



Информация для цитирования:

Алюнина Ю. М. Методические основы анализа киноречи на материале корпусов: технический, социальный и культурно-национальный аспекты / Ю. М. Алюнина // Научный диалог. — 2024. — Т. 13. — № 2. — С. 199—225. — DOI: 10.24224/2227-1295-2024-13-2-199-225.

Alyunina, Yu. M. (2024). Methodological Foundations of Film Speech Analysis Using Corpora: Technical, Social, and Cultural-National Aspects. *Nauchnyi dialog*, 13 (2): 199-225. DOI: 10.24224/2227-1295-2024-13-2-199-225. (In Russ.).



Перечень рецензируемых изданий ВАК при Минобрнауки РФ

Методические основы анализа киноречи на материале корпусов: технический, социальный и культурно-национальный аспекты

Алюнина Юлия Матвеевна

orcid.org/0000-0002-4970-5584

Scopus Author ID: 57217070423

кандидат филологических наук,

Doctorat en lexicologie et terminologie multilingues, traduction (кандидат наук в области лексикологии, многоязычной терминологии и перевода), ассистент кафедры иностранных языков филологического факультета
aliunina-yum@rudn.ru

Российский университет дружбы народов им. Патриса Лумумбы (Москва, Россия)

Благодарности:

Исследование выполнено при финансовой поддержке Российского научного фонда, проект № 23-78-01208

Methodological Foundations of Film Speech Analysis Using Corpora: Technical, Social, and Cultural-National Aspects

Yulia M. Alyunina

orcid.org/0000-0002-4970-5584

Scopus Author ID: 57217070423

PhD in Philology,

Ph.D. in Lexicology and Multilingual Terminology and Translation,

Assistant Professor, Department of Foreign Languages, Faculty of Philology
aliunina-yum@rudn.ru

Peoples' Friendship University of Russia named after Patrice Lumumba (Moscow, Russia)

Acknowledgments:

The study is supported by Russian Science Foundation, project number 23-78-01208

ОРИГИНАЛЬНЫЕ СТАТЬИ

Аннотация:

Цель исследования — описать возможности и ограничения корпусного подхода к анализу киносубтитров с учётом их технических, социальных и культурно-национальных особенностей. Предлагается обзор существующих корпусов киноконента в сопровождении их критического осмысления, во-первых, с точки зрения качества субтитров как эмпирических данных для лингвистического исследования, во-вторых, с точки зрения качества корпусов как среды для выполнения количественного анализа эмпирического материала. Далее с опорой на сложившуюся практику субтитрования предлагается авторская точка зрения на субтитры как материал для изучения киноречи с учётом технических (длина строки, подгонка под размер экрана), социальных (субтитры для глухих и слабослышащих) и культурно-национальных (субтитровочные страны VS страны дубляжа) особенностей субтитрования на разных языках и в разных странах. В ходе работы было выявлено, что наибольшее количество субтитров, доступных в открытых банках, являются переводными эквивалентами звучащей речи. Отмечается, что субтитрованию отдаётся предпочтение в США, Великобритании, Индии, Китае и Японии. Охарактеризованы особенности подготовки субтитров в таких странах, как Сербия, Финляндия, Россия. Сделан вывод о значимости названных особенностей для качества лингвистического исследования на базе переводной киноречи, представленной субтитрами.

Ключевые слова:

дальнее чтение; киноречь; классификация субтитров; скрипт; субтитры.

ORIGINAL ARTICLES

Abstract:

The aim of this study is to describe the possibilities and limitations of the corpus approach to the analysis of film subtitles, taking into account their technical, social, and cultural-national features. It offers a review of existing film content corpora accompanied by their critical interpretation, firstly from the perspective of subtitle quality as empirical data for linguistic research, and secondly from the perspective of corpus quality as an environment for quantitative analysis of empirical material. Furthermore, based on established subtitling practices, the author proposes her own viewpoint on subtitles as material for studying film speech, considering technical (line length, screen size adaptation), social (subtitles for the deaf and hard of hearing), and cultural-national (subtitling countries vs dubbing countries) aspects of subtitling in different languages and countries. In the course of the work, it was found that the largest number of subtitles available in open repositories are translated equivalents of cinema speech. It is noted that subtitling is preferred in the USA, UK, India, China, and Japan. The features of subtitling in countries such as Serbia, Finland, and Russia are described. The author concludes that these features are important for the quality of linguistic research on the basis of translated movie speech represented by subtitles.

Key words:

distant reading; film speech; subtitle classification; script; subtitles.



УДК 81'42:791.44.024

DOI: 10.24224/2227-1295-2024-13-2-199-225

Научная специальность ВАК
5.9.8. Теоретическая, прикладная и
сравнительно-сопоставительная лингвистика
5.9.9. Медиакоммуникации и журналистика

Методические основы анализа киноречи на материале корпусов: технический, социальный и культурно-национальный аспекты

© Алюнина Ю. М., 2024

1. Введение = Introduction

В 2006 году Клайв Хамби, британский математик, сказал, что «данные — это новая нефть» [Smolan et al., 2006, p. 11], имея в виду, что они похожи на сырьё и непригодны для использования в необработанном виде. Эта точка зрения получила широкое распространение в индустрии и разных направлениях науки, о чём свидетельствует простой поиск в браузере и количество обнаруженных результатов по данному запросу. С одной стороны, разнообразие данных, которые сегодня находятся в открытом доступе, могут стать богатой эмпирической базой исследования. С другой, увеличение количества данных непременно ставит вопрос об их качестве. Что (ещё) может стать источником данных? Каким данным можно доверять? Каковы критерии выборки данных? В настоящей статье речь пойдёт о языковых данных аудиовизуального контента, которые являются основой для лингвистического исследования.

Развитие производства аудиовизуального контента способствовало появлению богатой эмпирической базы для лингвистических исследований, на материале которой можно решать новые задачи или по-новому подходить к решению традиционных исследовательских вопросов. Как известно, аудиовизуальные произведения (художественные и документальные фильмы, сериалы, телевизионные передачи и шоу, видеоигры, театральные постановки, рекламные ролики и др. [Асташина и др., 2021, С. 5; Cintas et al., 2007, p. 12]) выступают богатым источником лингвокультурного материала, который оказывается в фокусе многих наук, включая лингвистику в различных её направлениях: дискурс-анализ [Сухая, 2010; Уртминцева и др., 2022; Kozloff, 2000], лингвокультурология [Анисимов и др., 2022], социолингвистика [Bednarek, 2018, p. 3; Bednarek, 2023, p. 1, 18—21], аудиовизуальный перевод [Малёнова, 2017; Малёнова, 2021; Cintas et al., 2007; Lukić, 2016;



Is Subtitling Equally ..., 2016], локализация (культурная адаптация контента к принимающей аудитории) [Анисимов и др., 2019; Малёнова, 2021] и др.

В лингвистическом анализе аудиовизуальных произведений сегодня преобладает качественная методология, но уже прослеживается внедрение количественных методов к поиску и систематизации эмпирического материала, а также сочетание этих подходов [Bednarek, 2023, p. 3]. Под качественной методологией в данном случае понимается так называемый пристальный или медленный анализ аудиовизуального контента (по аналогии с пристальным чтением в литературоведении — от англ. *close reading* [Махов, 2018, с. 32; Моретти, 2016, с. 11]), а под количественным — дальний анализ (по аналогии с дальним чтением — от англ. *distant reading* [Там же]). Благодаря возможностям компьютерного, в частности, корпусного анализа текста [Там же, с. 12, 329] в литературоведении дальнейшее чтение позволяет проследить «не конкретные черты стиля того или иного автора, а некие абстрактные закономерности, характеризующие сразу многие тексты» [Моретти, 2016, С. 11]. Такое противопоставление справедливо и для анализа аудиовизуального контента: фильмов и сериалов, видеоигр, пьес.

При медленном, пристальном анализе аудиовизуального контента тщательно изучаются отдельные произведения и решаются локальные исследовательские задачи: описание полного спектра визуальных и аудиальных средств, которые используются, например, для аутентичной «репрезентации тувинской этнокультуры» в одном фильме — «Планета Тыва» [Анисимов и др., 2022, С. 183].

Дальний анализ, напротив, предполагает исследование большого массива аудиовизуальных произведений, который может быть выполнен на базе корпусов. При изучении вербальной составляющей кино корпусный подход позволяет выполнять масштабный, а не фрагментарный анализ аудиовизуального контента [Bednarek, 2018, p. 4] через обращение к речи персонажей посредством субтитров («subtitles»), скриптов («dialogue scripts») или расшифровок киноречи («on-screen dialogue») [Bednarek, 2023, p. 2]. Корпусный подход позволяет обеспечить объективность и надёжность получаемых результатов с количественной точки зрения.

Однако при обращении к корпусу необходимо учитывать не только преимущество массового охвата языкового материала, но и качество исходных данных, особенно если это касается субтитров к фильмам и сериалам, доступ к которым получить проще всего (например, через открытые банки субтитров, как *OpenSubtitles* или *YIFY Subtitles*). Важно принимать во внимание то, что субтитры могут различаться по своим техническим, социальным и культурно-национальным характеристикам [Perego et al., 2016, p. 225].

