единого понимания этого термина и его места в языковой системе. Иногда термины *функциональный стиль*, *подъязык* и *язык для специальных целей (ЯСЦ)* или *Language for Special Purposes (LSP)* сопоставляются, однако, по нашему мнению, их следует различать. В данном разделе мы попытаемся выяснить, что такое подъязык и его особенности в оперативных и информативных текстах, сравнивая взгляды разных учёных.

2.1.1. Что такое подъязык?

Из определения Б. Н. Головина следует, что язык – это «исторически возникший и развивающийся сложный знаковый механизм общения, работающий в единстве и

взаимодействии с сознанием и мышлением человека» (Головин 1983, 10). Это определение подтверждает коммуникативную функцию языка во всех сферах общества. Язык следует рассматривать как совокупность подсистем, для этого выделяют понятие подъязыка.

В современной лингвистике существует много дискуссий по поводу структуры языковой системы и содержания подсистем. Например, Т. Н. Хомутова подчёркивает, что многие лингвисты уравнивают такие подсистемы, как функциональный стиль, подъязык и ЯСЦ (Хомутова 2007, 55–59).

Для начала выясним, чем отличаются термины *функциональный стиль*, *подъязык* и *ЯСЦ*.

Понятие функционального стиля языка зарождалось во второй половине прошлого столетия, хотя изучение функционального стиля началось гораздо раньше. Существовало множество дискуссий и споров по вопросам стилистики, особенно оживлённые споры разгорелись в 1954–1955 гг. Это подтверждается появлением различных взглядов, на которые В. В. Виноградов ссылается в статье «Итоги обсуждения вопросов стилистики» (Виноградов 1955).

По мнению автора статьи, расхождение взглядов заключалось в разном «понимании сущности «языка» художественной литературы и его места в системе стилей литературной речи» (Виноградов 1955, 85). В своей статье автор анализирует и сравнивает взгляды участников дискуссии по стилистике: Ю. С. Сорокиной, А. Н. Гвоздевой, Р. А. Будагова, И. Р. Гальперина, В. Д. Левина и др. Виноградов использует также материалы работ В. Гумбольдта, Л. В. Щербы, А. И. Смирницкого и др. Именно тогда, в 1955 году, Виноградов даёт определение функционального стиля (там же, 60–87):

> Стиль – это общественно осознанная и функционально обусловленная, внутренне объединенная совокупность приемов употребления, отбора и сочетания средств речевого общения в сфере того или иного общественного, общенационального языка, соотносительная с другими такими же способами выражения, которые служат для иных целей, выполняют иные функции в речевой общественной практике данного народа (Виноградов 1955, 73).

В современной лингвистике функциональный стиль трактуется по-разному. В данной работе приводятся примеры определений более традиционного подхода изучения стиля.

Например, Т. А. Воронцова высказывает мысль о том, что функциональный стиль базируется на особой системе языковых средств, и типичные средства выражения способствуют эффективности общения в типичных, повторяющихся ситуациях. Именно необходимость более эффективного общения в социально значимых сферах

коммуникации способствовала выделению функциональных стилей. По мнению автора, функциональный стиль – «разновидность литературного языка, обусловленная различием функций, выполняемых языком в определенной сфере общения». Под социально значимой сферой общения автор подразумевает «область коммуникации, соотносимую с видом общественной деятельности и формой общественного сознания, в которую так или иначе вовлекаются все носители данного языка» (Воронцова 2008, 41–42).

М. Н. Кожина также считает, что сфера общения является одним из важнейших понятий функциональной стилистики. По мнению автора статьи, сфера общения объединяет вид деятельности и форму общественного сознания, которые «дают возможность определить понятие функционального стиля и классифицировать стили на едином основании» (Кожина 2002, 84–89).

Из вышеприведённых определений можно сделать вывод, что функциональный стиль определяется в зависимости от определённой деятельности людей. Как известно, лексика, которую мы используем в повседневном общении, отличается от языка специальной сферы коммуникации, т. к. для понимания и передачи информации нам не требуются специальные знания. Так для языка специальной сферы коммуникации появляется новое название – подъязык. Выделение и использование лексических единиц, характерных для определённой сферы влечёт к появлению языковой подсистемы. Понятно, что прогресс во всех областях научного знания обуславливает развитие языка, соответственно появление подъязыков.

З. С. Хасанова утверждает, что в отечественном языкознании термин *подъязык* появился в 60-е гг. XX века в работах Н. Д. Андреева, который первым ввёл термин подъязыка, выделив в научном стиле более «дробную» единицу. Хотя ещё раньше данный термин использовался в английском языке. Хасанова подчёркивает важность разделения подъязыков специальных сфер коммуникации и языка сферы массовой и индивидуальной коммуникации (Хасанова 2013, 64–67).

Согласно определению Н. Д. Андреева, подъязык – это «набор языковых элементов и их отношений в текстах с однородной тематикой». В своей работе он утверждает о необходимости выделения «подъязыковой типологии и отраслевой лингвистики в отдельные разделы науки о языке». Андреев считает, что существует типология

подъязыков, и выделяет ее «экстралингвистические и внутрилингвистические особенности» (Андреев 1965, 91).

Итак, мы соглашаемся с Хасановой, что подъязык связан с определённой предметной областью, а функциональный стиль определяется в зависимости от определённой деятельности людей (Хасанова 2013, 67). Также мы поддерживаем мнение Воронцовой, что стиль реализуется через текст и является одним из его свойств, т. е. у каждого текста есть свой стиль. От того, в каком регистре будет произнесено то или иное выражение, будет зависеть его смысл (Воронцова 2008, 18).

Чтобы ещё лучше понять термин *подъязык*, мы также рассмотрим термин *ЯСЦ*, у которого также существует много вариантов определения из-за расхождений мнений в вопросе об отличительных признаках ЯСЦ. Например, И. В. Кудашев подчёркивает, что зачастую языки для специальных целей выделяют в составе национальных языков и рассматриваются многими исследователями «как функциональные разновидности национальных языков» (Кудашев 2007, 73).

Т. Н. Хомутова отмечает, что ЯСЦ появился в результате исторического разделения труда, поэтому коммуникация с помощью ЯСЦ стала общественно-исторической необходимостью. Сам термин *ЯСЦ* или *LSP* появился в результате изучения коммуникации людей на те или иные профессиональные темы. Автор подчёркивает, что «текст на LSP – это всегда текст, в котором аккумулировано и сохранено специальное знание» (Хомутова 2007, 55).

И. А. Забросаева и М. Э. Конурбаев пишут, что авторами термина LSP являются Т. Хатчинсон (Hutchinson) и А. Уотерс (Waters), которые использовали его в работе English for Specific Purposes: A Learning-Centered Approach в 1987 году. К функциям LSP авторы статьи причисляют «описание концептуальной и языковой картины данной области знания, осуществление профессиональной нормированности и адекватности лексических средств, хранение терминологий, номенклатур, специальных речевых оборотов и синтаксических конструкций» (Забросаева и Конурбаев 2014, 27–39).

ЯСЦ служит для передачи объективной научной информации об окружающем мире. М. В. Лейчик (1982, 1986) считает, что характерными признаками языков для специальных целей являются имена нарицательные, собственные и номенклатурные единицы, а также унифицированность, т. е. использование определённых лексических единиц. Использование в текстах иноязычных морфем, словосочетаний, образующих

термины, и сжатость информации также являются признаками ЯСЦ (цит. по: Забросаева и Конурбаев 45–47).

Лейчик (2011) не исключает наличия в ЯСЦ разговорного стиля, профессионализмов и профессионального жаргона, т. к. языки для специальных целей широко используются в повседневном общении (цит. по: Забросаева и Конурбаев, 50).

По мнению Кудашева, ЯСЦ – «это совокупность естественных или естественно-искусственных языковых средств, использующаяся в какой-либо области знаний и/или деятельности главным образом для передачи предметной информации и отражающая понятийный аппарат, не являющийся достоянием большинства носителей данного национального языка» (Кудашев 2007, 74).

А. В. Раздуев также подчёркивает значимость терминологии в ЯСЦ, заявляя, что она «является смысловым ядром и становится своего рода визитной карточкой и правом входа в мир профессиональной коммуникации» (Раздуев 2010, 90–95). Автор утверждает, что ЯСЦ «представляет собой высокоспециализированный динамически функционирующий лексический пласт, моделирование которого позволяет упорядочить коммуникативные отношения, имеющие место в специализированной предметной области…» (Раздуев 2013, 50).

Так же А. А. Новикова говорит об языке для специальных целей, который обслуживает определенную область знаний и содержит большое количество терминов и ключевых слов. Зачастую содержащий огромный терминологический пласт лексики, LSP называют «профессиональным подъязыком» (Новикова 2020, 73).

Таким образом, понятие ЯСЦ является очень близким к понятию подъязыка профессиональной коммуникации, который обеспечивает эффективное общение специалистов в определённой предметной области.

Подводя итог об анализе выше рассмотренных терминов, мы соглашаемся с определением подъязыка, предложенным Хасановой: «подъязык – это малая лингвистическая подсистема, которая содержит единицы как профессиональных, так и непрофессиональных уровней языковой системы, которые используются в текстах определённой предметной тематики» (Хасанова 2013, 70).

2.1.2. Место подъязыка в языковой системе

Исходя из определений ранее рассмотренных подсистем, функциональный стиль, подъязык и ЯСЦ (см. главу 2.1.1.), выяснилось, в чём заключаются их отличия. Хотя, действительно, в определениях данных подсистем существует много общего. Мы соглашаемся с лингвистами, разделяющими эти подсистемы между собой, и попытаемся найти место подъязыка в единой системе языка.

В процессе анализа теоретического материала выяснилось, что в классификации функциональных стилей также наблюдается расхождение во взглядах. Например, Т. А. Воронцова утверждает, что в русской лингвистике традиционно выделяют пять функциональных стилей: «научный, публицистический, официально-деловой, разговорный и художественный». Однако автор уточняет, что по поводу разговорного и художественного стиля среди лингвистов бытуют разные мнения. У каждого стиля в свою очередь могут быть свои подстили. Например, в научном стиле выделяют учебно-научный и научно-популярный подстили, которые выполняют разные задачи. Тексты могут включать в себя элементы разных стилей, поэтому выделение стиля в текстах происходит по тому, какие черты того или иного стиля преобладают (Воронцова 2008, 45–46).

Как упоминалось ранее, о принципах выделения функциональных стилей велось много дискуссий. Среди учёных, пытавшихся выделить функциональные стили по определённым признакам или функциям, Воронцова выделяет А. М. Пешковского, В. В. Виноградова, Д. Н. Шмелева и М. Н. Кожину. Большинство учёных считает, что выделение базовых функциональных стилей должно происходить на основе разделения социально значимых сфер коммуникации (там же, 46–47).

М. Н. Кожина, Л. Р. Дускаева и В. А. Салимовский также считают, что в результате появления новых сфер выделяются новые подстили, например, «законодательный и административно-хозяйственный, технический и патентный деловые стили» (Кожина и др. 2011, 412–427).

Итак, выяснилось, что типы функциональных стилей, являющихся разновидностью литературного языка, функционируют в той или иной социально значимой сфере общества, а подъязык связан с определённой предметной областью. В отличие от того, какую задачу выполняют тексты, происходит выделение подстилей, язык которых

отличается друг от друга, появляется необходимость в использовании подъязыка. Соответственно понятие подъязыка у́же понятия функциональный стиль.

Наше предположение об узости понятия подъязыка по отношению к функциональному стилю подтверждается высказыванием Хасановой, которая утверждает, что понятия подъязык и функциональный стиль могут быть соотнесены только в том случае, если какой-нибудь подъязык соответствует только одному определённому стилю, что пока не обнаружено. Сходство подъязыка и функционального стиля заключаются в том, что 1) лексические языковые единицы заимствуются из научных источников, 2) обе языковые разновидности могут применяться в публицистике (Хасанова 2013, 67–68).

Т. Н. Хомутова, хоть и разделяет понятия *подъязык и ЯСЦ*, но считает ЯСЦ подъязыком и сравнивает вышеупомянутые понятия в сравнении с ЯСЦ (напр., ЯСЦ *vs*. функциональный стиль, ЯСЦ *vs*. жанр, ЯСЦ *vs*. регистр). Автор также высказывает мнение о широте понятия функциональный стиль, в котором, в силу постоянного развития социальных и производственных отношений, а также появлению новых сфер деятельности, рождается разнообразие подстилей (Хомутова 2007, 57–58).

Хомутова считает, что *подъязык* является более широким понятием, чем *ЯСЦ*. Автор выделяет отличительные черты этих двух понятий, например, в подъязыке присутствует более простая лексика, понятная как специалистам, так и неспециалистам той или иной области знаний. Информация может передаваться с помощью различных поясняющих механизмов и с использованием неспециализированной лексики. Специальное знание может быть общедоступным для понимания. Автор рассматривает подъязык «как часть естественного языка, описывающую определённую область» (там же, 56–57).

Сравнивая понятия *ЯСЦ* и *регистр*, Хомутова определяет их разницу. «В отличие от LSP регистр относится к употреблению языка (речи) отдельного индивида, его способности переключать роли в зависимости от ситуативного контекста речевого акта». Автор утверждает, что регистр описывает дискурс с помощью трех категорий: содержательная (использование подъязыка), функциональная (использование определённого стиля) и формальная (использование определённого канала общения и жанра). В сравнении ЯСЦ и жанра также разница в спектре их функций. Жанр является более широким понятием по отношению к ЯСЦ, как подъязыку. Разделение на жанры определяется языковой вариативностью на уровне дискурса (там же, 58–59).

Подводя итог о разнице между вышеприведёнными понятиями *функциональный стиль, жанр, регистр, подъязык и ЯСЦ*, мы позволим себе высказать своё мнение о месте подъязыка в языковой системе. Мы соглашаемся с учёными, считающими ЯСЦ отдельной языковой подсистемой и отделяющими его от подъязыка, т. к. ЯСЦ используется для специальной коммуникации между специалистами-профессионалами в определенной области, характеризующейся своей терминологической лексикой.

В тоже время нельзя отрицать мнения учёных, которые называют ЯСЦ подъязыком. Как выяснилось подъязык имеет место везде, где существуют какие-либо отклонения от литературного языка, связанные с предметной деятельностью. На подъязыке могут общаться между собой профессионалы определённой специальности, используя ЯСЦ, а также специалисты и неспециалисты определённой сферы деятельности.

Итак, совокупность языковых средств с наличием какой-либо специальной лексики (язык моряков, юристов, бухгалтеров, жаргоны, язык людей с одинаковыми интересами, вкусами или работающих на одном и том же предприятии и др.) может считаться подъязыком.

2.1.3. Особенности подъязыка информативных и оперативных текстов

Как выяснилось в подразделе 2.1.2., подъязык может использоваться в разных функциональных стилях. Часть текстов данного исследования относятся к публицистическому стилю, который, по мнению Н. А. Купиной и Т. В. Матвеевой, «посвящен обсуждению общественно значимых вопросов» и предназначен для воздействия на широкий круг читателей (Купина и Матвеева 2013, 234–235).

Как правило, все тексты имеют прагматическую составляющую, в информационном отношении публицистический стиль служит для передачи актуальной и новой информации для широких масс. В зависимости от информативных или экспрессивно-воздействующих смыслов, по мнению Купиной и Матвеевой, публицистический стиль делится на подстили: информационный, аналитический и художественно-публицистический (там же, 234–235).

Однако, некоторые исследователи, например, Г. Я. Солганик и Т. С. Дроняева, не выделяют в нём подстилей, а называют публицистический стиль стилем массовой

коммуникации, который реализуется в разных жанрах, например интервью, репортаж, статья, очерк, обозрение и др. (Солганик и Дроняева 2005, 21–50).

Ранее мы выяснили, что каждый стиль служит для выполнения какой-либо функции, задача публицистического стиля – проинформировать и оказать воздействие на адресата. Купина и Матвеева отмечают, что в отличие от цели воздействия (убеждение, агитация, развлечение) в публицистике могут применяться самые разнообразные языковые средства (Купина и Матвеева 2013, 234–253).

Для данного исследования использовались тексты с правилами, оповещениями и с новостями, которые призваны выполнять информативную и оперативную функции[2]. По мнению Купиной и Матвеевой, тексты информативного жанра (информационное сообщение, научная статья, научно-популярный текст, инструкция и др.) обладают «объективной тональностью» (там же, 265–287).

Что касается оперативных текстов, к которым относятся, например, инструкции, в них также существуют определённые языковые особенности, влияющие на восприятие информации адресатом. О. В. Хорохордина выделяет «типологические характеристики текста-инструкции:

1) Типовая интенция, целью которой является научить, как действовать в какой-либо ситуации. Тексты приобретают директивную (императивную) тональность.
2) Типовое содержание и его композиция. Для достижения определённой цели используется достоверная статистическая информация.
3) Типовая завершённость в ответной реакции адресата, которая в идеальном сценарии предполагает исполнение им включённых в инструкцию предписаний.
4) Типовые средства языкового выражения» (Хорохордина 2013, 7–9).

Автор статьи разделяет тексты-инструкции «по формально-семантическим типам» на собственно прескриптивный, ограничительно-прескриптивный, наглядно-прескриптивный и вопросительно-прескриптивный (там же, 9).

---

[2] «Под оперативной функцией понимается побуждение к действию» (Алексеева 2004, 129). Как отмечает И. С. Алексеева, оперативные тексты (реклама, проповедь, инструкция) посвящены одной определённой теме и воздействуют на адресата, информативные тексты (информационное сообщение, научная статья, научно-популярный текст, оповещения) выполняют коммуникативную функцию и служат для передачи актуальной информации (там же, 247).

К типовым средствам языкового выражения в текстах-инструкциях собственно прескриптивного типа, в котором в наибольшей степени проявляется директивная тональность, Хорохордина относит использование различных морфологических форм волеизъявления. Они реализуются с помощью следующих конструкций: инфинитивные, императивные, совместно-императивные, с косвенным императивом, с прошедшим временим волеизъявления, а также «представляющие собой номинативные конструкции с существительными со значениями действия/состояния» (Хорохордина 2013, 9).

Ограничительно-прескриптивный тип, по мнению автора, также окрашен директивной тональностью, но в текстах-инструкциях данного типа присутствует информативная нацеленность. В текстах этого типа присутствуют синтаксические конструкции со значением необходимости, (не) допустимости, разрешения/запрета, которые выражаются с помощью модальных глаголов (необходимо, (не) должен, (не) следует и др.); форм настоящего или будущего времени глагола, включая пассив (там же, 10).

Как отмечает далее Хорохордина (2013, 10), в наглядно-прескриптивный типе директивная тональность менее заметна. В текстах-инструкциях этого типа используется изъявительное наклонение глагола (преимущественно настоящего времени). Функцией такого текста является сообщение о порядке совершаемых (совершённых) действий для достижения определённого результата. «Делай, как я (мы, все); сделаю/-ем, -ют и др.» (примеры автора статьи).

В вопросительно-прескриптивном типе, по мнению Хорохординой, присутствует вопросительная интонация, которая сглаживает императивную тональность и переводит «жёсткую директиву в статус более мягкой прескрипции – предложения действовать определённым образом». Тексты-инструкции такого типы характеризуются наличием в них диалогической речи. Также автор подчёркивает, что прототипическим для текстов-инструкций является инструктивный текст, который основан на инфинитивных конструкциях (там же, 10).

По утверждению Хорохординой, инструкция как тип текста является наджанром, обладающий планом содержания и планом выражения. В статье автор также описывает характерные черты для данного наджанра (там же, 12–14).

Выяснилось, что в зависимости от стиля, типа и функции текста язык может приобретать те или иные особенности. До сих говорилось о языковых особенностях информативных и оперативных текстов в дискурсе. Н. К. Гарбовский определяет дискурс «как

совокупность текстов, построенных в соответствии с определёнными для различных коммуникативных ситуаций нормами выбора и использования форм выражения, построения текстов определяет адекватность переводного текста, т. е. соответствие текстов перевода ожиданиям получателей» (Гарбовский 2011, 18).

Частью материала данного исследования являются также переводные тексты, то далее мы попытаемся выяснить особенности языка в переводах.

2.1.4. Подъязык переводных текстов

Как выяснилось ранее, языковые особенности текстов могут зависеть от многих факторов, например от стиля, жанра, регистра, предметной области и т. д. Те же самые факторы влияют на языковые особенности переводных текстов. Перевод является деятельностью переводчика, а также результатом этой деятельности. Существует множество вариантов определения понятия *перевод*, по нашему мнению, вариант, предложенный И. С. Алексеевой, наиболее точно описывает этот термин.

> Перевод — это деятельность, которая заключается в вариативном перевыражении, перекодировании текста, порожденного на одном языке, в текст на другом языке, осуществляемая переводчиком, который творчески выбирает вариант в зависимости от вариативных ресурсов языка, вида перевода, задач перевода, типа текста и под действием собственной индивидуальности; перевод — это также и результат описанной выше деятельности (Алексеева 2012, 7).

Исходя из этого определения, становится понятной сложность этого процесса и то, какую важную задачу выполняет в нём переводчик. И хотя язык даёт переводчику свободу выбора, иногда ему приходится ограничиваться ресурсами языка, например, при переводе географических названий, терминов, особенностей грамматики переводимого языка и т. д. (Алексеева 2012, 5–7).

В зависимости от типа перевода (устный или письменный) и от задачи перевода появляются определённые требования к переводчику. Для того, чтобы начать переводить текст, переводчику необходимо понять, к какому типу текстов он относится. Определив типологические характеристики текста, переводчик выбирает стратегию, методы и приёмы перевода.

До того, как переводоведение обрело статус самостоятельной науки в конце XX века, оно рассматривалось в рамках языкознания. Именно поэтому, как утверждает О. В. Кафискина, переводчики длительное время прибегали к функционально-стилистической классификации текстов. О классификации текстов было рассказано ранее (см. подраздел 2.1.2.). Мы убедились, что классификация функциональных стилей

представляет собой очень сложную и неоднозначную картину. Также выяснилось, что в некоторых жанрах можно наблюдать смешение стилей, образование подстилей и т. д. По мнению Кафискиной, эта неоднозначность также влияла на анализ и практику перевода (Кафискина 2017, 414–415).

Кафискина отмечает, что переводчику необходимо учитывать «транслатологию текста» (теорию текста), которая ориентирована на перевод (там же, 416). Алексеева даёт определение транслатологии текста: «это теория, которая ориентируется на перевод, рассматривает текст и как объект предстоящего перевода, и как результат выполненного перевода». Автор характеризует тексты по следующим параметрам: «1) коммуникативное задание и вид передаваемой текстом информации; 2) характер источника и характер адресата перевода; 3) объективные меры переводимости текста» (Алексеева 2012, 9).

В своей статье Кафискина описывает разные типы классификации текстов, например, П. Ньюмарка (Newmark), Х. Глинца (Glinz), Ф. Люкса (Lioux) и Р. Глазера (Glazer), А. Нойберта (Neubert), а также К. Райс (Reiss). Автор статьи отмечает, что, например Райс, опираясь на коммуникативную функцию текста, выделила четыре основные группы текстов: информативные, экспрессивные, оперативные и аудиомедиальные. Как правило, расхождение во взглядах на классификацию или определение понятия объясняется различием в подходах изучения (Кафискина 2017, 418).

Кафискина считает, что на основе транслатологической классификации типов текстов, которую предложила Райс, Алексеева сформировала свою. Автор статьи утверждает, что при переводе необходимо учитывать вид информации, т. к. определённый вид информации реализуется с помощью устоявшегося набора языковых средств (там же, 418).

Алексеева пишет, что языки отличаются между собой на фонетическом, грамматическом, морфологическом и лексическом уровне, но обладают одинаковой коммуникативной функцией. По мнению автора, например медицинские тексты, рекламы, статьи и др. в разных языках выглядят одинаково, они различаются только формальными признаками на фонетическом, лексическом и грамматическом уровнях языка (Алексеева 2012, 212–214).

В зависимости от состава информации Алексеева выделяет её виды: когнитивная, оперативная, эмоциональная и эстетическая. А в зависимости от преобладающего вида информации в тексте автор выделяет группы текстов:

1) примарно-когнитивные тексты: научно-учебный, научно-популярный, объявления, искусствоведческий, музыковедческий, философский, документы, деловое письмо
2) примарно-оперативные тексты: законодательный, религиозный, проповедь, инструкция, рецепт
3) примарно-эмоциональные тексты: траурное объявление, некролог, беллетристика, публичная речь, реклама, мемуарный текст
4) примарно-эстетические тексты: художественный текст, художественная публицистика (Алексеева 2012, 254–266).

Однако, как отмечает Алексеева, распределение текстов по группам является условным, т. к. информационный состав текста может расходиться с его коммуникативной функцией (там же, 266).

По мнению Кафискиной, именно транслатологическая типология, направленная на перевод, помогает осуществлению перевода. Например, выявив заранее проблемные места будущего перевода, переводчик может выбрать правильную стратегию перевода (Кафискина 2017, 424).

Для того, чтобы переводчик смог наиболее точно передать информацию, заложенную в оригинальном тексте, ему приходится прибегать к различным способам перевода. Как известно, при переводе переводной текст приобретает определённые свойства, которые отличают его от оригинального текста. Далее будут рассмотрены некоторые способы перевода и отличительные черты перевода от оригинального текста.

2.1.5. Переводческие универсалии в корпусном переводоведении

Так как частью материала исследования являются переводные тексты, которые исследуются с помощью корпусов, то соответственно данное исследование относится к области корпусного переводоведения и автоматического анализа языка. Данное исследование переводных текстов основано на теории переводческих универсалий (англ.

translation universals)[3], которая является ведущим направлением корпусного переводоведения.

Как утверждают М. А. Куниловская, Т. А. Ильющеня и М. А. Ковязина, результаты исследования этого направления применяются в области описания и оценки качества текста, а также в дидактике перевода. Авторы статьи отмечают, что теория переводческих универсалий сформировалась с развитием «дескриптивных и социокультурологических идей» в переводоведении, по которым перевод изучается как самостоятельный текст, в отрыве от традиционного сравнения с оригиналом на эквивалентность (Куниловская и др. 2017, 19).

Авторы статьи выделяют работы Г. Тури (Toury 1995), который считает, что перевод, как особая категория текстов, оказывает влияние «на принимающую культуру» и формируются «под воздействием широкого круга социокультурных факторов». К таким факторам Г. Тури относит, например «издательскую политику и особенности действующей нормы перевода, характеризующейся, например, той или иной степенью толерантности к интерференции» (цит. по: Куниловская и др. 2017, 19).

Как известно, язык перевода традиционно является предметом дискуссий, что связано с различием в подходах изучения языка переводных текстов. Авторы статьи отмечают, что другой основой теории переводческих универсалий является представление о «переводческом языке». Авторы также напоминают, что именно с развитием корпусной лингвистики переводоведы могут объективно изучать и описывать особенности «переводческого языка» (Куниловская и др. 2017, 19).

Например, диссертационное исследование Е. С. Краснопеевой, по её словам, выполнено «в русле дескриптивного корпусного подхода и когнитивно-дискурсивной парадигмы в переводоведении». Автор работы опирается на теорию переводческих универсалий и продолжает работу таких исследователей, как М. Бейкер (Baker), С. Бернардини (Bernardini), А. Честерман (Chesterman), Д. Кенни (Kenny), С. Лавиоса (Laviosa), А. Мауранен (Mauranen), Дж. Мандэй (Munday), М. Олохан (Olohan), Р. Сяо (R. Xiao) и др. (Краснопеева 2016, 3).

---

[3] Мона Бейкер (Baker) к переводческим универсалиям относит такие лингвистические свойства переводных текстов, которые отличают их от оригинальных текстов. Причём эти переводческие универсалии применимы к любым языковым парам (Baker 1993, 243).

Краснопеева, наряду с Тури, отмечает Н. К. Гарбовского, В. Н. Комиссарова, Ж. Мунен, М. Геллерстама (Gellerstam), У. Фроули (Frawley), П. Ньюмарка (Newmark), которые высказывают идею о «переводческом» языке как о «третьем коде» и называющих перевод речевой деятельностью особого свойства. Автор статьи, также как и Куниловская, Ильющеня и Ковязина, отмечает широкое применение корпусного подхода в исследованиях дискурса и перевода. Она высказывает мнение, что на анализе количественных данных возможна «верификация статистических гипотез о свойствах перевода» (Краснопеева 2016, 3).

В данной работе будут рассмотрены только некоторые переводческие универсалии. По мнению Куниловской, Ильющеня и Ковязиной, такие переводческие универсалии, как симплификация (анг. simplification), экспликация/эксплиситация (англ. explicitation), нормализация (англ. normalisation) и интерференция (англ. interference,), наиболее исследованы и эмпирически подтверждены с помощью исследований материалов многих языковых пар (Куниловская и др. 2017, 19).

