

## АВТОМАТИЧЕСКИЙ АНАЛИЗ МЕДИЙНЫХ ТЕКСТОВ

УДК 81'322.4: 316.472.47

### Автоматический голосовой перевод в социальных медиа: анализ воспринимаемого качества перевода тревел-влогов платформы YouTube

*А. Ю. Калинин*

Московский государственный университет им. М. В. Ломоносова,  
Российская Федерация, 119991, Москва, Ленинские горы, 1

**Для цитирования:** Калинин А. Ю. (2023). Автоматический голосовой перевод в социальных медиа: анализ воспринимаемого качества перевода тревел-влогов платформы YouTube. *Медиа-лингвистика*, 10 (4), 568–597. <https://doi.org/10.21638/spbu22.2023.408>

Одной из функциональных задач социальных медиа как средств цифрового взаимодействия является обеспечение медиадоступности, т. е. создание условий для свободного доступа пользователей к мультимедийному контенту платформ компьютерно-опосредованной интеракции. В условиях глобальной сетевой коммуникации среди препятствий к такому доступу не последнее место занимают языковые ограничения, связанные с использованием при создании медиаконтента лингвистического кода, отличного от языка потенциального пользователя. С учетом безусловного доминирования английского языка в интернет-пространстве речь главным образом идет о расширении сетевой аудитории за счет индивидов, не владеющих указанным языком на уровне, достаточном для беспрепятственного пользования социальными медиа, в частности просмотра и понимания аудиовизуальных произведений, размещаемых на видеохостингах и прочих онлайн-платформах. Поскольку пользовательский видеоконтент отличается, как правило, индивидуальным характером производства, бесплатным доступом и ориентированностью на целевую аудиторию адресатов, проявляющих интерес к определенной тематике, независимо от их национальной и языковой принадлежности, аудиовизуальный перевод, выполненный профессиональным переводчиком, в таких условиях оказывается невозможным. По этой причине в качестве способа преодоления языковых барьеров наиболее передовые интернет-порталы все чаще начинают внедрять сетевые решения автоматического (машинного) перевода видео, ставшего возможным с развитием цифровых лингвистических технологий. Эмпирическую базу настоящего исследования составили данные двухступенчатого эксперимента по субъективной оценке воспринимаемого качества автоматического голосового перевода англоязычных тревел-влогов на русский язык. Эксперимент включал опрос

потенциальных потребителей такого перевода, а также экспертную оценку, выполняющуюся профессиональными аудиовизуальными переводчиками. Анализ полученных результатов позволяет сделать вывод о принципиальной приемлемости автоматического перевода как средства обеспечения языковой медиадоступности в социальных медиа и сформулировать некоторые гипотезы относительно путей дальнейшего повышения его качества.

*Ключевые слова:* медиадоступность, аудиовизуальный перевод, автоматический голосовой перевод, оценка качества перевода, тревел-влог.

## Постановка проблемы

За последние полтора десятилетия популярность видеоматериалов о путешествиях и приключениях, размещаемых на различных платформах социальных медиа, достигла рекордного уровня. За годичный период 2013–2014 гг. количество просмотров видео, посвященных путешествиям и туризму, которые представлены на каналах видеохостинга YouTube, выросло сразу на 118 % [Crowel, Gribben, Loo 2014], в 2017–2018 гг. этот показатель составил 41 %<sup>1</sup>. Именно YouTube, занимающий второе место по популярности среди интернет-сайтов в мире<sup>2</sup>, является одновременно крупнейшим ресурсом видеоконтента: платформа имеет 2 млрд активных пользователей ежемесячно и более 30 млн ежедневных пользователей, а средняя продолжительность просматриваемого видео составляет порядка 1 млрд часов<sup>3</sup>. Концептуальную особенность YouTube, отличающую его от ряда других социальных медиа, составляет конфигурация социального графа: в центре структуры располагается онлайн-контент, вокруг которого строятся коммуникационные связи между пользователями [Wattenhofer, Wattenhofer, Zhu 2012].

С точки зрения пользовательских предпочтений в общей структуре видеоконтента YouTube, посвященного путешествиям и туризму, лидирующие позиции принадлежат жанру **влога** (видеоблога): 48 % потребителей интернет-контента тревел-тематики являются подписчиками YouTube-каналов авторства независимых блогеров. Для сравнения: лишь 19 % подписаны на веб-каналы о путешествиях, принадлежащие специализированным или общетематическим СМИ, и 16 % — каналы туристических компаний и агентств [Crowel, Gribben, Loo 2014]. В основе подобных предпочтений, очевидно, лежат возможность компьютерно-опосредованной интеракции с автором канала, а также «желание подписчиков услышать индивидуальные истории и узнать о личном опыте “таких же как они людей”» [Crowel, Gribben, Loo 2014]. С другой стороны, среди всех специализированных видеоблогов, представленных во Всемирной сети, тревел-влоги входят в тройку наиболее популярных тематик (наряду с бьюти-влогами и гейм-влогами)<sup>4</sup>. Именно пользовательский (user-generated) видеоконтент социальных медиа побудил 86 %

<sup>1</sup> Views of travel-related videos increased 41 % in August and September 2018 compared to 2017. (2018). Think with Google. Электронный ресурс <https://www.thinkwithgoogle.com/consumer-insights/consumer-trends/travel-video-view-statistics/>.

<sup>2</sup> Top Websites Ranking. (2023). Similarweb. Электронный ресурс <https://www.similarweb.com/top-websites/crowdriff.com/resources/blog/travel-statistics>.

<sup>3</sup> 84 YouTube Statistics You Can't Ignore in 2023. (2023). Invideo. Электронный ресурс <https://invideo.io/blog/YouTube-statistics/>.

<sup>4</sup> What are the most popular types of vlogs? (2022). Adobe. Электронный ресурс <https://www.adobe.com/creativecloud/video/hub/ideas/most-popular-types-of-vlogs.html>.

зрителей (92 % среди поколения миллениалов) проявить интерес к тому или иному географическому объекту или туристическому направлению<sup>5</sup>. Как следствие — растет и личная популярность тревел-влогеров, создающих наиболее удачный, с точки зрения потребителей, контент: так, на YouTube-канал *Nas Daily* мирового лидера тревел-влогинга 2022 г. Нусейра Ясина подписаны более 9 млн чел.<sup>6</sup> В Рунете показатели немного скромнее: на момент написания статьи среди русскоязычных тревел-влогов лидирующую позицию занимал канал украинца Антона Птушкина с 5,5 млн подписчиков<sup>7</sup>.

Теоретически единственными ограничениями для ознакомления с содержанием видеоблогов, размещаемых на платформе YouTube, являются наличие доступа к сети Интернет и любого экранного устройства, с помощью которого к ней можно подключиться. На практике существует целый ряд ограничений политического, юридического и в том числе лингвистического характера. Казалось бы, YouTube имеет более 100 локальных версий по всему миру, что позволяет ориентироваться на платформе на более чем 80 различных языках. Например, в России, по данным ВЦИОМ, за 2019 г. приблизительно 85 млн чел. (58 % населения) просматривают видео, размещенные на означенном хостинге (25 % — ежедневно), около 13 % россиян — активные пользователи YouTube, которые подписаны на различные каналы (12 %), реагируют и участвуют в обсуждениях и комментировании контента (6 %), делятся видео с другими (7 %) или размещают здесь свои собственные материалы (2 %). Число российских пользователей платформы, проявляющих интерес к видео о путешествиях, составляет порядка 11 млн<sup>8</sup>.

В то же время русскоязычные видео, хотя и занимают 4-е место по продолжительности ежедневно загружаемого на YouTube контента, составляют лишь 4 % объема данных, тогда как на долю английского языка приходится около 52 % видеороликов<sup>9</sup>. С другой стороны, согласно рейтингу EF English Proficiency Index, Россия располагается на 51-м месте из 112 стран по количеству жителей, владеющих английским языком (около 7,5 млн чел.)<sup>10</sup>, при этом достоверно установить, является ли уровень владения языком среди указанной группы достаточным для свободного просмотра аутентичного видео, не представляется возможным. Среднее положение РФ в данном рейтинге однозначно указывает на то, что Россия — не единственная страна, где основная масса населения не является англоговорящей.

Вышеупомянутые факты имеют непосредственное отношение к проблеме **медиадоступности**, которая длительное время рассматривалась исследователями ис-

<sup>5</sup> Statistics That Prove Social Media Will Influence Your Next Travel Destination (2019). Nosto. Электронный ресурс <https://www.nosto.com/blog/social-media-influence-travel-decisions/>.

<sup>6</sup> Nasdaily. Bring people together (2023). Social Blade. Электронный ресурс <https://socialblade.com/YouTube/c/nasdaily>.

<sup>7</sup> Антон Птушкин (2023). YouTube. Электронный ресурс <https://www.YouTube.com/c/ptuxermann/about>.

<sup>8</sup> YouTube — «телевидение» XXI века (2019). ВЦИОМ. Электронный ресурс <https://wciom.ru/analytical-reviews/analiticheskii-obzor/YouTube-televidenie-xxi-veka?ysclid=l77nf74dov856472267>.

<sup>9</sup> Spicer, A. (2022). Top Languages on YouTube [All The Stats!]. Alan Spicer. Электронный ресурс <https://alanspicer.com/top-languages-on-youtube>.

<sup>10</sup> EF English Proficiency Index: Рейтинг 112 стран и регионов по уровню владения английским языком (2021). Education First. Электронный ресурс <https://www.ef.ru/assetscdn/WIBIwq6RdJvcD9bc8RMd/cefcom-epi-site/reports/2021/ef-epi-2021-russian.pdf>.

ключительно в аспекте патологий восприятия (меры, направленные на предоставление доступа к аудиовизуальному контенту индивидам с расстройствами зрения и слуха). В настоящий момент медиадоступность понимается как обеспечение «доступа к медиапродуктам, медиауслугам и медиасреде для всех, кто не может или не в состоянии должным образом получить доступ к ним в их оригинальной форме» [Greco 2016: 23] (перевод наш. — А. К.). Тем самым межъязыковой барьер должен квалифицироваться как существенное препятствие к гарантированной медиадоступности, требующее преодоления [Romero-Fresco 2018].