Технические характеристики включают, например, длину строки, обусловленную размером экрана устройства, под которое создаётся субтитр [Асташина и др., 2021, с. 18; Сухая, 2010, с. 73; Kozloff, 2000, p. 26; FSS; SDFT; STTSG; KTTSG]. Отдельные требования предъявляются к созданию субтитров для компьютерных игр [NGIR] и надтитров («supertitles», «supratitles») для театральных постановок и оперных представлений [Малёнова, 2017, с. 42; Cintas et al., 2007, p. 25].

К **социальным ограничениям** можно отнести субтитрование для глухих и слабослышащих, регламентируемое национальными стандартами и нормативными документами, такими как ГОСТ Р 57763-2017 и «Требования к субтитрованию и тифлокомментированию полнометражных национальных фильмов, создаваемых в художественной или анимационной форме» в России [ГОСТ; ТСТ] или «Twenty-First Century Communications and Video Accessibility Act of 2010» в США («Закон о доступности телекоммуникаций и видеоконтента XXI века», здесь и далее перевод наш. — Ю. М.) [CVAA].

Культурно-национальная специфика субтитрования в определённой мере сочетает в себе технические и социальные особенности. Создание субтитров подчиняется не только стандартам субтитрования для внутреннего рынка, но и исторически сложившимся традициям к кинопереводу. Так, с точки зрения перевода полнометражного кино, Сербия — это субтитровочная страна («subtitling country») [Lukić, 2016, p. 64], где фильмы в кинотеатре идут преимущественно на языке оригинала с сербскими субтитрами. Россия и Германия, напротив, являются странами дубляжа [Малёнова, 2021, с. 64] [Statista].

Итак, с одной стороны, текст субтитров зависит от разных языковых и внеязыковых факторов, которые влияют на степень точности передачи исходного аудиовизуального сообщения. С другой стороны, субтитры являются ценным источником эмпирических данных и позволяют с помощью корпусных методов проводить масштабный, а не единичный поиск релевантной для исследования информации, не прибегая к пристальному анализу аудиовизуального произведения.

На основе вышеизложенных фактов о потенциале субтитров как исходных данных для лингвистического исследования аудиовизуального произведения в настоящей статье ставится **цель** описать возможности и ограничения корпусного подхода к анализу киносубтитров с учётом их технических, социальных и культурно-национальных особенностей. В связи с тем, что предлагаемая статья носит методический и аналитический характер, в её **задачи** входит, во-первых, определить ключевые понятия исследования (киноречь, субтитры, скрипты, расшифровки); во-вторых, описать



достоинства и недостатки существующих корпусов аудиовизуального контента как основы для лингвистического исследования; в-третьих, проанализировать языковые особенности создания субтитров на разных языках и описать специфику субтитровочного материала как эмпирической базы для лингвистического анализа аудиовизуального контента, который в этой работе ограничен сериалами и фильмами.

2. Материал, методы, обзор = Material, Methods, Review

Для изучения существующих корпусов аудиовизуального контента необходимо определить, что мы будем понимать под аудиовизуальными произведениями, а также на каком материале для них создаются корпуса — субтитры, скрипты, расшифровки. Кроме того, уточнения требует соотношение названных понятий с понятием, обозначенным словом *киноречь*.

В России термин *аудиовизуальное произведение* закреплён законодательно. Под ним понимается произведение, которое состоит «из зафиксированной серии связанных между собой изображений (с сопровождением или без сопровождения звуком) и предназначено для зрительного и слухового (в случае сопровождения звуком) восприятия с помощью соответствующих технических устройств» [ГК РФ 1263]. К аудиовизуальным произведениям относятся кинематографические работы, «а также все произведения, выраженные средствами, аналогичными кинематографическим (теле- и видеофильмы и другие подобные произведения), независимо от способа их первоначальной или последующей фиксации» [ГК РФ 1263].

Согласно этому определению, к аудиовизуальному контенту могут быть отнесены игровые и документальные фильмы и сериалы, компьютерные игры, а также театральные и оперные представления, если они зафиксированы как видеоматериал средствами кинематографии. Однако с точки зрения аудиовизуального перевода спектакли и оперные постановки в их первоначальном представлении тоже относятся к аудиовизуальному контенту [Асташина и др., 2021, с. 18; Малёнова, 2017, с. 33], поскольку их наполнение включает как аудиальные элементы (речь, музыкальное сопровождение, песни и др.), так и визуальные (декорации, костюмы, сценические эффекты и др.). Кроме того, современные театральные представления всё чаще сопровождаются надтитрами и субтитрами для глухих и слабослышащих, что вписывается в концепцию цифровой или медиадоступности [Асташина и др., 2021, с. 23; Малёнова, 2017, с. 42].

Таким образом, понятие аудиовизуального произведения выступает гиперонимическим по отношению ко всем вышеназванным типам контента, вербальная часть которого фиксируется и передаётся в виде субтитров, скриптов, надтитров, реплик и в некоторых случаях киноречи.

Понятие киноречи по определению применимо исключительно к кино. *Киноречью* (англ. *film speech, filmic speech, cinematic speech* в работе [Kozloff, 2000]) называются все озвученные в фильме реплики. Это звучащая речь, которая включает только устные, вербальные способы передачи информации:

— реплики персонажей (все диалоги, монологи и внутренние монологи);

— гуры, или фоновые реплики, которые могут принадлежать группе говорящих: *Ура! Бежим!* [Асташина и др., 2021, С. 10];

— повествование, которым может быть закадровая речь рассказчика в игровом кино (речь К. Хабенского во франшизе «Ёлки», 2010—2021); закадровая речь повествователя в документальном фильме (закадровый голос в киноленте о биографии М. Горького [Уртминцева и др., 2022]); речь рассказчика в кадре и за кадром (повествователь в сериале «A Series of Unfortunate Events» (2017—2019), который ведёт историю главных героев и периодически появляется на экране с обращением к зрителю или как персонаж сериала, взаимодействующий с другими его героями);

— песни, исполняемые персонажами, как часть киноповествования в мюзиклах («Мэри Поппинс, до свидания», 1983; «Les Misérables», 2012; «Frozen», 2013; «La-la Land», 2016; «Aladdin», 2019); песни, значимые для сюжета в картинах других жанров («Я спросил у ясеня» в комедии «Ирония судьбы, или с легким паром!», 1973).

Понятие киноречи можно вписать в диегетическую концепцию К. Меца, который, вслед за Э. Сурио, под диегезисом понимает совокупность всех означающих элементов в фильме («the sum of a film's denotation» [Tröhler et al., 2018, p. 132]). Киноречь является одним из таких элементов, которая в терминах данной концепции может быть диегетической («diegetic» [Tröhler, Kirsten, 2018, С. 113]) и недиегетической («non-diegetic» [Ibid., 2018, p. 153]). В первом случае речь «слышима» или потенциально «слышима» персонажами киномира. Во втором — речь адресована только зрителю, но внутри экранного мира она не звучит и не воспринимается персонажами.

К киноречи не относятся незвученные тексты, попадающие в кадр: надписи на вывесках и зданиях, заголовки книг и газет, сообщения на экранах электронных устройств и т. п.; а также саундтреки, звучащие в фильме в фоновом режиме и задающие ритм (например, песня «Suddenly I See» в комедии «The Devil Wears Prada», 2006). Первые являются аудиовизуальным контентом и, соответственно, объектом аудиовизуального перевода, но не киноречью. Последние относятся к недиегетическим звукам в фильме, но речью не являются.

Для выполнения лингвистического исследования на материале киноречи необходимо обращение к языковому материалу в письменном виде, который первичен для любого аудиовизуального произведения. Театральные постановки разыгрываются на сцене по репликам в пьесе. Актёры кино говорят репликами своих персонажей, которые прописаны в сценариях, а актёры озвучивания и дубляжа озвучивают фильмы, сериалы и мультфильмы по диалоговым листам (речь персонажей с тайм-кодами, которая может содержать объяснение некоторых терминов, имён собственных, неологизмов и окказионализмов, использующихся в кинокартине [Cintas et al., 2007, С. 74]). Перевод аудиовизуального контента может осуществляться со скриптов, диалоговых листов или субтитров на языке оригинала.

Скриптами называются «прописанные по ролям реплики персонажей в кадре и титры» [Асташина и др., 2021, С. 17]. Помимо реплик и титров, скрипты могут содержать информацию о действиях и ощущениях персонажей и о сцене. Например, во фрагменте скрипта к мультфильму «Frozen» (2013) ниже (рис. 1) присутствуют **помета** CUT TO: EXT. THE FJORD FOREST — DAY — СМЕНА КАДРА: НА УЛИЦЕ. ЛЕСА ФЬОРДА — ДЕНЬ; **помета** *shivering* (*дрожит, дрожащим голосом*); и **строка** *Anna rides her horse through two feet of snow. She shivers* (*Анна скачет на лошади по глубокому снегу. Она дрожит*) [Frozen].

ELSA (CONT'D)
THAT PERFECT GIRL IS GONE.
HERE I STAND IN THE LIGHT OF DAY.
LET THE STORM RAGE ON!!
THE COLD NEVER BOTHERED ME ANYWAY.

She turns and slams her ice palace door on us.