В более ранней работе Куниловской (2015) представлены определения данных универсалий. Например, «симплификация – сознательная или бессознательная тенденция к упрощению лексического состава и структуры исходного текста. Она проявляется в более низких показателях лексического разнообразия и информационной плотности переводов, в их повышенной читабельности, упрощенном синтаксисе» (Куниловская 2015, 4).

Для рассмотрения данных переводческих универсалий на конкретных примерах мы обращаемся к диссертационной работе Е. С. Краснопеевой, которая провела корпусное исследование, сравнивая корпус переводных текстов с английского на русский язык с корпусом непереводных текстов на русском языке. В своей работе автор изучает особенности русскоязычного переводного дискурса[4] на основе выявления переводческих универсалий в переводах на русский язык с применением методов корпусного переводоведения.

Краснопеева также выделяет симплификацию и высказывает гипотезу своего исследования, опираясь на определение С. Лавиосы (1997), что в переводных текстах

[4] Н. К. Гарбовский считает, что комплексность переводных текстов является частью общего дискурсивного пространства, которую можно определить как переводной дискурс (Гарбовский 2011, 3–4).

(независимо от языка оригинала) на английский язык «а) показатель лексического разнообразия ниже, б) количество словоупотреблений, выраженных знаменательными частями речи меньше, чем количество словоупотреблений, выраженных служебными частями речи, в) предложения короче» (цит. по: Краснопеева 2016, 16).

Куниловская определяет, что «эксплиситация – «разъяснительные» тенденции в переводе, призванные сократить возможные риски непонимания. Типичными лингвистическими индикаторами являются структуры с опциональными элементами (напр., that после глаголов, вводящих косвенную речь в английском языке, почти все виды дискурсивных маркеров и союзов)» (Куниловская 2015, 4–5).

В работе Краснопеевой отмечается, что по Ш. Блюм-Кулька (2005) экспликация также выражается в когезии, тем самым текст перевода приобретает избыточность. В своём исследовании Краснопеева высказывает гипотезу, что лексем, вводящих в текст эксплицирующие элементы в текст (напр., *который, что, а также значит, зато, причём, вернее, ведь, впрочем, наоборот, будто, кстати, например, потому, собственно, якобы, вдобавок, следовательно, поэтому, притом* и др.), будет больше в тексте перевода (Краснопеева 2016, 16).

По мнению Куниловской, «нормализация и стандартизация – сходные тенденции к однообразности/стереотипности языка переводов, выявляемые по отношению к сравнимым непереводным текстам или текстам оригинала». Автор статьи считает, что, например степень вариативности фразеологических сочетаний и частотность низкочастотных единиц в оригинальных текстах обычно выше, чем в текстах перевода, поэтому лингвистическая разница между оригинальными текстами также выше, чем между переводными текстами (Куниловская 2015, 5).

Объясняя нормализацию, Краснопеева ссылается на М. Бейкер (1993), которая считает, что в текстах перевода будут преобладать «традиционные, характерные, подчёркнуто типичные для языка перевода конструкции». Например, Краснопеева предполагает, что глагол *to say*, встречающийся в оригинальном тексте, не всегда будет переводиться глаголом *сказать* (Краснопеева 2016, 16–17).

Интерференция и перенос, по мнению Куниловской – «тенденция сохранять в переводе черты оригинала» (Куниловская 2015, 5). Краснопеева пишет, что по Г. Тури (1995) интерференция – это перенос в текст перевода феноменов, которые определяют организацию оригинального текста (там же, 16).

Из отечественных учёных В. В. Алимов очень подробно исследовал интерференцию в переводческом аспекте на материале разных языков. Автор определяет лингвистическую интерференцию как «вмешательство элементов одной языковой системы в другую, которое может быть как конструктивным, так и деструктивным» (Алимов 2005, 35).

Исследователь классифицирует интерференцию на звуковую (фонетическая, фонологическая и звуковая-репродукционная); орфографическую; грамматическую (морфологическая, синтаксическая, пунктуационная); лексическую; семантическую; стилистическую и внутриязыковую. Автор детально описывает каждую из них, подтверждая их примерами из русского, английского и французского языков (там же, 63–151).

В своей работе Алимов уделяет особое внимание проблемам, обусловленным интерференцией, исследованию языковой интерференции, а также её проявлению, преодолению и использованию «в профессионально ориентированной межкультурной коммуникации и в профессионально ориентированном переводе» (Алимов 2005).

Последняя из переводческих универсалий, которую мы хотели бы рассмотреть – гипотеза уникальных единиц. Куниловская утверждает, что это «сравнительно невысокая частотность в переводах единиц, типичных для языка перевода и не свойственных исходному языку (ex. распределение и частотность приставочных глаголов в русских переводах, соотношение видов глагола, для переводного английского – частотность предложений со сложным дополнением)» (Куниловская 2015, 5).

Как отмечается в работе Краснопеевой, по определению С. Тиркконен-Кондит (2004), специфичные для языка перевода единицы будут преобладать в оригинальных текстах в сравнении с переводными (Краснопеева 2016, 16–17).

Итак, в данном подразделе рассмотрена только часть переводческих универсалий, но, по нашему мнению, наиболее важных, так как наша работа основана на применении методов корпусного переводоведения. Также выяснилось, что применение корпусного подхода позволяет наиболее точно изучить особенности переводного дискурса.

В данном разделе нам удалось выделить и разъяснить основные понятия, необходимые для понимания содержания и цели нашего исследования. Далее читатель познакомится с инструментами, которые были применены для данного исследования, и которые непосредственно связаны с развитием современных технологий.

2.2. Лингвистическое исследование с помощью сопоставимых корпусов

Сейчас трудно себе представить повседневную жизнь без сотового телефона или компьютера, хотя ещё четверть века назад дело обстояло абсолютно иначе. Тогда далеко не у всех были сотовые телефоны или компьютеры, к тому же связь была достаточно дорогой, а о безлимитном пользовании Интернетом даже не могли и подумать. В наши же дни практически каждый человек сталкивается с необходимостью ежедневного использования различных цифровых устройств.

Инструментом данного исследования является корпус, создание и использование которого стало возможным благодаря компьютерным технологиям. В разделах данной главы рассматриваются основные термины, относящиеся к корпусной лингвистике. Корпусная лингвистика началась в англистике, поэтому в нашей работе мы знакомим читателя с терминами как русском, так и на английском языке.

2.2.1. Появление корпуса и развитие корпусной лингвистики

Ещё до появления компьютерных технологий, по словам многих лингвистов (А. Д. Майоровой, В. П. Захарова, М. Н. Михайлова и др.), проводились лингвистические исследования на основе корпусного материала, который собирался вручную и выглядел в виде карточек.

Ссылаясь на работу У. Фрэнсиса (Francis), М. Н. Михайлов пишет, что до XIX одной из главных задач науки о языке было составление словарей и грамматик, для составления которых использовались корпуса текстов. Ещё задолго до появления компьютерной техники лексикографы собирали картотеки примеров, с помощью которых возможно было детально и точно описать с разных сторон изучаемое языковое явление, а также найти «оригинальные структурные или семантические особенности языковых единиц». Например, за период с 1858 по 1933 гг. для «Оксфордского толкового словаря английского языка» была собрана картотека, которая насчитывает 11 миллионов примеров (Francis 1992, цит. по: Михайлов 2003, 12).

Становится понятно, что эта работа отнимала очень много различных ресурсов. Как утверждает А. Д. Майорова, исследования зачастую длились на протяжении нескольких лет, а сбор исследовательского материала считали важной частью научной работы. Всё изменилось с появлением вычислительной техники, сейчас лингвисты обрабатывают

материал с помощью специальных программ, также появилась научное направление, изучающее корпусы. С появлением языковых корпусов были расширены возможности изучения языка (Майорова 2017, 42).

По мнению Майоровой, корпусная лингвистика, инструментом которой являются корпусы текстов, также зародилась с появлением вычислительной техники и программного обеспечения. Невозможно сомневаться в важности корпусной лингвистики, так как в ней объединены гуманитарная и техническая науки. Так как она появилась сравнительно недавно, то корпусная лингвистика не имеет статуса самостоятельной науки (там же, 42).

С. А. Шаров пишет, что корпусная лингвистика объединяет многие положительные свойства гуманитарных и технических наук, поэтому её развитие имеет важное значение. Как и в любом научном исследовании перед исследователем в области корпусной лингвистики ставится две задачи: создание рабочих инструментов и использование этих инструментов для изучения лингвистических явлений. В создание рабочих инструментов для анализа входит построение моно- и многоязыковых корпусов текстов интересующей исследователя области, а также способы разметки корпусов, поиск и обработка результатов поиска (Шаров 2003, 9).

В. П. Захаров и С. Ю. Богданова в учебнике «Корпусная лингвистика» определяют, что корпусная лингвистика окончательно сформировалась как новое лингвистическое направление в первой половине 1990-х годов. Авторы описывают направления, которые предвосхитили появление корпусной лингвистики и называют технологии, которые повлияли на создание корпусов. Они также выделяют три основные области лингвистических исследований, которые вошли в основу корпусной лингвистики. Корпусное исследование получило широкое применение в исторической лингвистике (например, изучение изменений в языке и воссоздание праязыков); написании грамматик, составлении словарей, обучении, а также в социолингвистике (например, изучение языкового многообразия, диалектов и т. д.) (Захаров и Богданова 2020, 16–20).

Как считает А. А. Шунейко, на данный момент корпусная лингвистика занимает одно из ведущих мест среди лингвистических дисциплин. Термин *корпусная лингвистика* (corpus linguistics) претерпевал различные изменения. Первыми начали употреблять этот термин англисты Ян Свартвик (Svartvik) и Джефри Лич (Leech), до них эта область развивалась под другими названиями: статистическая лингвистика, математическая

лингвистика, компьютерная и квантитативная. Шунейко объясняет причины появления различных терминов: «постоянная смена актуальных научных парадигм, моды и то, что принято называть научно-техническим прогрессом» (Шунейко 2020, 9–10).

В своём учебном пособии Шунейко дает следующее определение термину *корпусная лингвистика*: «корпусная лингвистика – комплексная научная дисциплина, которая использует компьютерные технологии для выяснения природы языка, оптимизации коммуникации и решения иных лингвистических задач, а параллельно с этим создает объекты, помогающие решать эти задачи – корпуса данных» (там же, 11).

С появлением компьютерного корпуса, по мнению Шарова, деятельность лингвиста не сильно изменилась. Например, «Словарь английского языка» Сэмюэля Джонсона (Johnson) появился в середине 18-го века, «Грамматика» Отто Есперсена (Jespersen) – в первой половине 20-го века, также академические грамматики и словари русского языка создавались на основе анализа реальных примеров словоупотребления, которые были изначала записаны на карточках. Но как только лингвисты получили возможность использовать компьютеры для хранения и обработки больших объемов информации, появились первые проекты по созданию коллекций текстов и корпусов (Шаров 2003, 12).

Майорова отмечает, что «на данный момент использование корпусов играет ведущую роль при проведении большинства лингвистических исследований». Благодаря корпусам «структурировать и синхронизировать материал стало гораздо легче и быстрее, а затраты резко снизились» (Майорова 2017, 42).

Майорова упоминает, что первый, Брауновский корпус (Brown Corpus), был создан уже в 1963 году сотрудниками Брауновского университета (штат Рот-Айленд, США), У. Френсисом (Francis) и Г. Кучерой (Cuchera). Объём первого лингвистического корпуса составляет 1 млн словоупотреблений из текстов разных жанров, изданных в 1961 году в США. Создание этого корпуса стало прорывом в прикладной лингвистике (там же, 43).

В учебном пособии Захарова и Богдановой отмечается европейский корпусный проект, который получил своё имя по названиям британского и двух норвежских университетов, Ланкастер-Осло-Берген (The Lancaster-Oslo-Bergen corpus или кратко LOB). В него вошли 500 текстов 15 жанров по 2 000 слов, т. е. 1 млн слов британского варианта английского языка (Захаров и Богданова 2020, 25).

Ссылаясь на работу МакЭнери (McEnery) и Уилсона (Wilson), Михайлов отмечает ещё об одном корпусе, появившимся в эру корпусной лингвистики. В конце 1960-х годов

появился корпус Роберто Буса (Roberto Busa), объёмом 10 600 000 словоупотреблений, в который вошли тексты св. Фомы Аквинского и других средневековых философов. Эта работа зародила новою традицию использования электронных корпусов текстов для исследований в области гуманитарных наук (McEnery & Wilson 2001, цит. по: Михайлов 2003, 13).

Среди современных корпусов Захаров и Богданова выделяют Британский национальный корпус (British National Corpus – BNC), Международный корпус английского языка (International Corpus of English – ICE), Корпус современного американского английского (Corpus of Contemporary American English – COCA, Национальный корпус русского языка (НКРЯ) и др. В настоящее время существует множество корпусов на разных языках мира (Захаров и Богданова 2020, 25–26).

Шаров описывает тенденции, которые повлияли на зарождение корпусной лингвистики. Он также упоминает, что большинство исследований в области корпусной лингвистики начиналось на материале английского языка. Помимо активного развития компьютерной техники в США, в научной среде главенствовал рационалистический подход, который был основан на лингвистической интуиции, не требующей наличия корпусных данных. Считалось, как пишет Шаров, что «провести различие между правильными и неправильными конструкциями можно с помощью любого носителя изучаемого языка». В противовес такому подходу появляется «эмпирический подход, который предлагает рассматривать язык как ресурс, обеспечивающий набор возможности для коммуникации». Как правило, этот набор реализуется в дискурсе, соответственно объектом исследования в лингвистике являются слова и конструкции, употребленные в тексте (Шаров 2003, 16).

В отличие от США, в британской лингвистике более популярными были эмпирические тенденции, результатами которых были исследования Джона Фёрса (Firth) и его последователей Грегори (Gregory), Синклера (Sinclair), Хэллидея (Halliday) и др. В русской лингвистике также присутствовал эмпирический подход, так как лингвистический анализ практически всегда сопровождался примерами реального словоупотребления (Шаров 2003, 16). Майорова утверждает, что именно развитие эмпирического подхода послужило началом создания корпусной лингвистики (Майорова 2017, 42).

Итак, мы выяснили, что изучает корпусная лингвистика, и какое место она занимает в лингвистике. Теперь остаётся выяснить, что представляет собой корпус.

2.2.2. Понятие корпус и классификация корпусов

М. Н. Михайлов отмечает, что 1992 году Уильям Фрэнсис (Francis), один из первых исследователей современной корпусной лингвистики, определяет корпус текстов как «собрание текстов, считающееся репрезентативным по отношению к данному языку, диалекту или иной части языка и предназначенное для использования в лингвистических исследованиях» (Francis 1992, цит. по: Михайлов 2003).

В настоящее время у понятия корпус существует множество различных определений. Например, Захаров и Богданова определяют корпус текстов как «большой, представленный в машиночитаемом формате, унифицированный, структурированный, размеченный, филологически компетентный массив языковых знаний, предназначенный для решения конкретных лингвистических задач» (Захаров и Богданова 2020, 11).

Авторы подчёркивают прагматичность создания и использования корпусов, например, получение типичности данных и широкого спектра языковых явлений достигается с помощью большого и сбалансированного объёма корпуса; для более детального и объективного изучения данных мы можем исследовать также их контекст, представленный в корпусе и, наконец, корпусами можно пользоваться неоднократно в разных исследованиях (там же, 12).

По мнению авторов, собрание текстов в корпусе должно быть «конечного фиксированного размера». Если меняется состав и объём корпуса, то это не должно повлиять на его структуру, либо изменения должны быть обоснованными. В этом заключается представительность корпуса, репрезентативность. Второе важное условие для создания продуктивного корпуса – это его прагматическая направленность, т. е. корпус создается для решения определённой задачи (там же, 21–22).

В определении корпуса говорится об унификации и разметке, которые являются объектами стандартизации. Это необходимо для использования корпусов более широким кругом пользователей. Захаров и Богданова подчёркивают о сложности стандартизации транскрибирования устной речи и исторических корпусов. На данный момент существуют международные стандарты аннотирования текстов корпуса, например, Text Encoding Initiative (TEI), Corpus Encoding Standard (CES), International

Standards for Language Engineering (ISLE) и др., которые обобщены международной организацией по стандартизации ISO/TS 37 (Захаров и Богданова 2020, 26–27).

Корпусы могут быть размеченными (аннотированными) и неразмеченными. Для лингвистических исследований более полезными являются размеченные корпусы, при создании которых текстам приписываются специальные метки (англ. *tag*). Метки могут содержать полезную для исследователя информацию, например сведения об авторе, название текста, год публикации, жанр и др., а также лексические, грамматические и прочие характеристики элементов текста. С помощью метаразметки исследователь может выявить взаимосвязь языка и условий его существования, а также отобрать и исследовать отдельные языковые явления. Чаще всего для разметки текстов используется формат XML (там же, 34–55).

При анализе материала выяснилось, что существует множество вариантов классификации корпусов, далее будет рассмотрена классификация корпусов, предложенная Захаровым и Богдановой. Как отмечают авторы, корпусы делятся на разные классы по двум основаниям: «1) противопоставление корпусов (корпусы относятся к определённому жанру, стилю, языку определённой профессиональной деятельности и т. д., 2) разделение корпусов по типу лингвистической разметки (морфологической, синтаксической и др.)» (там же, 56).

Как упоминалось выше, корпусы создаются в зависимости от поставленной цели исследования, поэтому их разделяют на: «а) многоцелевые корпусы, которые содержат тексты различных жанров, например национальные корпусы; б) специализированные, ограниченные одним жанром или группой жанров». Например, терминологический корпус служит для составления терминологического словаря. При разделении корпусов по жанрам выделяют литературные, фольклорные, драматургические и другие корпусы (там же, 56–58).

Захаров и Богданова отмечают о разделении корпусов также по назначению. Например, с помощью исследовательских корпусов можно изучать различные аспекты функционирования языка, т. е. решать широкий класс лингвистических задач, а иллюстративные корпусы помогают подтвердить и обосновать уже полученные в ходе исследования результаты. По динамичности авторы выделяют «а) динамические корпусы (постоянно пополняющиеся), которые помогают выявить изменения языковых явлений, зачастую их используют для составления словарей, и б) статистические

(коллекции текстов писателей), которые отображают состояние языковой системы в определённый отрезок времени». Существуют и другие критерии деления корпусов, например, разметка, доступность, объём, параллельность, тип языковых данных и назначение (Захаров и Богданова 2020, 58–60).

Авторы выделяют два типа корпусов по параллельности: параллельные корпусы (translation corpora), которые состоят из оригинальных текстов и их переводов, и сопоставимые корпусы (comparable corpora), тексты которых объединены по какому-либо признаку, например, по тематике, диалекту, жанру и т. д. Параллельные и сопоставимые корпусы широко используются в технологиях машинного перевода, для составления терминологических словарей, в различных областях лингвистики для сравнительных исследований языков и т. д. (Захаров и Богданова 2020, 61).

По типу языковых данных Захаров и Богданова выделяют письменные (только письменные тексты), устные (только устная речь) и смешанные корпусы (например, НКРЯ и BNC, в которых язык представлен в разных формах в определённый период времени) (там же, 64).

Существуют также и так называемые учебные корпусы текстов (learner corpora – LC), которые используются для изучения языков. Авторы говорят о широком использовании учебных корпусов для выявления и анализа распространённых ошибок в изучаемых языках (там же, 67).

Ранее констатировался факт, что существуют и другие варианты классификации корпусов. Это связано с разницей взглядов на критерии, по которым происходит классификация. Например, С. В. Манукянц, сопоставив и проанализировав научную литературу, синтезировал единую классификацию, в которой выделяются мономодальные и мультимодальные, полнотекстовой и фрагментированный корпусы и др. (Манукянц 2021, 70–71).

Как утверждают Захаров и Богданова, основными пользователями корпусов являются лингвисты, их больше интересует «метатекстовая информация и примеры тех или иных языковых элементов», чем само содержание текста. Изначально корпусные исследования сводились к использованию подсчёта частот различных языковых элементов, которые необходимы для составления словарей и грамматик, проверки орфографии и грамматики, а также в машинном переводе. Например, с помощью таких статистических данных можно выделить, какие слова систематически встречаются

вместе и в результате могут быть отнесены к разряду устойчивых словосочетаний (Захаров и Богданова 2020, 132).

Авторы отмечают, что корпусы очень значимы в исследованиях в области лексикологии и семантики, т. к. помогают установить определённые признаки языковой единицы, характеризующие её семантику. Для лингвистов-теоретиков корпусы служат «экспериментальной базой для проверки гипотез и доказательства своих теорий» (там же, 132–133).

О различных корпусных исследованиях переводных текстов на финский язык рассказано в сборнике статей «Käännössuomeksi» под редакцией Мауранен и Янтунен. В него вошли статьи А. Мауранен (Mauranen) и Л. Тииттулы (Tiittula), М. Н. Михайлова, С. Невалайнена (Nevalainen), Х. Кемппанена (Kempanen), С. Эсколы (Eskola) и др. (Mauranen & Jantunen 2005).

Корпусы нашли широкое применение у прикладных лингвистов (преподаватели, переводчики) и компьютерных лингвистов, которые используют корпусы для создания компьютерных моделей языка, а также у других специалистов по языку (литературоведы, редакторы и др.). Для анализа текстов по жанру, периоду времени или по автору корпусы могут быть полезны специалистам по общественным наукам (историкам, социологам). Конечно же, корпусы просто незаменимы в компьютерной лингвистике и информационных технологиях, в которые корпусы служат для создания и усовершенствования различных автоматизированных систем, например, машинного перевода, распознавания речи и информационный поиск (Захаров и Богданова 2020, 132–133).

Захаров и Богданова подчёркивают значимость корпусов в работе исследователей в том, что корпусы «дают специалистам богатый репрезентативный эмпирический материал». Эмпирическая поддержка обеспечивается репрезентативностью и сбалансированностью языкового материала корпуса, которая подтверждается статистической информацией (частотность слов, фраз и конструкций). Предоставляющая корпусами метаинформация (возраст или пол говорящего, жанр текста и время его написания) позволяет исследовать разные языковые явления более всесторонне. Авторы отмечают, что с помощью корпусов были опровергнуты многие теории, считавшиеся верными длительное время (там же, 133–136).

2.2.3. Применение сопоставительных корпусов в лингвистических исследованиях

В предыдущем подразделе упоминалось о параллельных и сопоставительных корпусах и об их отличиях друг от друга. Процесс создания параллельных корпусов более трудоёмкий, т. к. включает в себя выравнивание текста (aligment) по предложениям, клаузам (грамматическим конструкциям), словосочетаниям или словам. Создание сопоставительных корпусов более прост.

Захаров и Богданова отмечают, что для исследования сопоставительных корпусов могут применяться инструменты автоматической классификации лексики, которая осуществляется программой с помощью вычисления «понятий и количества семантических связей между словами (леммами) и словоформами». Сравнительный анализ помогает выявить распределение языковых единиц в разных текстах. Например, корпусы семейства Aranea, которые совместно созданы Братиславским университетом Коменского и Словацкой академией наук, состоит из сопоставительных веб-корпусов для 22 языков (Захаров и Богданова 2020, 63–94).

Авторы подчёркивают «единство технологических решений при создании корпусов, приблизительную одновременность их создания, унификацию результатов разметки для разных языков». Корпусы получили своё название в зависимости от представленного в них языка, например, Araneum Russicum, Araneum Anglicum и т. д., которые загружены в корпусный менеджер NoSketch Engine (там же, 94).

Захаров и Богданова пишут, что наравне с параллельными корпусами в переводной лексикографии используются сопоставительные корпусы. С помощью выявления однословных терминологических единиц и сопоставления их с вариантами переводных эквивалентов происходит выравнивание по терминам. Для дальнейшего сопоставительного анализа терминов потребуется использование переводных словарей, которые помогут подтвердить эквивалентность терминов. Также при составлении терминологических баз данных машинного перевода принято опираться на распространенность лексической единицы в конкретном языке для специальных целей (там же, 187–188).

Как уже упоминалось в предыдущем подразделе, корпусы предназначены для решения конкретных лингвистических задач, т. е. для выполнения поставленной задачи необходимо чётко определить объект и цель исследования. Возможность использования

сопоставительных корпусов повлияло на успешные результаты многих лингвистических исследований. Для сопоставления или сравнения исследуемого материала современный лингвист может создавать свои сопоставительные корпусы или работать с уже готовыми корпусами. Существует множество примеров корпусных исследований, результаты которых повлияли на реорганизацию каких-либо процессов в других областях науки. В учебном пособии Захарова и Богдановой рассматриваются лексикографические, грамматические и исследования дискурса, основанные на корпусах (Захаров и Богданова 2020, 137–178).

Например, О. Н. Камшилова рассказывает о корпусном исследовании речи русских школьников на английском языке с помощью сопоставления Подкорпуса диалогов школьников (Learner Dialogue Corpus) и Корпуса диалогов-образцов, созданного из учебных пособий по английскому языку. В ходе работы был сделан «количественный анализ на основании сравнений числа словоупотреблений (токенов) и типов словоформ (типов токенов) грамматических классов слов в диалогах школьников и диалогах-образцах». Исследователи выявили степень связи, корреляции сопоставляемых диалогов, с помощью коэффициента Пирсона[5], а также распределение грамматических классов слов по сопоставляемым корпусам. В результате исследования были выяснены причины выбора употребления определённых языковых единиц учащихся, а также получено представление о процессе усвоения чужого языка, что помогает понять, какие изменения необходимо внести в дидактический материал и как эффективнее его организовать (Камшилова 2017, 32–39).

К примеру, И. Н. Болушевская в результате корпусного исследования изучила и описала некоторые лингвистические особенности дискурса криминальных интернет-новостей. Для исследования был составлен «корпус текстов интернет-новостей 1990–2018 гг. из американской газеты The New York Times и британской газеты The Guardian, содержащий 99 248 слов». Для сравнения исследуемого корпуса использовались упомянутые нами ранее Корпус американского английского (Brown Corpus) и Корпус британского английского (LOB), размер каждого составляет 1 миллион слов.

В результате работы Болушевской были выявлены ключевые слова, также выяснилось, что они по-разному употребляются в сопоставлявшихся корпусах. Затем исследователь

---

[5] «Коэффициент корреляции Пирсона характеризует существование линейной зависимости между двумя величинами» (MachineLearning.ru 2022).

распределила ключевые слова по тематическим группам, которые отражали криминальную тематику новостей. По частотным спискам было проанализировано наличие негативной лексики, характерное для данной темы, а также «параметры эвиденциальности (указание на источник информации, которую говорящий излагает в результате личного опыта или, цитируя сообщение другого лица)». По мнению автора, эти лингвистические особенности способствуют формированию у читателя определённой интерпретации новостей. С помощью корпусного исследования Болушевской удалось выделить основные инструменты воздействия на читателя и привлечения его внимания к важной, по мнению автора статьи, информации (Болушевская 2019, 184–188).

Итак, выяснилось, что с помощью корпусов, в том числе сопоставительных, можно выявить частоту языковых единиц и распределение их по текстам, даже по разным регистрам. Также с помощью конкорданса корпуса у нас есть возможность исследовать значения слов и выявить их коллокаты (collocates), т. е. слова, с которыми исследуемое слово часто встречается. Под конкордансом в данном контексте подразумевается «список всех употреблений данного слова в контексте со ссылками на источник» (Захаров и Богданова 2020, 12).

Захаров и Богданова в своём учебном пособии отмечают, что с помощью корпусов можно анализировать также удалённые коллокаты и использование слов, которые кажутся синонимами, но на самом деле таковыми не являются. Авторы описывают самые разные исследования на основе корпусов. Например, к ним относятся исследования морфологических характеристик слов и номинализаций (образование абстрактного существительного от глагола, напр., сущ. *civilization* от глаг. *civilize*), их распределение по регистрам; функций номинализаций и функции суффиксов номинализаций; исследования распределения грамматических категорий, а также исследования дискурса (там же, 137–179).

2.2.4. Система Sketch Engine для работы с корпусами

Ранее уже говорилось о том, что работы с корпусами необходимо иметь определённые навыки. Для создания собственных корпусов или использования уже готовых необходимо уметь пользоваться программными средствами и информационными ресурсами корпусной лингвистики. Неотъемлемой частью корпуса является поисковая система, корпусный менеджер (corpus manager), только при грамотном его

использовании можно получить необходимую информацию. Существует много программ для работы с корпусами, особенный интерес представляют программы автоматической разметки. Например, Sketch Engine является очень популярной программой для работы с корпусами текстов.