Очевидно, что устранить означенные лингвистические ограничения призван аудиовизуальный перевод. Несмотря на все различия, обусловленные ситуациями, видами и методами аудиовизуального перевода [Gambier 2013], их многообразие может быть сведено к двум основным **формам предъявления переводного контента** зрительской аудитории: графической (субтитры) и голосовой (переозвучивание) [Калинин 2019]. В зависимости от характера оригинального материала, бюджета проекта и пожеланий инициатора перевода аудиовизуальные переводчики работают по заказу кинокомпаний, телеканалов и других провайдеров аудиовизуальных произведений над созданием межъязыковых субтитров или скриптов для различных видов переозвучивания (дубляж, закадровое озвучивание, синхронный перевод). Принципиально иная ситуация возникает в социальных медиа. Даже если гипотетически независимые влогеры могли бы позволить себе профессиональный перевод своих видео в финансовом плане, количество языков потенциальных адресатов непрерывно пополняющегося видеоконтента, который размещается в открытом доступе, совершенно исключает такой вариант. На некоторых социальных платформах фрагментарно реализуется концепция переводческого краудсорсинга и любительского («пользовательского») перевода [Desjardins 2017]. Она, однако, не способна справиться с колоссальными объемами иноязычного контента, который может вызвать интерес пользователей. Тем самым практически единственная возможность языковой локализации пользовательского видеоконтента в социальных медиа — автоматический (машинный) перевод «по запросу». Анализу воспринимаемого качества голосовой формы именно такого перевода англоязычных тревел-влогов, адресованного русскоговорящим пользователям, и посвящено настоящее исследование.

## История вопроса

Несмотря на то что видеоблоги в целом и тревел-влоги в частности представляют собой социокультурное явление последних 10–15 лет, данная проблематика уже успела найти свое отражение в ряде отечественных и зарубежных работ. Среди исследований, посвященных влогосфере, можно усмотреть несколько концептуальных направлений и методологических подходов.

В рамках **эпистемолого-таксономического** направления исследователи пытаются осмыслить эпистемологический статус объекта изучения, определить место видеоблога среди родственных форм компьютерно-опосредованной коммуникации [Щипицина 2017] в социальных медиа и разработать жанрово-тематическую типологию разнородного пользовательского видеоконтента [Текутьева 2016]. Так, В. А. Лушиков и М. В. Терских полагают, что тревел-видеоблоги образуют отдель-

ный тип сетевого видеоконтента, а в жанровом плане относятся к категории **влоггов**. Причем влог в данном случае не просто сокращенный вариант обозначения видеоблога, а «...**жанр** видео, в котором блогер рассказывает о своей повседневной жизни, не находясь при этом в статичном положении перед камерой» [Лущиков, Терских 2018: 65]. Авторы отмечают также, что сфера видеоблогинга активно развивается, заимствуют жанры из тележурналистики и других экранных искусств, и по этой причине «...жанровая классификация видеоблогов постоянно будет нуждаться в дополнении» [Лущиков, Терских 2018: 57]. Этот вывод перекликается с позицией С. Херринг, которая считает жанры цифровой видеокommunikации интуитивными, реконфигурируемыми и подвижными [Herring 2013].

Представителей **коммуникативно-социологического** подхода интересует феномен видеоблога в аспекте онтологии мотивов его становления и развития как формы социальной интеракции [Микрюков, Саркисова 2020], а также природа деятельности влогера и его роль в динамике социальных процессов. Так, И. В. Показаньева отмечает, что рост популярности видеоблогов связан с фактором смещения внимания потребителей медиаконтента в сторону «всего личного», частной истории, которая стала лейтмотивом существования в сетевом пространстве [Показаньева 2015]. Особое внимание в рамках данного подхода уделяется вопросам соотношения видеоблогинга как формы активности в социальных медиа и видеожурналистики как профессиональной деятельности. Е. В. Медведева утверждает, в частности, что «...влогер, как и журналист, обычно не является единоличным создателем отдельных сообщений и влога в целом <...> [что. — А. К.] сближает традиционную авторскую программу, например, на радио или телевидении, и влог» [Медведева 2021]. Представляется, что данный вывод сделан на основе анализа видеоинтервью и подобных ему форм видеоблогинга, фактически перекочевавших в интернет-пространство с телеканалов, и не учитывает жанровой вариативности видеоблогов. Например, в жанре тревел-влога автором, как правило, выступает индивидуальный пользователь интернета, который делится с аудиторией своими собственными мыслями, чувствами и опытом в ходе посещения определенного места и может затрагивать любые темы по своему усмотрению, тогда как тревел-журналист в своей работе изначально ограничен полученным заданием и этическим кодексом СМИ, которое он представляет [Показаньева 2015].

Еще одним важным аспектом рассмотрения является изучение специфики аудитории потребителей влог-контента и характера их активности во влогосфере. Говоря об особенностях, отличающих телевизионную аудиторию от адресатов YouTube-каналов, Е. В. Медведева справедливо отмечает, что последние имеют «...возможность формирования личного программного потока: за счет подписки на каналы по собственному выбору...» [Медведева 2021]. В свою очередь Д. Азария, анализируя сущность коммуникативной интеракции в социальных медиа, подчеркивает, что такая платформа, как YouTube, представляет собой не только видеохостинг, но и «видеошеринг» [Azariah 2016], т. е. пространство сетевого мультипликативного обмена данными. Не меньшее внимание исследователей социологии влогосферы привлекают вопросы воздействия формы и содержания видеоконтента на аудиторию и их влияния на потребительские стратегии пользователей [Lee, Watkins 2016; Choi, Lee 2019]. Применительно к тревел-влогам констатируется корреляция между объемом культурно и социально значимой информации в блоге, а также потенциалом

идентификации зрителей с влогером, с одной стороны, и вероятностью выбора пользователями представленного в видеоблоге региона в качестве направления собственных путешествий — с другой [Chen, Guo, Pan 2021; Xu et al. 2021].

**Лингвосемиотическое** направление в изучении видеоблогов рассматривает видеоконтент социальных медиа как массив особых текстов поликодовой и полимодальной природы [Benson 2015]. В рамках направления исследуются как характер взаимодействия различных знаковых систем в структуре влога [Гребенев, Шаюк 2018], так и специфика вербального компонента видеонаратива. М. Фробениус изучала дискурсивные стратегии, используемые влогерами в монологической речи для установления и поддержания виртуального контакта с аудиторией: приветствия, обращения, языковые маркеры смыслового членения высказываний и т. п. [Frobenius 2011]. Э. Вернер фокусирует внимание на разнообразии риторических практик видеоблогов, объединенных эксплицируемой «псевддиалогичностью» [Werner 2012]. С. В. Мурсекаева, анализируя дискурсивный статус влога в целом, характеризует его как разновидность бытийного персонального дискурса [Мурсекаева 2017]. М. Йохансон отмечает, что YouTube-влоги открывают перед лингвистами широкое и пока малоизученное поле для исследований прагматики типичных коммуникативных актов, поскольку в них представлены актуальные тенденции неформального монологического и диалогического дискурса в сочетании с реализацией паттернов невербального поведения [Johansson 2017]. Основную лингвистическую ценность влога автор усматривает в обилии и доступности материала для изучения разговорной или квазиразговорной спонтанной речи, которые в реальной жизни эфемерны и сложно фиксируются.

**Транслатологический** аспект изучения социальных медиа наименее разработан. Основной корпус работ в рамках данного направления составляют исследования так называемого «любительского» субтитрирования и дублирования (fansub & fandub) игровых, анимационных фильмов и сериалов, размещаемых на видеохостингах [Nord, Khoshsaligheh, Ameri 2015; Vazquez-Calvo, Shafirova, Cassany 2019; Baños 2020]. Подробный обзор литературы, посвященной любительскому аудиовизуальному переводу, представлен в [Pérez-González 2019].

В монографии Р. Дежарден [Desjardins 2017] перевод в социальной среде рассматривается в теоретической, дидактической и практической плоскостях. Автор отмечает, что в рамках социальных медиа профессиональное переводческое сообщество задействовано преимущественно в сфере локализации информационного и рекламного контента, размещаемого по заказу крупных медиаигроков, стремящихся к расширению своего присутствия на рынках разных стран. Пользовательский же контент находится вне сферы их внимания, что вызвало к жизни концепцию так называемого «**пользовательского перевода**» (user-generated translation), т. е. практику, субъектами которой являются индивид или виртуальное краудсорсинговое сообщество, выполняющие функции переводчиков на волонтерских началах [Desjardins 2017]. Такие «переводчики-пользователи» выступают в качестве инициаторов, заказчиков и исполнителей перевода онлайн-контента, авторство которого принадлежит либо им самим, либо другим пользователям той или иной медиаплатформы. Данная модель подробно изучена в эмпирическом исследовании Е. С. Краснопеевой [Krasnopereyeva 2018], в рамках которого анализировались генезис, эволюция и принципы функционирования переводных каналов

русскоязычного сегмента видеохостинга YouTube. Исследовательнице удалось показать, что модель пользовательского перевода в полной мере отражает онтологические свойства социальных медиа как пространства интерактивной сетевой коммуникации, в котором пользователи не только потребляют видеоконтент, произведенный другими, и реагируют на него посредством апробационного функционала (лайков и комментариев), но могут также выступать в качестве индивидуальных или коллективных (со)авторов новых языковых версий исходного видеоматериала.

В то же время производственный потенциал таких каналов крайне лимитирован: перевод здесь представляет для их создателей, по сути, рекреационную деятельность, с чем связаны, в частности, нерегулярность или относительно низкая периодичность размещения новых переводных материалов, равно как и тот факт, что выбор объекта для субтитрирования или переозвучивания отражает, как правило, личные интересы и предпочтения пользователя<sup>11</sup>. Кроме того, возможны ограничения размещения уже переведенных видео на подобных каналах, связанные с наличием авторских прав у создателей оригинальных аудиовизуальных произведений. И, наконец, главная проблема — кардинальная асимметрия между сравнительно небольшим числом волонтеров-переводчиков, развивающих подобные проекты, и поистине колоссальным объемом видеоконтента, ежедневно заливаемого на платформу YouTube<sup>12</sup>. С учетом всего вышеизложенного единственным рациональным решением, направленным на обеспечение лингвистической доступности видео на онлайн-платформах, подобных YouTube, следует признать **автоматический (машинный) перевод**.

В литературе представлен ряд работ, посвященных языковой политике отдельных соцсетей и той роли, которую призван сыграть в ней автоматический перевод [Lenihan 2014; Almahasees, Jaccomard 2020; и т. д.]. Х. Каррера и соавторы полагают, что машинный перевод является оптимальным вариантом межъязыкового трансфера пользовательского контента в социальных медиа, принимая во внимание фундаментальные принципы их функционирования: доступность, открытость, высокую активность участников сетевой коммуникации и необходимость оперативной подачи информации [Carrera, Beregovaya, Yanishevsky 2009].

Исходя из соображений лингвистической доступности видео, размещаемых на своей платформе, YouTube в настоящее время предоставляет для многих роликов опцию **скрытых внутриязыковых субтитров** (closed captioning) с последующим переводом по запросу на 300 языков. Функционал основан на технологии нейронного машинного перевода текста в реальном времени **Google translate**. В данной технологии всплывающие на экране переводные субтитры представляют собой продукт машинного преобразования «текста в текст» (**T2T**), которому предшествует стадия автоматического распознавания содержания звуковой дорожки (**ASR**). Каково бы ни было качество таких переводных субтитров, предлагаемое решение, по всей вероятности, существенным образом расширяет аудиторию зрителей пользовательского видеоконтента за пределами англоязычных

---

<sup>11</sup> Показательным представляется авторское описание одного из каналов перевода иноязычных видео: «Перевожу интересный мне зарубежный контент в понятную нашему уху форму». Zébra. (2016). YouTube. Электронный ресурс <https://www.youtube.com/@zyobr/about>.