CUT TO:

EXT. THE FJORD FOREST -- DAY

Anna rides her horse through two feet of snow. She shivers.

ANNA
(shivering)
Elsa! Elsa! It's me, Anna...your
sister who didn't mean to make you
freeze the summer. I'm sorry. It's
all my f-f-f-f-f-f-fault.

Рис. 1. Скриншот фрагмента скрипта к мультфильму «Frozen», 2013

Подобная дополнительная информация является вспомогательной для режиссёра и актёра, но киноречью не является. К киноречи, диегетической и недиегетической, относятся только реплики в скриптах, поскольку они подлежат озвучиванию и субтитрованию. На скриншоте выше это песня Эль-

зы (выделено жирным шрифтом в скрипте) и реплика Анны. Если перевод выполняется со скриптов, он оформляется в виде диалогового листа (перевод под за кадр или дубляж) или субтитров (перевод под субтитрование).

Субтитры представляют собой «текстовое сопровождение фильма» [Сухая, 2010, С. 73] на языке оригинала или перевода, которое дублирует или дополняет звуковую дорожку. Они появляются на экране одновременно с соответствующим речевым событием. Помимо текстового представления речи персонажей, в кино субтитры могут содержать значимые для понимания сюжета надписи, которые попадают в кадр (вывески, тексты писем, сообщений на экране телефона и др.), а также описание шумо-музыкального ряда [Асташина и др., 2021, с. 17; Уртминцева и др., 2022, с. 340].

Поскольку присутствие субтитров на экране должно быть синхронно соответствующему речевому событию, субтитры не могут передавать речь персонажей дословно. Они передают «лишь основную мысль одной или нескольких реплик» [Сухая, 2010, с. 73]. При этом за субтитровщиком остаётся решение, какая информация должна быть отображена в субтитрах [Kozloff, 2000, С. 26] или какими языковыми единицами должна быть передана основная идея речевого сообщения. В субтитрах звучащая речь персонажа подвергается упрощению и перефразированию, чтобы её можно было вписать в тайм-код — «числовое обозначение начала и конца субтитра относительно видеоряда» [Асташина и др., 2021, с. 18]. Поэтому речь озвученная и речь в субтитрах не совпадают, если последние создаются по всем правилам субтитрования. Условное совпадение возможно, только когда субтитры создаются автоматически, как в *YouTube*, и не подвергаются ручному редактированию. Такое совпадение называется нами условным, так как сгенерированным субтитрам присущи речевые ошибки, обусловленные несовершенством технологий распознавания устной речи и её машинного стенографирования.

Наиболее близкой к звучащей киноречи формой её письменной фиксации являются **расшифровки к фильмам**. Это, условно говоря, тексты, записанные под диктовку актёров и / или закадрового повествователя. Они соответствуют диететической и недиеетической речи в киноленте; например, созданные вручную расшифровки к фильму о биографии М. Горького, составленные для анализа закадрового голоса в документальном кинодискурсе [Уртминцева и др., 2022, с. 341]. Подготовка такого языкового материала является наиболее трудозатратной для исследователя, но этот недостаток компенсируется надёжностью получаемой эмпирической базы.

Киноречь, имеющая любую форму письменной фиксации (субтитры, скрипты, расшифровки к фильмам), может стать материалом для лингвистического исследования и текстовой базой корпуса.

3. Результаты и обсуждение = Results and Discussion

3.1. Корпусы киноречи

Современная корпусная лингвистика предоставляет широкий спектр возможностей для анализа киноконтента на основе аннотированных корпусов. Для примера можно назвать **(1)** *The Movies Corpus* (199 479 302 слов¹) [TMC], созданный по материалам англоязычных субтитров из коллекции *OpenSubtitles* [OPUS] и соотнесённый с метаданными из базы *Internet Movie Database* [IMDb]; **(2)** *Film Corpus 2.0* [FC], созданный на основе скриптов из *The Internet Movie Script Database* [IMSDb] и соотнесённый с метаинформацией из IMDb; **(3)** многоязычная коллекция корпусов на материале субтитров и один корпус на основе скриптов в *Sketch Engine* [SE]. Рассмотрим подробнее контент, предлагаемый последним ресурсом.

Наиболее крупная коллекция корпусов киноконтента представлена в менеджере *Sketch Engine*. Он содержит корпуса к фильмам и сериалам на 58 языках (английском, армянском, греческом, корейском, русском, сербском, турецком, фарси, финском, чешском, эстонском и др.). Эти корпуса созданы **на материалах** открытой **базы** параллельных **субтитров** *OpenSubtitles* [OS PC], поэтому название каждого корпуса включает название этой базы и соответствующий язык: *OpenSubtitles 2018 parallel — Afrikaans*² (417 700 слов — 341 349 токенов), *OpenSubtitles 2018 parallel — English* (1 594 437 718 токенов — 1 211 666 401 слово), *OpenSubtitles 2018 parallel — Korean* (9 250 140 токенов — 7 432 927 слов) и т. д.

В *Sketch Engine* все корпуса представляют собой своеобразное семейство корпусов: выбирая в качестве фокусного какой-то один корпус, пользователь автоматически выбирает все связанные с ним параллельные корпуса на разных языках. Так, описание к корпусу англоязычных субтитров гласит следующее: «*English OpenSubtitles 2018* is a corpus from the collection of 60 sentence-aligned parallel corpora made up of movie subtitles available at www.OpenSubtitles.org) («*English OpenSubtitles 2018* — это корпус субтитров с сайта www.OpenSubtitles.org, который включает 60 параллельных корпусов, каждое предложение которых выровнено с английским») [SE]. Ниже приведён пример результата поиска в этом корпусе в языковой паре «русский — английский» по запросу *fatherland* (рис. 2).

В первом столбце (рис. 2) указаны названия фильмов и годы их выпуска. Во втором — контекст, на языке которого выполняется поиск. В третьем — контекст на языке, выбранном в качестве параллельного при настройке фильтров поиска. Контексты с субтитрами во втором столбце

1 Разработчики не предоставляют данных о количестве токенов.

2 Год 2018 в названии корпуса означает, что в него входят субтитры к фильмам и сериалам, которые были сняты и выпущены в прокат до 2018 года включительно.

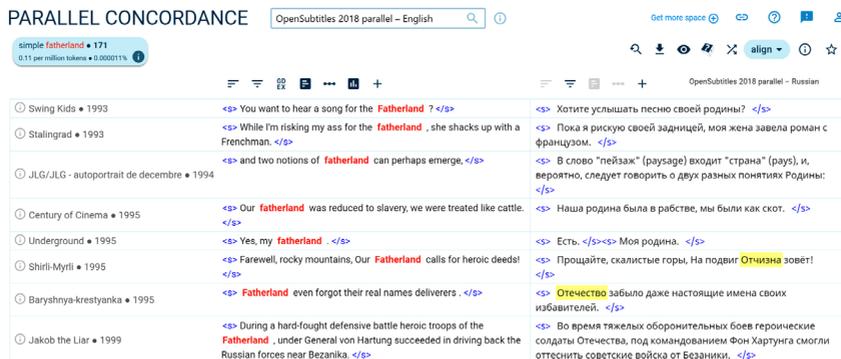


Рис. 2. Скриншот с результатом поиска по запросу *fatherland* в *OpenSubtitles 2018 parallel — English*

не всегда соответствуют языку оригинала киноматериала, а во втором — языку его перевода. Как видим на скриншоте, в базе обнаружены контексты из субтитров к фильмам, которые в оригинале были сняты не только на английском языке, но ещё на русском («Stalingrad» — «Сталинград», 1993; «Shirli-Myrli» — «Ширли-мырли», 1995; «Baryshnya-krestyanka» — «Baryshnya-krestyanka», 1995) и на французском («JLG/JLG, autoportrait de decembre», 1994). Для лингвистического исследования, выполняемого на материале такого корпуса, подобное качество эмпирической базы означает возможность анализировать параллельные субтитры, подвергая контексты ручной сортировке на основе предполагаемого или самостоятельно определяемого языка оригинала. Это обстоятельство, очевидно, снижает целесообразность обращения к таким данным и возможностям их механизированной обработки, хотя именно на это ориентированы инструменты дальнего анализа в целом и корпусные технологии в частности.

Безусловно, в корпусном менеджере предусмотрены различные возможности фильтрации контента, в том числе по стране, в которой был снят фильм — *Countries the film was shot it*, и языку, на котором в нём говорят — *Languages spoken in the film* (рис. 3). Однако пояснения к этим фильмам в *Sketch Engine* отсутствуют.