В помощь исследователям выходит много учебных пособий с инструкциями по работе с корпусами. В учебных заведениях проводятся курсы по ознакомлению с корпусами и корпусной лингвистикой, а также с программами для использования корпусов. Например, М. А. Куниловская и М. А. Ковязина отмечают появление в 2016 году второго издания учебного пособия Джеймса Томаса (Thomas), «Изучаем английский язык со Sketch Engine: корпусный подход» (Discovering English with Sketch Engine: A Corpus-Based Approach), а также первое издание рабочей тетради «Изучаем английский язык со Sketch Engine. Рабочая тетрадь и глоссарий» (Discovering English with Sketch Engine. Workbook and Glossary). Учебник прежде всего ориентирован на преподавателей английского языка, но также адресован студентам, лингвистам и переводчикам (Куниловская и Ковязина 2018, 332–333).

Дж. Томас пишет, что проект системы Sketch Engine был начат британским лексикографом, корпусным лингвистом А. Килгарриффом (Kilgarriff), в 2004. Основные функции данной системы разрабатывались Килгарриффом в сотрудничестве с П. Рыхлым (Rychly), сотрудником Лаборатории обработки естественного языка (Natural Language Processing Laboratory) факультета информатики Университета им. Масарика в г. Брно, Чехия (Thomas 2016, 13).

На ресурсе Sketch Engine предоставляется доступ примерно к 500 корпусам для более чем 90 языков. К примеру, для английского языка Sketch Engine предлагает 70 корпусов, один из них – English Web 2015 (enTenTen15), содержащий 15 млрд слов, тексты которого собраны из публикаций в Интернете. Для немецкого языка представлено 17 корпусов, из них самый большой веб-корпус German Web 2013 (deTenTen13), который содержит 16,5 млрд словоформ (Sketch Engine 2022). В некоторых лингвистических исследованиях именно размер корпуса играет важную роль, например при изучении идиом (Кротова 2018, 107).

Кротова отмечает, что ряд исследований основывается на так называемых сбалансированных корпусах, содержащих «по возможности все типы письменных и устных текстов данного языка (художественные разных жанров, публицистические,

учебные, научные, деловые, разговорные, диалектные и т. п.), и которые входят в корпус по возможности пропорционально их доле в языке соответствующего периода». К таким корпусам относятся Британский национальный корпус (BNC), содержащий 100 млн слов или DWDS-Kerncorpus, в состав которого входит 120 млн словоформ. Однако большинство корпусов являются несбалансированными. Как правило, это корпусы, содержащие определенный тип текстов и/или тексты, созданные для исследования конкретного языкового явления. Примером таких корпусов в Sketch Engine являются: Корпус медицинских текстов (Medical Web Corpus), Корпус предлогов английского языка (English Preposition Corpus) и др. (Кротова 2018, 108).

Кротова считает, что немаловажное значение имеет состав корпуса и имеющаяся в нём разметка при исследовании синтаксиса. Примерами таких корпусов является Синтаксический корпус НКРЯ, BNC и др. Например, при работе с текстами BNC через Sketch Engine можно указать типы текстов, дату и место публикации, имя автора и т. п. (Кротова 2018, 108).

Захаров и Богданова подтверждают данные Кротовой о составе корпусов в Sketch Engine и выделяют корпус русского языка RuTenTen, объем которого составляет 14,5 млрд слов и который создан из текстов Интернета по технологии WaC (поиск текстов из веба). По мнению авторов, программа Sketch Engine очень хорошо подходит для решения задачи по извлечению терминологии, так как с её помощью можно «получить данные о сочетаемости слов; создать списки семантически связанных лексем, а также извлечь ключевые слова и термины». По мнению авторов, Sketch Engine является на данный момент самой мощной системой в области корпусной лингвистики (Захаров и Богданова 2020, 78–93).

Корпусный менеджер Sketch Engine очень универсален. Захаров и Богданова определяют его универсальность по степени распространенности и по функциям создания использования корпусов. При загрузке текстов система формирует корпус, проводя морфологическую разметку и преобразовывая файлы на входе, например путем разархивации (там же, 111–113).

Ещё ранее, в 2012 году, Захаров и Хохлова говорили об уникальности системы: «В системе Sketch Engine имеются специальные инструменты, позволяющие измерять силу не только синтагматических, но и парадигматических связей на основе дистрибуции

лексем в корпусе: тезаурус (thesaurus), кластеризация (clustering) и дифференциация (differences)» (Захаров и Хохлова 2012, 3).

В учебном пособии Захарова и Богдановой подробно описаны возможности программы, например поиск по корпусу в системе Sketch Engine осуществляется как с помощью простого поиска (по конкретной словоформе или по точной фразе), поиска по лемме, с помощью шаблонов (wild cards), так и с помощью специального языка (CQL – Corpus Query Language) для более сложных запросов. К большому количеству разных корпусов и языков применяется один и тот же язык запросов, что помогает пользователю экономить время, которое он мог бы потратить на изучение нового языка запросов и интерфейса. Ещё одна из полезных для исследования функций Sketch Engine, по мнению авторов, является возможность создания скетчей (word sketches). Скетчем в данном случае является описание речевого поведения слова, полученное автоматическим путем обобщения информации обо всех контекстах, в которых исследуемое слово встретилось в корпусе (Захаров и Богданова 2020, 127–128).

Как мы уже упоминали ранее, в программе существует возможность использования готовых корпусов и создания своих корпусов, а также подкорпусов. Авторы отмечают, что программа Sketch Engine предоставляет возможность получения различных данных внешней и внутренней статистики. Как правило, это информация о корпусах или подкорпусах, частотные списки слов или n-грамм, количество токенов и слов, входящих в корпус, кластеризация лексических единиц, выявление коллокаций, ключевых слов и т. д. (там же, 127–128).

Авторы утверждают, что Sketch Engine относится к инструментам четвертого поколения, которые предлагают «лучшую масштабируемость за счет хранения корпуса в базе данных веб-сервера и предварительной индексации данных для обеспечения быстрого поиска» (там же, 113).

Однако, авторы напоминают, что несмотря на свою универсальность, инструменты четвёртого поколения имеют некоторые проблемы. Например, платный или ограниченный доступ, размытость границы между данными и инструментом. Так как данные на внешнем сервере хранятся в индексированной форме, то у пользователя нет возможности сразу обратиться к исходным данным и соотнести их с результатами поиска. Но несмотря на вышеперечисленные ограничения, по результатам опроса лингвистов о частоте использования компьютерных средств для научных исследований

(Tribble, 2008), Sketch Engine занимает одно из ведущих мест (Захаров и Богданова 2020, 113–115).

В данном разделе было рассмотрено, что такое корпус, предметом изучения какой науки он является, типы корпусов и инструменты для их использования. С нашей точки зрения, универсальность корпусов состоит ещё и в том, что мы можем исследовать интересующие нас темы, а именно абсолютно любые языковые явления в различных областях лингвистики, также работа с корпусами интересна и эффективна. К тому же доступ к большому количеству корпусов абсолютно свободный. До знакомства с корпусами мы даже не подозреваем о их существовании, узнав о них, мы не перестаём удивляться возможностям их применения.

3. КОРПУСНОЕ ИССЛЕДОВАНИЕ ЯЗЫКОВЫХ ОСОБЕННОСТЕЙ ТЕКСТОВ О COVID-19

Как уже упоминалось ранее, целью данного исследования является анализ и сравнение языковых особенностей оригинальных и переводных текстов с помощью сопоставимых корпусов. Также мы попытаемся выяснить, выполняют ли тексты свои функции, в первую очередь – информационную и оперативную. Так как материалом данного исследования являются тексты на русском и финском языках, а также переводы на эти языки, то нам будет интересно сравнить тексты разных корпусов между собой. Например, тексты на русском языке с переводными текстами на русский язык, тексты на финском языке с переводными текстами на финский язык, а также тексты на языках оригинала (русского и финского) и тексты переводов на русский и финский языки.

3.1. Материал и методы исследования

Как пишут М. В. Копотев и А. Мустайоки, «учёный должен определить, на чём основываются его суждения об изучаемом объекте». Авторы отмечают, что в научной практике употребляется слово *evidence* (подтверждение) (Копотев и Мустайоки 2008, 8). В качестве доказательной базы нашего исследования используются результаты квантитативного анализа, т. е. статистические данные корпусного исследования. Для осуществления анализа используется программа Sketch Engine. Основной задачей на данном этапе работы было создать репрезентативные и сбалансированные корпусы, поэтому мы собирали схожие по тематике и жанру тексты.

3.1.1. Отбор материала исследования

В корпусы вошли тексты о коронавирусе, размещенные в свободном доступе в Интернете. Тексты содержат инструкции, оповещение граждан и правила поведения в условиях пандемии, которые были опубликованы на официальных сайтах государственных учреждений и газет в период с марта 2020 до сентября 2021. Как в России, так и в Европе первые данные об опасности распространения нового коронавируса SARS-Cov-2, который способен вызвать респираторное заболевание COVID-19 появились в начале 2020 года. Уже в марте того же года Всемирная организация здравоохранения (ВОЗ) объявила о пандемии COVID-19. Именно тогда начала появляться информация о данном заболевании и соблюдении определенных мер

предосторожности. Таким образом, временные рамки нашего корпуса составляет примерно полтора года.

Поиск оригинальных текстов на русском и финском языках не составил трудностей, так как с начала пандемии и по настоящее время эта тема является одной из актуальных во всем мире. Например, используемая в исследовании информация о коронавирусе для граждан России была собрана с сайтов Министерства здравоохранения Российской Федерации и Федеральной службы по надзору в сфере здравоохранения, а также с других сайтов, на которых публикуется официальная информация о коронавирусе в России.

Процесс поиска переводных текстов был более трудоёмкий. Тексты для корпуса с переводными текстами на русском языке были собраны с сайтов государственных учреждений разных стран, в которых проживает русскоязычное население. В данный корпус также вошли также тексты о COVID-19 с сайта ВОЗ.

Оригинальные тексты на финском языке были собраны с сайтов Министерства социального обеспечения и здравоохранения (Sosiaali- ja terveysministeriö), Национального института здравоохранения и социального благосостояния Финляндии (Terveyden ja hyvinvoinnin laitos), Таможенной службы Финляндии (Tulli), Государственного Совета Финляндии (Suomen valtioneuvosto) и других. Все ссылки на сайты, с которых были получены тексты для исследования, перечислены в списке источников (см. Приложение).

Бóльшая часть переводных финских текстов состоит из переводов с русского, английского и испанского языков. Информация была собрана с сайтов газет для финноязычного населения, проживающего в России (в первую очередь – в республике Карелия), США и Испании.

Хотя задачей ВОЗ является предоставление международных рекомендаций в области здравоохранения на разных языках, но на сайте данной организации нет информации на финском языке. Официальными языками ВОЗ являются английский, арабский, испанский, китайский, русский и французский. Переводы текстов с английского языка на финский были собраны с сайта Еврокомиссии.

Отбор текстов был произведён вручную, что является удобным способом для контроля качества и содержания текстов. Именно достаточное количество текстов, соответствующих критериям содержания и качества, является обязательным условием

для создания репрезентативного корпуса. Сначала тексты были сохранены в виде текстовых файлов формата doc и затем были загружены на сайт Sketch Engine.

3.1.2. Описание корпусов

Для исследования были созданы две пары одноязычных сравнительно-сопоставительных корпусов[6], которые охватывают оригинальные и непараллельные переводные тексты на русском и финском языках. Для улучшения сопоставимости нами были созданы одинаковые по объёму корпусы, примерно 100 000 словоупотреблений. Для удобства работы с корпусами мы дадим им названия: 1) корпус с оригинальными текстами на русском языке – РУСО; 2) корпус с переводными текстами на русский язык – РУСП; 3) корпус с оригинальными текстами на финском языке – ФИНО и 4) корпус с переводными текстами на финский язык – ФИНП.

В РУСП вошли переводные тексты с немецкого, английского, финского языков и иврита (см. Рис. 1).

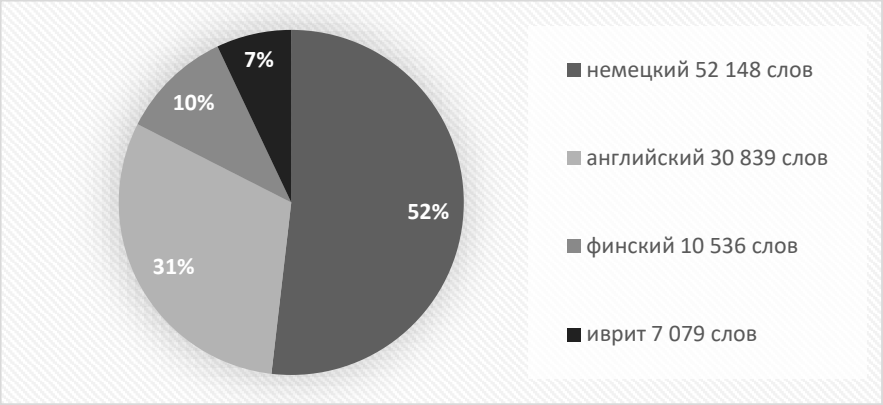


Рис. 1. Состав корпуса РУСП по языкам-источникам его текстов.

Корпус ФИНП содержит переводы текстов с испанского, английского, русского и французского языков. Состав языков, с которых переведены тексты корпуса ФИНП, также продемонстрирован в виде диаграммы на рисунке 2 (см. ниже).

---

[6] В данном случае используется термин, предложенный Богоявленской, которая разграничивает сопоставительные корпусы «по принципу количества языков, учитывая сферы их применения и функции». Например, она выделяет «одноязычные сравнительно-сопоставительные корпусы в отдельную группу, корпусы оригинальных и непараллельных переводных текстов на одном языке…» (Богоявленская 2017, 1). По нашему мнению, это определение очень хорошо подходит к составленным нами корпусам.

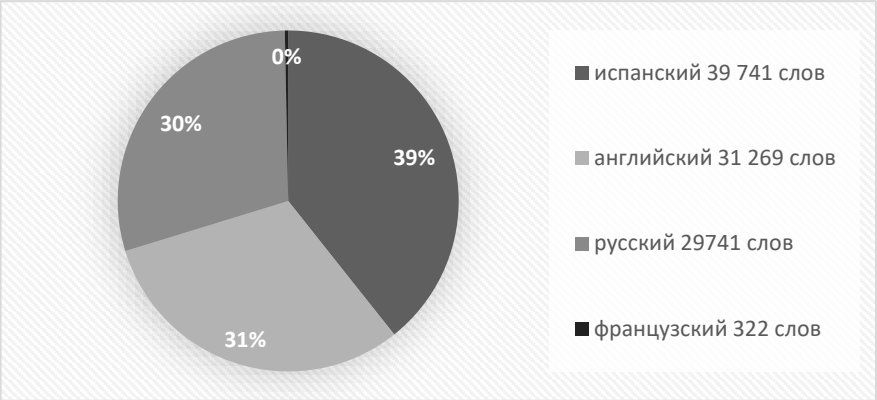


Рис. 2. Состав корпуса ФИНП по языкам-источникам его текстов.

Корпус РУСО содержит 10 файлов с текстами, в которых 100 007 словоупотреблений и 4 200 предложений. В 21 файл корпуса РУСП входят 100 187 словоупотреблений и 4 001 предложений. Корпус ФИНО состоит из 51 файла, в текстах которых 100 089 словоупотреблений и 6 953 предложений. В ФИНП вошли 5 файлов с текстами, в которых 105 074 словоупотреблений и 7 341 предложений. Данные корпусов продемонстрированы также в таблице 1 (см. ниже).

Таблица 1. Состав корпусов РУСО, РУСП, ФИНО и ФИНП.

	РУСО	**РУСП**	**ФИНО**	**ФИНП**
кол-во файлов	10	21	51	5
кол-во словоупотреблений	100 007	100 187	100 089	105 074
кол-во предложений	4 200	4 001	6 953	7 341

Корпусы содержат примерно одинаковое количество словоупотреблений, но количество предложений разное – 4 200 (РУСО), 4 001 (РУСП), 6 953 (ФИНО) и 7 341 (ФИНП). Из этого можно заключить, что в текстах на русском языке предложения длиннее, чем в текстах на финском языке. Например, количество слов в предложении в корпусе РУСО примерно 23,8 слова, в РУСП – 25 слов, в ФИНО – 14,4 слова и в ФИНП – 14,3 слова. Мы посчитали количество слов в предложении, разделив количество словоупотреблений в корпусе на количество предложений.

Большую разницу в количестве слов в предложении можно объяснить тем, что в финском языке много сложных слов, при переводе которых на русский эквивалентом может быть словосочетание. Также частое использование вводных слов и конструкций влияет на длину предложения.

3.1.3. Инструменты и методы исследования

С помощью корпусных инструментов в программе Sketch Engine можно обрабатывать большие объёмы текстов и находить достаточно легко и точно необходимые для исследования данные. Например, используя Wordlist, возможно получить частотные списки слов, которые играют важную роль при анализе ЯСЦ (Захаров и Богданова 2019, 187–188). Мы использовали абсолютные (F) и относительные частоты лексических единиц. Относительная частотность, или частотность слов на миллион (ipm), – это абсолютная частота, делённая на размер корпуса и умноженная на миллион.

Утилита N-grams выделяет списки коллокаций лексических единиц, которые могут играть большую роль при исследовании лингвистических явлений, и которые могут остаться незамеченными при использовании других инструментов.

KeyWords знакомит исследователя с тематикой материала и привлекает его внимание к потенциально интересным для исследования языковым единицам. В данной утилите программа сравнивает частоту единиц корпуса с их частотой в референтном корпусе. В нашем случае, в корпусах на русском языке, Sketch Engine использует по умолчанию корпус Russian Web 2011 (ruTenTen11), объём которого составляет более 14,5 млрд слов. В корпусах на финском языке программа соотносит данные с корпусом Finnish Web 2014 (fiTenTen14) объёмом 1,4 млрд слов. Корпусы составлены на базе веб-текстов. При сравнении частот программа присваивает наиболее частотным ключевым словам очки (keyness score), которые вычисляются по следующей формуле:

$$\frac{fpm_{rmfocus} + N}{fpm_{rmref} + N}$$

где $fpm_{rmfocus}$ – относительная частота слова в исследуемом корпусе, fpm_{rmref} – относительная частота слова в референтном корпусе, N – сглаживающий параметр (по умолчанию N=1). (Sketch Engine 2022, Simple match).

Утилиту Thesaurus также можно использовать для извлечения ключевых единиц, т. к. для любого слова Thesaurus определяет схожие по дистрибуции единицы. Данные могут быть представлены в виде таблицы, данные можно также визуализировать и получить облако слов (word cloud). Чем выше коэффициент статистической меры слова, которая отражает степень семантической близости данного слова к ключевому, тем крупнее шрифт (Sketch Engine 2022, Thesaurus – synonyms, antonyms and similar words).

Concordance демонстрирует контексты, которые содержат элемент, являющийся объектом поиска. Данная утилита позволяет достаточно быстро обнаружить коллокации интересующего нас слова, исследовать контекст, определить источник текста и т. д.

В нашем исследовании мы постарались максимально использовать возможности утилит Sketch Engine для выявления количественных данных в каждом из четырёх корпусов. Как уже упоминалось ранее, в исследовании используется квантитативный метод анализа, а именно анализ ключевых слов и частотных списков.

3.2. Работа с сопоставительными корпусами в программе Sketch Engine

В данном разделе рассказывается о ходе нашей работы с корпусами в системе Sketch Engine. В подразделах 3.2.1., 3.2.2., 3.2.3. и 3.2.4. будут описаны полученные нами результаты для корпусов на русском, в подразделах 3.2.5., 3.2.6., 3.2.7. и 3.2.8. – для корпусов на финском языке. Описания данных предоставлено в одинаковой последовательности. На первом этапе корпусного исследования текстов о COVID-19 была проведена оценка частотности языковых единиц корпусов с целью выделить термины, которые являются ключевыми в данном дискурсе. Далее были выделены и проанализированы н-граммы, а также получены списки ключевых слов. Затем был произведён анализ коллокаций и кластеров слова *коронавирус* в оригинальных и переводных текстах для русского и финского языка.

При работе с разными утилитами выделяются 15 примеров для каждого корпуса. Примеры слов и словосочетаний на финском языке для удобства сравнения мы глоссировали на русский язык. Некоторые слова могут иметь несколько значений, поэтому использовались слова и выражения, соответствующие эквивалентам перевода для текстов нашего исследования.

3.2.1. Частотные списки корпусов РУСО и РУСП

С помощью утилиты Wordlist мы получили частотный список корпуса РУСО и выбрали из него 15 первых лемм. Данные представлены в таблице 2 (см. ниже).

При рассмотрении полученного частотного списка мы наблюдаем, что из знаменательных частей речи в список вошли существительные и прилагательные, но

бóльшую часть списка составляют служебные слова (предлоги *в, с, при, на, для, по, от*; отрицательная частица *не* и союзы *и или*).

Таблица 2. Частотности слов по лемме в корпусе РУСО.

Слово	F	ipm
в	3 179	23 269,43
и	2 810	20 568,45
с	1 879	13 753,78
при	1 213	8 878,84
на	1 037	7 590,56
covid-	970	7 100,14
для	961	7 034,26
пациент	915	6 697,56
медицинский	903	6 609,72
или	903	6 609,72
по	666	4 874,72
не	584	4 274,2
заболевание	527	3 857,50
инфекция	498	3 645,23
от	459	3 359,76

Для сравнения в частотный список корпуса РУСП вошло только одно существительное *covid* и два глагола *быть* и *мочь*, все остальные слова – служебные части речи (предлоги *в, на, с, по, для, от, при*; отрицательная частица *не*; союзы *и, или, также* и местоимение *который)*. Данные продемонстрированы в таблице 3 (см. ниже). Существительное *covid* входит в частотные списки обоих корпусов.

Таблица 3. Частотности слов по лемме в корпусе РУСП.

Слово	F	ipm
в	4 148	32 451,63
и	3 601	28 172,21
на	1 654	12 939,97
с	1 548	12 110,69
или	1 108	8 668,37
по	1 065	8 331,96
для	1 001	7 831,26
не	894	6 994,16
covid-	645	5 046,12
от	608	4 756,65
быть	553	4 326,36
при	551	4 310,72
который	500	3 911,72
мочь	494	3 864,78
также	452	3 536,20

Далее были выделены наиболее частотные существительные корпуса РУСО: *пациент (913), заболевание (527), инфекция (498), версия (360), исследование (316), помощь (315), организация (290), случай (289), день (267), препарат (266), лечение (258), риск (255),*

терапия (252), метод (219), пневмония (215). Здесь и далее в скобках дана абсолютная частота слова.

Из перечисленных высокочастотных существительных большая часть относится исключительно к медицинской теме. Из этого списка нас заинтересовало слово *день*, поэтому мы проверили его использование в текстах с помощью Word Sketch и Concordance. При исследовании выяснилось, что *день* используется в связке со словами, относящимся к медицинской теме. Чаще всего *день* встречается в виде оборота *раз в день*, например: *лекарство применять 2 раза в день, в день проведения вакцинации, в день обращения к врачу* и т. д.

Как и для корпуса РУСО был составлен список частотных существительных корпуса РУСП. Из наиболее частотных существительных в корпусе переводных текстов на русский язык мы выделили: *лицо (450), инфекция (445), человек (392), пациент (363), мера (355), страна (329), результат (318), время (315), случай (311), защита (289), заболевание (286), карантин (269), день (262), тестирование (252), зона (251).*

В данном списке нас заинтересовало значение слова *лицо*, поэтому мы решили проверить коллокации этого слова в Concordance. Из 450 случаев 24 раза слово *лицо* означает переднюю часть головы человека, во всех остальных случаях *лицо* означает самого человека. Было найдено 19 словосочетаний *маска для лица* и 5 словосочетаний *защита для лица.*

Сравнивая списки частотных существительных корпусов РУСО и РУСП, можно понять, что тематика текстов связана непосредственно с медициной. Эта тематика ярче проявляется в оригинальных текстах на русском языке. Если в корпусе РУСО человек выступает в качестве пациента, то в частотном списке существительных корпуса РУСП присутствуют слова *человек*, *пациент* и *лицо*. В данном корпусе человек выступает в качестве пациента всего 363 раза, а в корпусе РУСО 913 раз.

Из высокочастотных глаголов корпуса РУСО мы выделили первые 15: *быть (349), мочь (327), использовать (178), являться (168), рекомендоваться (140), вызывать (137), следовать (127), проводить (119), проводится (115), иметь (113), рекомендовать (101), представлять (74), обеспечивать (74), получать (68), находить (67).*

Глагол *быть* чаще всего выступает в связке с модальными глаголами *должен* (111) и *мочь* (97). Примеры (1)–(3) были найдены с помощью утилиты Word Sketch.

(1) Лечение должно быть начато как можно раньше, что в большей степени обеспечивает выздоровление.

(2) Вопрос о прекращении ИВЛ может быть поставлен только в условиях регресса дыхательной недостаточности пациента.

(3) При неэффективности или выраженных симптомах могут быть использованы различные растворы с антисептическим действием.

В связке с глаголом *быть* в качестве субъекта выступают существительные: *коронавирус, опасность*, *угроза*, *риск*. Это наблюдается в примерах (4) и (5).

(4) Люди, которые живут с ними, тоже должны оставаться дома, но могут выходить в ближайший магазин или аптеку, выгуливать собаку, выносить мусор, покидать квартиру, если им нужна экстренная медицинская помощь или есть прямая угроза жизни и здоровью.

(5) Возможно, новый коронавирус был передан через летучих мышей неизвестному промежуточному хозяину, а затем людям.

Глагол *использовать* используется в связке с глаголами *рекомендуется*, *следовать*, *рекомендовать* и *мочь*, которые характерны для информативных и оперативных текстов, целью которых является оказать влияние на читателя и побудить к действию (6)–(8).

(6) Обязательно следует использовать телеметрический мониторинг ЭКГ.

(7) Не рекомендуется использовать гипотонические растворы, декстраны или растворы гидроксиэтилкрахмала.

(8) Для экспресс-диагностики могут использоваться экспресстесты по выявлению пневмококковой и легионеллезной антигенурии.

Высокочастотные глаголы корпуса РУСП: *быть (552), мочь (493), иметь (169), являться (167), следовать (146), соблюдать (118), проводить (114), принимать (101), находиться (101), использовать (99), находить (99), действовать (98), носить (97), разрешаться (93), получать (89).* Примеры (9) и (10) приводятся со словами *следовать* и *разрешаться*.

(9) Вы обязаны предпринять следующие шаги как можно скорее: самоизолироваться, оставаться дома в течение минимум 10 дней и следовать рекомендациям по изоляции домохозяйств с возможной коронавирусной инфекцией.

(10) В просторных открытых спортивных комплексах разрешается проведение спортивных занятий несколькими группами согласно требованиям § 9 абзац 1 при исключении контакта между ними.

Глагол *быть* также является самым высокочастотным в данном корпусе, в утилите Word Sketch мы проверили, в связке с какими глаголами он сочетается чаще всего: *должен быть* (103) и *может быть* (86). Примеры (11) и (12) демонстрируют роль данных глаголов.

(11) Все работники должны быть обеспечены соответствующими средствами индивидуальной защиты и пройти инструктаж по их правильному использованию.

(12) Собрания могут быть запрещены, если принятием иных мер, в частности, назначением дополнительных требований, надлежащая защита от инфекций не сможет быть обеспечена.

Можно наблюдать, что список высокочастотных глаголов корпуса РУСП очень схож со списком глаголов корпуса РУСО. В обоих списках присутствует модальный глагол *мочь* и другие глаголы, которые типичны для текстов инструкций (*рекомендоваться*, *следовать*, *соблюдать*, *использовать*, *разрешаться)*, а также глаголы информативных текстов (*являться*, *вызывать*, *проводится*, *принимать).* Можно сделать вывод, что тексты соответствуют своему предназначению – информировать и призывать к действию.

К высокочастотным прилагательным корпуса РУСО относятся следующие: *медицинский (903), клинический (266), должный (255), тяжёлый (224), респираторный (218), инфекционный (212), новый (203), государственный (197), российский (198), федеральный (186), дыхательный (184), высокий (178), лабораторный (157), острый (154), коронавирусный (148).*

Из данного списка нас заинтересовали слова: *новый, государственный, российский и федеральный*, которые не относятся к медицинской теме. Были найдены наиболее частые словосочетания, в которых встречается прилагательное *новый*: *новая инфекция (97), новый коронавирус (33), новый коронавирусный (68).* Прилагательные *государственный, российский и федеральный* входят в состав названий медицинских учреждений и чаще всего входят в одно и то же предложение: *Федеральное государственное учреждение, медицинский университет Российской Федерации, министерство здравоохранения Российской Федерации* и др.