<sup>12</sup> Как результат — большинство каналов пользовательского перевода, фигурировавших в исследовании Е. С. Краснопеевой, в настоящее время закрыты или малоактивны.

стран<sup>13</sup>. Вместе с тем статистика показывает, что более 70 % просмотров видео на YouTube осуществляются с использованием мобильных устройств, из которых 97 % — смартфоны<sup>14</sup>. При этом в специальной литературе имеется ряд указаний на то, что малые габариты экранов мобильных устройств не только существенно ограничивают размеры окна субтитров, но и могут значительно осложнить восприятие субтитрированной вербальной информации пользователями даже в случае так называемых «вшитых» субтитров [Gerber-Morón, Soler-Vilageliu, Castellà 2020], не говоря уже о всплывающих окнах субтитров автоматических. При высоком темпе речи, свойственном «разговорным» видеожанрам, смена субтитров на экране может происходить почти молниеносно, что также не способствует ни полноценной когнитивной обработке поступающей зрительно-вербальной информации, ни комфортному восприятию видеоролика в целом. Кроме того, еще со времен, когда фактически единственным видом аудиовизуального перевода являлся перевод фильмов для последующего проката в кинотеатрах и трансляций на телеканалах, во многих странах сложилась своего рода традиция предпочтений к той или иной форме предъявления переводного контента зрителям. Россия в этом отношении — «страна дубляжа» [Bogucki, Díaz-Cintas 2020]. Другими словами, отечественный зритель в ситуациях просмотра иноязычных аудиовизуальных произведений скорее скорее слушать озвученный перевод, нежели читать субтитры.

В этой связи крупнейшие компании-провайдеры сетевых информационных технологий ведут разработки систем автоматического перевода видео, в том числе в режиме реального времени (синхронный перевод интернет-трансляций), в которых переводной контент предъявляется реципиентам в звучащей форме. В 2021 г. интернет-портал «Яндекс» запустил сетевое решение **автоматического голосового перевода** (АГП) видео, в настоящее время доступное всем пользователям одноименного браузера. В качестве объекта перевода могут выступать видеоролики, размещенные на наиболее популярных хостингах открытого доступа, таких как YouTube, Vimeo и т.п. Указав адрес ссылки в адресной строке браузера «Яндекс» и выбрав опцию «перевод с помощью нейросетей», пользователь через считанные минуты (длительность подготовки перевода определяется размером исходного видеофайла) может перейти к просмотру видеоролика, сопровождающегося звучащим переводом на русский язык. Переводная аудиодорожка полностью синхронизирована с оригинальной, равно как и с видеорядом. Перевод может быть одно- или двухголосным, при этом автоматический выбор синтезированного голоса для озвучивания осуществляется на основе анализа пола протагонистов в кадре: «мужской» голос — для протагонистов-мужчин, «женский» — для женщин. Немаловажно также, что АГП является виртуальным продуктом: переводная аудиодорожка хранится на сервере «Яндекс», а не размещается на альтернативном канале видеохостинга, что позволяет избежать нарушений авторских прав создателей оригинального контента.

<sup>13</sup> Примечательно, что даже в такой в целом англоязычной стране, как США, 21 % населения используют в частной жизни и предпочитают смотреть видео на иных языках, нежели английский (из них 62 % — на испанском): Get More YouTube Views with Foreign Language Subtitles (2014). Mini Matters. Электронный ресурс <https://www.minimatters.com/get-more-YouTube-views-with-foreign-language-subtitles/>.

<sup>14</sup> YouTube Statistics You Can't Ignore in 2023 (2023). Invideo. Электронный ресурс <https://invideo.io/blog/YouTube-statistics/>



Модель АГП, развиваемая компанией «Яндекс», схематично представлена нами на рисунке и относится к числу так называемых **каскадных** алгоритмов машинного перевода. Центральным звеном модели, так же как и в случае скрытых субтитров, является **T2T-преобразование**, т. е. собственно межтекстовый перевод, однако с учетом того, что как на входе, так и на выходе системы располагается звуко-речевой сигнал, такую модель можно условно обозначить как **S2T2T2S** (speech-to-text-to и text-to-speech)<sup>15</sup>.

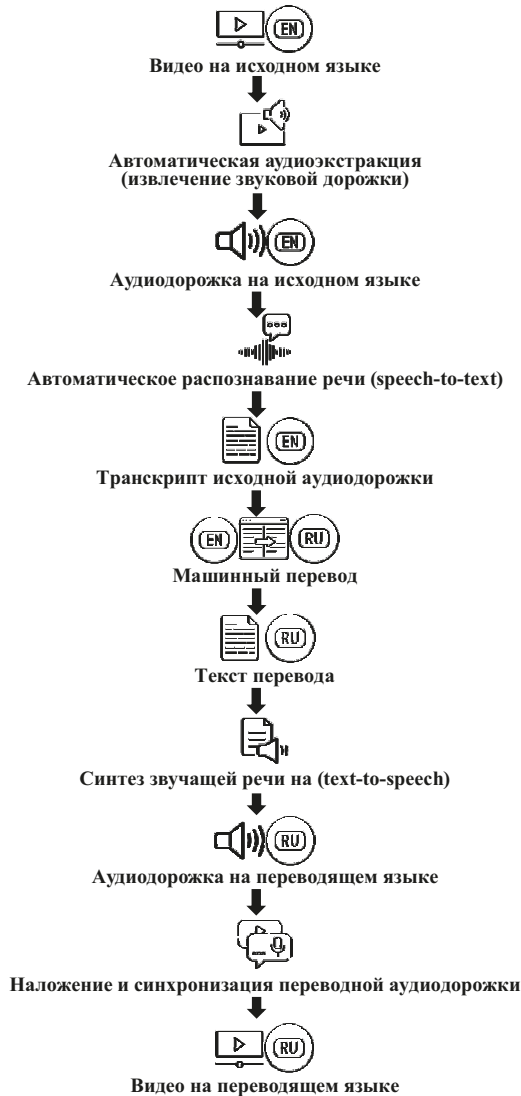


Рис. Типовая схема осуществления «каскадного» голосового перевода

<sup>15</sup> Компания Google разрабатывает альтернативную «сквозную» модель машинного перевода «речи в речь» на основе анализа широкополосных спектрограмм и минуя фазу текстовых преобразований [Jia et al. 2019], что не позволяет нам использовать аббревиатуру S2S для обозначения АГП от «Яндекса» как варианта каскадной модели.

Разумеется, здесь, как и во всех сферах использования автоматического перевода, ключевым является вопрос качества. В ситуациях машинного перевода функциональных мономодальных текстов, огрехи нивелируются или минимизируются с помощью постредактирования, выполнение которого возлагается на переводчика-редактора. В случае АГП, инкорпорируемого в структуру социомедийной коммуникации, постредактирование не представляется возможным. АГП сетевых видео является конечным виртуальным продуктом, качественные показатели которого в заданном формате не могут быть оптимизированы редактором, что в известном смысле ужесточает требования к **воспринимаемому адресатом качеству перевода**.

Общетеоретическим и прикладным вопросам оценивания качества перевода посвящено внушительное количество работ (см. подробный обзор в [Нап 2020]). В том, что касается машинного перевода, его качество в принципе может оцениваться как экспертным методом, так и с использованием автоматических метрик. Поскольку первый путь часто рассматривается как ресурсозатратный и субъективный, в большинстве эвлюативных исследований применяется автоматическая оценка: систематическое сопоставление оцениваемого перевода с референтным («эталонным» текстом на переводящем языке) на предмет лексико-синтаксического подобия [Castilho et al. 2018], а наиболее популярным инструментом оценивания выступает N-граммный алгоритм BLEU (обзор и принципы работы различных метрик приведены в: [Соснин, Балакина, Кацкихин 2022]). В то же время применение подобных метрик к оценке качества голосового перевода аудиовизуальных произведений в целом и АГП видеоконтента социальных медиа в частности представляется малоперспективным по целому ряду причин.

Во-первых, алгоритмы автоматического оценивания предназначены для анализа исключительно вербально-текстового компонента в терминах поверхностной структуры (лексические совпадения, близость синтаксической организации), они не позволяют оценить паралингвистические параметры звучащего перевода (ритм, интонацию, качество голоса и т. п.). Во-вторых, математические метрики оценивания не имеют доступа к визуальному контексту объекта перевода, что затрудняет вынесение системой суждений о корректности ситуативного выбора межъязыкового соответствия, опорой для которого мог бы служить видеоряд [Burchardt et al. 2016]. В этом смысле и экспертный метод может оказаться малорепрезентативным в тех случаях, когда непосредственным объектом оценивания выступает не целостное аудиовизуальное произведение, а лишь его транскрипт (вторичная письменная фиксация вербального компонента) [Gambier 2019]. Наконец, специфика лексикона и в особенности синтаксической организации устнопорождаемой разговорной речи может существенно затруднять автоматическую сегментацию транскриптов и приводить к деградации релевантности сопоставления текстовых сегментов [Burchardt et al. 2016]. Последнее препятствие отчасти преодолимо путем длительного обучения системы оценивания качества транслята на обширных корпусах разговорных текстов, однако такие мероприятия требуют значительных временных и информационных ресурсов. Все эти соображения заставляют ученых при оценке качества машинного перевода аудиовизуальных произведений использовать комплексный экспертно-математический подход [Fernández, Matamala 2015]

или разрабатывать собственные многомерные и громоздкие метрики оценивания [Burchardt et al. 2016].

«Ручная» (экспертная) оценка предполагает сопоставление текстов оригинала и перевода в терминах эквивалентности (на лексико-семантическом, стилистическом, реже — синтаксическом уровнях), а также анализ языковых качеств перевода, прежде всего его соответствие норме и узусу переводящего языка [Rivera-Trigueros 2022]<sup>16</sup>. В то же время Дж. Хаус полагает, что лингвистический анализ не может выступать в качестве единственного инструмента оценивания перевода. Без учета социальной (межличностной) функции, которую переводной текст выполняет по отношению к реальным или потенциальным получателям, оценка его качества окажется однобокой, абстрактной и выхолощенной. Иными словами, для формирования полноценного представления о качестве перевода важны не только оценки экспертов-лингвистов, но и субъективные суждения не обладающих профессиональными компетенциями пользователей, даже если последние ограничиваются упрощенной дихотомией «плохой/хороший перевод» [House 2015].