Предположительно названные возможности настройки поискового фильтра (рис. 3) в корпусе соответствуют тем группам метаданных, которыми сопровождается каждый фильм или сериал на своей странице на сайте *OpenSubtitles*. Если это так, то страной, в которой был снят фильм, может являться страна-производитель или страна, в которой происходили съёмки, что не всегда одно и то же и далеко не всегда что-либо говорит

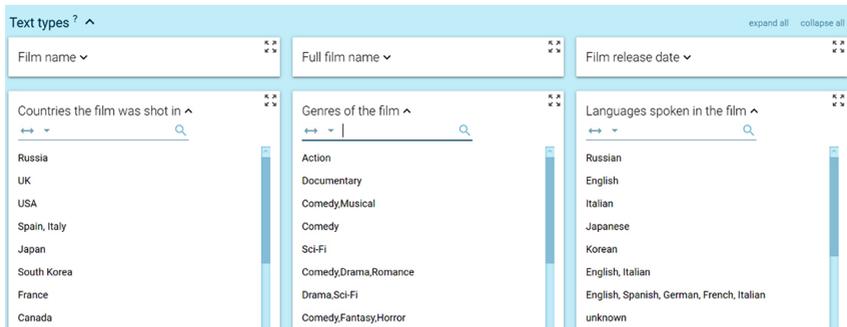


Рис. 3. Скриншот с изображением фильтров поиска по корпусу
OpenSubtitles 2018 parallel — English

о языке, на котором снимается кино. Фильтр *Languages spoken in the film* потенциально может ограничивать поисковую выборку теми фильмами, (1) которые были сняты на определённом языке, (2) которые были переведены на определённый язык (3) или в которых вне зависимости от языка оригинала есть речь на других языках. Например, сериал «War and Peace» (2007) был снят на английском языке, но в нём, как и в романе Л. Толстого «Война и мир», присутствуют реплики на французском. Съёмки фильма проходили на киностудии в Вильнюсе (Литва), а странами-производителями являются Россия, Франция, Германия, Италия, Испания, Польша [BM]. Получается, что, настраивая фильтр для поиска фильмов, снятых, к примеру, только во Франции, у исследователя нет возможности автоматически исключить проекты, выполненные несколькими странами, а выбирая французский в качестве языка, на котором говорят в кино, исследователь определяет в выборку все кинокартины, в которых есть французская речь, даже если там присутствуют реплики на других языках, которых в сравнении с французской киноречью может оказаться больше. В результате формальный объём эмпирического материала будет расходиться с фактическим. Это в свою очередь означает недостоверность количественных данных об исследовательском корпусе и, как следствие, получаемых качественных результатов лингвистического исследования.

Отдельный корпус разработан в *Sketch Engine* на материале скриптов к англоязычным фильмам — *Film Corpus* (27 292 710 — токенов, 21 661 806 — слов). В состав корпуса входят скрипты к 1 068 фильмам из базы IMDb, которые, согласно пояснениям разработчиков, классифицированы по 304 жанрам [FCMS]. На самом деле это количество жанров соответствует не только собственно жанрам (drama; thriller; comedy; action;

crime; romance), но и их группам (comedy, drama; drama, romance; crime, drama, thriller; action, adventure, fantasy, sci-fi и другие комбинации) (рис. 4).



Рис. 4. Скриншот с составом и метаданными корпуса *Film Corpus* в *Sketch Engine*

Метаданные к кинокартине собираются методом краулинга [Алюнина, 2022, с. 16] сайта *IMSDb*, где к каждому фильму указывается своя жанровая категория. Поскольку современные кинопроекты принадлежат, как правило, нескольким жанрам, то и напротив фильма и сериала в корпусе *Sketch Engine* с определённого момента автоматически появляется уникальная жанровая принадлежность, являющаяся, по сути, комбинацией жанров. Получается, что при выборе, например, триллеров (№ 2 в таблице на рис. 4) для изучения их языковых особенностей в кино из выборки автоматически исключаются все кинокартины, которые являются не только триллерами, но принадлежат ещё каким-то жанрам (№ 35, 37, 38 в таблице на рис. 4). Это приводит к тому, что в эмпирическую базу попадают не все сведения, релевантные цели и задачам исследования.

Помимо анализа киноконтента на материале размеченных корпусов, существует возможность создавать **пользовательские корпуса** в различных менеджерах: *AntConc*, *IRaMuTeQ*, *Orange*, *Sketch Engine*, *Voyant tools*, *WordSmith Tools* и др. Для этого требуется собственная коллекция отобранных субтитров и скриптов или вручную созданных расшифровок к фильмам. Также можно воспользоваться опцией веб-краулинга [Алюнина, 2022, с. 16], которую предлагает *Sketch Engine*. Корпусные менеджеры позволяют механизировать анализ языкового материала, превращая коллекцию

текстов в автоматически размеченный корпус или предоставляя возможности его машинного анализа: составление списка ключевых и частотных слов, выявление лексической сочетаемости по коллокациям или n-gram, создание облаков слов, выполнение поиска лексической сочетаемости на языке CQL и др. Спектр возможностей разных менеджеров может отличаться, но все они равно применимы как к предустановленным корпусам, так и к создаваемым в них пользовательским корпусам.

Основным преимуществом пользовательского корпуса перед предустановленным является чистота эмпирического материала, которая обеспечивается ручной выборкой контента и позволяет исключить тот шум, который характеризует, например, корпуса в *Sketch Engine*. Однако для проведения ручной выборки киноречи, зафиксированной в письменной форме, необходимо принимать во внимание некоторые экстралингвистические факторы, которые отражаются на результатах исследования. Поскольку большинство размеченных корпусов киноконента создано на материале субтитров, внимание следующего раздела будет сосредоточено на их качественных характеристиках.

3.2. Качество субтитров

Среди письменных источников киноречи в отечественном языкознании субтитры являются «самым малоисследованным» [Сухая, 2010, с. 73]. При этом, согласно Европейской федерации аудиовизуального перевода (*Audiovisual Translation Europe*), в последние годы отмечается рост популярности фильмов и сериалов с субтитрами, обусловленный распространением стриминговых платформ [FTU] (*Netflix, Hulu, Wink, Okko, ivi, Кинопоиск* и др.). В этой связи закономерным видится теоретическое осмысление особенностей субтитрования в разных языках на основе сложившейся практики.

Наличие субтитров к тому или иному киноматериалу во многом связано с его аудиовизуальным переводом (межъязыковым или внутриязыковым), который подчиняется своим правилам. При переводе аудиовизуального произведения во внимание принимаются такие факторы, как тип озвучивания (закадр или дубляж), тип субтитров (классические, для глухих и слабослышащих), целевая аудитория (возраст, пол, социальный статус и др.), тип киноматериала (экранизация, в основе которой лежит художественное произведение, или режиссёрский фильм; полнометражный фильм или сериал), особенности локализации и адаптации киноконента к принимающему рынку, требования заказчика и др. [Асташина и др., 2021, с. 22—23; Малёнова, 2017, с. 37; Cintas et al., 2007, p. 13, 34—35]. Вне зависимости от того, выполняется ли перевод под озвучивание или субтитрование, все перечисленные факторы влияют на качество субтитров. В рамках данной

статьи качество субтитров рассматривается через технические, социальные и культурно-национальные особенности субтитрования.

Говоря о технических особенностях субтитрования, мы не будем рассматривать техническую специфику двухмерных и трёхмерных субтитров [Малёнова, 2017, с. 36], вшитых / жёстких, программных / мягких, а также пререндеренных субтитров (pre-rendered subtitles) [Асташина и др., 2021, с. 10] [Flussonic]. Внимание будет сосредоточено только на общих технических характеристиках классических субтитров и их влиянии на качество субтитров как эмпирического материала исследования.

Технические особенности классических субтитров касаются в первую очередь количества знаков в строке субтитра и места его расположения, что свойственно как межъязыковому субтитрованию, так и внутриязыковому. Эти ограничения определены стандартами субтитрования, но они могут различаться в зависимости от страны и соответствующего языка.

Согласно общемировым стандартам, «нижняя строка субтитра должна отступать» от нижней границы экрана на 1/12 высоты кадра [ОСИС], а по обе стороны от текста должно оставаться расстояние в 1/12 от ширины экрана. Следовательно, при изменении размера экрана (экран кинотеатра, если субтитрование выполняется для показа в кинотеатре; экран ноутбука и компьютера, если субтитрование выполняется для стриминговых интернет-платформ; экран, встроенный в кресло впереди стоящего сидения на борту самолёта или в поезде, если субтитрование выполняется для мультимедиа-платформы компании-перевозчика) меняется размер пространства для размещения субтитров, следовательно, это влияет на допустимое количество знаков в них. Так, согласно *Ассоциации перевода и адаптации аудиовизуального контента (Association des traducteurs adapateurs de l'audiovisuel)*, длина строки французских субтитров для кинотеатра составляет 40—41 знак, для стриминговых платформ — 40—42 знака, для телеэкрана — 36—37 знаков [FSS].

Выравнивание текста субтитра настраивается по ширине, поскольку большая часть действий совершается в центре кадра. Максимальное количество строк в субтитрах — две [ОСИС]. Если текст укладывается в одну строчку, то он занимает нижнее положение. Количество знаков в строке составляет в среднем 35—37 символов, которые эквивалентны 5—6 секундам звучащей киноречи [ОСИС] [Cintas et al., 2007, p. 63].

Количество знаков в строке субтитра может варьироваться в зависимости не только от размера экрана, но также от страны и языка. Так, согласно стандартам платформы *Neflix* (США), количество символов в строке для субтитров на сербском и финском языках не должно превышать 42 знаков

[STTSG; FTTSG], на русском языке — 39 знаков [RTTSG], на корейском — 16 [KTTSG], на японском — 13 [JTTSG].