В список высокочастотных прилагательных корпуса РУСП вошли следующие слова: *должный (371), новый (327), медицинский (296), коронавирусный (234), высокий (172), отрицательный (142), федеральный (129), настоящий (121), социальный (121), тяжёлый (119), соответствующий (117), инфекционный (117), северный (105), общественный (96), общий (93).*

Интерес представляет прилагательное *северный*. С помощью Word Sketch мы определили, что наиболее часто северный встречается со словом *Азия* и *центральный.* В Concordance во всех случаях эти слова входят в состав одного и того же предложения, например: *в Северной и Центральной Азии, в странах Северной и Центральной Азии.*

Сравнивая списки высокочастотных прилагательных, также можно сделать наблюдение, что в корпусе РУСО больше медицинских терминов (*медицинский, клинический*,

респираторный, *инфекционный*, *дыхательный*, *лабораторный*, *коронавирусный)*, чем в РУСП (*медицинский*, *коронавирусный*, *инфекционный*).

3.2.2. Сравнение N-грамм корпусов РУСО и РУСП

Для исследования языковых особенностей была также использована утилита N-grams, с помощью которой были найдены часто повторяющиеся в текстах словосочетания из 3–4 слов. Примеры N-грамм и их частота продемонстрированы в таблице 4.

Таблица 4. N-граммы корпусов РУСО и РУСП

РУСО	F	РУСП	F
Федерального государственного бюджетного учреждения	104	новой коронавирусной инфекции	161
в соответствии с	95	в соответствии с	123
набор реагентов для	90	по профилактике и	89
в том числе	87	лечению новой коронавирусной инфекции	87
Министерства здравоохранения Российской Федерации	61	руководство по профилактике	86
новой коронавирусной инфекции	57	класса защиты FFP	59
скорой медицинской помощи	56	северной и центральной Азии	56
Национальный медицинский исследовательский центр	51	с высокой заболеваемостью	49
пациентов с COVID-	49	стратегии более эффективного восстановления	48
у пациентов с	48	меры и стратегии более	48
с подозрением на	47	ответные меры и	48
главный внештатный специалист	43	с вариантом вируса	44
для выявления РНК	41	защите от инфекционных	40
фтизиопульмонологии и инфекционных	35	в связи с	40
оказание медицинской помощи	33	в том числе	38

Данные таблицы 4 свидетельствуют о наличии специальной лексики в текстах. Например, в текстах РУСО присутствуют названия учреждений (*Федеральное государственное бюджетное учреждение, Министерство здравоохранения Российской Федерации, Национальный медицинский исследовательский центр*, словосочетание *фтизиопульмонологии и инфекционных* входит также в состав названия *Национальный медицинский исследовательский центр фтизиопульмонологии и инфекционных заболеваний*). В список N-грамм корпуса РУСО входят словосочетания с медицинскими терминами (*набор реагентов для, новой коронавирусной инфекции, главный внештатный специалист, оказание медицинской помощи* и т. д.).

В корпусе РУСП нет названий учреждений, но присутствует название географического объекта (*северная и центральная Азия*) и словосочетания, образующие термины (*новая коронавирусная инфекция, класс защиты FFP, с вариантом вируса* и т. д.).

В текстах обоих корпусов присутствуют также словосочетания *в соответствии с, в том числе, в связи с*, наличие которых характерно для специальных текстов. Таким образом, данные особенности текстов исследования являются также отличительными для подъязыка в дискурсе о коронавирусе.

3.2.3. Ключевые слова корпусов РУСО и РУСП

Воспользовавшись утилитой KeyWords мы получили список ключевых слов, их абсолютную и относительную частотность, а также показатель ключевого слова. Данные занесены в таблицу 5.

Таблица 5. Частотности ключевых слов корпуса РУСО.

Ключевое слово	**F**	**ipm**	**keyness score**
пациент	915	6 697,56	77,3
медицинский	903	6 609,72	44,1
заболевание	527	3 857,50	24,7
инфекция	498	3 645,23	101,3
версия	360	2 635,10	20,9
быть	340	2 488,71	0,3
мочь	327	2 393,55	0,9
исследование	316	2 313,04	8,7
помощь	315	2 305,72	4,6
больной	307	2 247,16	16,6
случай	289	2 115,40	2,8
наличие	279	2 042,21	11,2
препарат	268	1 961,69	17,7
клинический	266	1 947,05	52,1
лечение	258	1 888,49	10,1

В список ключевых слов не включены имена собственные, местоимения и служебные слова. Как видно из таблицы, большая часть слов относится к медицинской тематике. У слов *инфекция, пациент, клинический* и *медицинский* довольно высокий показатель ключевого слова. В Concordance было найдено несколько примеров предложений с данными словами.

(13) Создание модели эпидемии вируса лихорадки Эбола, вспышек коронавирусных инфекций, определение факторов, способствующих распространению инфекций, позволило выделить наиболее значимые меры по предотвращению распространения опасных инфекций.

(14) У большинства пациентов с тяжелым течением COVID-19 на первой неделе заболевания развивается пневмония.

(15) Интерферон бета-1b (ИФН–β1b) обладает антипролиферативной, противовирусной и иммуномодулирующей активностью. В текущих клинических исследованиях ИФН–β1b используется в комбинации с лопинавиром.

(16) Мероприятия, направленные на механизм передачи возбудителя инфекции: – соблюдение правил личной гигиены (мыть рук с мылом, использовать одноразовые салфетки при чихании и кашле, прикасаться к лицу только чистыми салфетками или вымытыми руками); – использование одноразовых медицинских масок, которые должны сменяться каждые 2 часа; – использование защитной одежды для медработников; – проведение дезинфекционных мероприятий; – утилизация медицинских отходов класса В; – эвакуация больных специальным транспортом.

Из вышеперечисленных примеров (13)–(16) также видно, что корпус РУСО содержит тексты с явно выраженной медицинской тематикой.

Далее был составлен список ключевых слов корпуса РУСП. Абсолютная и относительная частота, а также показатель ключевых слов продемонстрированы в таблице 6.

Таблица 6. Частотности ключевых слов корпуса РУСП.

Ключевое слово	**F**	**ipm**	**keyness score**
быть	553	4 326,36	0,5
мочь	494	3 864,78	1,5
лицо	450	3 520,55	7,3
инфекция	445	3 481,43	96,7
человек	392	3 066,79	1,6
должный	374	2 925,97	2,9
пациент	363	2 839,91	32,8
мера	355	2 777,32	10,1
страна	329	2 573,91	3,9
новый	327	2 558,27	1,9
результат	318	2 487,85	4,3
время	315	2 464,38	1,4
случай	311	2 433,09	2,2
согласно	309	2 417,44	16,8
медицинский	296	2 315,74	15,4

Как и с корпусом РУСО мы исключили из списка имена собственные, местоимения и служебные слова. Из данного списка к медицинским терминам можно отнести слова *инфекция*, *пациент* и *медицинский*. У слов *инфекция* и *пациент* высокий коэффициент ключевого слова, поэтому далее приводятся примеры (17) и (18) с данными словами.

(17) При отсутствии соответствующих материальных приспособлений для защиты от инфекции гражданам рекомендуется соблюдать минимальную дистанцию в 1,5 метра по отношению к другим лицам.

(18) Под изоляцией понимают обособление уже зараженного инфекционным заболеванием пациента от здоровых людей.

Слова *пациент*, *медицинский*, *инфекция*, *быть*, *мочь* и *случай* встречаются в обоих списках ключевых слов, но с разной частотой (см. таблицу 7).

Частота слова *пациент* в корпусе РУСО почти в три раза выше, чем в корпусе РУСП, а частота слова *медицинский* – почти в четыре раза. Показатель ключевого слова *инфекция* является самым высоким в обоих корпусах, что показывает фокус содержания текстов в данных корпусах.

Таблица 7. Общие ключевые слова и их частоты в корпусах РУСО и РУСП.

Ключевое слово	F (РУСО)	keyness score (РУСО)	F (РУСП)	keyness score (РУСП)
пациент	915	77,3	363	32,8
медицинский	903	44,1	296	15,4
инфекция	498	101,3	445	96,7
быть	340	0,3	553	0,5
мочь	327	0,9	494	1,5
случай	289	2,8	311	2,2

В списке ключевых слов корпуса РУСП, помимо медицинских терминов, появились слова, относящиеся к жизни человека и общества: *человек, страна, согласно.* Примеры (19)–(22) также свидетельствуют, что в текстах корпуса РУСП присутствует информация о коронавирусе и о правилах поведения в обществе.

(19) В мероприятиях в закрытых помещениях с одновременным присутствием более 20 человек могут принимать участие только лица с отрицательным результатом тестирования.

(20) У пожилых людей с сопутствующими заболеваниями чаще наблюдается тяжелое течение болезни.

(21) Обязательное тестирование необходимо пройти до въезда. Никаких исключений нет. Это гарантирует, что въезжающие лица сделают тест на коронавирус SARS-CoV-2 до въезда в страну.

(22) На проводимые мероприятия распространяется запрет на посещение и участие согласно § 7.

3.2.4. Слова *covid* и *коронавирус* в корпусах РУСО и РУСП

Так как данная работа основана на исследовании подъязыка в дискурсе о COVID-19, то интересно знать частотность слов *covid* и *коронавирус.* Частота этих слов в корпусе РУСО составила: 970 для *covid* и 82 для слова *коронавирус*.

Далее, воспользовавшись утилитой Thesaurus, мы определили, какие слова семантически связаны со словом *коронавирус*, т. к. по слову *covid* данных найдено не было. Thesaurus предоставляет данные в разных формах, так нами было получено облако слов (см. Рис 3).

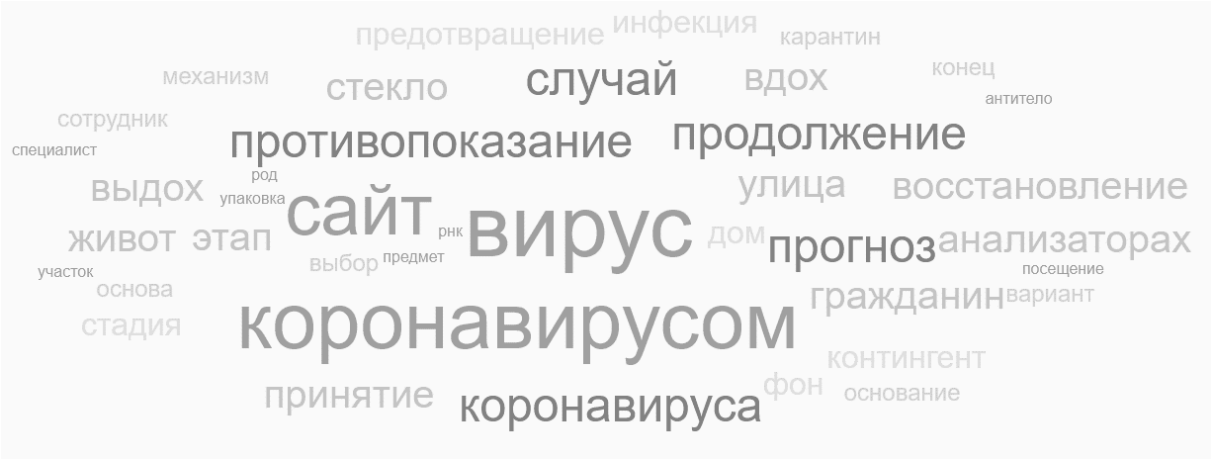


Рис. 3. Облако слов для *коронавирус* в корпусе РУСО.

Как можно наблюдать на рисунке 3, не все слова стоят в форме именительного падежа (*анализаторах*, *коронавируса*, *коронавирусом)*. Большинство слов, схожих по дистрибуции, относятся к медицинской теме, например: *вирус*, *прогноз*, *инфекция, восстановление*, *прогноз*, *выдох*, *живот*, *карантин*, *РНК*, *противопоказание* и др. Слова *выдох*, *вдох* и *живот* относятся к словам, описывающих физическую составляющую жизни человека. Само понятие *человек* определяется словами *сотрудник, гражданин* и *специалист.*

По данному облаку слов можно предположить, что в текстах идёт речь о вирусе, который опасен для граждан и требует особых мер предосторожности как дома, так и на улице. Понятно также, что данным вирусом занимаются специалисты и данные о вирусе можно получить на определённых сайтах. К такому выводу мы пришли, проанализировав примеры (22)–(27), полученные из Concordance по поиску слов: *сайт*, *сотрудник и специалист*.

(23) Список медицинских организаций, в которых можно сделать прививку, можно уточнить на сайте регионального министерства здравоохранения.

(24) Данные о странах с неблагополучной эпидситуацией размещены на сайтах ВОЗ и Роспотребнадзора.

(25) Коронавирусная инфекция – острое вирусное заболевание с преимущественным поражением верхних дыхательных путей, вызываемое РНК-геномным вирусом рода Betacoronavirus семейства Coronaviridae.

(26) Мойте руки с мылом и водой тщательно после возвращения с улицы, контактов с посторонними людьми.

(27) Появление COVID-19 поставило перед специалистами здравоохранения задачи, связанные с быстрой диагностикой и оказанием медицинской помощи больным.

Далее были выделены семантически близкие слова для *коронавирус* в корпусе РУСП. Облако слов представлено на рисунке 4 (см. ниже).

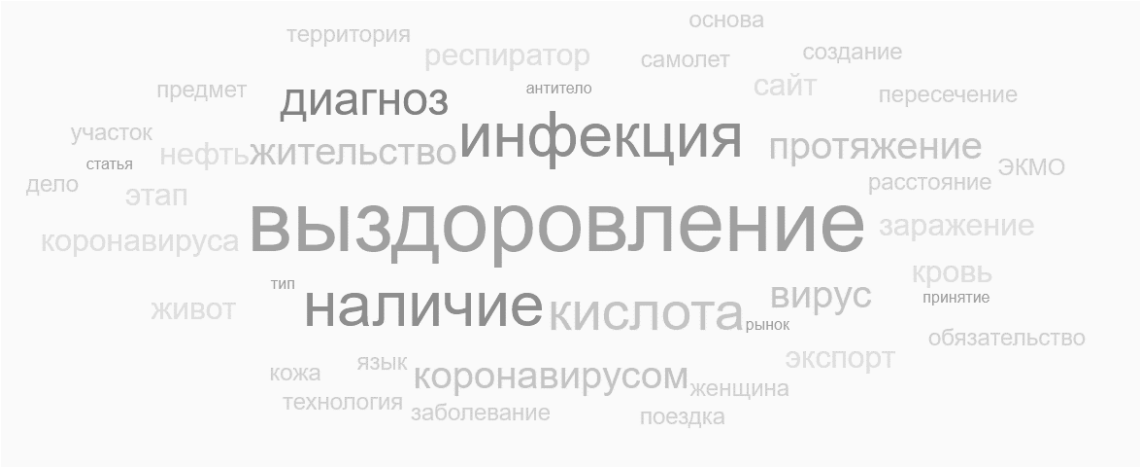


Рис. 4. Облако слов для *коронавирус* в корпусе РУСП.

В облаке слов корпуса РУСП для слова *коронавирус* меньше слов, относящихся к медицине, чем в корпусе РУСО. Появляются слова: *территория*, *поездка*, *женщина*, *рынок*, *жительство*, *самолет*, *экспорт, нефть*. Становится понятно, что кроме медицинской темы в дискурсе о COVID-19 затрагивается тема социальной жизни человека. У слова *выздоровление* самый высокий коэффициент семантической близости со словом *коронавирус*, что придает текстам положительный оттенок и сообщает читателю о возможности борьбы с инфекцией (28).

(28) Пациент имеет большие шансы на выздоровление при активной противовирусной и антицитокиновой терапии и поддерживающем лечении.

Однако для продолжения нормальной социальной жизни (общение, путешествие, работа и т. д.) необходимо соблюдать меры предосторожности по распространению коронавируса. Как известно, вызванная коронавирусом болезнь, превратилась в пандемию, которая повлияла на все сферы деятельности человека, также на экономическую жизнь всех стран. Примеры (29)–(32) из текстов корпуса РУСП подтверждают это.

(29) Более 14 процентов опрошенных женщин потеряли работу, у 41 процента было сокращено оплачиваемое рабочее время.

(30) До поездки: медицинская страховка, тест на коронавирус, приложение Corona-Warn-App. Проверьте и при необходимости продлите Вашу медицинскую страховку для выезжающих за рубеж.

(31) Отмена ограничений на пересечение внешней границы в отношении Ватикана. На внешних границах движение между Финляндией и Ватиканом разрешено без ограничений через все пункты пропуска с соблюдением правил въезда в страну для иностранных граждан.

(32) Экономика стран субрегиона постепенно восстанавливалась после кризиса цен на нефть, однако по мере наступления последствий пандемии пространство для маневра правительств в бюджетно-финансовой сфере стало различаться в плане оказания чрезвычайной поддержки и принятия мер по защите домашних хозяйств и предприятий.

Примеры (29)–(32) были найдены из Concordance по словам: *выздоровление, женщина, поездка, пересечение* и *нефть*.

3.2.5. Частотные списки корпусов ФИНО и ФИНП

Как и для корпусов на русском языке, были получены частотные списки слов (по лемме), существительных, глаголов и прилагательных для корпусов на финском языке. Например, 15 наиболее частотных лемм с глоссами, а также их частоты продемонстрированы в таблице 8 (см. ниже). В данный список входят: одно существительное (*työntekijä* ’работник’), четыре глагола (*olla* ’быть’, *voida* ’мочь’, *tulla* ’прийти’ и *saada* ’получить’), местоимения (*muu* ’другой’, *se* ’то, тот’, *joka* ’который’), остальную часть составляют служебные слова (*ja* ’и’, *tai* ’или’, *että* ’что’, *jotta* ’чтобы’, *sekä* ’так и’, *ei* ’не’ и *myös* ’тоже’).

Таблица 8. Частотности слов по лемме в корпусе ФИНО.

Слово и глосса	**F**	**ipm**
olla ’быть’	4 447	35 313,55
ja ’и’	4 087	32 454,80
ei ’не’	1 202	9 545,06
voida ’мочь’	1 175	9 330,65
tai ’или’	1 136	9 020,96
se ’то, тот’	931	7 393,05
että ’что’	699	5 550,75
tulla ’прийти’	671	5 328,40
jotta ’чтобы’	629	4 994,88
joka ’который’	612	4 859,88
myös ’тоже’	590	4 685,18
saada ’получить’	526	4 176,96
muu ’другой’	499	3 962,55
sekä ’так и’	487	3 867,26
työntekijä ’работник’	425	3 374,92

В список данного корпуса вошли абсолютно общеупотребительные слова. Правда, частота существительного *työntekijä* ’работник’ вызвало удивление. Поэтому были найдены примеры (33)–(35) с данным словом.

(33) Jos <u>työntekijä</u> saa koronavirustartunnan, tartuntatautilääkäri ja hoitava taho ohjeistavat jatkotoimenpiteistä.

(34) On suositeltavaa, että <u>työntekijä</u> kertoo työnantajalleen koronavirustartuntaepäilystä, vaikka hänellä ei ole velvollisuutta kertoa terveydentilaansa liittyvää tietoa.

(35) Jos <u>työntekijä</u> jää oma-aloitteisesti pois työstä, palkanmaksuvelvollisuutta ei ole.

Из примеров (33)–(35) следует, что в текстах речь идёт об инструкциях и рекомендациях для заразившихся коронавирусом, а также о правах работников и работодателей.

В список частотных существительных корпуса ФИНО вошли слова: *työntekijä* ’работник’ (425), *lapsi* ’ребёнок’ (355), *käyttö* ’использование’ (346), *henkilö* ’лицо/личность’ (323), *työ* ’работа’ (321), *työnantaja* ’работодатель’ (315), *suositus* ’рекомендация’ (308), *tieto* ’сведение’ (298), *kunta* ’муниципалитет’ (296), *ohje* ’инструкция’ (289), *aika* ’время’ (258), *tilanne* ’ситуация’ (255), *esimerkki* ’пример’ (245), *paikka* ’место’ (239), *toiminta* ’действие’ (227).

Как мы видим, в списке существительных присутствуют слова, которые относятся к социальным и трудовым отношениям, а также указывающих на функцию текстов, наставлять, давать инструкции и информировать. Примеры (36)–(39) со словами *työntekijä* ’работник’, *työnantaja* ’работодатель’, *suositus* ’рекомендация’, *ohje* ’инструкция’.

(36) Etätöitä organisoitiin myös huomioiden <u>työntekijöiden</u> tarpeet.

(37) Lomake, joka <u>työnantajan</u> on täytettävä, löytyy myös samalta verkkosivustolta.

(38) <u>Suositukset</u> ovat voimassa kesäkuun alkuun asti.

(39) Tampereen kaupungin ympäristöterveyden laatimat <u>ohjeet</u> koskien koronavirustilanteen huomioimista elintarvikehuoneistoissa ja muissa asiakaspalvelutiloissa

Из высокочастотных глаголов корпуса ФИНО были выделены: *olla* ’быть’ (4 140), *voida* ’мочь’ (1 162), *tulla* ’приходить’ (637), *saada* ’получить’ (343), *käyttää* ’использовать’ (268), *tehdä* ’делать’ (265), *pitää* ’держать/быть необходимым’ (205), *ottaa* ’брать’ (193), *tarvita* ’нуждаться’ (178), *noudattaa* ’соблюдать’ (157), *antaa* ’дать’ (147), *suositella* ’рекомендовать’ (145), *koskea* ’трогать, касаться’ (142), *välttää* ’избегать’ (112), *arvioida* ’оценивать’ (110).

В списке высокочастотных глаголов присутствуют модальные глаголы (*voida, pitää, tarvita*) и другие глаголы, используемые в информирующих и призывающих к действию текстах (*käyttää, noudattaa, suositella, välttää*). Это можно наблюдать из примеров (40)–(42).

(40) Harjoitustapahtumissa <u>suositellaan</u> välttämään pukuhuoneiden ja sisätilojen käyttöä.

(41) <u>Noudata</u> huolellista käsihygieniaa ja hyviä yskimiskäytäntöjä myös matkan aikana.

(42) Kaikilla asiakkailla <u>pitää</u> olla oma istumapaikka pöydän tai tason ääressä.

В группу высокочастотных прилагательных вошли: *mahdollinen* ’возможный’ (258), *oma* ’свой’ (206), *julkinen* ’общественный’ (165), *uusi* ’новый’ (160), *yleinen* ’общий’ (153), *mukainen* ’подобный’ (127), *seuraava* ’следующий’ (102), *vastaava* ’соответствующий’ (102), *hyvä* ’хороший’ (100), *alueellinen* ’территориальный’ (98), *tärkeä* ’важный’ (96), *eri* ’разный’ (96), *suuri* ’большой’ (95), *usea* ’частый’ (78), *yksityinen* ’частный’ (78).

Можно заметить, что в список высокочастотных прилагательных вошли слова относящиеся к жизни общества (*julkinen* ’общественный’, *yleinen* ’общий’, *alueellinen* ’территориальный’, *yksityinen* ’частный’).

Наиболее частотные слова корпуса ФИНП с глоссами на русском языке, а также их частоты, продемонстрированы в таблице 9 (см. ниже).

В состав вошли существительные (*Espanja* ’Испания’, *Karjala* ’Карелия’, *rokote* ’прививка’), глаголы (*olla* ’быть’, *voida* ’мочь’, *saada* ’получить’), местоимения (*joka* ’который’, *se* ’это’, *tämä* ’это’) и служебные слова (*ja* ’и’, *ei* ’не’, *mukaan* ’по, с собой’, *että* ’что’, *myös* ’тоже’, *tai* ’или’).

Таблица 9. Частотности слов по лемме в корпусе ФИНП.

Слово и глосса	**F**	**ipm**
olla ’быть’	5 301	40 710,85
ja ’и’	3 613	27 747,27
voida ’мочь’	1 032	7 925,60
se ’это’	924	7 096,17
ei ’не’	870	6 681,46
että ’что’	803	6 166,91
joka ’который’	738	5 667,72
saada ’получить’	696	5 345,17
mukaan ’по, с собой’	635	4 876,70
Espanja ’Испания’	630	4 838,30
Karjala ’Карелия’	612	4 700,06
myös ’тоже’	566	4 346,79
tämä ’это’	548	4 208,55
tai ’или’	515	3 955,12
rokote ’прививка’	507	3 893,68

В данном списке, кроме слов *Espanja* ’Испания’ и *Karjala* ’Карелия’, также содержатся общеупотребительные слова. Правда, появляется слово *rokote* ’прививка’, которого не было в частотных списках в корпусе ФИНО.

В список частотных существительных вошли слова: *rokote* ’прививка’ (507), *maa* ’страна’ (344), *tartunta* ’заражение’ (315), *määrä* ’количество’ (307), *prosentti* ’процент’ (293), *viikko* ’неделя’ (290), *käyttö* ’использование’ (285), *rajoitus* ’ограничение’ (275), *päivä* ’день’ (252), *alue* ’территория’ (236), *tieto* ’сведение’ (229), *ihminen* ’человек’ (227), *aika* ’время’ (216), *henkilö* ’лицо/личность’ (213), *koronavirus* ’коронавирус’ (205).

Как мы видим, существительное *rokote* ’прививка’ является самым высокочастотным, значит в текстах много говорится о вакцинации. Также можно наблюдать, что в список высокочастотных существительных вошли слова разных тематик, медицинская (*rokote* ’прививка’, *tartunta* ’заражение’, *koronavirus* ’коронавирус’), социума (*ihminen* ’человек’,

henkilö ’лицо/личность’), а также слова, относящиеся к расстоянию, количеству и времени (*maa* ’страна’, *alue* ’территория’, *määrä* ’количество’, *prosentti* ’процент’, *aika* ’время’, *viikko* ’неделя’).

К высокочастотным глаголам корпуса ФИНП относятся: *olla* ’быть’ (5,007), *voida* ’мочь’ (1 018), *saada* ’мочь, получить’ (402), *kertoa* ’рассказать’ (339), *tulla* ’приходить’ (269), *pitää* ’держать/быть необходимым’ (225), *tehdä* ’делать’ (220), *aloittaa* ’начать’ (179), *ottaa* ’брать’ (177), *käyttää* ’использовать’ (150), *koskea* ’касаться’ (137), *antaa* ’дать’ (131), *lähteä* ’отправиться’ (127), *tarvita* ’нуждаться’ (126), *sanoa* ’сказать’ (118). Было найдено несколько предложений (43)–(45) с высокочастотными глаголами *aloittaa* ’начать’, *kertoa* ’рассказать’ и *voida* ’мочь’.

(43) Petroskoin valtionyliopiston johto ei sulje pois mahdollisuutta, että opiskelijat aloittavat lähiopetuksessa.

(44) Jotkut tartunnan saaneista ovat myös kertoneet kokeneensa ihon kihelmöintiä tai kuumuuden tunnetta iholla.

(45) Tällaiset sopimukset ja päätökset voivat olla voimassa enintään kuusi kuukautta.

Как и в корпусе ФИНО, высокочастотными являются модальные глаголы (*voida, saada, pitää, tarvita*), но нет глаголов, которые конкретно указывали бы на ограничение действий, как мы это наблюдали в корпусе ФИНО (*noudattaa* ’соблюдать’, *välttää* ’избегать’, *suositella* ’рекомендовать’).

Из высокочастотных прилагательных были выделены следующие: *uusi* ’новый’ (359), *mahdollinen* ’возможный’ (200), *suuri* ’большой’ (184), *viime* ’прошлый’ (147), *oma* ’свой’ (137), *eri* ’разный’ (127), *tärkeä* ’важный’ (102), *julkinen* ’общественный’ (96), *usea* ’частый’ (91), *seuraava* ’следующий’ (90), *viimeinen* ’последний’ (88), *turvallinen* ’безопасный’ (77), *pakollinen* ’обязательный’ (77), *korkea* ’высокий’ (76), *vastaava* ’соответствующий’ (74).

Нас заинтересовало слово *pakollinen* ’обязательный’, поэтому мы решили выяснить, в коллокациях с какими существительными оно выступает. Были найдены такие примеры: *käyttö* ’использование’, *desinfiointi* ’дезинфицирование’, *turvaväli* ’безопасное расстояние’, *terveystarkastus* ’медицинский осмотр’, *karanteeni* ’карантин’ и др. Эти словосочетания побуждают читателя к обязательному действию, что является основной задачей информативных и оперативных текстов. Примеры (46)–(50) очень хорошо демонстрируют функции текстов.

(46) Maskien käyttö on pakollista kaikissa yleisissä sisätiloissa kuten joukkoliikenteessä ja ostoskeskuksissa.

(47) Selväksi on tullut, että pakollinen karanteeni voi vaikeuttaa monien ihmisten elämää, Parfentšikov selittää.

(48) Lisäksi Lappeenrannan Nuijamaan rajanylityspaikalla alkavat rajanylittäjien pakolliset terveystarkastukset torstaina 1. huhtikuuta.

(49) Pakolliset terveystarkastukset koskevat matkustajaliikennettä, eivät tavarankuljettajia Venäjällä on rekisteröity eläinkoronarokote, raportoi TASS.