Отдельным аспектом проблемы является форма предъявления переводного контента реципиентам. Говоря о пользователях аудиовизуального перевода, И. Гамбье отмечает, что в подобной ситуации целесообразно учитывать не только характер «рецепции», но и особенности «перцепции» материала [Gambier 2009]. Очевидно, что в подавляющем большинстве работ, посвященных воспринимаемому качеству перевода, анализировался письменный перевод. Значительно реже изучалось восприятие устного, главным образом синхронного, перевода [Kurz 2001]. Однако эти исследования показали важную роль паралингвистических параметров (ритм, интонация, тембр голоса и т. п.) в создании впечатления о качестве перевода при восприятии на слух [Aís, Spinolo, Garwood 2016]. В этом отношении АГП видео приближается к устному переводу, поскольку здесь переводной контент также предъявляется пользователю в виде акустического сигнала. Вместе с тем аудиовизуальное восприятие является по определению бимодальным и многоканальным и, хотя переводу подвергается исключительно аудиодорожка, зрители имеют возможность сопоставлять содержание звучащего вербального компонента с динамическим изображением (видеорядом). Отсюда, с одной стороны, наличие визуальных каналов информации облегчает восприятие видеоролика и дает переводчику-человеку «подсказки», способствующие адекватной интерпретации исходного вербального компонента, а с другой, может выступать в качестве фактора, осложняющего восприятие, в тех случаях, когда изображение и результат машинного перевода конфликтуют между собой.

Методология исследований АГП чрезвычайно мало разработана. К. Ортис-Бойш и А. Матамала, изучавшие в контрастивном аспекте восприятие профессионального и машинного перевода с постредактированием научно-популярных фильмов, предложили трехуровневую модель оценивания (рейтингования) с привлечением в качестве участников зрителей, переводчиков-экспертов и специалистов по дубляжу и закадровому озвучиванию [Ortiz-Voix, Matamala 2015]. Поскольку настоящее исследование имело основной задачей анализ **воспринимаемого качества**

---

<sup>16</sup> В ситуациях звучащей речи в число параметров оценивания включаются также и паралингвистические характеристики, о которых мы упоминали выше.

АГП выпусков англоязычных тревел-влогов, мы при разработке его методов во многом опирались на процедуру, предложенную в указанной работе, остановив свой выбор на методах двойного субъектно-реципиентного и экспертно-аналитического оценивания.

## Описание методики исследования

Материалом для компиляции корпуса настоящего исследования послужили три выпуска тревел-влогов, размещенных на личных каналах видеохостинга YouTube и принадлежащих авторству англоязычных видеоблогеров (по одному видео-сюжету из каждого влога):

1) “Kandy to Ella train”, видео канала PsychoTraveller британки Aly; продолжительность — 12:21<sup>17</sup>;

2) “Why should you visit Norway — Unspoken paradise”, видео канала Lost LeBlanc канадца Christian Leblanc; продолжительность — 13:30<sup>18</sup>;

3) “I took an extreme Russian bath and I lost my mind”, видео, размещенное американцем Drew Binsky на одноименном канале; продолжительность — 11:31<sup>19</sup>.

Ссылки на видео были загружены в веб-браузер «Яндекс» и переведены в автоматическом режиме. Оригинальные и переводные видеодорожки были записаны на электронный носитель и затем транскрибированы с использованием процедуры полуавтоматической транскрибации (автоматическая транскрибация с постредактированием)<sup>20</sup>. Один из видеороликов (№ 3) был также переведен профессиональным аудиовизуальным переводчиком для последующего закадрового озвучивания. Текст перевода был преобразован в звучащую речь с использованием системы T2S открытого доступа («Яндекс») и записан с на виртуальном диктофоне **Моо0**. При этом синтезированный голос совпадал с параметрами «голоса», созданного на основе речевых сэмплов чат-бота **Алиса**. Эта манипуляция была необходима в связи с тем, что именно указанным голосом озвучиваются говорящие-женщины в голосовом видеопереводчике «Яндекс». Полученный аудиофайл был сохранен в качестве звуковой дорожки одного из аудиоканалов и синхронизирован с оригинальным изображением, при этом звук другого аудиоканала (оригинальное аудио) был приглушен. Таким образом, корпус исследования составили три видео, семь аудиодорожек (три исходные, три автоматически переведенные и одна переведенная профессиональным переводчиком), шесть транскриптов (три оригинальных звучащих текста и три переведенных с использованием ресурса «Яндекс.Переводчик»).

Эксперимент, проводившийся в форме опроса, имел своей целью охарактеризовать воспринимаемое качество АГП англоязычных тревел-влогов на русский язык в холистической и аналитической перспективе.

<sup>17</sup> PsychoTraveller (2019). Kandy to Ella train (2<sup>nd</sup> Class). *YouTube*. Электронный ресурс <https://www.YouTube.com/watch?v=rP1Pdnntgw0>.

<sup>18</sup> Lost LeBlanc (2019). Why you should visit Norway — Unspoken paradise. *YouTube*. Электронный ресурс <https://www.YouTube.com/watch?v=lkepF2uK1sg>.

<sup>19</sup> Drew Binsky (2022). I took an extreme Russian bath and I lost my mind. *YouTube*. Электронный ресурс <https://www.YouTube.com/watch?v=JY-PNYkYccs&t=2s>.

<sup>20</sup> Оригинальные аудиодорожки (на английском языке) к видео всех трех каналов на платформе YouTube снабжены скрытыми автоматическими субтитрами, которые были отредактированы с целью устранения расхождений между звучащей формой текста и транскриптами.

В качестве участников к эксперименту были привлечены две группы информантов:

1) 48 носителей русского языка, не имеющих лингвистического образования и не занимающихся переводом как профессиональной деятельностью;

2) 3 профессиональных аудиовизуальных переводчика, работающих в языковой комбинации «английский — русский» (опыт работы от 7 до 16 лет).

Отбор информантов первой группы проходил по результатам предварительного анкетирования. Первоначально в нескольких социальных сетях было размещено объявление, приглашающее к участию в исследовании на тему видеоблогов о путешествиях. Были получены заявки на участие в качестве респондентов от 83 человек. Анкетирование имело своей целью отобрать из числа добровольцев лиц, проявляющих интерес к теме индивидуальных путешествий, регулярно или sporadически просматривающих выпуски тревел-влогов в социальных медиа и не владеющих английским языком на уровне, достаточном для просмотра и уверенного понимания англоязычных видеороликов в оригинале. Указанные критерии могут рассматриваться как ключевые при идентификации целевой аудитории голосового перевода видеоблогов. Всего было отобрано 55 человек, свое участие в эксперименте окончательно подтвердили 48, из которых 27 — мужчины и 21 — женщины. Возраст информантов первой группы колеблется в пределах 18–49 лет при медиане 26 лет.

В качестве экспериментального задания участникам первой группы предлагалось просмотреть в режиме реального времени два выпуска тревел-влогов, сопровождаемых автоматическим голосовым переводом на русский язык от «Яндекса», и один видеоролик, перевод которого был выполнен профессиональным исполнителем, а затем озвучен с помощью сетевого решения T2S (см. выше). Информанты не посвящались в технологические особенности создания переводных аудиодорожек. По окончании просмотра каждого видеоролика участникам предлагалось дать ответы на первый раздел вопросов анкеты. Ответы на оставшиеся вопросы давались после просмотра всех трех видео. Данный этап эксперимента проводился с использованием средств видео-конференц-связи и инструментов онлайн-опросов в три сессии, продолжительностью около 90 мин каждая.

Опросник разрабатывался с целью получения данных, характеризующих три основных аспекта рецепции аудиовизуальных произведений: понимание, интерес / комфортность просмотра и предпочтения пользователей [Gambier 2019] — и содержал три категории вопросов:

1) вопросы с альтернативным выбором ответов, например: «*мне доводилось смотреть выпуски влогов на английском языке с русскими субтитрами*», «*я просматриваю только выпуски видеоблогов в русскоязычном сегменте Сети*» и т. п.;

2) вопросы закрытого типа, представляющие собой утверждения с множественным выбором из пяти вариантов по шкале Ликерта от «*абсолютно не согласен*» до «*полностью согласен*», например: «*мне было интересно смотреть данное видео*», «*я с удовольствием смотрел(-а) бы выпуски тревел-влогов с закадровым переводом на русский язык, если бы такая технология была доступной*»;

3) вопросы открытого типа и раздел дополнительных комментариев в свободной форме, например: «*Были ли в переводе моменты, которые остались вам непонятны? (Какие именно?)*». Ответы на вопросы данного типа в дальнейшем систематизировались с целью выявления рекуррентных впечатлений.

Второй группе информантов (профессиональные аудиовизуальные переводчики) в качестве стимулов были предложены не только видеоролики с переводными аудиодорожками, но и транскрипты оригинальных звуковых дорожек и их переводы на русский язык. Помимо ответов на вопросы анкеты, раскрывающих воспринимаемое качество перевода, экспертам предлагалось также отредактировать тексты переводов для последующего озвучивания, т. е. внести необходимые, на их взгляд, исправления. Оценивание качества перевода видео группой экспертов проводилось в два этапа. На первом этапе информанты осуществляли аналитическую оценку звучащих переводов в качестве «профессиональных зрителей», отвечая на вопросы анкеты. Участникам группы экспертов, в отличие от респондентов первой группы, предъявлялись последовательно видео с оригинальной и переводной аудиодорожками. Данный этап проводился в онлайн-режиме в три сессии, продолжительностью около 100 мин каждая. Опросник, предложенный группе аудиовизуальных переводчиков, содержал матрицу оценивания, которая была разработана с учетом профессионального статуса участников, и предполагал оценку параметров качества голосового перевода на основе перцептивных впечатлений по пятибалльной шкале.

На втором этапе эксперты осуществляли редактирование текстов в индивидуальном режиме в пределах недельного срока и представляли результаты своих правок в электронном виде. В завершение работы экспертам предлагалось сопроводить выполненные правки комментариями в свободной форме. Исправления, предложенные переводчиками для каждого текста, в дальнейшем анализировались и систематизировались в соответствии с категориями и параметрами, которые были сформулированы в матрице оценивания, использованной на предыдущем этапе.

## **Анализ материала и результаты исследования**

### **1. Социологический аспект.**

Как показал опрос представителей аудитории тревел-влогов (табл. 1), данный жанр видеоконтента социальных медиа может оказать значительное влияние на предпочтения зрителей в выборе направления собственных путешествий и их практическое планирование (77 и 67 % опрошенных соответственно). При этом большинство тех, кто интересуется тревел-влогами, считают себя активными пользователями социальных медиа (85 %), подписаны хотя бы на один тревел-канал (73 %), обращаются в поисках информации не только к видеохостингу YouTube, но и к другим платформам, на которых размещается видео (77 %). В то же время 58 % респондентов являются зрителями исключительно русскоязычных тревел-каналов. Оставшиеся 42 % при просмотре иноязычного контента пользуются скрытыми субтитрами на русском языке.