Таким образом, технические ограничения, накладываемые на субтитры, обусловлены не только «пространством в нижней части экрана», в который необходимо вписать суть реплики, «максимально точно передав смысл» высказывания [Сухая, 2010, с. 73], но и усреднённой способностью языкового знака передавать основное содержание диегетической речи ограниченным количеством символов.

Следующий тип ограничений, рассматриваемый в статье, касается социального аспекта.

Социальные ограничения субтитрования связаны с особенностями целевой аудитории, для которой производится адаптация медиаконтента, в том числе фильмов и сериалов. Эти особенности обусловлены «когнитивными потребностями» зрителей, которые относятся к следующим категориям: «(слепые, слабовидящие, глухие, слабослышащие и прочие зрители)» [Асташина и др., 2021, с. 23]. Такие субтитры маркируются аббревиатурой SDH «subtitling for the deaf and the hard-of-hearing» [Асташина и др., 2021, с. 23; Cintas et al., 2007, p. 12]. Субтитры для глухих и слабослышащих (то есть аудитории с особыми когнитивными потребностями, способной читать) содержат дополнительную информацию, которая не является диегетической речью, но включается в субтитры (рис. 5).

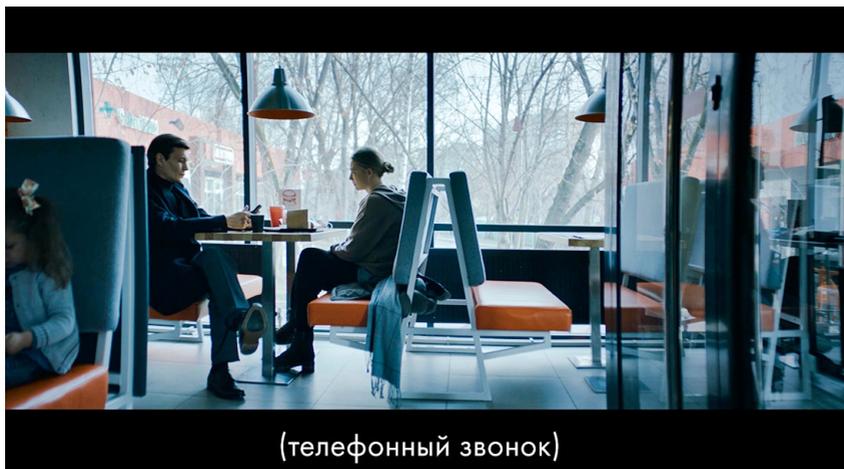


Рис. 5. Кадр с субтитрами из сериала «Контейнер» (источник скриншота: <https://start.ru/journal/zachem-smotret-rossiyskie-serialy-s-subtitramy>)

На изображении выше (рис. 5) приведён пример включения в кадр фразы, относящейся к субтитрам для глухих и слабослышащих: *телефонный звонок*. Подобные фразы описывают звуки в кадре, чтобы у зрителя, имеющего проблемы с восприятием информации на слух, была возможность прочесть, что в данном месте используются дополнительные звуковые эффекты, которые входят «в канву сюжета» [ГОСТ] и важны для понимания. К таким эффектам могут относиться звуки смеха (*громко смеётся, ехидно усмехается*), плача (*громко рыдает, мычит в слезах*), музыки (*играет тихая мелодия, звучит торжественная музыка*), природы (*скрип снега, воет ветер, барабанит дождь*), интерьера (*хлопнула дверь, скрипит пол, скрежет замочной скважины*) и т. п. Также в SDH, в отличие от классических субтитров, присутствует обозначение говорящего, чтобы зритель смог понять, кто произносит реплику. Это дополнение является актуальным ввиду того, что рассматриваемая целевая аудитория не всегда может определить на слух, кому принадлежит закадровая речь в кино: персонажу мужского рода или женского, ребёнку или взрослому и т. д.

Необходимость включения дополнительной информации в SDH приводит к необходимости расширения пространства для субтитров. Поэтому в субтитрах для глухих и слабослышащих допустимо «использование третьей строки» [ГОСТ]. При этом по умолчанию количество символов в строке составляет 38 знаков [ГОСТ]. Для создания корпуса по материалам субтитров, в который могут попасть SDH, это значит увеличение количества слов и токенов. Такое увеличение в каком-то смысле является искусственным, поскольку фразы, описывающие звуки в кадре и обозначающие говорящего, не являются киноречью, то есть объектом исследования, которое направлено на анализ звучащей речи в кино и сериалах. Эти текстовые данные не являются релевантными и могут рассматриваться как шум в эмпирическом материале.

Последний тип особенностей субтитров, значимый для принятия во внимание при создании корпуса киноречи на материале субтитров, связан с **культурно-национальной спецификой** субтитрования.

Прежде всего, необходимо отметить, что субтитрование как вид межязыкового аудиовизуального перевода в разных странах в разной степени предпочтителен. Так, согласно данным статистического агентства *Statista* на 2021 год (Рис. 6), субтитрованию отдаётся предпочтение в США, Великобритании, Индии, Китае и Японии. Примерно на одном уровне субтитрование и дубляж находятся в Бразилии и Франции, а в Италии и Германии в большей степени предпочитают дублированный контент, но при этом разница с субтитрованием не настолько велика, как в предпочтениях зрителей США и Великобритании [Statista].

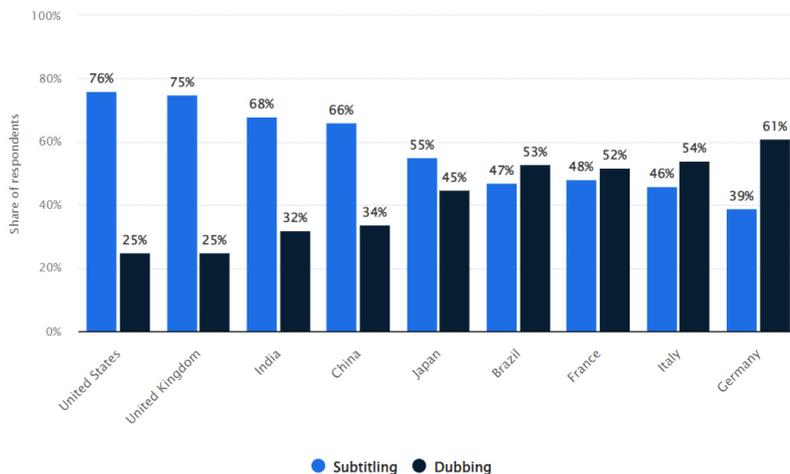


Рис. 6. Распределение предпочтений зрителей по потреблению зарубежного контента на иностранном языке с субтитрами или в дубляже на ноябрь 2021 года

Эти сведения говорят о том, на каких языках с большей долей вероятности есть возможность сформировать эмпирическую базу исследования на материале **переводных субтитров** к киноконенту. В данном случае речь идёт исключительно о межъязыковом субтитровании, так как язык звуковой дорожки и язык субтитров одного и того же киноматериала разные.

Рассмотрим подробнее культурно-национальные особенности субтитрования на примере трёх стран: Сербии, Финляндии и России. Выбор стран обусловлен их разными подходами к адаптации киноконента через субтитры.

Сербия является субтитровочной страной, поскольку фильмы в кинотеатрах и на некоторых телеканалах Сербии показывают на языке оригинала в сопровождении **сербских субтитров** [Lukić, 2016, p. 64—65]. Исключение составляют детские мультфильмы, традиционно дублируемые на сербский язык [Сербские хроники]. См., например, трейлеры к фильмам («Аквaмен», 2023, США; «Холоп 2» (серб. «Слуга 2» / «Sluga 2»), 2024, Россия) на сайте сербского кинотеатра *Cineplexx* (<https://www.cineplexx.rs/>). Такой подход к субтитрованию означает, что сербские субтитры могут лечь в основу лингвистического корпуса для изучения переводной речи в фильмах для взрослой аудитории, но не речи в детских мультфильмах, поскольку последние переводятся под дубляж, а не субтитры. Субтитрование для глухих и слабослышащих в Сербии пока ещё недостаточно разви-

то [Lukić, 2016, p. 65], поэтому традиция внутриязыкового субтитрования для сербского кино не сформирована.

Таким образом, сербские субтитры к иностранным фильмам являются переводом зарубежного аудиовизуального контента. Они формируют богатую эмпирическую базу для изучения вопросов языковой компрессии сербского субтитрования, локализации, а также лингвокультурных аспектов адаптации зарубежного кино к сербской аудитории, но не для изучения киноречи сербских кинокартин.

В кинотеатрах **Финляндии** фильмы идут на языке оригинала с субтитрами **на финском и шведском одновременно** (рис. 7). См., например, скриншот сайта кинотеатра *Finnkino* (<https://www.finnkino.fi/>), где напротив каждого сеанса отмечается язык фильма в прокате и язык субтитров, которые обязательно сопровождают зарубежное кино в Финляндии (рис. 7). Как и в Сербии, исключение делается для детских мультфильмов: их переводят на шведский или финский языки [ЯПФ].