(50) Kasvomaskin käyttö on pakollinen.

Хотя в корпусе ФИНП меньше высокочастотных глаголов, призывающих к действию или ограничивающих их, появилось высокочастотное прилагательное *pakollinen* ’обязательный’, которое указывает на необходимость каких-либо действий.

3.2.6. Сравнение N-грамм ФИНО и ФИНП

Как и в корпусах на русском языке, в корпусах ФИНО и ФИНП с помощью утилиты N-grams были выделены словосочетания 3–4 слов. Данные продемонстрированы в таблице 10 (см. ниже).

В корпусе ФИНО наблюдается присутствие номенклатур, например название учреждения (*terveyden ja hyvinvoinnin laitos* ’институт здравоохранения и социального благополучия’), название должности (*tartuntataudeista vastaava lääkäri* ’врач-инфекционист’) и словосочетания с терминами (*korona ja rajoitustoimi* ’корона и ограничительные меры’, *alueellisesti rajoitettu ulkotila* ’территориально ограниченная уличная площадка’. Словосочетание *korona tuli kylään* ’корона пришла в гости’, которое, на первый взгляд, кажется обычным выражением, оказалось названием проекта о влиянии коронавируса на дошкольное образование. Это было выяснено с помощью примеров (51) и (52) из Concordance.

(51) Korona tuli kylään – tutkimushankkeen loppuraportti: Koronakriisin vaikutuksia varhaiskasvatukseen.

(52) Tutkimus on tehty yhteistyössä seuraavien toimijoiden kanssa: ’Korona tuli kylään’ -hankkeen tutkijat (Helsingin yliopiston kasvatustieteellinen tiedekunta), Helsingin yliopiston Koulutuksen arviointikeskus, Tutkivat ja kehittävät harjoittelupäiväkodit -hanke sekä Ebeneser-säätiö.

Среди N-грамм корпуса ФИНП также можно выделить названия учреждений и органов власти (*terveyden ja hyvinvoinnin laitos* ’институт здравоохранения и социального благополучия’, *Euroopan komissio* ’Европейская комиссия’), географического объекта (Karjala ’Карелия’) и названия должностей (*pääministeri*

’премьер-министр’, *puheenjohtaja* ’председатель’, terveysministeri ’министр здравоохранения’).

Таблица 10. N-граммы корпусов ФИНО и ФИНП.

ФИНО	F	ФИНП	F
terveyden ja hyvinvoinnin laitoksen ’института здравоохранения и социального благополучия’	68	Ursula von der Leyen ’Урсула фон дер Ляйен’	54
korona tuli kylään ’корона пришла в гости’	57	Karjalan pääministeri Artur Parfentšikov ’премьер-министр Карелии Артур Парфенчиков’	45
ja rajoitustoimen vaikutukset liikkumiseen ’и влияние ограничительных мер на передвижение’	36	puheenjohtaja Ursula von der ’председатель Урсула фон дер’	37
koronan ja rajoitustoimen ’короны и ограничительных мер’	32	Euroopan komissio on ’Европейская комиссия’	30
ei ole mahdollista ’не представляется возможным’	26	terveysministeri Mihail Ohlopkov ’министр здравоохранения Михаил Охлопков’	23
tartuntataudeista vastaava lääkäri ’врач, лечащий инфекционные заболевания’	26	terveyden ja hyvinvoinnin ’здравоохранения и социального благополучия’	21
lasten ja nuorten ’детей и подростков’	17	kahden viikon aikana ’в течение двух недель’	18
Someron kaupungin tilannekatsastus koronarajoituksiin ’обзор на ограничения в связи с коронавирусом в г. Сомеро’	15	on tällä hetkellä ’в данный момент’	16
ja yleisiä kokouksia ’и общие собрания’	15	tartuntojen määrä on ollut ’количество заражений было’	13
ja sen jälkeen ’и после этого’	15	on saanut ensimmäisen ’получил первый’	13
kunnan tai sairaanhoitopiirin ’муниципалитета или медицинского округа’	15	von der Leyenin mukaan ’по фон дер Ляйен’	13
yleisötilaisuuksia ja yleisiä ’общественные мероприятия и общие’	15	vastaavan määräyksen allekirjoitti ’соответствующее указание подписал’	12
viihdepalveluissa ja erikoisliikkeissä ’в развлекательных услугах и специализированных заведениях’	14	voivat avata ovensa ’могут открыть свои двери’	12
jos sinulla on ’если у тебя есть’	13	Karjalassa on todettu ’В Карелии подтверждены’	12
alueellisesti rajatuissa ulkotiloissa ’на территориально ограниченных уличных площадках’	13	asukasta kohden viimeisen ’на жителя последнего’	12

В списке также присутствуют словосочетания, характерные для специальных текстов, например *von der Leyen mukaan* ’по фон дер Ляйен’, *vastaavan määräyksen allekirjoitti* ’соответственный указ подписал’ и т. д.

Данные N-граммы также, как и в корпусах на русском языке, подтверждают наличие в текстах специальной лексики, характерное для подъязыка.

3.2.7. Ключевые слова корпусов ФИНО и ФИНП

Из списка ключевых слов для корпусов ФИНО и ФИНП были исключены собственные имена, местоимения и служебные слова. Мы обратили внимание на то, что в данный список, также как и в список высокочастотных существительных, вошло слово *työntekijä* ’работник’ с показателем ключевого слова 18,3. Однако более высокий показатель ключевого слова у слов *työnantaja* ’работодатель’ (32,1) и *suositus* ’рекомендация’ (61,3). Ключевые слова корпуса ФИНО с глоссами на русском языке, а также их частоты и показатель ключевого слова продемонстрированы в таблице 11 (см. ниже).

Модальный глагол *voida* ’мочь’, обозначающий возможность или способность совершить действие имеет самую высокую частоту.

Таблица 11. Частотности ключевых слов корпуса ФИНО.

Ключевое слово и глосса	F	ipm	keyness score
voida ’мочь’	1 175	9 330,66	1,9
muu ’другой’	499	3 962,55	1,6
työntekijä ’работник’	425	3 374,92	18,3
käyttää ’использовать’	353	2 803,17	2,3
henkilö ’личность/лицо’	323	2 564,94	7,5
antaa ’дать’	318	2 525,23	2,3
työantaja ’работодатель’	315	2 501,41	32,1
suositus ’рекомендация’	308	2 445,82	61,3
tieto ’сведение’	298	2 366,41	3,0
ohje ’инструкция’	289	2 294,94	12,4
mahdollinen ’возможный’	258	2 048,77	3,6
tilanne ’ситуация’	255	2 024,95	5,5
tila ’место’	240	1 905,84	3,4
toiminta ’действие’	227	1 802,60	3,3
päätös ’решение’	221	1 754,96	6,9

Ключевые слова корпуса ФИНО демонстрируют, что в текстах данного корпуса говорится об общественных и трудовых отношениях. Это можно заметить также из ниже приведённых примеров (53)–(56) со словами *työnantaja* ’работодатель’, *työntekijä* ’работник’, *mahdollinen* ’возможный’ и *päätös* ’решение’.

(53) Noudata <u>työnantajasi</u> turvallisuudesta, hygieniasta, työvaatetuksesta, suojainten ja kasvomaskien käytöstä antamia määräyksiä ja ohjeita.

(54) <u>Työntekijät</u> voidaan lomauttaa toistaiseksi.

(55) Suunnitelmassa on huomioitava turvavälien noudattaminen erityisesti: - yleisön saapuminen ottelustadionille, turvatarkastusten suorittaminen sekä lipunmyynti - yleisön ohjaaminen osoitetuille istumapaikoille <u>mahdollisten</u> myyntipisteiden organisointi sekä hygienian säilyttäminen WC-tilojen puhtaus ja riittävyys yleisön poistuminen stadionilta porrastaen.

(56) Lounais-Suomen aluehallintovirastolla on valmius tehdä kiihtymisvaiheen mukaiset <u>päätökset</u> koskien muun muassa yleisötilaisuuksia ja tapahtumia.

Ключевые слова корпуса ФИНП с глоссами на русском языке, а также их частоты и показатель ключевого слова продемонстрированы в таблице 12.

Таблица 12. Ключевые слова корпуса ФИНП.

Ключевое слово и глосса	F	ipm	keyness score
olla ’быть’	5 301	40 710,85	0,9
voida ’мочь’	1 032	7 925,60	1,7
saada ’получить’	696	5 345,17	1,3
rokote ’прививка’	507	3 893,68	238,7
kaikki ’все’	469	3 601,85	1,1
muu ’другой’	368	2 826,18	1,1
kertoa ’рассказать’	347	2 664,91	2,9
tehdä ’делать’	346	2 657,23	0,8
maa ’страна’	344	2 641,87	3,4
tulla ’прийти’	328	2 518,99	0,7
tartunta ’заражение’	315	2 419,15	180,4
määrä ’количество’	307	2 357,71	5,4
prosentti ’процент’	293	2 250,19	8,6
viikko ’неделя’	290	2 227,15	2,8
käyttö ’использование’	285	2 188,76	3,8

Из таблицы 12 видно, что самые высокие частоты у глаголов *olla* ’быть’ и *voida* ’мочь’, а у слов *rokote* ’прививка’ и *tartunta* ’заражение’ очень высокие показатели ключевого слова – 238,7 (*rokote*) и 180,4 (*tartunta*). Список ключевых слов корпуса ФИНП отражает наличие информативных текстов. Например, слова *tartunta* ’заражение’, *määrä* ’количество’, *prosentti* ’процент’ и *viikko* ’неделя’ использованы в данных текстах для презентации статистики. Глагол *kertoa* ’рассказать’ выступает в коллокациях с *tiedote* ’пресс-релиз’, *viranomainen* ’орган власти’, *terveysministeri* ’министр здравоохранения’, что свидетельствует о важности и достоверности сообщений. Это доказывают и примеры (57)–(61).

(57) Terveysministeri kertoi, että laumasuoja syntyy, kun riittävän suuri määrä ihmisiä on vastustuskykyisiä taudille.

(58) Rajaliikenne on ollut viime viikkoina vähäisintä koko pandemian aikana, kerrotaan tiedotteessa.

(59) Kuluvan vuoden 25. viikon aikana koronatartuntojen määrä kasvoi 82 prosenttia.

(60) Kesään 2021 mennessä vähintään 70 prosenttia EU-maiden aikuisväestöstä pitäisi olla rokotettuina.

(61) Kesällä 2021 tartunnat lähtivät nousuun uuden virusvariantin lähdettyä leviämään.

Сравнивая списки ключевых слов, показатели ключевых слов и примеры из текстов корпусов ФИНО и ФИНП, можно заключить, что в текстах ФИНО содержатся инструкции для работодателей и работников, а в ФИНП – информация о прививке и статистические данные о заражениях.

Списки объединяют всего два слова *voida* ’мочь’ и *muu* ’другой’, у которых низкий показатель ключевого слова. Однако, модальный глагол *voida* ’мочь’ имеет высокую абсолютную частоту в обоих корпусах (см. таблицу 13). Частое использование данного ключевого слова говорит о том, что в текстах о коронавирусе говорится о возможности действий.

Таблица 13. Общие ключевые слова и их частоты в корпусах ФИНО и ФИНП.

Ключевое слово	**F (ФИНО)**	**keyness score (ФИНО)**	**F (ФИНП)**	**keyness score (ФИНП)**
voida ’мочь’	1 175	1,9	1 032	1,7
muu ’другой’	499	1,6	368	1,1

3.2.8. Слова *covid* и *коронавирус* в корпусах ФИНО и ФИНП

Слово *covid* ’ковид’ встречается в текстах корпуса ФИНО 66 раз, а *koronavirus* ’коронавирус’ – 169.

Воспользовавшись утилитой Thesaurus, также, как и в корпусах на русском языке, мы определили, какие слова находятся в семантической близости со словом *коронавирус* в корпусе ФИНО. Облако слов для *koronavirus* ’коронавирус’ представлено на рисунке 5.

Рис. 5. Облако слов для *koronavirus* ’коронавирус’ в корпусе ФИНО.

В облако слов схожих по дистрибуции со словом *koronavirus* ’коронавирус’ вошли слова: *COVID-19*, *koodi* ’код’, *nuori* ’молодой’, *oppilas* ’ученик’, *muunnos* ’вариант’, *virus* ’вирус’, *ohjeistus* ’инструкция’, *linjaus* ’линия’ и др. Слово *maskeja* ’маски’ находится в партитиве мн. ч., обозначающим часть чего-либо. В данном облаке, также, как и в облаке корпуса РУСО, из слов с высоким коэффициентом семантической близости преобладают

слова медицинской тематики: *COVID-19, tauti* ’заболевание’*, virus* ’вирус’*, muunnos* ’вариант’*, maski* ’маска’. У слова *COVID-19* самый высокий коэффициент семантической близости со словом *koronavirus*. Примеры (62) и (63) со словами *tauti* ’заболевание’ и *muunnos* ’вариант’.

(62) Testi kertoo vain, onko sinulla sillä hetkellä tautia.

(63) Kyseinen muunnos on jo levinnyt useimpiin Euroopan maihin sekä esimerkiksi Yhdysvaltoihin ja leviää kotoperäisesti myös Ruotsissa

Для определения *человек* используются определения *henkilökunta* ’персонал’, *huoltaja* ’опекун’, *oppilas* ’ученик’, что говорит о прагматической направленности текста. В текстах данного корпуса, помимо медицинской темы, также затрагивается тема социальных отношений, т. к. появляются слова: *henkilökunta* ’персонал’, *oppilas* ’ученик’, *nuori* ’молодой’, *perhe* ’семья’. Примеры (64)–(67) из текста с данными словами.

(64) Oppilaille jaetaan uusia maskeja koulupäivän aikana.

(65) Tarvittaessa henkilökuntamme opastaa käsihygienian toteuttamisessa, suojainten käytössä ja turvavälien toteutumisen ohjaamisessa.

(66) Jos nuori ei ole kykenevä tai halukas päättämään rokotuksestaan itse, rokottamiseen kysytään suostumusta huoltajilta.

(67) Toisinaan huoltajat kyseenalaistivat ohjeistuksia tai rajoitustoimenpiteitä, jolloin johtajat toimivat työntekijöidensä tukena.

Слово *covid* ’ковид’ в корпусе ФИНП встречается 75 раз, а *koronavirus* ’коронавирус’ – *205*. В облако слов данного корпуса вошли слова в именительном падеже (*säännös* ’правило’, *virus* ’вирус’, *ohjeistus* ’инструкция’, *kampanja* ’кампания’), также присутствуют слова в партитиве (*koronatartuntaa* ’заражение короной’, *koronavirustartuntaa* ’заражение коронавирусом’, *koronarajoituksia* ’ограничения по коронавирусу’). По-видимому, данные слова недостаточно часто повторяются, поэтому парсер ошибается. Облако слов представлено на рисунке 6 (см. ниже).

Со словами из облака были найдены примеры из текстов (68)–(72).

(68) SARS-CoV-2 on virus, joka aiheuttaa covid-19-taudin.

(69) Venäjän alueella purettiin koronarajoituksia ikäihmisille.

(70) Karjalassa etenee koronarokotuskampanja.

(71) Andalusian aluehallinto on päättänyt höllentää koronarajoituksia ja uudet säännökset astuvat voimaan heti torstaista alkaen.

(72) Euroopan pandemia hiljeni kesällä, kun Amerikan mantereen pandemia saavutti huippunsa. 4. maaliskuuta 2021 maailmassa oli yli 116 miljoonaa koronavirustartuntaa.

Рис. 6. Облако слов для *koronavirus* ’коронавирус’ в корпусе ФИНП.

Анализируя состав облака слов, которые дистрибутивно близки слову *koronavirus*, можно сделать вывод, что авторы текстов акцентирует внимание читателя на инструкциях и ограничениях, связанных с инфекцией коронавируса.

При сравнении облаков слов выяснилось, что в корпусах ФИНО и ФИНП присутствуют одинаковые слова (*virus* ’вирус’ и *ohjeistus* ’инструкция’), которые семантически близки слову *koronavirus* ’коронавирус’, а также слова близкие по смыслу *linjaus* ’линия’ (ФИНО), säännös ’устав’(ФИНП). На основании этого можно определить, что тема коронавируса переплетается с темой инструктажа во время пандемии. Однако, в облаке слов корпуса ФИНО присутствуют слова *oppilas* ’ученик’, *perhe* ’семья’, *henkilökunta* ’персонал’ и *huoltaja* ’опекун’, поэтому можно предположить, что данные тексты направлены конкретным адресатам.

3.3. Общие и отличительные черты текстов исследования

В данном разделе сравниваются данные, полученные в ходе работы с корпусами. В подразделе 3.3.4. небольшое внимание уделяется использованию переводческих универсалий в переводных текстах корпусов РУСП и ФИНП.

3.3.1. Частотность слов

Анализируя частотные списки слов (по лемме) корпусов, в которых преобладают служебные слова, мы замечаем, что в корпусах на русском языке предлоги являются высокочастотными единицами, т. к. они служат для связи слов в предложении в русском языке. В финском языке эта связь выражается с помощью окончания слова или

послелога. Из существительных в список слов корпуса РУСО входят *covid, пациент, заболевание* и *инфекция*, в РУСП – *covid*, в ФИНО – *työntekijä* ’работник’ и в ФИНП *rokote* ’прививка’. В частотных списках всех корпусов присутствуют одинаковые союзы *и* и *или*, что говорит о схожести построения предложений.

По частотности существительных можно определить, о какой информации в текстах идёт речь, и к какой предметной области эти тексты относятся. Во всех корпусах среди частотных существительных оказались слова, связанные с медициной. В корпусах РУСО и РУСП их оказалось больше всего, например: *пациент, заболевание, инфекция, препарат, лечение, терапия, пневмония* и др. Также становится ясно, что все тексты содержат информативную и оперативную информацию, сообщающую данные и призывающею к действию. Об этом можно судить по словам: *исследование, случай, день, человек, организация, риск, время, карантин, тестирование, инструкция,* а также глаголы: *мочь, рекомендовать, быть необходимым, использовать, избегать, соблюдать, следовать* и др.

Мы обратили внимание на высокочастотные существительные, обозначающие человека. В корпусе РУСО человек выступает в качестве *пациента* (913), в корпусе РУСП кроме существительного *человек* (392), появляется ещё *лицо* (426) и *пациент* (363). В корпусах текстов на финском языке из высокочастотных существительных появляются: *työntekijä* ’работник’ (425), *lapsi* ’ребёнок’ (355), *henkilö* ’личность/лицо’ (323) и *työnantaja* ’работодатель’ (315) в корпусе ФИНО; *ihminen* ’человек’ (227) и *henkilö* ’личность’ (213) в корпусе ФИНП. Наблюдая, какую роль выполняют разные существительные в семантической структуре категории предметности, и сравнивая эти данные, можно заключить, что в текстах РУСО более всего выражена медицинская тематика, в корпусах переводных текстов РУСП и ФИНП более нейтральная тематика, а в текстах ФИНО акцент смещается на трудовые отношения.

В корпусе ФИНО присутствуют и другие слова, относящиеся к теме социальных и трудовых отношений, например, *työ* ’работа’, *kunta* ’муниципалитет’, *julkinen* ’общественный’, *yleinen* ’общий’, *yksityinen* ’частный’. С помощью Concordance мы определили, что слово *työ* встречается как самостоятельно, так и в составе сложных слов в корпусе ФИНО 2 246 раз, а в корпусе ФИНП – 510 раз. Таким образом, можно предположить, что в текстах корпуса ФИНО присутствует идеология труда. Например, Ханну Кемппанен (2008) исследовал идеологические признаки в исторических текстах путем анализа ключевых слов (Kempanen 2008).

Существительное *rokote* ’прививка’ вошло в список 15 самых высокочастотных слов только в корпусе переводных текстов на финский язык (ФИНП). В текстах этого корпуса чаще говорится о важности вакцинации. Воспользовавшись утилитой Word Sketch, были выделены следующие словосочетания: *tehokas rokote* (11) ’эффективная вакцина’, *turvallinen rokote* (6) ’безопасная вакцина’, *lupaava rokote* (2) ’перспективная вакцина’, *rokote antaa* (2) ’вакцина даёт’, *rokote osoittautui* (9) ’вакцина доказала’, *rokotteen turvallisuus* (18) ’безопасность вакцины’, *saada rokotetta* (6) ’получить вакцину’ , а также *hyväksytty rokote* (3) ’одобренная прививка’, *toimiva rokote* (2) ’функциональная вакцина’ и др. Таким образом становится понятно, что авторы текстов корпуса ФИНП убеждают читателя в безопасности и эффективности вакцинации. Это можно наблюдать из примеров (73) и (74).

(73) Rokote osoittautui tehokkaaksi kaikkia uusia virusmuunnoksia vastaan.

(74) Yhdysvaltalainen Moderna ilmoitti maanantaina tiedotteessaan, että lääkeyhtiön kehittämä rokote antaa koronasuojan 94,5 prosentin varmuudella.

Среди существительных корпуса ФИНП присутствуют слова: *määrä* ’количество’, *prosentti* ’процент’, *viikko* ’неделя’, *aika* ’время’, *tieto* ’сведение’, которые могут быть использованы при описании статистических данных.

Глагол *быть* является высокочастотным во всех корпусах, но частотность этого глагола в корпусах финских текстов значительно выше, чем в корпусах русских текстов. Это можно объяснить тем, что в русском языке в настоящем времени глагол-связка *быть* опускается, тогда как в финском языке его наличие обязательно.

3.3.2. N-граммы

С помощью утилиты N-grams были выделены N-граммы, состоящие из 3–4 слов для каждого корпуса. Как мы убедились, N-граммы также демонстрируют языковые особенности текстов и наличие в них специальной лексики.

Во всех корпусах были обнаружены признаки ЯСЦ, например присутствие терминов, названия объектов (нарицательные, собственные имена и номенклатуры), а также словосочетания, часто используемые в специальных текстах (*в соответствии с, в том числе, ei ole mahdollista* ’не представляется возможным’), словосочетания, образующие термины и используемые в текстах одной тематики (*новая коронавирусная инфекция, пациенты с COVID-19, скорая медицинская помощь*) и т. д.

Список N-грамм корпуса ФИНП подтверждает наше предположение о том, что тексты содержат больше статистической информации. Об этом можно судить по словосочетаниям: *kahden viikon aikana* ’в течение двух недель’, *on tällä hetkellä* ’в данный момент’ и *asukasta kohden* ’на одного жителя’.

Во всех корпусах текстов присутствуют имена собственные (*Ursula von der Leyen* ’Урсула фон дер Ляйен’, *Artur Parfentšikov* ’Артур Парфенчиков’), названия должностей (*главный внештатный специалист, tartuntataudeista vastaava lääkäri* ’врач-инфекционист’), названия организаций (*Министерство здравоохранения Российской Федерации, Национальный медицинский исследовательский центр, terveyden ja hyvinvoinnin laitos* ’институт здравоохранения и социального благополучия’), которые являются авторитетными источниками информации о коронавирусе. Таким образом, информация, содержащаяся в текстах, является проверенной и достоверной, что положительно влияет на её восприятие.

Как уже упоминалось ранее, N-граммы демонстрируют особенности текстов о коронавирусе. Примеры, найденные с помощью N-grams, ещё раз убеждают нас, что все тексты данного исследования относятся к подъязыку, в котором отсутствуют художественные приёмы, характерные для литературного языка. Зато присутствуют признаки ЯСЦ, определение которого схоже с подъязыком, о чём говорилось в теоретической части работы. Если раньше мы знали об этом только из теории, то анализ результатов данной работы также подтвердил высказывания исследователей о подъязыке.

3.3.3. Сопоставление списков ключевых слов

При сопоставлении показателя относительной частоты и показателя ключевого слова нам удалось определить, какие слова отличаются более высокими статистическими данными.

В корпусах с русскими текстами (РУСО и РУСП) у слов *инфекция* и *пациент* оказались более высокие показатели; в корпусе ФИНО – у слова *suositus* ’рекомендация’ и в корпусе ФИНП – *rokote* ’прививка’ и *tartunta*’заражение, инфекция’. По полученным данным можно судить, что в текстах на русском языке больше говорится о заражении людей, в текстах корпуса ФИНО – о правилах и рекомендациях, а в текстах корпуса ФИНП – о прививке от вируса.

Сопоставив списки ключевых слов всех корпусов, мы выделили повторяющиеся слова и внесли данные в таблицу 14.

Таблица 14. Одинаковые ключевые слова в корпусах РУСО, РУСП, ФИНО и ФИНП.

Ключевое слово рус. / фин.	РУСО	РУСП	ФИНО	ФИНП
мочь/voida	X	X	X	X
быть/olla	X	X		X
пациент/potilas	X	X		
случай/tapaus	X	X		
страна/maa		X	X	
инфекция/tartunta	X	X		
человек/ihminen		X	X	
другой/muu			X	X
медицинский/lääketieteellinen	X	X		

Существительные *пациент, инфекция и случай* является ключевыми словами в корпусах на русском языке, РУСО и РУСП, что ещё раз доказывает принадлежность этих текстов к медицинской теме. Корпусы РУСП и ФИНО объединяют существительные *страна/maa* и *человек/ihminen*. В данных корпусах акцент темы смещается на человека и общество.

Так как в состав ключевых слов корпусов кроме существительных вошли также прилагательные, то можно предположить, что они имеет отношение к терминам. Например, прилагательное *медицинский* является ключевым словом в корпусах РУСО и РУСП. Мы обратились к функции Word Sketch и определили, какие существительные сочетаются с этим прилагательным. Из существительных в словосочетании со словом *медицинский* и с частотностью не ниже 10 были выделены следующие: *помощь* (РУСО, РУСП), *работник* (РУСО, РУСП), *персонал* (РУСО, РУСП), *маска* (РУСО, РУСП), *организация* (РУСО), *центр* (РУСО), *реабилитация* (РУСО), *изделие* (РУСО), *университет* (РУСО), *наблюдение* (РУСО), *обслуживание* (РУСП). По данным примерам можно сделать вывод, что прилагательные входят в состав терминов.

Ключевое местоимение *muu* ’другой’ в корпусах ФИНО и ФИНП не образует терминов, а используются только для указания на признак предмета, не обозначая его. Это местоимение сочетается с такими существительными, как *ihminen* ’человек’, *maa* ’страна’. В данном случае мы также привели примеры с частотой не ниже 10.

Как видно из таблицы 14 слово *мочь/voida* является ключевым во всех корпусах. Данное слово является модальным глаголом как в русском, так и в финском языке и обозначает возможность действия (Калинина 2010, 19–20; VISK § 1562). В корпусах на русском

языке глагол *мочь* сочетается с глаголами: *быть* (РУСО, РУСП), *применять* (РУСО), *проводить* (РУСО), *использовать* (РУСО), *получать* (РУСП), *возникать* (РУСП), *приводить* (РУСП). В данном случае представлены только те глаголы, частотность которых выше 10.

В корпусах на финском языке *voida* ’мочь’ является одним из самых высокочастотных ключевых слов и сочетается со многими глаголами. Например, в корпусах ФИНО и ФИНП этот глагол сочетается с: *olla* ’быть’, *saada* ’получить’, *tehdä* ’делать’, *käyttää* ’использовать’, *hakea* ’запрашивать’, *tulla* ’приходить’, *toteuttaa* ’исполнять’, *antaa* ’дать’; в корпусе ФИНО с: *pitää* ’держать’, *määrätä* ’назначать’, *arvioida* ’оценивать’, *järjestää* ’организовывать’, *maksaa* ’платить’, *valita* ’выбирать’, *vähentää* ’сокращать’, *hyödyntää* ’использовать с пользой’, *sopia* ’договориться’, *lähteä* ’отправляться’ и в корпусе ФИНП с глаголами: *saapua* ’прибывать’, *aiheuttaa* ’причинять, вызывать’ и др.

Можно наблюдать, что спектр глаголов, используемых в связке с модальным глаголом *voida* ’мочь’, в корпусах на финском языке намного шире, чем в корпусах на русском языке. Что говорит о том, что тексты о COVID-19 на финском языке более ярко отражают возможность действия, чем тексты на русском языке.

3.3.4. Особенности использования слов *covid-19* и *коронавирус* в корпусах данного исследования

Данные об употреблении слов *covid-19* и *коронавирус* в текстах на русском и финском языках оказались очень различными. Мы выяснили, что в текстах, объединённых одним дискурсом, употребление одного и того же термина может сильно отличаться. Так в текстах на русском языке 1) частота терминов *covid-19* и *коронавирус* намного выше, чем в текстах на финском языке; 2) использование *covid-19* предпочтительнее термина *коронавирус*, а в текстах на финском языке – наоборот.