Разумеется, подобные данные во многом обусловлены методикой отбора участников опроса, о которой мы говорили выше. Больше половины опрошенных — 54 % мотивировали свой выбор недостаточным уровнем понимания английской речи на слух. Еще 40 % указали, что считают чтение субтитров фактором, осложняющим восприятие видеоряда аудиовизуального произведения. Сопоставляя вышеуказанную статистику с данными холистической оценки голосового перевода тревел-влогов (табл. 2), по итогам которой от 41 до 87 % информантов заявили, что

Таблица 1. Вовлеченность респондентов в тревел-влогосферу (n = 48)

Индикатор вовлеченности	Количество респондентов	
	чел.	%
Считают себя активными пользователями социальных медиа	41	85
Просматривают видеоконтент на каналах видеохостинга YouTube	48	100
Просматривают видеоконтент в иных социальных медиа	37	77
Регулярно просматривают выпуски тревел-влогов и подписаны хотя бы на один из них	32	73
Спорадически просматривают выпуски тревел-влогов	16	33
Оставляют комментарии, размещают ссылки на внешний контент на страницах тревел-влогов	14	29
Ориентируются на информацию тревел-влогов при выборе направлений собственных уже осуществленных или планируемых путешествий	37	77
Черпают из тревел-влогов практическую информацию по организации путешествий	32	67
Планируют в будущем стать тревел-влогерами	7	15
Являются зрителями влогов только русскоязычного сегмента социальных медиа	28	58
Просматривают выпуски влогов на английском языке с использованием внутриязыковых субтитров	9	19
Просматривают выпуски влогов на английском языке с использованием переводных субтитров	20	42
Не смотрят выпуски англоязычных тревел-влогов, поскольку не владеют английским языком на уровне понимания звучащей речи в достаточной степени	26	54
Не смотрят выпуски англоязычных тревел-влогов с субтитрами на русском языке, поскольку чтение субтитров отвлекает от визуального восприятия видеоряда	19	40
Слышали или читали о возможностях общедоступного голосового перевода видео от «Яндекса»	29	60
Просматривают выпуски влогов на английском языке с использованием голосового перевода	2	4

с удовольствием смотрели бы видео иноязычных блогеров с голосовым переводом, можно констатировать, что социальный запрос на перевод такого рода сформирован или, по крайней мере, находится в стадии своего формирования. Однако при том, что 60 % участников опроса известно о существовании общедоступного голосового перевода интернет-видео от «Яндекса», лишь двое (4 %) из 48 респондентов до проведения эксперимента использовали данную технологию при просмотре видео о путешествиях.

## 2. Холистическая оценка.

Примечательно, что по ходу эксперимента информанты несколько меняли свое отношение к автоматическому голосовому переводу: после просмотра первого видео количество тех, кто заявил, что скорее всего или однозначно будет

Таблица 2. Холистическая оценка качества голосового перевода (n = 48), %

Утверждение	Видео	1	2	3	4	5
Мне было интересно смотреть это видео	АГП1	–	–	15	47	18
	АГП2	–	6	23	52	19
	ПАО	–	–	–	35	65
Мне было понятно все, что говорилось влогером	АГП1	19	56	–	25	–
	АГП2	15	37	–	33	15
	ПАО	–	2	10	46	42
В основном все было понятно, но многие реплики звучали странно	АГП1	–	–	10	71	19
	АГП2	–	10	15	54	21
	ПАО	65	25	8	2	–
Голос «переводчика» звучал неестественно, «как у робота»	АГП1	–	31	17	40	12
	АГП2	8	40	15	21	16
	ПАО	35	52	10	3	–
У меня сложилось впечатление, что переведено не все, что говорил влогер	АГП1	–	25	71	4	–
	АГП2	10	25	63	2	–
	ПАО	19	23	33	–	–
Если бы все иноязычные видео можно было перевести таким образом, я без ограничений смотрел(-а) бы их на YouTube	АГП1	10	23	27	28	13
	АГП2	10	15	25	31	19
	ПАО	–	3	10	51	36

*Примечания.* АГП1 — автоматический перевод 1-го выпуска; АГП2 — автоматический перевод 2-го выпуска; ПАО — перевод, выполненный профессиональным переводчиком с последующим автоматическим озвучиванием (синтезом речи); 1 = «абсолютно не согласен»; 2 = «скорее не согласен»; 3 = «затрудняюсь ответить»; 4 = «скорее согласен»; 5 = «полностью согласен».

пользоваться данным решением для просмотра тревел-влогов составило суммарно 41 %, после второго — 50 % и, наконец, после просмотра третьего видео (профессиональный перевод, озвученный с использованием технологий синтеза речи) — 87 %. Здесь напрашиваются как минимум два предположения о причинах такой динамики. Во-первых, с просмотром каждого нового видео, вероятно, увеличивался пользовательский опыт, «привыкание» зрителей к данному виду предъявления переводной информации. Во-вторых, автокоррекция мнений респондентов о приемлемости для них такого перевода отчасти коррелирует с их перцептивными впечатлениями о его качестве. Так, для первого и второго видео процент информантов, заявивших о полном понимании всего, о чем говорилось в ролике, составил 25 и 48 % соответственно. Для третьего ролика этот показатель вырос уже до 88 %. Обратная зависимость (тенденция к регрессии воспринимаемого признака) наблюдалась при характеристике впечатлений о лакунах в переводе (4, 2 и 0 %) и непривычных, с точки зрения зрителей, речевых конструкций (90, 76 и 2 %).

Отдельного внимания заслуживает вопрос о воспринимаемом качестве озвучивания перевода, точнее той степени, в которой синтезированная переводная



дорожка приближалась по звучанию к естественной человеческой речи, которую мы слышим, например, от русскоязычных видеоблогеров или с экранов телевизора. Напомним, что все три перевода были озвучены с использованием одной и той же системы синтеза речи, с той лишь разницей, что ролик, переведенный профессиональным переводчиком, ведущий которого — мужчина, был озвучен «женским» голосом чат-бота Алиса. Тем не менее звучание закадрового перевода в некоторых местах показалось неестественным и «роботоподобным» 52 % информантов в первом видео, 37 % — во втором и лишь 3 % — в третьем. В целом эти впечатления совпадают с оценками адекватности ритмико-интонационной организации звучащего перевода, высказанными экспертами на втором этапе эксперимента (см. ниже). С учетом идентичности модуля речевого синтеза логично предположить, что текст перевода, выполненного профессиональным переводчиком, обладал более «прозрачной» синтаксической структурой, которая при озвучивании проявила себя более естественной, т. е. логичной с точки зрения слушающих, просодической организацией звучащей речи.

Некоторые участники эксперимента в свободных комментариях в конце анкеты высказывали свои собственные предположения о различиях в восприятии звуковой дорожки видеороликов: *Мне показалось, что перевод последнего видеоблога сильно отличался от первых двух. Голос больше походил на человеческий. Я слышала, что нейронные сети «Яндекса» очень быстро развиваются, и, может быть, первые 2 перевода были сделаны раньше по времени, а 3-й — позже. Так было с гугл-переводчиком.*

Другие отмечали, что были готовы к тому, что «голос переводчика» может звучать несколько механически: *Когда нам сказали, что переводить будет робот, я сразу подумал про Алису. Так и вышло. Я иногда разговариваю с Алисой, но не знал, что голосовой помощник может еще и синхронно переводить.*

В целом по итогам просмотра всех видео большинство респондентов первой группы охарактеризовали качество звучащей речи как приемлемое, хотя и не всегда естественное. 42 человека (87,5 % опрошенных) заявили о своих намерениях в дальнейшем использовать голосовой перевод от «Яндекса» для просмотра иноязычного видео в интернете.

### 3. Аналитическая оценка.

Поскольку вторая группа информантов, задействованных в настоящем исследовании (эксперты), выносила свои суждения о качестве звучащего перевода на основе его сопоставления с оригинальной аудиодорожкой видеотекста и следуя структурированной матрице оценивания, мы в целом можем рассматривать данную процедуру как аналитическую. За основу матрицы экспертного оценивания были взяты параметры, предложенные в [Ortiz-Voix, Matamala 2015] (табл. 3).

Единственными параметрами, по которым эксперты остались полностью удовлетворены качеством всех трех переводов, являются разборчивость звучащих текстов и отсутствие неоправданных добавлений. Последнее обстоятельство в случае с АГП говорит, по всей вероятности, о невозможности добавлений, в силу того что практически все алгоритмы машинного перевода ориентированы на максимальную изоморфность исходного и переводного текстов. Применительно к переводу, выполненному профессиональным переводчиком, данное наблюдение отражает скорее противоположную тенденцию аудиовизуального трансфера к сокращению

Таблица 3. Аналитическая оценка качества голосового перевода (n = 3)

Категория оценивания	Параметр оценивания	Текст	Средняя оценка
Содержательные компоненты перевода	Отсутствие смысловых искажений	АГП1	3,67
		АГП2	4,33
		ПАО	5,00
	Адекватность выбора межъязыкового соответствия	АГП1	3,33
		АГП2	3,67
		ПАО	4,67
	Отсутствие опущений релевантной информации	АГП1	5,00
		АГП2	5,00
		ПАО	4,67
	Отсутствие неоправданных добавлений	АГП1	5,00
		АГП2	5,00
		ПАО	5,00
Лингвопрагматические аспекты перевода	Эквивалентность речевого регистра	АГП1	3,67
		АГП2	4,33
		ПАО	4,66
	Стилистическая конгруэнтность	АГП1	3,00
		АГП2	3,67
		ПАО	5,00
	Грамматическая корректность	АГП1	4,33
		АГП2	4,67
		ПАО	5,00
	Идиоматичность высказываний	АГП1	2,33
		АГП2	3,33
		ПАО	4,67
	Контекстуальная когерентность	АГП1	3,33
		АГП2	4,00
		ПАО	5,00

переводного текста, предназначенного как для закадрового озвучивания, так и для субтитрирования<sup>21</sup>. Такое стремление со стороны аудиовизуального переводчика, возможно, повлияло на решение экспертов о том, что часть информации была опущена именно в переводе, осуществленном человеком.

<sup>21</sup> При автоматическом голосовом переводе соблюдается лишь одно из требований жанра voice-over к сохранению воспринимаемой аутентичности материала: оригинальная аудиодорожка не заменяется полностью переводной, а лишь приглушается. Согласно другой жанровой традиции данной разновидности аудиовизуального перевода в том виде, в котором она сформировалась в кинопроизводстве и аудиовизуальных СМИ, звучание переводного аудио должно начинаться через несколько (хотя бы 1–2) секунд после начала звучания оригинальной дорожки и заканчиваться за несколько секунд до ее окончания [Matamala 2019]. В автоматическом же переводе оригинал и перевод максимально синхронизированы.

По всем остальным аспектам оценивания переводы АГП1 и АГП2, по мнению экспертной группы, проигрывают переводу ПАО. Это впечатление лишь усилилось на этапе внесения правок в транскрипты переводов: двое из трех экспертов эксплицитно выразили сомнение в том, что третий перевод идентичен предыдущим по технологии выполнения, однако, в отличие от первой группы респондентов, связали такой контраст с фактом постредактирования. Тем не менее эксперты внесли некоторые исправления и в данный текст перевода (см. табл. 4).