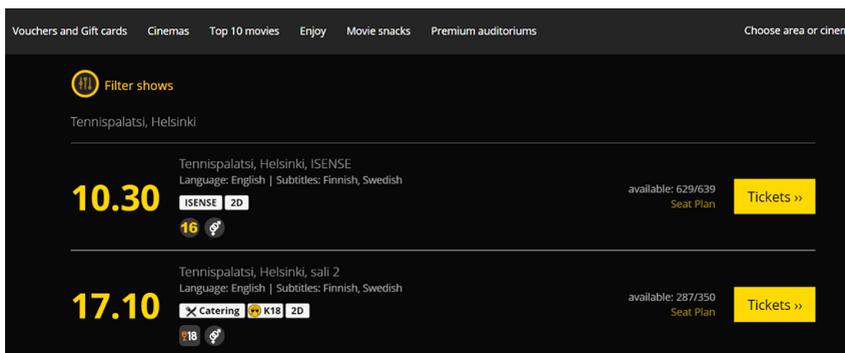


Рис. 7. Скриншот сайта финского кинотеатра *Finnkino*

Если к фильмам на экране даются субтитры на двух языках, это значит, что место для текста на каждом из языков сокращается. При этом субтитры на каждом из языков необходимо вписывать в более жёсткие пространственные рамки. Следовательно, субтитры, создаваемые на финском и шведском для параллельного сопровождения англоязычных фильмов, укорачиваются вдвое или практически вдвое. Это в свою очередь уменьшает объём языкового материала, который потенциально может являться эмпирической базой для анализа переводной киноречи на финском и шведском, если текстовое сопровождение фильма выполнено по всем канонам субтитрования.

В **России** сложилась **традиция дубляжа** зарубежного киноконента [Малёнова, 2021, с. 64], а межязыковое субтитрование является феноме-

ном, который идёт в ногу с ростом популярности стриминговых платформ [FTU]. Внутриязыковое субтитрование обусловлено в первую очередь развитием технологий медиадоступности и созданием SDH [Асташина и др., 2021, с. 23]. Также, согласно материалам *OpenSubtitles*, в России распространены так называемые **пользовательские субтитры**, которые создаются потребителями киноконтента на добровольных основаниях и распространяются на безвозмездной основе. В том числе поэтому к некоторым фильмам можно найти несколько файлов с разными русскоязычными субтитрами. Формат пользовательских субтитров далёк от канонического, поскольку они публикуются, как правило, не профессиональными субтитровщиками и без проверки редактора. Нередко среди пользовательских субтитров встречается подобие расшифровок с проставленными тайм-кодами.

Одним из проявлений феномена пользовательского субтитрования являются фанатские субтитры, или фансабы, которые выпускаются поклонниками фильмов и сериалов независимо от их официального распространения. В основном они создаются для аниме и японских фильмов и размещаются на фанатских платформах (*Doramaland.org*, *funsub.ru* и др.). Русскоязычные фансабы отличает то, что их авторы зачастую не владеют языком оригинала — японским, поэтому субтитры создаются через язык-посредник — английский (рис. 8) [Фансабы]. Однако как английский язык, так и русский у фансаберов не всегда достигает должного уровня, поэтому в фанатских субтитрах обнаруживается немалое количество переводческих (рис. 8) и орфографических (рис. 9) ошибок.



Рис. 8. Пример переводческой ошибки в фанатских субтитрах

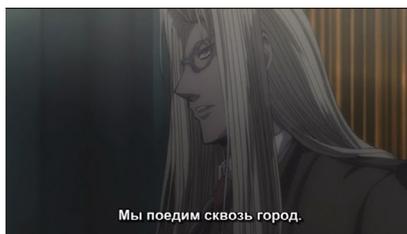


Рис. 9. Пример орфографической ошибки в фанатских субтитрах

Таким образом, при формировании корпуса субтитров на русском языке необходимо принимать во внимание следующие факторы. Во-первых, традиция дубляжа в России ограничивает популярность и, следовательно, количество русскоязычных субтитров. Во-вторых, платформы с субтитрами нередко предлагают материалы, созданные не профессиональными субтитровщиками, а пользователями, которые не всегда соблюдают правила суб-

титрования. В-третьих, некоторые пользовательские субтитры содержат ошибки, которые могут негативно отразиться на результате исследования.

Рассмотренные культурно-национальные особенности субтитрования, безусловно, не являются универсальными, а иллюстрирующие их примеры допустимо назвать единичными случаями адаптации киноконента посредством субтитров. Описанные случаи говорят о том, что при составлении исследовательского корпуса субтитров для изучения переводной киноречи на том или ином языке необходимо проанализировать и учесть специфику и, возможно, историю субтитрования в определённой стране, чтобы обеспечить качество эмпирического материала.

4. Заключение = Conclusions

Рассмотренные в данной статье возможности и ограничения корпусного подхода к анализу кино на материале субтитров с учётом технических, социальных и культурно-национальных особенностей субтитрования являются значимыми с фундаментальной точки зрения в первую очередь для создания эмпирической базы исследования. В ходе работы было выявлено, что наибольшее количество субтитров, доступных в открытых банках, являются переводными эквивалентами звучащей речи. Внутриязыковое субтитрование актуально прежде всего для аудитории с особыми когнитивными потребностями, поэтому материал SDH подходит только для решения некоторых и достаточно специфических исследовательских задач в анализе киноречи.

Поскольку любое аудиовизуальное произведение по определению содержит аудиальную и визуальную информацию, исключение одной из них из фокуса внимания неизбежно ведёт к снижению достоверности и верифицируемости полученных результатов. В этой связи эффективное использование корпусного подхода в анализе аудиовизуального контента возможно для решения такой задачи, для которой допустимо исключение визуальной составляющей. Например, на стадии первичного количественного анализа материала с целью его дальнейшего качественного осмысления:

— в лингвокультурологии — обнаружение фильмов и сериалов, которые содержат определённые аудиовизуальные образы, подкреплённые вербальными отсылками;

— в дискурс-анализе и межкультурной коммуникации — выявление коммуникативных событий, в которых реализуются какие-либо стереотипы или стигмы;

— в ономастике — поиск имён собственных, которые формируют образ реального или вымышленного пространства аудиовизуального произведения;



— в маркетинге — извлечение частных случаев перевода для изучения стратегий локализации киноконтента к различным аудиториям и национальным рынкам.

Это лишь несколько примеров исследовательских вопросов, в работе над которыми может быть полезен дальний анализ аудиовизуального контента на материале корпуса субтитров. Настоящее исследование показывает, что при составлении корпуса киносубтитров качество прогнозируемых результатов во многом зависит от понимания технической, социальной и культурно-национальной специфики эмпирического материала. Перспектива данного исследования видится в изучении аналогичных особенностей других типов аудиовизуального контента, таких как пьесы и видеоигры, корпуса которых на сегодняшний день только начинают появляться.

Автор заявляет об отсутствии конфликта интересов.	The author declares no conflicts of interests.
---	--

Источники и принятые сокращения

1. ВМ — *Война и мир*. — Кино-театр.ру [Электронный ресурс]. — Режим доступа : <https://www.kino-teatr.ru/kino/movie/euro/10167/annot/> (дата обращения 28.12.2023).
2. ГК РФ 1263 — *ГК РФ Статья 1263*. Аудиовизуальное произведение. — Консультант Плюс [Электронный ресурс]. — Режим доступа : https://www.consultant.ru/document/cons_doc_LAW_64629/bcd3bad76a5cc627f036c9e23135e25579d3635/ (дата обращения 26.12.2023).
3. ГОСТ — *ГОСТ Р 57763-2017*. Национальный стандарт Российской Федерации. Скрытые субтитры для инвалидов по слуху. — Электронный фонд правовых и нормативно-технических документов [Электронный ресурс]. — Режим доступа : <https://docs.cntd.ru/document/1200156999> (дата обращения 21.12.2023).
4. Сербские хроники — *Кино в Шабце* [Электронный ресурс]. — Режим доступа : <https://www.serbia-home.com/archives/2228> (дата обращения 28.12.2023).
5. ТСТ — *Требования к субтитрованию и тифлокомментированию полнометражных национальных фильмов, создаваемых в художественной или анимационной форме*. — Консультант Плюс [Электронный ресурс]. — Режим доступа : https://www.consultant.ru/document/cons_doc_LAW_282550/82d02de07f5851896fb255d9a6c9844df995584c/ (дата обращения 21.12.2023).
6. Фансаб — *Русский фансаб такой русский* [Электронный ресурс]. — Режим доступа : https://nya.sh/files/epic_fansubs/ (дата обращения 02.01.2024).
7. ЯПФ — *Шведский язык и языковая политика в Финляндии* [Электронный ресурс]. — Режим доступа : <https://fennica.pohjoiseen.fi/ru/2019/01/13/finland-swedish/> (дата обращения 02.01.2024).
8. CVAA — *Twenty-First Century Communications and Video Accessibility Act of 2010* [Electronic resource]. — Access mode : <https://www.fcc.gov/consumers/guides/21st-century-communications-and-video-accessibility-act-cvaa> (accessed 22.12.2023).
9. FC — *Film Corpus 2.0* [Electronic resource]. — Access mode : <https://nlds.soe.ucsc.edu/fc2> (accessed 26.12.2023).