Слово *covid*, написанное на русском языке (*ковид*) в текстах корпусов РУСО и РУСП не обнаружено. А слово *корона*, производное от *коронавирус* в русских текстах встретилось всего 3 раза в РУСО и 4 раза в РУСП. Зато в текстах на финском языке слово *korona* употребляется чаще, 96 раз в ФИНО и 20 в ФИНП. Наличие в текстах лексики разговорной речи не удивляет, т. к. это также свойственно текстам подъязыка. Об этом говорилось в теоретической части работы (см. подраздел 2.1.1.). Результаты данных представлены в таблице 15 (см. ниже).

Таблица 15. Частотность слов *covid-19* и *коронавирус* в корпусах.

Слово/корпус	РУСО	РУСП	Слово/корпус	ФИНО	ФИНП
covid-19	970	645	*covid-19*	66	75
коронавирус	61	100	*koronavirus*	169	205
корона	3	4	*korona*	96	20
Всего	1034	749	**Всего**	331	300

Е. С. Громенко А. С. Павлова и М. Н. Приёмышева освещают процессы в лексико-семантической системе русского языка, которые произошли в период начала пандемии. Авторы статьи отмечают, что термин *коронавирус* используется в СМИ с 2000 года и обозначает род вируса, так и SARS-CoV-2 является коронавирусом нового типа. Но так как с начала 2020 года использование этого термина участилось, то его значение сузилось – коронавирусом начали называть сам вирус, явившийся причиной пандемии. Данное слово начали употреблять в значении названия болезни, отсюда выражения: *заболеть коронавирусом, симптомы коронавируса* (Громенко и др. 2021, 54).

Громенко и др. также говорят об изменении в сфере употребления слова. Если раньше термин *коронавирус* употреблялся чаще в медицинской области, то после начала пандемии он перешёл в разряд общеупотребительной лексики (там же, 54).

Существительное *ковид* является неологизмом, который появился в 2020 г. и образован от английской аббревиатуры COVID-19 (от COronaVIrus Disease2019). *Ковид* обозначает инфекцию, вызываемую новой разновидностью коронавируса – SARS-CoV-2. Как отмечают авторы статьи, изначально этот термин использовался в профессиональном жаргоне медицинских работников, но затем быстро проник в общеупотребительную лексику. Термин *ковид* до сих пор является медицинским термином (там же, 55).

Авторы подчёркивают, что, используя термин *ковид*, мы говорим о *коронавирусе* и наоборот, хотя у этих слов немного разные значения. Говоря о *ковид/COVID-19*, правильно использовать словосочетания: *новый коронавирус, коронавирус нового типа, новый вариант коронавируса.*

Мы соглашаемся с Громенко и др., что термины *ковид* и *коронавирус* на данном этапе функционирования в русском языке являются синонимами.

В текстах данного исследования эти термины зачастую используются параллельно, но в русских текстах используется более правильная терминология для этого названия. Например, в корпусе РУСО со словом *коронавирус* используется прилагательное *новый* 33 раза из 61, т. е в других случаях просто *коронавирус*. В корпусе РУСП 3 из 100, в

остальных случаях *ковид* называют *коронавирусом*. Также используются словосочетания *новая коронавирусная инфекция* (в РУСО 97 раз, в РУСП 221 раз), *коронавирус нового типа* (в РУСП 21 раз) или *новый вариант коронавируса* (в РУСО 2 раза, в РУСП 2 раза)

В корпусе ФИНО *uusi koronavirus* ’новый коронавирус’ встречается 9 раз из 169, в остальных случаях просто *koronavirus* и в корпусе ФИНП – 2 из 205, в некоторых случаях для названия *COVID-19* используются словосочетания *uusi variantti* ’новый вариант’ (в ФИНО 5 раз), *uusi tiuunnos* ’новая версия’ (в ФИНО 5 раз, в ФИНП 12 раз). Таким образом в текстах ФИНО термин koronavirus ’коронавирус’ используется правильнее, чем в текстах ФИНП. Зато термин *korona* ’корона’, который больше используется в разговорной речи, чаще встречается в ФИНО, чем в ФИНП. Можно сделать вывод, что авторы текстов ФИНО пытаются использовать понятный обывателю язык.

Громенко и другие исследователи рассматривают семантику прилагательных *коронавирусный* и *ковидный* и приходят к выводу, что неологизмы *ковид* и *ковидный*, появившиеся с началом пандемии, отличаются по функционированию и частотности употребления от *коронавирус* и *коронавирусный*. Из словосочетаний, в составе которых присутствует *коронавирусный*, в официальных документах закрепилось только *коронавирусная инфекция*. Громенко и др. объясняют это тем, что это словосочетание являлось термином ещё до приобретения им нового значения (Громенко и др. 2021, 55–56).

Термины *ковид* и *ковидный* являются более частотными в официальных документах на русском языке. Авторы статьи предполагают, что на это могло повлиять: «а) сила действия закона экономии языковых средств и б) частотность употребления термина в профессиональной речи медицинскими работниками». Также авторы статьи отмечают, что ковид и ковидный получили широкое распространение в разговорной речи (Громенко и др. 2021, 56).

В текстах на русском языке нашего исследования используется только прилагательное *коронавирусный* (в РУСО 179 раз, в РУСП 280 раз). Например, в корпусе РУСО из 179 случаев *коронавирусный* сочетается с существительными: *инфекция* (171), *болезнь* (3), *пневмония* (3) и *синдром* (2).

Из 280 случаев в корпусе РУСП *коронавирусный* сочетается с существительными: *инфекция* (166), *заболевание* (11), *пациент* (1), *пандемия* (1), *болезнь* (1) и *пневмония* (1). Это говорит о том, что в наших текстах используется тот же термин, что и в официальных

документах. Возможно, что на момент создания текстов нашего исследования слова *ковид* и *ковидный* не были ещё так распространены.

В финском языке термины *koronavirus* ’коронавирус’ и *covid* также являются синонимами. К этому выводу мы пришли, анализируя коллокации данных терминов, воспользовавшись функцией Word Sketch и Concordance. Лишь в некоторых случаях для обозначения *COVID-19* используется словосочетание *uusi koronavirus* ’новый коронавирус’: в корпусе ФИНО 3 раза, в ФИНП – 2, в остальных случаях используется *koronavirus*’коронавирус’.

Почему же в финском языке предпочтительнее длинное слово *koronavirus* чем *covid*? Образование новых слов путём соединения двух или нескольких слов является очень продуктивным способом словообразования в финском языке (Vartiainen 2018). Суви Исохелла и Анита Нуоппонен рассматривают критерии ”хорошего термина”, который не должен быть слишком длинным. В то же время термин должен отвечать определённым требованиям языка, например написание и произношение. Возможность создания производных является также важным критерием (Isohella & Nuopponen 2016, 229–230).
Авторы статьи пишут также о предназначении термина, т. е. кто и как будет его использовать. Например, является ли термин абсолютно новым словом для пользователя, легко ли его запомнить и использовать на письме и в устной речи (Isohella & Nuopponen 2016, 231).
Термин *covid-19* является заимствованным словом из английского языка и больше относится к профессиональной лексике, а термин *koronavirus* ’коронавирус’ используется чаще в повседневной жизни, его легко писать и воспринимать на слух. В финском языке существует много сложных слов, в составе которых присутствует *koronavirus*, что было бы практически невозможно с *covid-19*. Очень часто заимствованные из других языков слова при склонении являются проблемными для финского языка, т. к. образуют исключения из правил.
Исходя из вышеперечисленных признаков термина, можно предположить, что *koronavirus* является ”хорошим термином” для говорящих на финском языке. Возможно поэтому частотность термина *koronavirus* ’коронавирус’ выше частотности *covid-19* в текстах на финском языке. Мы просмотрели примеры предложений, в которых встречаются данные термины и пришли к выводу, что *koronavirus* используется в общеупотребительной речи, а *COVID-19* – в профессиональной. Примеры (75) и (76) из корпуса ФИНО, а примеры (77) и (78) из ФИНП.

(75) Uusi koronavirus tarttuu ensisijaisesti pisaratartuntana, jos olet tartunnan saaneen lähellä.

(76) Ohjauksella ei syrjäytetä sitä yksilöllistä ja lääketieteellistä harkintaa joka päätöksiin sisältyy, vaan täsmennetään rajatusti sen käyttämisen yleisesti hyväksyttyjä kokemusperäisiä kriteerejä covid-19- taudin osalta ja sen luomassa poikkeuksellisessa erityistilanteessa.

(77) Käsien pesu on tunnetusti tärkeää koronaviruksen torjunnassa.

(78) Turvallisuuden lisäksi covid-19:n seuranta ja valvonta edellyttävät tehostettuja EU:n tason valvontajärjestelmiä, joihin kerätään tietoa paitsi taudin epidemiologiasta myös rokotuskattavuudesta kohderyhmien keskuudessa.

3.3.5. Особенности переводных текстов в корпусах РУСП и ФИНП

В данной подразделе кратко рассматриваются особенности переводных текстов и сравниваются между собой. Так как мы не имеем параллельных текстов, то возможно только в общих чертах описать особенности использования переводческих универсалий.

В подразделе 2.1.5. упоминалось, что мы остановим своё внимание на универсалиях, которые наиболее исследованы и эмпирически подтверждены с помощью многих исследований для различных пар языков. К ним относятся симплификация, экспликация, нормализация и интерференция.

Анализ симплификации переводных текстов нашего исследования показал, что показатель лексического разнообразия в корпусе РУСП незначительно ниже, чем в РУСО, а в корпусах ФИНО и ФИНП – наоборот. Например, 0,620 (РУСО), 0,616 (РУСП), 0,800 (ФИНО) и 0,871 (ФИНП). Поэтому по данным результатам трудно сделать какие-либо выводы.

Показатель лексического разнообразия TTR (англ. type-token ratio) в данной работе был вычислен по формуле: $TTR = \frac{V}{N}$, где V – количество уникальных лексических единиц (а именно существительных, прилагательных, глаголов и наречий), N – общее количество словоформ в тексте. Существуют и более сложные формулы вычисления данного показателя (Захарова и Савина 2020, 20–34).

Зато соотношение количества слов и предложений в корпусах показало, что в корпусе ФИНП предложения немного короче, чем в ФИНО, а в корпусах русских текстов наоборот. Как раньше уже упоминалось, количество слов в предложении корпуса РУСО составляет примерно 23,81, в РУСП – 25,04, в ФИНО – 14,39, а в ФИНП – 14,31. Разница между длинами предложений в текстах на русском, также как и в текстах на финском языке, незначительная. Разница же между длинами предложений в текстах на русском и финском языке является существенной, поэтому далее приводятся несколько примеров из текстов со словом *мочь/voida*, которое является ключевым во всех корпусах

исследования. Примеры (79) и (80) из корпуса РУСО, а примеры (81) и (82) из корпуса РУСП.

(79) Наступление летального исхода от других тяжелых ранее существовавших заболеваний, при наличии («присоединении») инфекции, вызванной SARS-CoV-2, но без таких ее клинико-морфологических проявлений, которые могли бы стать самостоятельной причиной смерти (в т. ч. при не тяжелой пневмонии и других смертельных проявлений).

(80) Кандидатами для продленной профилактики ТГВ (вплоть до 45 дней после выписки) могут быть больные пожилого возраста, больные, лечившиеся в блоке интенсивной терапии, с активным злокачественным новообразованием, ТГВ/ТЭЛА в анамнезе, сохраняющимся выраженным ограничением подвижности, концентрацией D-димера в крови > 2 раз выше верхней границы нормы.

(81) Сбор и сохранение данных может также производиться с использованием метода сквозного шифрования в форме, отвечающей современным техническим требованиям и не читаемой лицом, которому поручена обработка данных, при условии, что компетентное ведомство здравоохранения получит по требованию от лица, которому поручена обработка данных, указанные данные в форме, читаемой ведомством здравоохранения.

(82) Требуемые отступления от ограничений, изложенных в настоящем Постановлении, касательно обучения, получения образования, повышения квалификации, подготовки и проведения экзаменов, могут определяться в отношении Высшей школы полиции БаденВюртемберга, включая управление образования Высшей школы полиции БаденВюртемберга, Министерством внутренним дел, а в отношении Института юриспруденции г. Шветцинген - Министерством юстиции.

Примеры (83) и (84) из корпуса ФИНО, а примеры (85) и (86) из корпуса ФИНП.

(83) Suomesta voi matkustaa Viroon vapaa-ajan merkeissä ja ilman koronan ennakkotestiä tai Virossa karanteeniin vetäytymistä.

(84) Totsinameraania voidaan antaa yli 16-vuotiaille, Moderna ja AstraZeneca Covid-19-rokotteita yli 18-vuotiaille. Rokotteiden suojavaikutuksen kestosta ei voi tehdä vielä pitkäaikaista arviota.

(85) Näin EU voi havainnoida ja arvioida covid-19-rokotteen hyödyistä ja turvallisuudesta saatavat uudet tiedot viipymättä.

(86) EU:n talousarvio on hyväksi todettu luotettava väline, jonka avulla elvytysrahaston edellyttämät toimet voidaan toteuttaa.

Таким образом, для исследования симплификации переводов нельзя опираться на данные только одного коэффициента. Для более точного анализа необходимо учитывать разные факторы.

Для исследования экспликации мы проверили частотность нескольких лексем, которые могут выполнять разъяснительную функцию в тексте. На примере лексем (*который/joka, например/esimerkiksi, что/mikä, поэтому/joten* и *потому/siksi*) были получены данные, подтверждающие наличие экспликации в переводных текстах. Данные продемонстрированы в таблице 16 (см. ниже). В корпусах переводных текстов (РУСП и ФИНП) частотность лексем выше, чем в корпусах оригинальных текстов. Исключением является только слово *esimerkiksi* ’например’ в корпусах ФИНО и ФИНП. Тем самым

данные о частотности лексем подтверждают теорию об избыточности переводного текста.

Таблица 16. Количество эксплицирующих элементов в корпусах.

Русская лексема	**РУСО**	**РУСП**	**Финская лексема**	**ФИНО**	**ФИНП**
который	261	500	*joka*	612	738
например	46	138	*esimerkiksi*	232	189
что	25	87	*mikä*	134	190
поэтому	23	49	*joten*	32	43
потому	1	3	*siksi*	26	37

Нами были найдены также примеры интерференции в переводных тестах. Например, в корпусе РУСП встречается много заимствованных слов, которые со многих языков могут переводиться одинаково, например, *стратегия (национальная стратегия тестирования, разрабатывать стратегию вакцинации), зона (зона заражения, зона риска), изоляция (домашняя изоляция, рекомендации по изоляции), дистанция, контроль, концепция* и др. В оригинальных текстах частота данных слов ниже.

В корпусе ФИНП также были выделены заимствованные слова: *karanteeni* ’карантин’ (*omaehtoinen karanteeni, joutua karanteeniin*), *aktiivinen* ’активный’ (*aktiivinen tautipesäke, aktiivinen käyttäjä*), *positiivinen* ’позитивный’ (*positiivinen vaikutus, positiivinen mieli*), *data* ’данные’ (*avoin data, jakaa dataa ja tietoa*), *maski* ’маска’, частота которых в переводных текстах также выше, чем в оригинальных.

То, что в переводных текстах на русский язык предложения немного длиннее, чем в оригинальных текстах, тоже может указывать на влияние другого языка, т. е. интерференцию. Хотя на данное явление может влиять и экспликация. Так как переводческие универсалии не являются основной темой данного исследования, то было рассмотрено лишь несколько случаев. Однако, мы уверены, что можно найти множество примеров, которые имеют отношение к переводческим универсалиям.

Анализируя результаты сравнения корпусов, мы пришли к выводу, что тексты одного и того же стиля, затрагивающие одинаковую тему, и изначально, предназначенные выполнять одну и ту же функцию, могут отличаться друг от друга по разным параметрам. Даже в текстах на оригинальном языке и в переводах на этот язык было найдено много различий, в том числе идеологические признаки. Несмотря на то, что терминология должна быть последовательна и однозначна, а термины точно соответствовать понятию, мы нашли разницу употребления терминов (*коронавирус* и *COVID-19*) в русском и финском языках. В ходе работы выяснилось, что все тексты объединяет подъязык,

который имеет признаки ЯСЦ. В теоретической части (см. подраздел 2.1.1.) были подробно рассмотрены эти два понятия, и высказано мнение о схожести этих терминов. В результате исследования языковых особенностей текстов о коронавирусе это мнение было подтверждено.

В данном разделе были выделены общие черты и различия текстов исследования на лексическом, морфологическом и частично на синтаксическом уровнях, отдельно были рассмотрены особенности переводных текстов, также было определено, какую функцию тексты выполняют более эффективно. С помощью различных утилит программы Sketch Engine были проанализированы частоты и примеры использования терминов *covid* и *коронавирус.*

На этом этапе исследование закончено, но его можно было бы продолжить, т. к. в работе использованы далеко не все возможности корпусов и программы Sketch Engine.

4. ЗАКЛЮЧЕНИЕ

В данной главе подводятся итоги работы, а именно в общих чертах рассказывается о полученных главных результатах, о подтверждении гипотезы и об удачных и неудачных моментах исследования. В последнем разделе данной главы высказывается мнение о возможной полезности работы и о перспективах дальнейшего исследования.

4.1. О результатах исследования

В результате корпусного исследования информативных и оперативных текстов дискурса о COVID-19 нам удалось подробно проанализировать особенности подъязыка оригинальных и переводных текстах на русском и финском языках, а также функционирование этих текстов. Для работы была использована программа Sketch Engine и применён анализ ключевых слов.

На первом этапе работы были выделены частотные списки всех слов, существительных, прилагательных и глаголов и приведены 15 примеров для каждого случая. Затем были получены списки N-грамм и ключевых слов и также приведены 15 примеров с наиболее высокой абсолютной и относительной частотностью. После анализа этих списков мы убедились в правильности отбора текстов и получили общую картину дискурса COVID-19 в текстах на русском и финском языках. Затем был проведён анализ и сравнение полученных данных.

Было выяснено, что в оригинальных и переводных текстах на русском языке медицинская тематика выражена ярче, чем в оригинальных и переводных текстах на финском языке. В текстах на русском языке больше медицинских терминов, также термины *COVID-19* и *коронавирус* используется по-разному в текстах на русском и финском языках. В русском языке предпочтительнее термин *COVID-19*, а в финском – *koronavirus* ’коронавирус’. Термин *коронавирус* используется правильнее в текстах на русском языке с медицинской точки зрения, т. к. в данных текстах этот термин чаще встречается с прилагательным *новый.*

В результате анализа также выяснилось, что в текстах для обозначения человека используются разные категории предметности, например *пациент, человек, лицо, работодатель, персонал, работник*. Сравнивая эти данные, можно заключить, что в текстах РУСО более всего выражена медицинская тематика, т. к. среди высокочастотных

существительных оказалось слово *пациент*. В корпусах переводных текстов РУСП и ФИНП более нейтральная тематика, в них появляются слова *человек* и *лицо*, в текстах ФИНО акцент смещается на трудовые отношения, т. к. присутствуют слова *työnantaja* ’работодатель’ и *työntekijä* ’работник’.

В теоретической части упоминалось о том, что термины подъязык и ЯСЦ близки по значению. В процессе данного исследования мы убедились в правильности этого мнения, например при анализе N-грамм. Данные, полученные из корпуса, демонстрируют особенности подъязыка и принадлежность текстов к специальной сфере.

По частотным спискам и спискам ключевых слов, а также после анализа примеров, нам удалось определить, какую функцию тексты выполняют более эффективно. По нашему мнению, оригинальные тексты выполняют как информативную, так и оперативную функции, в переводных текстах более ярко проявляется информативная функция.

Если в теоретической части нашей работы были описаны возможности и методы корпусных исследований, то в практической части мы применили некоторые методы и сами выяснили, какие данные возможно получить в ходе работы с корпусом. Уже на первом этапе работы были получены очень важные статистические данные, которые невозможно было бы получить без помощи корпуса и инструментов работы с ним.

Наша работа продвигалась по мере выявления нами новых, интересных данных из корпусов. В последующих этапах мы анализировали уже полученные данные и продолжали искать новые. Тем самым именно данные, полученные из корпусов, двигали наше исследование вперёд. Т. е. мы не пытались искать данные, исходя из теории, мы анализировали полученные данные применяя теорию.

4.2. Вопрос и гипотеза исследования

В данной работе нам удалось выполнить поставленную перед нами задачу и ответить на интересующие нас вопросы о подъязыке. Были выделены и сравнены языковые особенности оригинальных и переводных текстов, а также рассмотрены функционирование этих особенностей в текстах о COVID-19.

В начале работы были выдвинуты две гипотезы, которые подтвердились в результате исследования. Во-первых, мы убедились, что переводные тексты могут отличаться от оригинальных не только на лексическом, но и на морфологическом и синтаксическом уровнях. Во-вторых, функционирование стиля зависит от языковых особенностей. В

данном исследовании нам удалось выполнить бонусную задачу, а именно, на практике убедится в том, что понятия подъязыка и ЯСЦ очень близки по значению.

4.3. Всё ли нам удалось?

Данная работа является для нас первым самостоятельным лингвистическим исследованием с помощью корпусов. В данном исследовании использовано только часть ресурсов корпусов и программы Sketch Engine. Из теоретической части стало известно о практически безграничных возможностях исследователей в корпусной лингвистике с помощью компьютеров. Однако, применение разных программ и инструментов требует особых навыков. Всё же нам удалось собрать необходимый для работы материал и применить необходимые инструменты для его исследования. Хотя отбор текстов вручную требует особого внимания и определённой затраты времени, создав репрезентативные и сбалансированные корпусы текстов, был получен хороший материал для исследования. Применив полученные ранее навыки работы с программой Sketch Engine, были найдены необходимые данные из текстов и применены в данном исследовании.

Мы собрали достаточно много материала для исследования, и хотелось бы предоставить читателю ещё больше примеров из текстов, но из-за ограничений по объёму работы это не удалось. Всё же, хочется надеяться, что у нас получилось передать общую мысль работы.

4.4. Перспективы и польза исследования

Данное исследование можно продолжать в разных направлениях. Например, применить новые программы для выявления и визуализации ключевых слов всего корпуса. Также можно применить разные формулы для вычисления коэффициентов, сравнение которых даст новые версии для исследования. Более глубокое изучение подъязыка на лексическом, морфологическом и синтаксическом уровнях несомненно принесёт новые результаты.

Можно более подробно продолжить исследование переводных текстов, к примеру, определить и объяснить когнитивные особенности переводческой деятельности или выявить особые свойства переводов, отличающие их от оригинальных текстов на русском и финском языках. Например, проанализировать виды информации в текстах

(когнитивная, оперативная, эмоциональная и эстетическая) и определить, как она воздействует на читателя. Данные задачи возможно решить в рамках корпусного переводоведения или дидактики перевода.

Так как пандемия продолжается, с каждым днём появляются новые термины, соответственно происходит развитие языка. Интересно было бы продолжить исследование также в диахроническом аспекте. Мы предполагаем, что результаты данной работы могут помочь будущим исследователям этой темы, например при анализе ключевых слов. Хотя пандемия продолжается по сей день, сегодня люди относятся к этому явлению иначе, чем в самом начале появления нового коронавируса. Содержание текстов меняется, т. к. некогда новые термины становятся знакомыми, появляются новые, также акцентирование тем в текстах меняется. Временной отрезок собранных нами текстов составляет полтора года. Понятно, что информативные и оперативные тексты, появившиеся в следующие полтора года, могут отличаться от наших, соответственно и языковые особенности текстов будут другими.

В данном исследовании мы обратили внимание на присутствие идеологии труда в оригинальных текстах на финском языке, интересно было бы узнать, присутствовало ли это явление в подобных текстах до пандемии и будет ли присутствовать после.

Нам хотелось бы надеяться, что пандемия закончится и останется только в истории, тогда будущий исследователь может исследовать и сравнивать свои и наши тексты, например по эмоциональной окраске.

Возможно, результаты анализа данного исследования могут пригодиться составителям и переводчикам текстов данного жанра для достижения эффективности их функционирования. Проанализировав полученные нами результаты, специалисты могут определить эффективность работы данных текстов и, опираясь на эти сведения, получить помощь в создании новых текстов. Это относится также и к терминологии, которая была использована с начала появления нового коронавируса, и к которой читатели уже привыкли. Как мы сами убедились, корпусная лингвистика открывает новые перспективы изучения языка. После овладения определённых навыков для работы с корпусами исследование ограничивается только воображением исследователя.

СПИСОК ЛИТЕРАТУРЫ

Алексеева, И. С. 2012. *Введение в переводоведение*. М.; СПб.: Академия.

Алимов, В. В. 2005. *Интерференция в переводе.* М.: Ком Книга.

Богоявленская, Ю. В. 2017. ”Сопоставительный объектно-ориентированный корпус: определение понятия и принципы формирования.” *Многоязычие в образовательн ом пространстве.* № 9: 8–19. https://f-iiyl.udsu.ru/udsu_Lingua/sbornik_lingva/nauchnyj-zhurnal-mnogoyazychie-v-obrazovatelnom-prostranstve-vyp-9-2017

Болушевская, И. Н. 2019. ”Корпусное исследование лингвистических особенностей дискурса криминальных интернет-новостей на примере новостей о похищении (на материале английского языка.” *Грамота. Филологические науки. Вопросы теории и практики.* Том 12. Выпуск 10: 184–188. http://www.gramota.net/editions/2.html

Виноградов, В. В. 1955. ”Итоги обсуждения вопросов стилистики.” *Вопросы языкознания*. № 1: 60–87. https://vja.ruslang.ru/archive/1955-1.pdf

ВОЗ – Всемирная организация здравоохранения. 2022. «Чрезвычайные ситуации». https://www.who.int/ru (Просмотрено 20.03.2022.)

Воронцова, Т. А. 2009. *Элементарная стилистика.* Ижевск. Удмуртский университет.

Гарбовский, Н. К. 2011. ”Перевод и «переводной дискурс».” *Вестник Московского университета.* Сер. 22. *Теория перевода* № 4: 3–19. http://www.vestnik-translation.ru/volumes/2011-4/

Громенко, Е. С., Павлова А. С. & Приёмышева, М. Н. 2021. ”О «ковидно-коронавирусных» процессах в русском языке.” *Studia Slavica Hung* 65 (1): 51–70. https://doi.org/10.1556/060.2020.00005

Головин, Б. Н. 1983. *Введение в языкознание*. 4-е изд., исп. и доп. М: Высшая школа.

Забросаева, И. А. & Конурбаев, М. Э. 2014. ”От LSP до специализированного дискурса: исторический срез.” В: *Язык, сознание, коммуникация*. Отв. ред. В. В. Красных, А. И. Изотов. М: Макс Пресс, 26–87. https://www.elibrary.ru/item.asp?id=23826411&pff=1

Захаров, В. П. & Хохлова, М. В. 2014. ”Автоматическое выявление терминологических словосочетаний.” *Структурная и прикладная лингвистика.* Выпуск 10. СПб: С. Петерб. ун-т. https://www.elibrary.ru/item.asp?id=26445141

Захаров, В. П. & Богданова, С. Ю. 2020. *Корпусная лингвистика.* Учебник 3. СПб: С. Петерб. ун-т.

Захарова, Е. Ю. & Савина О. Ю. 2020. ”Лексическое разнообразие текста и способы его измерения.” *Вестник Тюменского государственного университета. Гуманитарны е исследования. Humanitates.* Том 6. № 1 (21): 20–34. DOI: 10.21684/2411-197X-2020-6-1-20-34.

Интерфакс. 2022. «Пандемия коронавируса». https://www.interfax.ru (Просмотрено 20.03.2022.)

Калинина, Л. А. 2010. ”Модальные глаголы в современном русском языке.” *Вестник педагогического опыта.* № 32: 19–20. https://elibrary.ru/contents.asp?id=33729966

Камшилова, О. Н. 2017. ”Корпусное исследование иноязычной речи русских школьников: образец *vs* продукт.” *Известия Российского государственного педагогического университета им. А. И. Герцена* № 186: 32–39. https://www.herzen.spb.ru/main/nauka/1297772316/izvestia/

Кафискина, О. В. 2017. ”К вопросу о переводческой классификации типов текста.” *Преподаватель XXI век* № 4: 413–425. http://prepodavatel-xxi.ru/archive

Кожина, М. Н. 2002. «Речеведение и функциональная стилистика: вопросы теории». Избранные труды. Пермь: Перм. ун-т, ПСИ, ПССГК. https://1lib.sk/book/3242178/9d86b6

Кожина, М. Н., Дускаева, Л. Р. & Салимовский, В. А. 2011. *Стилистика русского языка.* М.: Наука.