В качестве наиболее существенных недостатков автоматического голосового перевода экспертами были отмечены: **неадекватность ритмико-интонационно-го оформления** (что уже обсуждалось выше); **низкая эквивалентность речевого регистра**, выбранного в переводе, и нарушения **стилистической конгруэнтности текста перевода** (заведомо разговорные элементы оригинального аудиоряда могли передаваться в переводе как сходными в функциональном плане, так и стилистически нейтральными, что создавало определенный диссонанс); **смысловые искажения**, которые могли быть результатом как некорректного распознавания исходной звуковой дорожки, так и следствием неудачного **выбора межъязыкового соответствия и деформации при передаче идиоматических выражений**.

Принимая во внимание, что многие из указанных недостатков традиционно рассматриваются как насущные проблемы машинного перевода в целом, а также с учетом того факта, что во всех случаях конечная форма предъявления переводного контента создавалась методом компьютерного преобразования текста в речь, логично предположить, что этап синтеза речи является наименее слабым звеном в цепи операций, осуществляемых в рамках данной модели голосового перевода. Отдельные сбои в модуле синтеза речи проявлялись в настоящем исследовании в виде игнорирования системой правил подвижности русского ударения и акцентных структур заимствованных слов и встречались относительно редко («домик на холме», «это то место, где мы остановимся на ночь», «турецкая баня хаммам»).

Часть ошибок, безусловно, могут возникать на стадии распознавания исходного аудиосигнала, однако здесь следует учесть, что коэффициент точности распознавания речи, который обычно заявляется разработчиком системы, может быть гарантирован в акустических условиях, приближенных к студийным. Тревел-влогеры чаще всего записывают свои видео в «полевых» условиях, далеких от студийных и изобилующих экзогенными шумами, которые сказываются на качестве аудио даже при использовании высокотехнологичных микрофонов и прочих устройств звукозаписи. Неравномерный темп и снижение четкости артикуляции, которые наблюдаются прежде всего при «экшн-съемках», равно как и тот факт, что даже опытные видеоблогеры редко являются профессионалами публичной телевизионной речи, также не способствуют сохранению высокой степени идентификации речевых звуков. Поэтому в определенном смысле сбой распознавания речи в таких роликах обусловлены имманентными характеристиками жанра, а значит, практически неизбежны.

Возможны, однако, и ошибки распознавания речи, спровоцированные отнюдь не качеством звукового сигнала, а лимитированным размером контекстного фрейма, который в исследуемом алгоритме, по всей вероятности, ограничен пределами фразы. В тревел-влогах подобные нарушения чаще всего проявляются при распознавании топонимов и прочих имен собственных. Например, название одного

Таблица 4. Сравнение фрагментов транскрипта автоматического и текста профессионального переводов видео № 3

Транскрипт оригинала	Транскрипт АГПЗ	Текст перевода ПАО	Исправления и комментарии экспертов
...Russian bath which is called banya	...традиционную русскую баню, которая называется <b>поняно</b>	...традиционную русскую баню	–
...I did a traditional hammam which is a turkish bath and the guy was slapping me	...в Турции я посетил традиционный <b>хám)мам</b> , который представляет собой турецкую баню и <b>парень дал мне пощечину</b>	...в Турции я был в традиционном хаммаме, и там банщик <b>шлепал</b> меня по всему телу	...в Турции я был в традиционном хаммаме, и банщик <b>хлопал</b> меня по всему телу
...then the process involves getting whacked by leafy and thorny tree branches while laying on a wooden bench	...тогда процесс включает в себя <b>удары по покрытым листвой и колючими ветвями деревьям</b> , когда вы лежите на деревянной скамье	...в русской бане вы лежите на лавке и <b>вас хлещут ветками</b> из лиственных и хвойных веток	...в русской бане вы ложитесь на лавку и вас бьют <b>ветками</b> лиственных и хвойных деревьев. [Блогер — американец, «веники» из его уст звучат притянуто.]
...this banya got third best in all of Novosibirsk. So, this place is legit	<b>этот кролик у меня третий лучший во всем пиве</b> . Значит, это место <b>законное</b>	...эта баня — одна из лучших в Новосибирске. Значит, <b>это я удачно зашел</b>	...эта баня <b>входит в тройку лучших в городе</b> . Короче, <b>правильная баня</b> . [Неуместная цитата из Гайдая]
I was not expecting that, no warning was given whatsoever	Я этого не ожидал, мои глаза не давали никаких предупреждений	Это было <b>неожиданно</b>	Меня никто об этом не <b>предупредил</b>
Watch the balls, please!	<b>Следите за мячами</b> , пожалуйста!	<b>Тут, пожалуйста, поаккуратней!</b>	<b>Между ног осторожнее!</b>
Holy crap! That was one of the most intense rains I've ever had in my life	<b>О святое дерьмо!</b> Это была одна из самых сильных дождей, которые я когда-либо испытывал в своей жизни	<b>Блин</b> , в жизни не было так больно	Не часто мне приходилось <b>испытывать</b> такую боль
Stay safe, be spontaneous and just go!	Оставайтесь в безопасности, будьте спонтанны и просто идите!	Берегите себя, проживайте каждый момент и двигайтесь <b>вперед!</b>	Берегите себя, будьте собой и <b>путешествуйте!</b>

и того же водопада в Норвегии *Langfoss* в переводе передано тремя отличными друг от друга вариантами трансфонации: *Лангфосс*, *Луфус* и даже *Лабс* (*мы добрались до подножия Лаоса*). В тех ситуациях, когда речь идет о малоизвестных, не несущих ключевого смысла в нарративе видеоблога онимах, такие ошибки распознавания не критичны, особенно с учетом того, что при монтаже видеоряда многие из них дублируются влогером в форме титров («логовизуализация» [Леви 2019]). В случаях же, когда сбой распознавания приводит к подмене понятий (*Лаос* — реально существующая страна, расположенная в Юго-Восточной Азии), такие искажения могут вызвать у реципиентов определенный когнитивный диссонанс. Учет в модуле распознавания более широкого контекста и, возможно, географической привязки сюжета позволил бы скорректировать результаты автоматического фонемного анализа и сократить количество подобных ошибок.

Однако наибольшее количество недочетов, снижающих воспринимаемое качество автоматического голосового перевода тревел-влогов, приходится, судя по всему, на этап собственно межтекстового трансфера (машинного перевода текстов). Проанализировав редакторские правки переводных транскриптов, выполненные экспертами, мы систематизировали наиболее частотные (не менее трех случаев одного типа) искажения в переводе, релевантные для языковой пары английский → русский. Они приведены ниже в порядке убывания частотности.

1. Конкретизация глаголов движения (*to go, to come*):

*We're going to the waterfall* → *Мы идем к водопаду* (на экране — блогер и его спутница едут на машине);

*If you do want to go first class* → *Если вы хотите лететь первым классом* (блогер рассказывает о поездке по ж/д, на экране — вагон поезда).

2. Конкретизация местоимений 2-го лица (*you* → *ты/вы*):

*Tell me how did you manage to get such a piece* → *Как вам удалось забронировать это место?* (блогер обращается к своей близкой подруге и спутнице);

*Norway, why are you so beautiful?!* → *Норвегия, почему вы так красивы?!*

3. Выбор эквивалента при переводе обращений к аудитории (*guys, man*):

*It was crazy, guys* → *Это сумасшествие, ребята*;

*It's good, man* → *Это хорошо, чувак*<sup>22</sup>.

4. Параллелизм при переводе предложений конструкции *It is (was)*:

*It's my first time doing this* → *Это в первый раз, когда я делаю это*;

*It's insane how everything has changed* → *Это безумие, как все изменилось*.

5. Передача коллоквиальных значений слов в конструкциях *It is + N, Adj + N*:

*We are driving through Bergen and it's insane!* → *Мы едем по Бергену и это — безумие!*;

*It's freakingly beautiful* → *Это чудаковато красиво*.

6. Конкретизация рода местоимений, соотносимых с неодушевленными существительными и зоонимами:

*Norway is known for its seafood being surrounded by the Atlantic ocean and I wanted to try some of its best...* → *Норвегия известна своими морепродуктами, он окружен Атлантическим океаном и я хотел попробовать кое-что из его лучшего*;

*Here is my cod. It looks pretty dead...* → *Вот моя треска, он довольно мертвый...*

<sup>22</sup> Точно так же (чувак) слово *man* передавалось в АГП в ситуациях, когда использовалось говорящим в функции междометия.

7. Сочетаемость при употреблении наречий меры и степени (*pretty, quite*). См. предыдущий пример.

8. Передача идиоматических и сленговых выражений<sup>23</sup>:

*Holly crap* → *Святное дерьмо*;

*Am I like the gun show?* → *Я что, как оружейное шоу?*

9. Передача редуцированных форм (*gonna, leggo*):

*Now I'm gonna to take a shower* → *Я ушел в душ*;

*Leggo!* → *Лего*.

10. Неконвенциональные кальки:

*downtown* → *нижний город*.

Приведенный список далеко не полный и при желании может быть продолжен. Большая часть эрратологических типов, представленных выше, по всей вероятности, обусловлены объективными трудностями выбора системой машинного перевода актуального межъязыкового соответствия, преодоление которых возможно, на наш взгляд, исключительно путем расширения объема контекстного фрейма. Не стоит забывать также, что в ситуации межъязыкового трансфера звукового ряда тревел-влога речь идет о переводе преимущественно спонтанной речи, как правило, разговорного стиля. В транслатологическом аспекте разговорная речь практически не изучалась по вполне понятным причинам: такой вид дискурса чрезвычайно редко является объектом профессионального перевода. Переводчикам художественной литературы иногда приходится иметь дело лишь со стилизацией диалогов «под разговорную речь». Устный последовательный и синхронный виды перевода используются в ситуациях двуязычной коммуникации главным образом формального характера, в силу чего даже устнопорождаемая речь публичных выступлений не может рассматриваться как разговорная [Rigual, Spinolo 2018]<sup>24</sup>. Тем не менее когда объектом перевода становится произведение, которое фиксирует или имитирует различные события жизни людей, развивающиеся в реальном времени, как это происходит в аудиовизуальных жанрах, разговорная речь обретает свое место среди предметов транслатологической рефлексии. Именно поэтому о необходимости адаптации переводческих стратегий и тактик к данному виду дискурса стали говорить лишь в момент становления аудиовизуального перевода как отдельной и полноправной области переводоведения. В центре внимания исследователей аудиовизуального трансфера находится прежде всего феномен **псевдоустности** (*prefabricated orality*) [Baños, Chaume 2009], рассматриваемый как образ результата в переводе диалогов кино- и телефильмов. Однако аутентичная разговорная речь может оказаться объектом деятельности переводчика только в контексте медиаперевода, т. е. межъязыкового трансфера вербального компонента произведений аудиовизуальных СМИ и социальных медиа, поэтому в данной разновидности аудиовизуального перевода учет качественных особенностей разговорного дискурса как никогда актуален. Здесь уместно напомнить, что главным отличием

<sup>23</sup> В то же время один из трех экспертов в комментариях особо отметил использование автоматическим переводчиком указательного местоимения *такой(-ая)* со значением грамматического оформителя в качестве удачного в стилистическом отношении эквивалента английского просторечного выражения *I am like: "...and I'm like oh I can't wait to shoot..."* → «...и я такой: не могу дождаться когда начну снимать...».