10. FCMS — *Film corpus of movie scripts*. — Sketch Engine [Electronic resource]. — Access mode : <https://www.sketchengine.eu/film-corpus/> (accessed 26.12.2023).
11. Flussonic — *Прожили* субтитров и для чего он нужен [Электронный ресурс]. — Режим доступа : <https://flussonic.ru/blog/news/why-do-we-burn-subtitles/> (дата обращения 04.01.2024).
12. *Frozen* — Frozen Disney Script. IMISDb [Electronic resource]. — Access mode : [https://imsdb.com/Movie%20Scripts/Frozen%20\(Disney\)%20Script.html](https://imsdb.com/Movie%20Scripts/Frozen%20(Disney)%20Script.html) (accessed 24.12.2023).
13. FSS — *French Subtitling Standards*. — ATAA [Electronic resource]. — Access mode : https://www.ataa.fr/documents/NORMES_SOUS-TITRAGE_FRANCAIS_ANGLAIS_EN-TETE.pdf (accessed 20.12.2023).
14. FTTSG — *Finnish Timed Text Style Guide*. — Netflix [Electronic resource]. — Access mode : <https://partnerhelp.netflixstudios.com/hc/en-us/articles/215087558-Finnish-Timed-Text-Style-Guide> (accessed 02.10.2024).
15. FTU — *Filmmakers and translators unite* [Electronic resource]. — Access mode : <https://avteurope.eu/2019/04/09/filmmakers-and-translators-unite/> (accessed 20.12.2023).
16. IMDb — *Internet Movie Database* — [Electronic resource]. — Access mode : <https://www.imdb.com/> (accessed 20.12.2023).
17. IMISDb — *The Internet Movie Script Database* [Electronic resource]. — Access mode : <https://imsdb.com/> (accessed 27.12.2023).
18. JTTSG — *Japanese Timed Text Style Guide*. — Netflix [Electronic resource]. — Access mode : <https://partnerhelp.netflixstudios.com/hc/en-us/articles/215767517-Japanese-Timed-Text-Style-Guide> (accessed 02.10.2024).
19. KTTSG — *Korean Timed Text Style Guide*. — Netflix [Electronic resource]. — Access mode : <https://partnerhelp.netflixstudios.com/hc/en-us/articles/216001127-Korean-Timed-Text-Style-Guide> (accessed 20.12.2023).
20. NGIR — *Netflix Games Internationalization Requirements*. — Netflix [Electronic resource]. — Access mode : https://partnerhelp.netflixstudios.com/hc/en-us/articles/4411631891859-Netflix-Games-Internationalization-Requirements#h_01FNS8D3C5VWTKVZ54736FNXG1 (accessed 20.12.2023).
21. OS PC — *OpenSubtitles parallel corpora*. — Sketch Engine [Electronic resource]. — Access mode : <https://www.sketchengine.eu/opensubtitles-parallel-corpora/> (accessed 24.12.2023).
22. SFTT — *Subtitling and Dubbing for Film and Television*. — ATAA [Electronic resource]. — Access mode : https://www.ataa.fr/documents/Guide_EN_2019.pdf (accessed 20.12.2023).
23. RTTSG — *Russian Timed Text Style Guide*. — Netflix [Electronic resource]. — Access mode : <https://partnerhelp.netflixstudios.com/hc/en-us/articles/215346638-Russian-Timed-Text-Style-Guide> (accessed 02.10.2024).
24. STTSG — *Serbian Timed Text Style Guide*. — Netflix [Electronic resource]. — Access mode : <https://partnerhelp.netflixstudios.com/hc/en-us/articles/115002813348-Serbian-Timed-Text-Style-Guide> (accessed 20.12.2023).
25. Statista — *Share of users who watch content in other languages and prefer subtitling or dubbing as of November 2021, by country*. — Statista [Electronic resource]. — Access mode : <https://www.statista.com/statistics/1289864/subtitles-dubbing-audience-preference-by-country/> (accessed 04.01.2024).
26. SE — *Sketch Engine* [Electronic resource]. — Access mode : <https://www.sketchengine.eu/> (accessed 24.12.2023).



27. TMC — *The Movies Corpus* [Electronic resource]. — Access mode : <https://www.english-corpora.org/movies/> (accessed 20.12.2023).

Литература

1. *Алюнина Ю. М.* Где живут чудовища? / Ю. М. Алюнина // Корпусный метод обнаружения англицизмов и их производных в русскоязычном Интернете // Вестник Томского государственного университета. Филология. — 2022. — № 80. — С. 5—29. — DOI: 10.17223/19986645/80/1.
2. *Анисимов В. Е.* Образы традиционной и современной Тувы в российском этнографическом кино (на примере фильма С. Родиной и А. Меньшова «Планета Тыва») / В. Е. Анисимов, А. С. Борисова, Е. Д. Калининкова // Новые исследования Тувы. — 2022. — № 1. — С. 183—197. — DOI: <https://doi.org/10.25178/nit.2022.1.12>.
3. *Анисимов В. Е.* Лингвокультурная локализация кинозаголовков / В. Е. Анисимов, А. С. Борисова, Г. Р. Консон // Russian Journal of Linguistics. — 2019. — Т. 23. — № 2. — С. 435—459. — DOI: 10.22363/2312-9182-2019-23-2-435-459.
4. *Асташина Д. А.* Специалист в области перевода и медиадоступности : рамка компетенций / Д. А. Асташина, И. С. Борщевский, Н. В. Гайдаш. — Казань : Издательство «Бук», 2021. — 46 с. — ISBN 978-5-00118-752-3.
5. *Малёнова Е. Д.* Теория и практика аудиовизуального перевода : отечественный и зарубежный опыт / Е. Д. Малёнова // Коммуникативные исследования. — 2017. — Т. 2. — № 12. — С. 32—46.
6. *Малёнова Е. Д.* Аудиовизуальный перевод в СССР и России : краткая историческая ретроспектива / Е. Д. Малёнова // Вестни НГЛУ. Язык и культура. — 2021. — Т. 3. — № 553. — С. 61—83. — DOI: 10.47388/2072-3490/lunn2021-55-3-61-83.
7. *Махов А. Е.* «Пристальное чтение» перед лицом «больших данных» : проблема метода в современном англоязычном литературоведении / А. Е. Махов // Социальные и гуманитарные науки. Отечественная и зарубежная литература. — 2018. — № 3. — С. 11—32.
8. *Моретти Ф.* Дальнее чтение / Ф. Моретти. — Москва : Изд-во Института Гайда-ра, 2016. — 352 с. — ISBN 978-5-93255-446-3.
9. *Сухая Е. В.* Типы источников киноречи как основа лингвистических исследований / Е. Ф. Сухая // Вопросы современной лингвистики. — 2010. — Т. 2. — С. 68—74.
10. *Уртминцева М. Г.* Закадровый текст в документальном кинодискурсе / М. Г. Уртминцева, В. Н. Скачкова // Научный диалог. — 2022. — Т. 11. — № 6. — С. 334—349. — DOI: 10.24224/2227-1295-2023-12-3-134-152.
11. *Bednarek M.* Language and Television Series : A Linguistic Approach to TV Dialogue / M. Bednarek. — Cambridge : Cambridge University Press, 2018. — 321 p.
12. *Bednarek M.* Language and Characterisation in Television Series : A corpus-informed approach to the construction of social identity in the media / M. Bednarek. — Amsterdam : John Benjamins, 2023. — 279 p.
13. *Cintas J. D.* Audiovisual Translation : Subtitling / J. D. Cintas, A. Remael. — London : Routledge, 2007. — 284 p.
14. *Is Subtitling Equally Effective Everywhere? A First Cross-national Study on the Reception of Interlingually Subtitled Messages* / E. Perego, M. Laskowska, A. Matamala, A. Remael, I. S. Robert // Across Languages and Cultures. — 2016. — Vol. 17. — № 2. — Pp. 205—229.
15. *Kozloff S.* Overhearing Film Dialogue / S. Kozloff. — Berkley — Los Angeles — London : University of California Press, 2000. — 335 p. — ISBN 0-520-22138-9.



16. Lukić N. Audiovisual Translation in Serbia : History and Professional Considerations / N. Lukić // Hieronymus. — 2016. — № 3. — Pp. 64—89.
17. Smolan R. The Human Face of Big Data / R. Smolan, J. Erwitz // Filmplatform [Electronic resource]. — Access mode : https://www.filmplatform.net/wp-content/uploads/2016/03/FINAL_HFOBDStudyGuide-1.pdf. (accessed 04.01.2024).
18. Tröhler M. Christian Metz and the Codes of Cinema. Film Semiology and Beyond / M. Tröhler, G. Kirsten. — Amsterdam : Amsterdam University Press, 2018. — 484 p.

Статья поступила в редакцию 18.01.2024,
одобрена после рецензирования 11.03.2024,
подготовлена к публикации 18.03.2024.