Копотев, М. В. & Мустайоки, Арто. 2008. ”Современная корпусная русистика.” *Slavica Helsingiensia 34. Инструментарий русистики: корпусные подходы* https://www.mv.helsinki.fi/home/kopotev/Kopotev_Mustajoki_2008.pdf

Краснопеева, Е. С. 2016. «Лексические особенности русскоязычного переводного дискурса (корпусное сравнительно-сопоставительное исследование на материале современной художественной прозы)». Автореф. дис. … канд. филол. наук. Тюмень: Челяб. гос. ун-т. https://elib.utmn.ru/jspui/handle/ru-tsu/13647

Кротова, Е. Б. 2018. ”Sketch Engine для лингвистических исследований”. *Германистика сегодня*: материалы Международной научно-практической конференции, Казань, 16–17 октября 2018 г. ред.: М. А. Кулькова, 107–112. Казань: Казанский ун-т. https://dspace.kpfu.ru/xmlui/handle/net/151490

Кудашев, И. С. 2007. «Проектирование переводческих словарей специальной лексики». Монография 3, Хельсинки: Университет г. Хельсинки. https://helda.helsinki.fi/bitstream/handle/10138/19272/proektir.pdf

Куниловская, М. А. 2015. ”Переводческие универсалии как центральное направление корпусного переводоведения.” Материалы конференции «Проблемы перевода региональных авторов, в том числе представителей коренных малочисленных народов Севера Западной Сибири, на иностранные языки» 18 ноября; 18 декабря 2015 г. Ханты-Мансийск, Тюмень.

Куниловская, М. А., Ильющеня, Т. А. & Ковязина, М. А. 2017. ”Перспективы исследования лингвистических признаков русскоязычного переводного дискурса”. *Вестник ПНИПУ. Проблемы языкознания и педагогики* № 2: 17–27. http://vestnik.pstu.ru/pedag/archives/?id=&folder_id=6800

Куниловская, М. А. & Ковязина, М. А. 2018. ”Языковые открытия своими руками: корпусы в изучении языка. Рецензия на учебник Дж. Томаса «Изучаем английский язык со Sketch Engine: корпусный подход к исследованию языка»”. *Актуальные вопросы переводоведения и практики перевода* Выпуск 8: 332–338. Нижний Новгород: Альба. https://elibrary.ru/item.asp?id=35457664&pff=1

Купина, Н. А. & Матвеева, Т. В. 2013. *Стилистика современного русского языка.* М.: Юрайт.

Майорова, А. Д. 2017. ”Корпусная лингвистика: Исторический и лингводидактический аспекты”. *Педагогические науки* 5 (59): 42–46. https://research-journal.org/pedagogy/korpusnaya-lingvistika-istoricheskij-i-lingvodidakticheskij-aspekty/

Манукянц, С. В. 2021. ”Классификация лингвистических корпусов”. *Филологические науки* № 31: 70–75. https://elibrary.ru/download/elibrary_46313813_90841195.pdf

Михайлов, М. Н. 2003. «Параллельные корпуса художественных текстов». Диссертация. Тампере: Университет г. Тампере. https://urn.fi/urn:isbn:951-44-5754-4

Раздуев, А. В. 2010. ”Основные характеристики языка для специальных целей.” *Университетские чтения-2010*. Материалы научно-методических чтений ПГЛУ, 90–95. http://www2.pglu.ru/upload/iblock/13e/uch_2010_vi_00017.pdf

Раздуев, А. В. 2013. «Современный английский подъязык нанотехнологий: структурно-семантическая, когнитивно-фреймовая и лексикографическая модели». Диссертация кандидата филологических наук. Пятигорск: ПГЛУ. https://elibrary.ru/item.asp?id=21416200

Солганик, Г. Я. & Дроняева, Т. С. 2005. *Стилистика современного русского языка и культура речи.* 3-е изд. М.: Академия.

Шаров, С. А. 2003. ”Представительный корпус русского языка в контексте мирового опыта.” *Научно–техническая информация* 2 (6): 12–16. http://www.viniti.ru/products/publications/pub-1817

Шунейко, А. А. 2020. *Корпусная лингвистика: учебник для вузов.* М.: Юрайт.

Хасанова, З. С. 2013. ”Подъязык и его функционально-системные характеристики.” *Lingua mobilis* 1 (40): 64–79. https://linguamobilis.ucoz.ru/2013/2013-1-40.pdf

Хомутова, Т. Н. 2007. ”Языки для специальных целей (LSP): Вопросы теории.” *Вестник ЮУрГУ* № 15. Серия: Лингвистика № 5: 55–62. https://cyberleninka.ru/article/n/yazyk-dlya-spetsialnyh-tseley-lsp-voprosy-teorii/viewer

Хорохордина, О. В. 2013. ”Инструкция как тип текста”. *Мир русского слова. Лингвистические заметки* № 4: 7–14. http://mirs.ropryal.ru/mirs-2013-4/

Baker, Mona. 1993. “Corpus linguistics and translation studies: Implications and applications”. *Text and technology: In honour of John Sinclair*, 1993, vol. 64, pp. 232-250. https://doi.org/10.1075/z.64

Isohella, Suvi & Anita Nuopponen. 2016. ”Terminologia kohtaa käytettävyyden. Terminologisen käytettävyyden ydintä rakentamassa”. *Teksti ja tekstuaalisuus.* Hirvonen, P., D. Rellstab & N. Siponkoski (toim.), Vaasa: VAKKI Publications 7, 226–237. http://www.vakki.net/publications/no7_fin.html

Kemppanen, Hannu 2008. «Avainsanoja ja ideologiaa: käännettyjen ja ei-käännettyjen historiatekstien korpuslingvistinen analyysi». Joensuu: Joensuun yliopisto.

MachineLearning.ru. 2022. “Коэффициент корреляции Пирсона”. http://www.machinelearning.ru (Просмотрено 12.04.2022.)

Mauranen, Anna & Jarmo Jantunen. 2005. *Käännössuomeksi. Tutkimuksia suomennosten kielestä.* Tampere: Tampereen yliopistopaino.

Sketch Engine. 2022. https://www.sketchengine.eu

Thomas, James. 2016. *Discovering English with Sketch Engine: A Corpus-based approach to language exploration*. 2nd ed. Brno: Versatile.

VISK – Hakulinen, Auli, Maria Vilkuna, Riitta Korhonen, Vesa Koivisto, Tarja Riitta Heinonen ja Irja Alho. 2004. *Iso suomen kielioppi*. Helsinki: Suomalaisen Kirjallisuuden Seura. http://scripta.kotus.fi/visk (29.1.2022.)

Vartiainen, Vilma. 2018. ”Pisin sana”. *Kysyttyä* 3. https://www.kielikello.fi/-/pisin-sana (30.1.2022.)

ПРИЛОЖЕНИЕ

Список веб-сайтов, с которых найдены тексты для исследования

Всемирная организация здравоохранения (ВОЗ). www.who.int (Просмотрено 20.08.2021.)

Департамент здравоохранения города Москвы. www.mosgorzdrav.ru (Просмотрено 05.06.2021.)

Информационный центр «Корона» Израиля. www.corona.health.gov.il (Просмотрено 02.07.2021.)

Министерство здравоохранения Израиля. www.govextra.gov.il (Просмотрено 02.07.2021.)

Oma media – национальный портал Карелии. www.omamedia.ru (Просмотрено 20.08.2021.)

Официальный портал мэра и правительства Москвы. www.mos.ru (Просмотрено 05.06.2021.)

Официальный интернет-ресурс для информирования населения по вопросам коронавируса (COVID-19). www.стопкоронавирус.рф (Просмотрено 06.06.2021.)

Портал непрерывного медицинского и фармацевтического образования Минздрава России. www.edu.rosminzdrav.ru (Просмотрено 05.06.2021.)

Посольство Франции в Хельсинки. www.fi.ambafrance.org (Просмотрено 20.08.2021.)

BBC NEWS Русская служба. www.bbc.com/russian (Просмотрено 02.07.2021)

Bundesministerium für Gesundheit. www.zusammengegencorona.de (Просмотрено 02.07.2021.)

Espanja.com – Avaimet Espanjaan. www.espanja.com (Просмотрено 20.08.2021.)

Euroopan komissio. www.ec.europa.eu (Просмотрено 20.08.2021.)

Eurooppalainen rokotusjärjestelmä. www.vaccination-info.eu (Просмотрено 20.06.2021.)

Kuntaliitto. www.kuntaliitto.fi (Просмотрено 23.06.2021.)

Pirkanmaan sairaanhoitopiiri. www.tays.fi (Просмотрено 23.06.2021.)

Terveyden ja hyvinvoinnin laitos (THL). www.thl.fi (Просмотрено 23.06.2021.)

Tulli. www.tulli.fi (Просмотрено 23.06.2021.)

Työterveyslaitos. www.hyvatyo.ttl.fi (Просмотрено 23.06.2021.)

Valtioneuvosto. www.valtioneuvosto.fi (Просмотрено 23.06.2021.)

Suomenkielinen lyhennelmä

COVID-19-diskursin kielellisten piirteiden korpustutkimus venäjän- ja suomenkielisissä informatiivisissa ja operatiivisissa teksteissä sekä käännöksissä näille kielille

1 Johdanto

Tämä opinnäytetyö perustuu korpustutkimukseen, jonka aineistona on koronavirusaiheisia tekstejä venäjäksi ja suomeksi sekä käännöksiä näille kielille. Tekstikorpuksiin kuuluu sekä informatiivisia että operatiivisia tekstejä, joten työn aihe on *COVID-19-diskurssin kielellisten piirteiden korpustutkimus venäjän- ja suomenkielisissä informatiivisissa ja operatiivisissa teksteissä sekä käännöksissä näille kielille*.

1.1 Mitä ja miksi tutkitaan?

Uusien teknologioiden kehittyessä lukemisen ja tiedonvälittämisen luonne muuttuu, ja lukijan täytyy nopeasti sopeutua muuttuvaan kontekstiin sekä hahmottaa tietoa asianmukaisesti. Maailmassa tapahtuu päivittäin erilaisia onnettomuuksia, hätätilanteita ja luonnonkatastrofeja. Kaikki nämä ilmiöt vaikuttavat muun muassa ihmisten jokapäiväiseen elämään sekä koko maan talouteen ja politiikkaan. Tieto erilaisista tapahtumista leviää salamannopeasti ympäri maailmaa erilaisten viestintäverkkojen kautta. Yksilöiden ja koko yhteiskunnan käyttäytyminen määräytyy pitkälti sen mukaan, miten he arvioivat ja omaksuvat tarjotun tiedon.

Tässä opinnäytetyössä tarkastellaan tekstuaalisia tekijöitä, jotka vaikuttavat tiedon havaitsemisen ja ymmärtämisen tehokkuuteen, sekä alkuperäisen ja käännetyn tekstin kielityylin toimivuutta.

Tutkimusaineistoa yhdistää aihe, joka on ollut ajankohtainen jo kolmatta vuotta ja joka huolestuttaa koko maailmaa. Ensimmäinen virallinen raportti uudesta koronaviruksesta SARS-CoV-2, joka aiheuttaa COVID-19-nimistä tautia, julkaistiin maailman tiedotusvälineissä 31. joulukuuta 2019. Virus lähti liikkeelle Kiinan Wuhanista, josta se levisi nopeasti koko maailmaan, ja koronavirusepidemia muuttui pandemiaksi.

Pandemian aikana joissakin kaupungeissa tai jopa kokonaisissa maissa jouduttiin turvautumaan täydelliseen eristämiseen ja julistamaan maanlaajuinen karanteeni. Sinä aikana muun muassa yritysten ja laitosten toiminta keskeytettiin sekä ihmisten liikkumisvapautta maan sisällä ja sen ulkopuolella rajoitettiin tuntuvasti. Silloin kansainvälinen lentoliikenne pysähtyi, kaikki

konferenssit, konsertit ja urheilutapahtumat peruttiin, koululaiset ja opiskelijat siirtyivät etäopetukseen ja monien työelämä oli mahdollista vain netin välityksellä.

Koronavirusta yritetään torjua aktiivisesti kaikkialla maailmassa jakamalla ohjeita ja tiedotuksia toiminnasta pandemian aikana. Vaikka koronavirustartuntojen ja -kuolemien määrä on katastrofaalinen, monissa maissa vastustetaan rajoituksia ja rokotuksia.

Uutistoimisto Interfaxin mukaan 11.5.2022 mennessä koronavirukseen on kuollut maailmanlaajuisesti yli 6 miljoonaa ihmistä ja yli 518 miljoonaa ihmistä on saanut tartunnan (Interfax 2022). Rokotteiden kehittäminen antaa ihmiskunnalle toivoa tilanteen parantumisesta, koska rokotteet pelastavat vuosittain miljoonia ihmishenkiä. Maailman terveysjärjestön (WHO:n) mukaan maailmassa on joulukuuhun 2021 mennessä ollut käytössä ainakin kymmenen eri rokotetta. Kliinisiä kokeita tehdään jatkuvasti, ja kehitteillä on yli 200 rokotekandidaattia (WHO 2022).

Pandemia jatkuu edelleen, ja kaikkien maiden hallitukset tukevat voimakkaasti toimia, joilla pyritään hillitsemään koronaviruksen leviämistä. Yksi näistä tehokkaista toimista on tiedottaa ajoissa ja antaa ohjeita siitä, miten ihmisen täytyy käyttäytyä pandemian aikana.

Nykyään on helppoa saada tietoa varotoimenpiteistä, esimerkiksi miten välttää koronavirustartunta, mitä pitäisi tehdä tartunnan saatuaan, mistä saa hoitoa tai miten ja missä rokotteen voi saada. Virallista tietoa sekä äidinkielellä että millä tahansa muulla kielellä voi saada esimerkiksi WHO:n, ministeriöiden ja terveyskeskusten verkkosivuilta, hallituksen portaaleista sekä terveys- ja sosiaalialan laitoksista.

Tämän tutkimuksen merkitys ja relevanssi perustuu siihen, että aihetta käsittelevien tekstien kirjoittajat ja kääntäjät voivat hyötyä tutkimuksen tuloksista.

1.2 Tutkimusaineisto ja -menetelmät

Tutkimusta varten olen kerännyt venäjän- ja suomenkielisiä tekstejä sekä eri kielistä venäjään ja suomeen tehtyjä käännöksiä, joissa puhutaan koronaviruksesta. Kaikki tekstit ovat samaa tekstilajia, ja niiden tarkoitus on opastaa, ohjeistaa ja informoida kansalaisia pandemian aikana. Nämä ovat esimerkiksi ohjeita eri alojen ammattilaisille, tiedotteita rokotuksista, rokotuspaikoista ja tartuntaluvuista sekä muita tekstejä, jotka sisältävät muun muassa erilaisia ohjeita ja opastuksia koronaviruksesta ja jotka on julkaistu eri viranomaisten nettisivuilla. Venäjäksi tietoa koronaviruksesta löytyy muun muassa venäjänkieliseltä Stopkoronavirus.rf-tiedotussivustolta ja Venäjän terveysministeriön nettisivuilta.

Alkuperäiset suomenkieliset tekstit on kerätty viranomaisten ja laitosten nettisivuilta, kuten sosiaali- ja terveysministeriö, Terveyden ja hyvinvoinnin laitos, Tulli ja Suomen valtioneuvosto. Venäjäksi ja suomeksi käännetyt tekstit, jotka on tarkoitettu muissa maissa asuville venäjää ja suomea puhuville, on kerätty muiden maiden viranomaisten nettisivuilta.

Aineistoon kuuluu tietyn aihepiirin ja tyylilajin tekstejä, mikä edellyttää tiettyä terminologiaa ja kielityyliä. Näin työn tarkoituksena on tutkia tekstien erikoiskieltä. Tieteen termipankin määritelmän mukaan *erikoiskieli* on ”tieteen-, ammatti- tai harrastealan kielimuoto, joka eroaa yleiskielestä lähinnä sanastonsa vuoksi”. Erikoiskieli poikkeaa yleiskielestä termistön lisäksi tyylillisillä ja syntaktisilla piirteillä (Tieteen termipankki: s. v. *erikoiskieli*).

Tässä tutkimuksessa, kuten kaikessa nykytutkimuksessa, ei ole vältytty tietotekniikan käytöltä. Tutkimus perustuu korpustutkimukseen, jota varten on koottu neljä tekstikorpusta: 1) tekstit venäjän kielellä 2) käännökset venäjäksi, 3) tekstit suomen kielellä ja 4) käännökset suomeksi. Korpuksien koko on noin 100 000 sanaa. Korpuslingvistiikka, joka on yksi kielitieteen osa-alueista ja jonka välineenä ovat korpukset, syntyi tietotekniikan ja ohjelmistojen myötä. Korpuslingvistiikan merkitystä on mahdotonta epäillä, sillä siinä yhdistyvät humanistiset ja tekniset tieteet (A. D. Maiorova 2017, 42).

Tutkimukseni on korpusvetoinen (corpus-driven). Jantunen (2009) luokittelee tutkimukset aineiston käytön mukaan kolmeen luokkaan: 1) aineistoesimerkein tuettu; 2) aineistopohjainen ja 3) aineistovetoinen. Näistä ensimmäisessä aineistolla on vähemmän painoarvoa, kun taas aineistopohjaisessa tutkimuksessa aineistolla on suuri rooli, koska sitä analysoidaan hyvin laajasti. Aineistovetoisessa tutkimuksessa aineiston rooli on vielä suurempi kuin aineistopohjaisessa, sillä aineistolla on merkitystä muun muassa tutkimushypoteesin luomisessa (Jantunen 2009, 101–102, 106).

Jantusen mukaan (2012) korpusaineistot sopivat hyvin tekstilajitutkimukseen, joissa voi analysoida sanastopiirteitä ja kieliopillisia piirteitä sekä kokonaisia tekstejä. Tekstintutkimuksen näkökulmasta korpusvetoisessa lähestymistavassa voidaan hyvin käyttää apuna frekvenssisanalistoja ja avainsana-analyysia (Jantunen 2012, 360–361). Tutkin aineistoa korpusvetoisella lähestymistavalla, mikä tarkoittaa myös sitä, että sovellan teoriaan korpuksesta saatuja tietoja. Tutkimuksessani käytän kvantitatiivista analyysia, koska sitä on helppoa toteuttaa korpuksien avulla. Kvalitatiivisen tutkimuksen avulla pyrin selittämään määrällisiä havaintoja. Kvantitatiivisen analyysin toteuttamiseen käytän Sketch Engine -ohjelmaa

(Kilgarriff 2013) ja sen työkaluja, kuten Wordlist, Thesaurus, Word Sketch, N-grams, Concordance ja Keywords. Nämä työkalut sopivat hyvin avainsana-analyysin tekemiseen.

Mielestäni tämän aiheen tärkeys ja ajankohtaisuus perustuvat myös siihen, että eri alojen edustajat, myös kielitieteilijät, osallistuvat aktiivisesti koronaviruksen torjuntaan. Kielitieteilijät osallistuvat tämän aiheen tekstien ja tiedon välittämisen nopeuteen vaikuttavien kielellisten keinojen tutkimiseen. Kielitutkijat ovat myös luoneet Sketch Engineen korpuksen COVID-19:stä. Englanninkielinen korpus Covid-19 sisältää yli 280 miljoonaa sanaa (Sketch Engine 2022).

1.3 Tutkimuksen tavoitteet ja päämäärät

Tutkimustyön tavoitteena on tutkia korona-aiheisten tekstien kielellisiä piirteitä leksikaalisella, morfologisella ja syntaktisella tasolla sekä analysoida ja vertailla tekstien toimintoja verrannollisten korpusten avulla. Työssäni pyrin tutkimaan, kummassa tehtävässä tekstit toimivat paremmin, informatiivisessa vai operatiivisessa. Tarkastelen myös kielellisiä tekijöitä, jotka vaikuttavat tiedon havaitsemisen ja ymmärtämisen tehokkuuteen. Kielitieteellisessä tutkimuksessa pyrin ratkaisemaan seuraavat teoreettiset tehtävät:

- löytämään keskeisiä kielellisiä piirteitä alkuperäisistä venäjänkielisistä teksteistä ja venäjänkielisistä käännöksistä sekä vertailemaan niitä
- löytämään keskeisiä kielellisiä piirteitä alkuperäisistä suomenkielisistä teksteistä ja suomenkielisistä käännöksistä sekä vertailemaan niitä
- vertailemaan kaikkien korpuksien kielellisiä ominaisuuksia keskenään.

Tutkimustyöhön kuuluu neljä osaa: johdanto, teoria, analyysi ja päätelmät. Johdannossa kerrotaan työn aiheesta ja sen ajankohtaisuudesta, tutkimuskysymyksistä sekä tutkimusmenetelmistä. Teoriaosassa tarkastelen työn keskeisiä käsitteitä muiden tutkijoiden tutkimustöiden ja tieteellisten artikkeleiden avulla. Analyysissa kuvaan ensin tarkemmin tutkimusaineistoa ja -työvälineitä, sitten kerron yksityiskohtaisesti tehdyistä havainnoista, joita kuvaan korpus kerrallaan ja lopuksi vertailen vielä niitä keskenään. Tutkimuksen tuloksia havainnollistetaan muun muassa esimerkeillä teksteistä. Yhteensä työhön kuuluu 72 virke-esimerkkiä. Päätelmissä kerron työni hyvistä ja huonoista puolista, esittelen mahdollisia työni hyötyjä ja ehdotan tuleville tutkijoille jatkotutkimuksia.

2 Teoriatausta

Teoriaosuudessa käsittelen termiä *erikoiskieli*, koska teksteissä on paljon sanoja, jotka kuuluvat lääketieteen alaan. Samalla yritän selvittää, mikä on erikoiskielen asema kielijärjestelmässä. Ennen kuin aloitetaan kielellisten ilmiöiden tutkiminen tietyissä teksteissä, on tiedettävä, mikä on tekstien tarkoitus. Tarkastelen muun muassa informatiivisten ja operatiivisten tekstien erikoispiirteitä sekä tekstintuottajien strategioita, joita he käyttävät vaikuttaakseen tehokkaammin lukijaan. Yritän etsiä vastauksia näihin kysymyksiin T. A. Voroncovan, M. N. Kožinan, Z. S. Hasanovan, I. V. Kudaševin, T. N. Homutovan, A. V. Razdujevin ja muiden tutkijoiden töistä.

Työni tutkimusaineisto sisältää autenttisia venäjän- ja suomenkielisiä tekstejä sekä käännöksiä venäjään ja suomeen muista kielistä, joten tarkastelen käännöskielen erikoispiirteitä ja käännösuniversaaleja, kuten yksinkertaistuminen (simplification), eksplisiittistyminen (explicitation), normalisointi (normalization) ja lähdekielen interferenssi (interference). Näistä aiheista kerron tutkimuksessa käyttäen muun muassa I. S. Aleksejevan, M. A. Kunilovskajan, M. B. Grol´manin, P. B. Rybinin, A. Maurasen, J. Jantusen, M. Bakerin, V. V. Alimovin tutkimustöitä.

Tutkimustyössäni käytän vertailevia korpuksia, joten katson tarpeelliseksi esitellä termit *korpus* ja *korpuslingvistiikka*. *Korpuksella* tarkoitetaan laajaa tekstikokoelmaa, joka on edustava ja strukturoitu sekä sopii hyvin monenlaisiin kielentutkimuksiin. Korpukset sopivat hyvin sekä kvantitatiivisiin tutkimuksiin, jotka perustuvat määrällisiin tietoihin, että kvalitatiivisiin tutkimuksiin, joissa aineistoa tarkastellaan syvällisemmin (Luodonpää-Manni ym. 2020). *Korpuslingvistiikka* on tutkimusta, jossa hyödynnetään kielellisiä korpuksia (Lounela & Heikkinen 2012, 121–122). Esittelen myös muutamia esimerkkejä korpustutkimuksista ja kerron Sketch Enginestä. Tätä varten olen tutustunut V. P. Zaharovin, S. J. Bogdanovan, M. N. Mikhailovin, A. A. Šunejkon, S. A. Šarovin, A. D. Majorovan töihin ja tutkimuksiin.

Suomalaisia tutkimuksia, joissa on käytetty avainsana-analyysia ovat tehneet muun muassa Sampo Nevalainen (2003 ja 2004), joka on tarkastellut leksikaalisia piirteitä käännetyissä kaunokirjallisuusteksteissä, Hannu Kemppanen (2002, 2004, 2005, 2006 ja 2008), joka on tarkastellut ideologian kuvausta käännetyissä ja alkuperäisissä historiateksteissä, sekä Svetlana Probirskaja-Turunen (2006), joka on vertaillut käännettyjä ja alkuperäisiä juridisia tekstejä avainsana-analyysilla. Jantunen (2015) on tutkinut muun muassa oppimisympäristön vaikutusta

käyttäen avainsana-analyysia. Hän mainitsee, että nykyiset tutkimukset, joissa halutaan tutkia kielenoppimista tai oppijan kieltä, pohjautuvat korpuksiin (Jantunen 2015, 106).

3 Analyysi

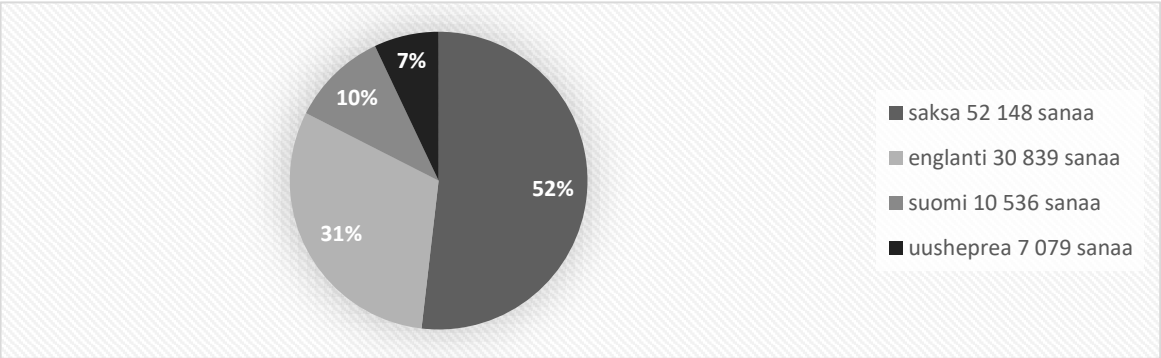
3.1 Korpukset

Olen laatinut kerätyistä teksteistä neljä vertailevaa korpusta Sketch Engine -ohjelmaan ja annoin niille omat nimet: 1) RUSO (korpus sisältää supivenäjän tekstejä); 2) RUSP (korpus sisältää käännösvenäjän tekstejä); 3) FINO (korpus sisältää supisuomen tekstejä) ja 4) FINP (korpus sisältää käännössuomen tekstejä). Kirjaimet RUS ja FIN tarkoittavat kielen nimeä (rus – venäjää, fin – suomea), O (venäjästä 'original'nyj') tarkoittaa alkuperäistä ja P (venäjästä 'perevodnoi') tarkoittaa käännettyä. Korpuksien koko ja sisältö näkyy taulukosta 1.

Taulukko 1. Tutkimuskorpuksien koko ja sisältö.

	RUSO	**RUSP**	**FINO**	**FINP**
koko/sanetta	100 007	100 187	100 089	105 074
tiedostot/kpl	10	21	51	5
virkkeet/kpl	4 200	4 001	6 953	7 341

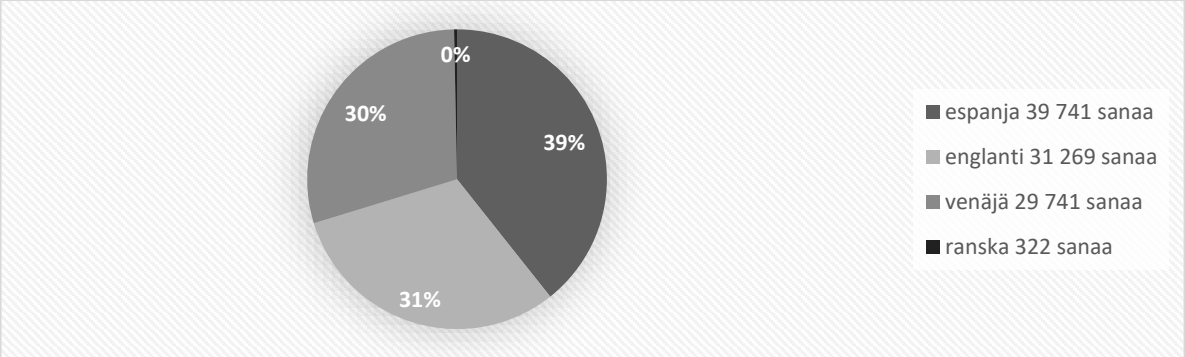
Taulukosta 1 huomaa, että tiedostojen koko on hyvin erilainen sen mukaan, paljonko tekstejä löytyi samalta nettisivustolta. Alkuperäisiä tekstejä venäjäksi löytyi muutamalta nettisivulta, käännettyjä tekstejä jouduin etsimään useammasta paikasta. Suurin osa venäjäksi käännetyistä teksteistä on käännetty saksasta ja englannista, vähemmän on suomesta ja pienin osa on uushepreasta. Eri lähdekielisten tekstien osuudet RUSP-korpuksessa on esitetty kuviossa 1.