<sup>24</sup> Исключение, возможно, составляет лишь сфера социально-сопроводительного перевода, которая пока тоже недостаточно изучена в лингвистическом аспекте.

разговорных форм речи от литературных является характер когезионных средств, которые в разговорной речи имеют преимущественно просодическую природу. В то же время специфика разговорного дискурса проявляется на самых различных уровнях лингвистического анализа: фонетический синкретизм, коллоквиальная лексика, диффузный синтаксис на грани паратаксиста, условность границ языковой нормы и т. д. Все эти факторы существенно осложняют перевод в автоматических системах, обученных преимущественно на письменных текстах, и имеющих в качестве центрального модуля преобразования текста в текст.

Так или иначе, несмотря на все отмеченные ошибки и недочеты, эксперты признали предъявленные им результаты автоматического перевода в целом приемлемыми для поставленной задачи и выразили мнение, что при определенных условиях такая практика может активно использоваться в бытовых ситуациях, к которым и относится индивидуальный просмотр видеоблогов.

## Выводы

Очевидно, что тревел-влоги как особый жанр социальных медиа вызывают значительный интерес многих интернет-пользователей, что, по всей вероятности, обусловлено более высоким потенциалом эмпатии и идентификации реципиента с индивидуальным видеоблогером по сравнению с журналистами или авторскими коллективами традиционных, в том числе аудиовизуальных, СМИ, а также возможностями коммуникативной интеракции на интернет-платформах размещения видеоконтента. Языковой барьер, который потенциально возникает между отправителем и получателем сообщения в гетеролингвальной ситуации, может быть преодолен за счет использования таких видов перевода, как автоматическое субтитрирование и автоматический голосовой перевод. При этом в пределах выборки настоящего исследования была выявлена тенденция к предпочтению аудиторией тревел-влогов перевода, предъявляемого в звучащей форме.

За счет отсутствия потребности в приобретении специального программного обеспечения и платных подписок (по крайней мере, применительно к интернет-порталу «Яндекс») сетевые решения автоматического голосового перевода «по запросу» повышают медиадоступность аудиовизуального контента и тем самым способствуют расширению пользовательской аудитории социальных медиа. По всей вероятности, такое расширение может быть ограничено лишь совместимостью платформы-сервиса перевода с внешними ресурсами, на которых размещаются видео, а также необходимостью соблюдения прав на лицензионную видеопroduкцию.

В то же время автоматический голосовой перевод видео, как и машинный перевод в целом, в том виде, в котором он к настоящему моменту представлен в открытом доступе, образует скорее обширное поле для деятельности по дальнейшему развитию и улучшению технологий его выполнения, нежели общепризнанную рутинную практику. Проблемы повышения качества данного вида межъязыкового перевода распространяются как на область создания условий для адекватного понимания реципиентами смыслового содержания передаваемых сообщений, так и на параметризацию приемлемой лингвистической и паралингвистической форм транслята, необходимых для обеспечения неосложненной коммуникации, опосредованной переводом. В этой связи одним из ключевых направлений развития

автоматического голосового перевода является совершенствование алгоритмов учета контекстной информации, что еще раз было продемонстрировано нами в ходе анализа редакторских правок переводных транскриптов и сопоставления последних с текстом перевода для закадрового озвучивания, выполненного профессиональным переводчиком. Впрочем, тот факт, что в основе использованного в настоящем исследовании сервиса перевода лежит принцип прогрессирующего обучения нейронных сетей на постоянно растущих объемах данных, вселяет определенные надежды на перспективы его качественного развития. Небезынтересным представляется также изучение восприятия качества так называемого «сквозного» автоматического перевода звучащей речи применительно к межкультурному трансферу видеоконтента социальных медиа. По заявлениям разработчиков, алгоритмы перевода «речи в речь», минуя фазу текстовых преобразований, позволяют добиться существенной экономии времени обработки сигнала и располагаемых объемов памяти, однако пока широкой аудитории представлены лишь образцы голосовых переводов, ограниченные отдельными фразами [Jia et al. 2019].

Наконец, если единицу видеоблога, т.е. аудиовизуальный текст, рассматривать как целостный полисемiotический и многоканальный конструкт, а автоматический перевод — как многоуровневую технологию искусственного интеллекта, математически моделирующую деятельность переводчика-человека, то ограничение алгоритма преобразований исключительно вербальным компонентом аудиовизуального произведения априори редуцирует возможности выбора варианта поверхностной структуры для адекватной передачи смысла исходного высказывания. Напротив, интеграция в алгоритм автоматического перевода видео модуля распознавания и описания изображения позволила бы оптимизировать процесс принятия системой переводческого решения с учетом дополнительной экстралингвистической информации, которую еще на заре развития машинного перевода И. И. Ревзин и В. Ю. Розенцвейг называли «обращением к описываемой ситуации» [Ревзин, Розенцвейг 1964]. Кроме того, без обращения к визуальному ряду межкультурной трансфер такого полимодального текста, как пост видеоблога, не может быть полноценным в силу того факта, что вербальный компонент аудиовизуальных произведений иногда содержит значительный объем «логовизуализированной» [Леви 2019] информации (титры, надписи, графическая реприза ключевых элементов звучащей речи влогера и других персонажей ролика и т. п.). В этой связи задействование модуля распознавания изображения позволило бы как оптимизировать качество перевода аудиодорожки, так и способствовать формированию у адресатов интегрального представления о вербальном содержании тревел-влога. Технические возможности для извлечения и имплементации такой информации имеются уже сегодня и используются, например, при локализации рекламных видео, однако, вероятно, потребуется еще целый ряд эмпирических исследований, прежде чем гипотеза об эффективности подобной многоуровневой модели сможет быть подтверждена или опровергнута.

## Литература

Гребенев, А. Н., Шаюк, А. Ю. (2018). Структура полисемiotического артефакта: мультимодальный анализ начальной веб-страницы видеоблог-канала на хостинге «YouTube». *Филологические науки. Вопросы теории и практики*, 5–2 (83), 316–322.



- Калинин, А. Ю. (2019). К вопросу о таксономии аудиовизуального перевода в СМИ. *Теория языка и межкультурная коммуникация*, 4 (35), 104–112.
- Леве, И. (2019). Логовизуализация — письменная разновидность слова в телевизионном дискурсе. *Медиалингвистика*, 1 (6), 19–34.
- Лущиков, В. А., Терских, М. В. (2018). Жанрово-тематические и языковые особенности видеоблогов. *Вестник Тамбовского университета*, 4 (14), 57–75.
- Медведева, Е. В. (2021). Специфика аудитории видеоблога и ее влияние на процесс ретивальной коммуникации. *Медиалингвистика*, 3 (8), 261–272.
- Микрюков, В. О., Саркисова, Д. А. (2020). Современные тенденции и развитие travel-блогинга в России. *Международный журнал гуманитарных и естественных наук*, 1 (8), 96–99.
- Мурсекаева, С. В. (2017). Некоторые особенности видеоблога как жанра риторического дискурса. *Преподаватель XXI век*, 4 (2), 342–347.
- Показаньева, И. В. (2015). Теоретическое осмысление основ трэвел-блогинга. Функциональные отличия трэвел-блогера и трэвел-журналиста. *Universum: Филология и искусствознание: электрон. науч. журн.*, 3–4 (17). Электронный ресурс <http://7universum.com/ru/philology/archive/item/2083>.
- Ревзин, И. И., Розенцвейг, В. Ю. (1964). *Основы общего и машинного перевода*. М.: Высшая школа.
- Соснин, А. В., Балакина, Ю. В., Кащихин, А. Н. (2022). Взаимосвязь экспертных категорий и автоматических метрик, используемых для оценки качества перевода. *Вестник Санкт-Петербургского университета. Язык и литература*, 1 (19), 125–148.
- Текутьева, И. А. (2016). Жанрово-тематическая классификация видеоблогинга. *Медиасреда*, 11, 107–113.
- Щипицина, Л. Ю. (2017). *Компьютерно-опосредованная коммуникация: лингвистический аспект анализа*. Москва: URSS.
- Aís, Á. C., Spinolo, N., Garwood, C. J. (2016). Quality assessment and intonation in simultaneous interpreting: evaluation patterns. *Translating Orality. MonTI Special Issue 3*, 1–24.
- Almahasees, Z., Jaccomard, H. (2020). Facebook\* translation service (FTS) usage among Jordanians during COVID-19 lockdown. *Advances in Science, Technology, Engineering Systems Journal*, 6 (5), 514–519.
- Azariah, D.R. (2016). *Tourism, travel, and blogging: A discursive analysis of online travel narratives*. New York: Routledge.
- Baños, R. (2020). Fandubbing. In Ł. Bogucki, M. Deckert (Eds), *The Palgrave Handbook of Audiovisual Translation and Media Accessibility* (pp. 209–226). Cham: Palgrave Macmillan.
- Baños, R., Chaume, F. (2009). Prefabricated Orality: A Challenge in Audiovisual Translation. *inTRAlinea*. Retrieved from <http://www.intralinea.org/specials/article/1714>.
- Benson, Ph. (2015). YouTube as text. In H. R. Jones, A. Chik, Ch. A. Hafner (Eds), *Discourse and Digital Practices. Doing Discourse Analysis in the Digital Age* (pp. 81–96). Oxon, New York: Routledge.
- Bogucki, Ł., Díaz-Cintas, J. (2020). An excursus on audiovisual translation. In Ł. Bogucki, M. Deckert (Eds), *The Palgrave Handbook of Audiovisual Translation and Media Accessibility* (pp. 11–32). Palgrave Macmillan.
- Burchardt, A., Lommel, A., Bywood, L., Harris, K., Popovic, M. (2016). Machine translation quality in an audiovisual context. *Target*, 2 (28), 206–221.
- Carrera, J., Beregovaya, O., Yanishevsky, A. (2009). Machine translation for cross-language social media. *PROMT Americas Inc*. Retrieved from [https://www.promt.ru/press/pdf/machine\\_translation\\_for\\_cross\\_language\\_social\\_media.pdf](https://www.promt.ru/press/pdf/machine_translation_for_cross_language_social_media.pdf).
- Castilho, S., Doherty, S., Gaspari, F., Moorkens, J. (2018). Approaches to human and machine translation quality assessment. In J. Moorkens, S. Castilho, F. Gaspari and S. Doherty (Eds), *Translation quality assessment* (pp. 9–38). Cham: Springer.
- Chen, Y., Guo, Zh., Pan, Q. (2021). Analysis on the Characteristics of Travel Vlog Video and Its Impact on Users' Travel Intention. In *Advances in Social Science, Education and Humanities Research. Vol. 554. Proceedings of the 7<sup>th</sup> International Conference on Humanities and Social Science Research (ICHSSR 2021)* (pp. 169–175). Qingdao: Atlantis Press.
- Choi, W., Lee, Y. (2019). Effects of fashion vlogger attributes on product attitude and content sharing. *Fashion and Textiles*, 1 (6), 1–18.
- Crowel, H., Gribben, H., Loo, J. (2014). *Travel content takes off on YouTube*. Электронный ресурс [https://www.thinkwithgoogle.com/\\_qs/documents/128/travel-content-takes-off-on-youtube\\_articles.pdf](https://www.thinkwithgoogle.com/_qs/documents/128/travel-content-takes-off-on-youtube_articles.pdf).