Material resources

- Civil Code of the Russian Federation 1263 — *Civil Code of the Russian Federation Article 1263. An audiovisual work. Consultant Plus*. Available at https://www.consultant.ru/document/cons_doc_LAW_64629/bcdb3bad76a5cc627f036e23135e25579d3635/ (accessed 12.26.2023). (In Russ.).
- CVAA — *Twenty-First Century Communications and Video Accessibility Act of 2010*. Available at <https://www.fcc.gov/consumers/guides/21st-century-communications-and-video-accessibility-act-cvaa> (accessed 22.12.2023).
- Funsub — *Russian funsub is so Russian*. Available at: https://nya.sh/files/epic_fansubs/ (accessed 02.01.2024). (In Russ.).
- FC — *Film Corpus 2.0*. Available at: <https://nlds.soe.ucsc.edu/fe2> (accessed 26.12.2023).
- FCMS — *Film corpus of movie scripts. Sketch Engine*. Available at: <https://www.sketchengine.eu/film-corpus/> (accessed 26.12.2023).
- Flussonic — *Burning subtitles and what is it for*. Available at: <https://flussonic.ru/blog/news/why-do-we-burn-subtitles/> (дата обращения 04.01.2024).
- Frozen — *Frozen Disney Script. IMSDb*. Available at: [https://imdb.com/Movie%20Scripts/Frozen%20\(Disney\)%20Script.html](https://imdb.com/Movie%20Scripts/Frozen%20(Disney)%20Script.html) (accessed 24.12.2023).
- FSS — *French Subtitling Standards. ATAA*. Available at: https://www.ataa.fr/documents/NORMES_SOUS-TITRAGE_FRANCAIS_ANGLAIS_EN-TETE.pdf (accessed 20.12.2023).
- FTTSG — *Finnish Timed Text Style Guide. Netflix*. Available at: <https://partnerhelp.netflixstudios.com/hc/en-us/articles/215087558-Finnish-Timed-Text-Style-Guide> (accessed 02.10.2024).
- FTU — *Filmmakers and translators unite*. Available at: <https://avteurope.eu/2019/04/09/filmmakers-and-translators-unite/> (accessed 20.12.2023).
- GOST — *GOST R 57763-2017. The national standard of the Russian Federation. Hidden subtitles for the hearing impaired. Electronic fund of legal and regulatory documents*. Available at: <https://docs.cntd.ru/document/1200156999> (accessed 12.21.2023). (In Russ.).
- IMDb — *Internet Movie Database*. Available at: <https://www.imdb.com/> (accessed 20.12.2023).
- IMSDb — *The Internet Movie Script Database*. Available at: <https://imdb.com/> (accessed 27.12.2023).
- JPF — *Swedish language and language policy in Finland*. Available at: <https://fennica.pohjoisen.fi/ru/2019/01/13/finland-swedish/> (accessed 02.01.2024). (In Russ.).

- JTTSG — *Japanese Timed Text Style Guide*. Netflix. Available at: <https://partnerhelp.netflixstudios.com/hc/en-us/articles/215767517-Japanese-Timed-Text-Style-Guide> (accessed 02.10.2024).
- KTTSG — *Korean Timed Text Style Guide*. Netflix. Available at: <https://partnerhelp.netflixstudios.com/hc/en-us/articles/216001127-Korean-Timed-Text-Style-Guide> (accessed 20.12.2023).
- NGIR — *Netflix Games Internationalization Requirements*. Netflix. Available at: https://partnerhelp.netflixstudios.com/hc/en-us/articles/4411631891859-Netflix-Games-Internationalization-Requirements#h_01FNS8D3C5VWTKVZ54736FNXG1 (accessed 20.12.2023).
- OS PC — *OpenSubtitles parallel corpora*. Sketch Engine. Available at: <https://www.sketchengine.eu/opensubtitles-parallel-corpora/> (accessed 24.12.2023).
- RTTSG — *Russian Timed Text Style Guide*. Netflix. Available at: <https://partnerhelp.netflixstudios.com/hc/en-us/articles/215346638-Russian-Timed-Text-Style-Guide> (accessed 02.10.2024).
- SDFT — *Subtitling and Dubbing for Film and Television*. ATAA. Available at: https://www.ataa.fr/documents/Guide_EN_2019.pdf (accessed 20.12.2023).
- SE — *Sketch Engine*. Available at: <https://www.sketchengine.eu/> (accessed 24.12.2023).
- Serbian Chronicles — *Cinema in Šabac*. Available at: <https://www.serbia-home.com/archives/2228> (accessed 12.28.2023). (In Russ.).
- Statista — *Share of users who watch content in other languages and prefer subtitling or dubbing as of November 2021, by country*. Statista. Available at: <https://www.statista.com/statistics/1289864/subtitles-dubbing-audience-preference-by-country/> (accessed 04.01.2024).
- STTSG — *Serbian Timed Text Style Guide*. Netflix. Available at: <https://partnerhelp.netflixstudios.com/hc/en-us/articles/115002813348-Serbian-Timed-Text-Style-Guide> (accessed 20.12.2023).
- TMC — *The Movies Corpus*. Available at: <https://www.english-corpora.org/movies/> (accessed 20.12.2023).
- TST — *Requirements for subtitling and typhocommenting full-length national films created in artistic or animated form*. Consultant Plus. Available at: https://www.consultant.ru/document/cons_doc_LAW_282550/82d02de07f5851896fb255d9a6c9844df995584c Russian Russian Fansab/ (accessed 12.21.2023). (In Russ.).
- VM — *War and Peace*. [Kino Teatr er](https://www.kino-teatr.ru/kino/movie/euro/10167/annot/). Available at: <https://www.kino-teatr.ru/kino/movie/euro/10167/annot/> (accessed 12.28.2023). (In Russ.).

References

- Alyunina, Yu. M. (2022). Where do the wild things live? Corpus method to detect anglicisms and their derivatives on Russian Internet. *Bulletin of Tomsk State University. Philology*, 80: 5—29. DOI: 10.17223/19986645/80/1. (In Russ.).
- Anisimov, V. E., Borisova, A. S., Kalinnikova, E. D. (2022). Images of Traditional and Modern Tuva in Ethnographic Cinema (based on the film of S. Rodnina and A. Menshov “Tyva Planet”). *New studies of Tuva*, 1: 183—197. DOI: <https://doi.org/10.25178/nit.2022.1.12>. (In Russ.).
- Anisimov, V. E., Borisova, A. S., Conson G. R. (2019). Linguocultural Localization of Movie Titles. *Russian Journal of Linguistics*, 23 (2): 435—459. DOI: 10.22363/2312-9182-2019-23-2-435-459. (In Russ.).



- Astashina, D. A., Borshchevsky, I. S., Gaidash, N. V. (2021). *Specialist in translation and media accessibility: a framework of competencies*. Kazan: Publishing house “Buk”. 46 p. ISBN 978-5-00118-752-3. (In Russ.).
- Bednarek, M. (2023). *Language and Characterisation in Television Series: A corpus-informed approach to the construction of social identity in the media*. Amsterdam: John Benjamins. 279 p.
- Bednarek, M. (2018). *Language and Television Series: A Linguistic Approach to TV Dialogue*. Cambridge: Cambridge University Press. 321 p.
- Cintas, J. D., Remael, A. (2007). *Audiovisual Translation: Subtitling*. London: Routledge. 284 p.
- Kozloff, S. (2000). *Overhearing Film Dialogue*. Berkley — Los Angeles — London: University of California Press. 335 p. ISBN 0-520-22138-9.
- Lukić, N. (2016). Audiovisual Translation in Serbia: History and Professional Considerations. *Hieronymus*, 3: 64—89.
- Makhov, A. E. (2018). “Close reading” in the face of “big data”: the problem of method in modern English-language literary criticism. *Social and humanitarian sciences. Domestic and foreign literature*, 3: 11—32. (In Russ.).
- Malenova, E. D. (2021). Audiovisual translation in the USSR and Russia: a brief historical retrospective. *Vestnik NGLU. Language and culture*, 3 (553): 61—83. DOI: 10.47388/2072-3490/lunn2021-55-3-61-83. (In Russ.).
- Malenova, E. D. (2017). Theory and practice of audiovisual translation: domestic and foreign experience. *Communicative research*, 2 (12): 32—46. (In Russ.).
- Moretti, F. (2016). *Distant reading*. Moscow: Publishing House of the Gaidar Institute. 352 p. ISBN 978-5-93255-446-3. (In Russ.).
- Perego, E., Laskowska, M., Matamala, A. (2016). Is Subtitling Equally Effective Everywhere? A First Cross-national Study on the Reception of Interlingually Subtitled Messages. *Across Languages and Cultures*, 17 (2): 205—229.
- Smolan, R., Erwit, J. The Human Face of Big Data. *Filmplatform*. Available at: https://www.filmplatform.net/wp-content/uploads/2016/03/FINAL_HFOBDSStudyGuide-1.pdf. (accessed 04.01.2024).
- Sukhaya, E. V. (2010). Types of sources of film speech as the basis of linguistic research. *Questions of modern linguistics*, 2: 68—74. (In Russ.).
- Tröhler, M., Kirsten, G. (2018). *Christian Metz and the Codes of Cinema. Film Semiology and Beyond*. Amsterdam: Amsterdam University Press. 484 p.
- Urtmintseva, M. G., Skachkova, V. N. (2022). Voice-Over Text in Documentary Film Discourse. *Nauchnyi dialog*, 11 (6): 334—349. <https://doi.org/10.24224/2227-1295-2022-11-6-334-349> (In Russ.).

*The article was submitted 18.01.2024;
approved after reviewing 11.03.2024;
accepted for publication 18.03.2024.*