Kuvio 1. RUSP-korpuksen lähdetekstien osuudet eri kielillä.

Alkuperäisiä suomenkielisiä tekstejä koronaviruksesta löytyi runsaasti eri lähteistä. Kun taas käännettyjä tekstejä löytyi paljon Espanjan ja Venäjän Karjalan viranomaisten nettisivuilta. Sen

voi selittää sillä, että Espanjassa ja Venäjän Karjalassa asuu paljon suomea puhuvia, ja näillä alueilla viestitään koronaviruksesta muun muassa suomen kielellä. Kuviossa 2 on esitetty, mistä kielistä suomenkieliset tekstit on käännetty ja kuinka suuret ovat lähdetekstien osuudet FINP-korpuksessa.



Kuvio 2. FINP-korpuksen lähdetekstien osuudet eri kielillä.

3.2 Analyysin havainnot

Käyttäen korpuksien tietoja (ks. Taulukko 1) voidaan laskea keskimääräinen sanojen määrä virkkeissä, jolloin RUSO-korpuksessa on noin 23,8 sanaa per virke, RUSP-korpuksessa – 25, FINO-korpuksessa – 14,4 ja FINP-korpuksessa – 14,3. Tämä kuvaa hyvin suomen ja venäjän kielen eroja, sillä esimerkiksi yksi suomen yhdyssana voi vastata venäjän kahden tai useamman sanan sanaliittoa. Myös sijamuotojen ja prepositioiden käytössä on eroja – venäjässä käytetään paljon prepositioita, kun taas suomessa sanoja useimmiten taivutetaan lisäämällä vain sijapääte.

Ensin valitsin Wordlist-toiminnolla jokaisesta korpuksesta frekvenssisanalistat lemman, substantiivin, adjektiivin ja verbin mukaan. Frekvenssisanalistat antavat yleiskuvan tekstien leksikosta ja sanojen taajuudesta. Sketch Engine laskee tutkimuskorpuksen absoluuttiset ja suhteelliset frekvenssit vertailemalla niitä referenssikorpukseen. Tutkimustyöni kohdalla ohjelma käyttää Russian Web 2011 (ruTenTen11) -referenssikorpusta, joka sisältää 14,5 mrd. sanetta ja Finnish Web 2014 (fiTenTen 14) -referenssikorpusta, joka sisältää 1,4 mrd. sanetta. Työssäni analysoin ja vertailen 15 sanaa kustakin korpuksesta.

Samat konjunktiot *ja* ja *tai* esiintyvät kaikkien korpusten frekvenssiluetteloissa, mikä viittaa samankaltaiseen lauserakenteeseen kaikissa teksteissä. Substantiivien frekvenssien perusteella voidaan päätellä, mihin tietoon tekstit viittaavat ja mihin aihepiiriin ne liittyvät. Kaikista korpuksista löytyi sanoja, jotka liittyvät lääketieteeseen, mutta korpuksissa RUSO ja RUSP

niitä oli enemmän (esim. *pacient* ('potilas'), *zabolevanie* ('tauti'), *infekciâ* ('tartunta'), *preparat* ('valmiste'), *lečenie* ('hoito'), *terapiâ* ('terapia', 'hoito'), *pnevmoniâ* ('keuhkokuume') jne.).

Esimerkeistä tulee esille myös se, että kaikki tekstit sisältävät informatiivista ja operatiivista tietoa, jossa välitetään tietoja ja kehotetaan toimimaan. Tämä näkyy muun muassa substantiiveista *issledovanie* ('tutkimus'), *slučaj* ('tapaus'), *den´*('päivä'), *čelovek* ('ihminen'), *organizaciâ* ('järjestäminen'), *risk* ('riski'), *vremâ* ('aika'), *karantin* ('karanteeni'), *testirovanie* ('testaus'), *instrukciâ* ('ohje') ja verbeistä *moč´* ('voida'), *rekomendovat´* ('suositella'), *byt´ neobhodimym* ('tarvita'), *ispol´zovat´* ('käyttää'), *izbegat´* ('välttää'), *soblûdat´* ('noudattaa') sekä suomenkielisistä sanoista: *aika, toiminta, suositus, rajoitus, prosentti, noudattaa, välttää, pitää* ja *tarvita.* Kaikki korpukset sisältävät modaaliverbejä, jotka ilmaisevat mahdollisuutta (*voida, saada*) ja välttämättömyyttä (*pitää, tarvita*).

N-grammit ovat merkkijonoja tai monisanaisia yksiköitä, jotka muodostuvat kahdesta tai useammasta sanasta. Tässä tutkimuksessa tarkastelin 3–4 sanan n-grammeja, jotka myös kuvasivat tekstien erikoiskielen piirteitä. Esimerkiksi kaikista korpuksista löytyy sanaliittoja, jotka muodostavat termejä, esimerkiksi *novaâ koronavirusnaâ infekciâ* ('uusi korovirustartunta'), *variant virusa* ('viruksen variantti'). Myös laitosten ja toimeenpanoelinten nimet, kuten *Ministerstvo zdravoohraneniâ Rossijskoi Federacii* ('Venäjän terveysministeriö'), *terveyden ja hyvinvoinnin laitos, Euroopan komissio*, ammattinimikkeet *tartuntataudeista vastaava lääkäri, pääministeri ja puheenjohtaja* ja yksiköt *v sootvetstvii s* ('jonkun mukaisesti'), *v svâzi s* ('jonkun tai jossakin yhteydessä'), *von der Leyen mukaan*, viittaavat siihen, että tekstit kuuluvat erikoiskieleen.

Sitten loin avainsanalistat KeyWord-toiminnon avulla, jossa ohjelma laskee myös taajuudet ja avainsanaindeksit (keyness score). Vertailemalla suhteellista frekvenssi-indeksiä ja avainsanaindeksiä pystyin määrittämään, mitkä sanat sisältävät enemmän tilastotietoa. Tällaisia ovat sanat *infekciâ* ('tartunta') и *pacient* ('potilas') RUSO:ssa ja RUSP:ssa, sana *suositus* FINO:ssa ja sanat *rokote* ja *tartunta* FINP:ssa. Saatujen tietojen ja muiden esimerkkien perusteella voidaan päätellä, että venäjänkielisissä teksteissä puhutaan enemmän tartunnasta, joka aiheuttaa tautia, FINO-korpuksen teksteissä säädöksistä sekä suosituksista ja FINP-korpuksen teksteissä rokottamisesta virusta vastaan.

Verbi *voida* kuuluu kaikkien korpuksien avainsanalistoihin. Tarkistaessani verbin yhdistelmiä WordSketch-toiminnolla havaitsin, että modaaliverbin *voida* yhteydessä muiden verbien valikoima on paljon laajempi (*olla*, *saada*, *käyttää*, *pitää*, *vähentää*, *hyödyntää*, *sopia* ym.)

suomen kielen korpuksissa (FINO:ssa ja FINP:ssa) kuin venäjän kielen korpuksissa. Tämä voi viitata siihen, että suomenkieliset tekstit COVID-19:stä heijastavat toiminnan mahdollisuutta elävämmin kuin venäjänkieliset tekstit.

Substantiivit *pacient* ('potilas'), *infekciâ* ('tartunta') и *slučaj* ('tapaus') ovat RUSO:n ja RUSP:n avainsanalistoissa, mikä vielä kerran viittaa siihen, että teksteissä käsitellään lääketieteellistä aihetta. Tämä näkyy myös seuraavista venäjänkielisten tekstien monista esimerkeistä (1)–(6).

(1) Вопрос о прекращении ИВЛ может быть поставлен только в условиях регресса дыхательной недостаточности пациента.

(2) При неэффективности или выраженных симптомах могут быть использованы различные растворы с антисептическим действием.

(3) Обязательно следует использовать телеметрический мониторинг ЭКГ.

(4) Не рекомендуется использовать гипотонические растворы, декстраны или растворы гидроксиэтилкрахмала.

(5) Для экспресс-диагностики могут использоваться экспресстесты по выявлению пневмококковой и легионеллезной антигенурии.

(6) Пациент имеет большие шансы на выздоровление при активной противовирусной и антицитокиновой терапии и поддерживающем лечении.

Korpuksissa RUSP ja FINO yhteisiä sanoja ovat *ihminen* ja *maa*, eli näissä teksteissä painopiste siirtyy ihmiseen.

Sama aihe, COVID-19 tai koronavirus, yhdistää tutkimukseni tekstejä, ja sen takia tarkastelin myös näiden termien käyttöä kaikissa korpuksissa. Tutkimuksessa tuli esille, että näiden termien käyttö on erilaista venäjän ja suomen kielissä. Hakutulokset näkyvät taulukosta 2.

Taulukko 2. Sanojen *COVID-19* ja *koronavirus* esiintymät korpuksien teksteissä.

	RUSO	**RUSP**		**FINO**	**FINP**
covid-19	970	645	*covid-19*	66	75
коронавирус	82	107	*koronavirus*	169	205
Yhteensä	1052	752	**Yhteensä**	235	280

Kuten taulukosta 2 huomataan, molempien hakusanojen käyttö on laajempaa venäjänkielisissä kuin suomenkielisissä teksteissä. Sana *covid-19* on suositumpi kuin *koronavirus* korpuksissa RUSO ja RUSP, kun taas suomenkielisissä teksteissä on päinvastoin.

Olen aikaisemmin huomannut, että puhutussa suomen kielessä *koronavirus* on suositumpi kuin *COVID-19*. Joskus näitä termejä käytetään toistensa vastineina, vaikka todellisuudessa ne eivät ole synonyymejä. Termiä *koronavirus* on käytetty tiedotusvälineissä vuodesta 2000 lähtien, ja se tarkoittaa viruksen lajia. Myös SARS-CoV-2 (COVID-19) on uusi koronavirustyyppi.

Termin käyttö on lisääntynyt vuoden 2020 alusta lähtien, ja sen merkitys on kaventunut: koronaviruksella tarkoitetaan nykyään itse virusta, joka aiheutti pandemian, eli termi *COVID-19* vastaa termiä *uusi koronavirus*.

Venäjänkielisissä teksteissä, varsinkin RUSO-korpuksessa, termejä *koronavirus ja covid-19* käytetään johdonmukaisemmin kuin suomenkielisissä teksteissä, koska ilmaisua *novyj koronavirus* ('uusi koronavirus') käytetään useimmiten. Esimerkiksi korpuksessa RUSO sanaliitto *novyj koronavirus* ('uusi koronavirus') esiintyy 33 kertaa mutta *koronavirus* 28 kertaa. Korpuksessa RUSP *koronavirus* esiintyy 100 kertaa ja kolme kertaa sanaliitossa *novyj koronavirus* ('uusi koronavirus'). FINO:ssa termi *koronavirus* esiintyy 169 kertaa, joista yhdeksän kertaa adjektiivin *uusi* kanssa ja FINP:ssa 205 kertaa ja näistä kaksi kertaa sanaliitossa *uusi koronavirus*. Jopa muutamasta esimerkistä huomataan, että termiä *koronavirus* käytetään yleiskielen teksteissä ja *COVID-19* erikoiskielen teksteissä. Esimerkit (7) ja (8) ovat FINO:sta ja esimerkit (9) ja (10) FINP:sta.

(7) Uusi koronavirus tarttuu ensisijaisesti pisaratartuntana, jos olet tartunnan saaneen lähellä.

(8) Ohjauksella ei syrjäytetä sitä yksilöllistä ja lääketieteellistä harkintaa joka päätöksiin sisältyy, vaan täsmennetään rajatusti sen käyttämisen yleisesti hyväksyttyjä kokemusperäisiä kriteerejä covid-19- taudin osalta ja sen luomassa poikkeuksellisessa erityistilanteessa.

(9) Käsien pesu on tunnetusti tärkeää koronaviruksen torjunnassa.

(10) Turvallisuuden lisäksi covid-19:n seuranta ja valvonta edellyttävät tehostettuja EU:n tason valvontajärjestelmiä, joihin kerätään tietoa paitsi taudin epidemiologiasta myös rokotuskattavuudesta kohderyhmien keskuudessa.

Thesaurusta voidaan käyttää myös avainsanojen poimimiseen, sillä ohjelma tunnistaa semanttisesti läheisiä sanoja. Tiedot voidaan esittää taulukkona tai visualisoida sanapilvenä, jossa sanan fontti on sitä suurempi mitä suurempi on sen tilastollinen kerroin, joka kuvastaa sen semanttista läheisyyttä avainsanaan. Etsin semanttisesti läheisiä sanoja sanalle *koronavirus*. RUSO-korpuksen sanapilvi osoitti, että koronavirus on vaarallinen ja vaatii asianmukaista hoitoa. Tämä sanapilvi sisältää paljon lääketieteellisiä termejä, minkä huomaa myös tekstien esimerkeistä (11)–(15).

(11) Список медицинских организаций, в которых можно сделать прививку, можно уточнить на сайте регионального министерства здравоохранения.

(12) Данные о странах с неблагополучной эпидситуацией размещены на сайтах ВОЗ и Роспотребнадзора.

(13) Коронавирусная инфекция – острое вирусное заболевание с преимущественным поражением верхних дыхательных путей, вызываемое РНК-геномным вирусом рода Betacoronavirus семейства Coronaviridae.

(14) Мойте руки с мылом и водой тщательно после возвращения с улицы, контактов с посторонними людьми.

(15) Появление COVID-19 поставило перед специалистами здравоохранения задачи, связанные с быстрой диагностикой и оказанием медицинской помощи больным.

RUSP-korpuksen sanapilvi näytti positiivisemmalta, sillä sieltä löytyy muun muassa ihmisen sosiaaliseen elämään kuuluvia sanoja. Sanalla *vyzdorovlenie* (’parantuminen’ tai ’toipuminen’) on korkein tilastollinen kerroin, koska tämä sana on sanapilven keskellä ja on kirjoitettu isoimmalla fontilla. Esimerkit (16)–(20) on löydetty RUSP-korpuksesta sanapilven sanojen mukaan.

(16) Пациент имеет большие шансы на выздоровление при активной противовирусной и антицитокиновой терапии и поддерживающем лечении.

(17) Более 14 процентов опрошенных женщин потеряли работу, у 41 процента было сокращено оплачиваемое рабочее время.

(18) До поездки: медицинская страховка, тест на коронавирус, приложение Corona-Warn-App Проверьте и при необходимости продлите Вашу медицинскую страховку для выезжающих за рубеж.

(19) Отмена ограничений на пересечение внешней границы в отношении Ватикана. На внешних границах движение между Финляндией и Ватиканом разрешено без ограничений через все пункты пропуска с соблюдением правил въезда в страну для иностранных граждан.

(20) Экономика стран субрегиона постепенно восстанавливалась после кризиса цен на нефть, однако по мере наступления последствий пандемии пространство для маневра правительств в бюджетно-финансовой сфере стало различаться в плане оказания чрезвычайной поддержки и принятия мер по защите домашних хозяйств и предприятий.

FINO:n sanapilvessä lääketieteen termien lisäksi on myös ihmisen sosiaalielämää kuvaavia sanoja. Esimerkit (21)–(25) ovat virkkeistä, joihin kuuluu muun muassa sanapilven sanoja.

(21) Oppilaille jaetaan uusia maskeja koulupäivän aikana.

(22) Tarvittaessa henkilökuntamme opastaa käsihygienian toteuttamisessa, suojainten käytössä ja turvavälien toteutumisen ohjaamisessa.

(23) Kyseinen muunnos on jo levinnyt useimpiin Euroopan maihin sekä esimerkiksi Yhdysvaltoihin ja leviää kotoperäisesti myös Ruotsissa

(24) Jos nuori ei ole kykenevä tai halukas päättämään rokotuksestaan itse, rokottamiseen kysytään suostumusta huoltajilta.

(25) Toisinaan huoltajat kyseenalaistivat ohjeistuksia tai rajoitustoimenpiteitä, jolloin johtajat toimivat työntekijöidensä tukena.

Analysoimalla FINP:n korpuksen sanapilveä päättelin, että tekstien tuottajat yrittävät välittää lukijalle tietoja viruksesta, rajoituksista ja säännöksistä, jotka liittyvät koronavirukseen. Tämän huomaa seuraavistakin esimerkeistä (26)–(30).

(26) SARS-CoV-2 on virus, joka aiheuttaa covid-19-taudin.

(27) Venäjän alueella purettiin koronarajoituksia ikäihmisille.

(28) Karjalassa etenee koronarokotuskampanja.

(29) Andalusian aluehallinto on päättänyt höllentää koronarajoituksia ja uudet säännökset astuvat voimaan heti torstaista alkaen.

(30) Euroopan pandemia hiljeni kesällä, kun Amerikan mantereen pandemia saavutti huippunsa. 4. maaliskuuta 2021 maailmassa oli yli 116 miljoonaa koronavirustartuntaa.

Saatujen tietojen mukaan totesin, että sekä alkuperäiset että käännetyt venäjänkieliset tekstit käsittelevät lääketieteellisiä aiheita laajemmin kuin alkuperäiset tai käännetyt suomenkieliset tekstit. Termejä *COVID-19* ja *koronavirus* käytetään eri tavoin suomeksi ja venäjäksi, ja lääketieteellisestä näkökulmasta niiden käyttö on oikeampi venäjänkielisissä teksteissä.

Tutkimustyössäni kiinnitin huomioni ihmistä tarkoittaviin substantiiveihin, joilla on korkea frekvenssi. RUSO-korpuksessa ihminen esiintyy *potilaana* (913), RUSP-korpuksessa substantiivin *ihminen* (392) lisäksi esiintyy *henkilö* (426) ja *potilas* (363), FINO:ssa ihminen on *työntekijä* (425), *lapsi* (355), *henkilö* (323) ja *työnantaja* (315), ja FINP:ssa *ihminen* (227) ja *henkilö* (213). Näistä esimerkeistä huomaa, mikä on eri persoonasubstantiivien rooli subjektikategorian semanttisessa rakenteessa. Vertailemalla näitä tietoja voidaan myös päätellä, että RUSO-teksteissä lääketieteellinen aihe on esillä voimakkaimmin, käännettyjen tekstien RUSP- ja FINP-korpuksissa aihe on neutraalimpi, ja FINO-teksteissä painopiste siirtyy työsuhteisiin.

FINO-korpus sisältää myös muita sosiaali- ja työelämän suhteisiin liittyviä sanoja, esimerkiksi *työ, suositus, kunta, ohje, julkinen, yleinen* ja *yksityinen*. Sana *työ* esiintyy itsenäisenä tai yhdyssanoissa FINO:ssa 2 246 kertaa ja 510 kertaa FINP-korpuksessa. Voidaan siis olettaa, että FINO-korpuksen tekstit pyrkivät ylläpitämään työideologiaa kansalaisissa. Tämä on vain olettamukseni, mutta sitä voidaan tutkia lisää. Esimerkiksi Hannu Kemppanen (2008) on tutkinut historiallisissa teksteissä esiintyviä ideologisia attribuutteja avainsana-analyysin avulla.

Analysoituani korpustutkimuksesta saatuja tietoja tulin siihen tulokseen, että samantyyliset, samaa aihetta käsittelevät ja alun perin samaan tehtävään tarkoitetut tekstit voivat poiketa toisistaan eri tavoin. Myös alkuperäiskielisissä teksteissä ja käännöksissä kyseiselle kielelle havaitsin monia eroja. Vaikka terminologian pitäisi olla johdonmukaista ja yksiselitteistä sekä termien pitäisi vastata täsmälleen käsitettä, löysin eroja termien (*koronavirus* ja *COVID-19*) käytössä venäjän ja suomen kielissä.

Tutkimukseni on korpusvetoinen ja työni eteni sitä mukaa, kun korpuksista löytyi uusia, mielenkiintoisia tietoja ja esimerkkejä. Toisin sanoen en yrittänyt etsiä teoriaan perustuvia tietoja vaan sovelsin teoriaa analysoimalla saatuja tietoja.

4 Loppupäätelmät

Tämä työ on ensimmäinen itsenäinen kielitieteellinen tutkimukseni, jossa käytän korpuksia. Tiedän, että käytin vain osaa korpuksien ja Sketch Engine -ohjelman resursseista. Teoreettisesta

osasta selvisi, että tietokoneiden avulla korpuslingvistiikan tutkijoilla on lähes rajattomia mahdollisuuksia. Eri ohjelmien ja työkalujen käyttö vaatii kuitenkin erityistaitoja.

Soveltamalla avainsana-analyysia korpustutkimukseen pystyin analysoimaan yksityiskohtaisesti supivenäjän ja supisuomen sekä käännöstekstien kielellisiä piirteitä. Sen jälkeen, kun olin poiminut korpuksista avainsanalistoja ja frekvenssilistoja sekä analysoinut niitä, sain yleiskuvan tutkimustekstien COVID-19-diskurssista. Vertailemalla korpuksista saatuja tietoja löysin tekstien yhtäläisyyksiä ja eroavaisuuksia. Vaikka tutkin vain lyhyesti tutkimustekstien käännösvenäjää ja käännössuomea, pystyin jonkin verran tarkastelemaan käännöskielten erikoispiirteitä.

Vastaamalla minua kiinnostaviin kysymyksiin pääsin työssäni päämäärään ja työn tavoitteisiin. Tarkastelin ja määrittelin muun muassa alkuperäisten ja käännettyjen tekstien toimintaa COVID-19-diskurssissa.

Tätä tutkimusta voitaisiin jatkaa eri suuntiin. Esimerkiksi käyttämällä uusia ohjelmia on mahdollista tunnistaa koko korpuksen avainsanoja ja visualisoida niitä. Uskon, että tekstien kielellisten piirteiden syvällisempi tutkiminen leksikaalisella, morfologisella ja syntaktisella tasolla tuo ehdottomasti uusia tuloksia.

Myös käännettyjen tekstien tutkimusta on mahdollista jatkaa yksityiskohtaisemmin, jolloin on mahdollista esimerkiksi tunnistaa ja selittää käännöstoiminnan kognitiivisia piirteitä sekä tunnistaa venäjän- ja suomenkielisten käännösten erityispiirteitä, jotka erottavat ne alkuperäisteksteistä. Voidaan esimerkiksi analysoida tekstien sisältämiä tietotyyppejä (kognitiivinen, toiminnallinen, emotionaalinen ja esteettinen) ja määritellä, miten ne vaikuttavat lukijaan. Näitä kysymyksiä voidaan tarkastella korpuskäännöstutkimuksen tai kääntämisen didaktiikan puitteissa.

Pandemian jatkuessa uusia termejä ilmaantuu joka päivä, ja kieli kehittyy sen mukaisesti. Olisi mielenkiintoista jatkaa tutkimusta myös diakronisesti. Haluaisin uskoa, että tutkimuksestani voi olla apua tuleville aiheesta kiinnostuneille tutkijoille esimerkiksi avainsanojen analysoinnissa. Vaikka pandemia jatkuu tänäkin päivänä, ihmiset suhtautuvat ilmiöön nyt eri tavalla kuin koronaviruksen alkuaikoina. Näin ollen tekstien sisältö muuttuu, uusia termejä ilmaantuu, ja myös tekstien aiheiden painotus voi muuttua.

Tässä tutkimuksessa kiinnitin huomiota työideologian läsnäoloon alkuperäisissä suomenkielisissä teksteissä. Olisi mielenkiintoista tietää, oliko tämä ilmiö läsnä tämäntyyppisissä teksteissä ennen pandemiaa ja onko se läsnä sen jälkeen. Toivottavasti

pandemia pian loppuu ja jää vain historiaan, jolloin tulevaisuudessa voitaisiin tutkia ja vertailla koronavirusaiheisia tekstejä esimerkiksi tunnevärityksen osalta.

Ehkäpä tutkimukseni analyysin tulokset, jotka kertovat tekstien tehokkuudesta, voivat olla hyödyksi tämän tyylin tekstien laatijoille ja kääntäjille. Tämä koskee myös termejä, joita on käytetty pandemian alusta lähtien.

Kuten olen itsekin huomannut, korpuslingvistiikka avaa uusia näkökulmia kielentutkimukseen. Kun korpuksien tietyt käsittelytaidot ovat hallussa, tutkimusta rajoittaa vain tutkijan mielikuvitus.

Lähteet

Jantunen, Jarmo Harri 2009. Ei pelkästään mielikuvituksen puutteen vuoksi – kieliaineistojen systemaattinen käyttö kielentutkimuksessa. *Virittäjä* 1, 101–113. https://journal.fi/virittaja/issue/view/672

Jantunen, Jarmo Harri 2012. Korpusvetoinen tekstilajianalyysi: sanalistat ja genreavainsanat. Teoksessa Vesa Heikkinen, Eero Voutilainen, Petri Lauerma, Ulla Tiililä & Mikko Lounela (toim.)., *Genreanalyysi – tekstilajitutkimuksen käsikirja.* Helsinki: Gaudeamus Oy, 360–371.

Jantunen, Jarmo Harri 2015. Oppimiskontekstin vaikutus oppijan pragmatiikkaan: astemääritteet leksikaalisina nallekarhuina. *Lähivõrdlusi. Lähivertailuja*, 25, 105–136. http://dx.doi.org/10.5128/LV25.05

Kemppanen, Hannu 2002. Ystävyyden tuontia venäjästä suomeen: avainsanat ideologian toteutumina käännetyissä ja alkuperäisissä historiateksteissä. Teoksessa Anna Mauranen & Liisa Tiittula (toim.), *Kieli yhteiskunnassa – yhteiskunta kielessä.* AFinLan vuosikirja n:o 60. Jyväskylä: Suomen soveltavan kielitieteen yhdistys AFinLa. https://journal.fi/afinlavk/issue/view/4205v

Kemppanen, Hannu 2004. Keywords and ideology in translated history texts: A corpusbased analysis. *Across Languages and Cultures* 5: 1, 89–106. https://doi-org.libproxy.tuni.fi/10.1556/Acr.5.2004.1.5

Kemppanen, Hannu 2005. Avainsanoja, ideologiaa ja lojaaliutta: venäjästä suomennettujen historiatekstien korpusanalyysi. Teoksessa Anna Mauranen & Jarmo Hannu Jantunen (toim.), *Käännössuomeksi. Tutkimuksia suomennosten kielestä*. Tampere Studies in Language, Translation and Culture 1. Tampere: Tampereen yliopisto, 185–208.

Kemppanen, Hannu 2008. Keywords revisited: lähdetekstin korpuspohjainen analyysi kääntäjän apuvälineenä. Kalle Konttinen, Minna Kumpulainen & Tuija Luokkakallio (toim.), *MikaEL Kääntämisen ja tulkkauksen symposiumin verkkojulkaisu* Vol.2. Turku: Näkymä kääntäjien koulutukseen. https://www.sktl.fi/liitto/seminaarit/mikael-verkkojulkaisu/arkisto-archive/vol-2-2008/

Lounela, Mikko & Vesa Heikkinen 2012. Korpusvetoinen tekstilajianalyysi: sanalistat ja genreavainsanat. Teoksessa Vesa Heikkinen, Eero Voutilainen, Petri Lauerma, Ulla Tiililä & Mikko Lounela (toim.), *Genreanalyysi – tekstilajitutkimuksen käsikirja.* Helsinki: Gaudeamus Oy, 120–121.

Luodonpää-Manni, Milla & Krista Ojutkangas 2020. Laadullinen aineistopohjainen kielentutkimus. Teoksessa Reetta Konstenius, Milla Luodonpää-Manni, Urpo Nikanne, Kaius Sinnemäki, Matti Miestamo & Markus Hamunen (toim.), *Kielitutkimuksen menetelmiä I–IV*. Helsinki: SKS. https://doi.org/10.21435/skst.1457

Nevalainen, Sampo 2003. Käännöskirjallisuuden puhekielisyyksistä – kaksinkertaista illuusiota? *Virittäjä* 107: 1, 2–26. https://journal.fi/virittaja/article/view/40235

Nevalainen, Sampo 2004. Colloquialisms in translated text. Double illusion? *Across Languages and Cultures* 5: 1, 67–88. https://doi-org.libproxy.tuni.fi/10.1556/Acr.5.2004.1.4

Probirskaja-Turunen, Svetlana 2005. *Korpusvetoinen juridinen kääntäminen.* Tampereen yliopisto, lisensiaatintutkimus. https://urn.fi/urn:nbn:fi:uta-1-15253

Tieteen termipankki: hakusana *erikoiskieli*. 2022. https://tieteentermipankki.fi/wiki/Terminologiaoppi:erikoiskieli (1.4.2022)

Sketch Engine 2022. https://app.sketchengine.eu/ (4.5.2022)