\* Мета признана на территории РФ экстремистской организацией.

- Desjardins, R. (2017). *Translation and Social Media*. London: Palgrave Pivot.
- Fernández, A., Matamala, A. (2015). Text-to-speech vs human voiced audio descriptions: A reception study in films dubbed into Catalan. *The Journal of Specialised Translation*, 24, 61–88.
- Frobenius, M. (2011). Beginning a monologue: The opening sequence of video blogs. *Journal of Pragmatics*, 43, 814–827.
- Gambier, Y. (2009). Perception and reception of audiovisual translation: Implications and challenges. In H. Che Omar, H. Haroon, A. Abd Ghani (Eds), *The sustainability of the translation field: The 12<sup>th</sup> international conference on translation* (pp. 40–57). Kuala Lumpur.
- Gambier, Y. (2013). The Position of Audiovisual Translation Studies. In C. Millan, F. Bartrina (Eds), *The Routledge Handbook of Translation Studies* (pp. 45–59). London: Routledge.
- Gambier, Y. (2019). Audiovisual translation and reception. *Slovo.ru: Baltijskij accent*, 1 (10), 52–68.
- Gerber-Morón, O., Soler-Vilageliu, O., Castellà, J. (2020). Effects of screen size on subtitle layout preferences and comprehension across devices. *Hermēneus. Revista de Traducción e Interpretación*, 22, 157–182.
- Greco, G. M. (2016). On Accessibility as a human right, with an application to media accessibility. In A. Matamala, P. Orero (Eds), *Researching audio description. New approaches* (pp. 11–33). London: Palgrave MacMillan.
- Han, Ch. (2020). Translation quality assessment: A critical methodological review. *Translator*, 26, 257–273.
- Herring, S. (2013). Discourse in Web 2.0: Familiar, reconfigured, and emergent. In D. Tannen, A. M. Trester (Eds), *Discourse 2.0. Language and New Media* (pp. 1–24). Washington, D. C.: Georgetown University Press.
- House, J. (2015). *Translation quality assessment: Past and present*. London: Routledge Press.
- Jia, Y., Weiss, R. J., Biadsy, F., Macherey, W., Johnson, M., Chen, Z., Wu, Y. (2019). Direct speech-to-speech translation with a sequence-to-sequence model. *INTERSPEECH. Электронный ресурс* <https://arxiv.org/abs/1904.06037>.
- Johansson, M. (2017). YouTube. In Ch. Hoffmann, W. Bublitz (Eds), *Pragmatics of Social Media* (pp. 173–200). Berlin, Boston: Mouton de Gruyter.
- Krasnopeyeva, E. (2018). Understanding the Dynamics of User-Generated Translation on YouTube: A Bourdieusian Perspective. *New Voices in Translation Studies*, 18, 38–83.
- Kurz, I. (2001). Conference Interpreting: Quality in the Ears of the User. *Meta*, 2 (46), 394–409.
- Lee, J. E., Watkins, B. (2016). YouTube vloggers' influence on consumer luxury brand perceptions and intentions. *Journal of Business Research*, 12 (69), 5753–5760.
- Lenihan, A. (2014). Investigating language policy in social media: Translation practices on Facebook\*. In P. Seargeant, C. Tagg (Eds), *The Language of Social Media* (pp. 208–227). London: Palgrave Macmillan.
- Matamala, A. (2019). Voice-over: Practice, research and future prospects. In L. Pérez-González (Eds), *The Routledge Handbook of Audiovisual Translation* (pp. 64–81). Milton Park, Regne Unit: Routledge.
- Nord, C., Khoshsaligheh, M., Ameri, S. (2015). Socio-Cultural and Technical Issues in Non-Expert Dubbing: A Case Study. *International Journal of Society, Culture & Language*, 2 (3), 1–16.
- Ortiz-Boix, C., Matamala, A. (2015). Quality assessment of post-edited versus translated wildlife documentary films: A three-level approach. In Sh. O'Brien, M. Simard (Eds) *Proceedings of 4<sup>th</sup> Workshop on Post-Editing Technology and Practice (WPTP4)*. Retrieved from [https://amtaweb.org/wp-content/uploads/2015/10/MTSummitXV\\_WPTP4Proceedings.pdf](https://amtaweb.org/wp-content/uploads/2015/10/MTSummitXV_WPTP4Proceedings.pdf).
- Pérez-González, L. (2019). Fan audiovisual translation. In M. Baker, K. Malmkjær (Eds), *Routledge encyclopedia of translation studies* (pp. 172–177). Routledge.
- Rigual, C. C., Spinolo, N. (2018). Translating and Interpreting orality. *МорТИ*, 3, 33–54.
- Rivera-Trigueros, I. (2022). Machine translation systems and quality assessment: A systematic review. *Language Resources & Evaluation*, 56, 593–619.
- Romero-Fresco, P. (2018). In support of a wide notion of media accessibility: Access to content and access to creation. *Journal of Audiovisual Translation*, 1 (1), 187–204.
- Vazquez-Calvo, B., Shafirova, L., Cassany, D. (2019). An overview of multimodal fan translation: fansubbing, fandubbing, fan translation of games, and scanlation. In M. del Mar Ogea Pozo, F.R. Rodríguez (Eds), *Insights into Audiovisual and Comic Translation. Changing Perspectives on Films, Comics and Video Games* (pp. 191–213). Córdoba: UCO Press.
- Wattenhofer, M., Wattenhofer, R., Zhu, Z. (2021). The YouTube Social Network. *Proceedings of the International AAAI Conference on Web and Social Media*, 1 (6), 354–361.

---

\* Meta признана на территории РФ экстремистской организацией.

Werner, E. A. (2012). *Rants, reactions, and other rhetorics: Genres of the YouTube vlog*. Dr. Sci. thesis. Chapel Hill.

Xu, D., Chen, T., Pearce, J., Mohammadi, Z., Pearce, Ph. (2021). Reaching audiences through travel vlogs: The perspective of involvement. *Tourism Management*, 86. Retrieved from <https://www.sciencedirect.com/science/article/abs/pii/S0261517721000455?via%3Dihub>.

Статья поступила в редакцию 1 марта 2023 г.;  
рекомендована к печати 8 сентября 2023 г.

Контактная информация:

Калинин Андрей Юрьевич — канд. филол. наук; a\_kalinine@mail.ru

## Machine voice-over translation in social media: Perceived quality of translation of YouTube travel vlogs

A. Yu. Kalinin

Lomonosov Moscow State University,  
1, Leninskie Gory, Moscow, 119991, Russian Federation

**For citation:** Kalinin A. Yu. (2023). Machine voice-over translation in social media: Perceived quality of translation of YouTube travel vlogs. *Media Linguistics*, 10 (4), 568–597. <https://doi.org/10.21638/spbu22.2023.408> (In Russian)

One of the challenges of social media as a means of interaction is to ensure media accessibility, i. e. to provide users with a free and unhindered access to the content of computer-mediated communication platforms. In the context of global networking, linguistic constraints due to the use of a verbal code different from the language of the potential user while creating media content, are not the least of the obstacles to such access. Given the obvious dominance of English in the online space, this is mainly a matter of expanding the audience by engaging individuals whose source language proficiency is not sufficient for unimpeded use of social media, in particular for viewing, listening and understanding video content uploaded on video hosting sites and other digital platforms. Since user-generated video content is generally characterised by individual and independent production, free accessibility and target audiences sharing a common interest for a specific subject matter, regardless of their residence area and language affiliation, human audio-visual translation is not considerable in such circumstances. For this reason, as a way to overcome language barriers and ensure media accessibility, the most advanced Internet portals are beginning to implement Web-based online machine translation solutions, made possible by the development of linguistic digital technologies. The survey conducted as a part of this study has shed light on the users' perception of machine voice-over translation of English-language travel vlogs into Russian, and formulated some hypotheses regarding the ways to further improve its quality. The empirical evidence for this study was obtained from a two-step survey involving users and experts (audiovisual translators) who assessed the quality of machine voice translation of English-language travel vlog into Russian. According to the results obtained machine translation seems to be conceptually acceptable as a means of enabling language media accessibility in social media.

*Keywords:* media accessibility, machine voice-over translation, translation quality assessment, video blogging, travel vlog.

## References

- Aís, Á. C., Spinolo, N., Garwood, C. J. (2016). Quality assessment and intonation in simultaneous interpreting: Evaluation patterns. *Translating Orality. MonTI Special Issue 3*, 1–24.
- Almahasees, Z., Jaccomard, H. (2020). Facebook\* translation service (FTS) usage among Jordanians during COVID-19 lockdown. *Advances in Science, Technology, Engineering Systems Journal*, 6 (5), 514–519.
- Azariah, D. R. (2016). *Tourism, travel, and blogging: A discursive analysis of online travel narratives*. New York: Routledge.
- Baños, R. (2020). Fandubbing. In Ł. Bogucki, M. Deckert (Eds), *The Palgrave Handbook of Audiovisual Translation and Media Accessibility* (pp. 209–226). Cham: Palgrave Macmillan.
- Baños, R., Chaume, F. (2009). Prefabricated Orality: A Challenge in Audiovisual Translation. *inTRAlinea*. Retrieved from <http://www.intralinea.org/specials/article/1714>.
- Benson, Ph. (2015). YouTube as text. In H. R. Jones, A. Chik, Ch. A. Hafner (Eds), *Discourse and Digital Practices. Doing Discourse Analysis in the Digital Age* (pp. 81–96). Oxon, New York: Routledge.
- Bogucki, Ł., Diaz-Cintas, J. (2020). An excursus on audiovisual translation. In Ł. Bogucki, M. Deckert (Eds), *The Palgrave Handbook of Audiovisual Translation and Media Accessibility* (pp. 11–32). Palgrave Macmillan.
- Burchardt, A., Lommel, A., Bywood, L., Harris, K., Popovic, M. (2016). Machine translation quality in an audiovisual context. *Target*, 2 (28), 206–221.
- Carrera, J., Beregovaya, O., Yanishevsky, A. (2009). Machine translation for cross-language social media. *PROMT Americas Inc.* Retrieved from [https://www.promt.ru/press/pdf/machine\\_translation\\_for\\_cross\\_language\\_social\\_media.pdf](https://www.promt.ru/press/pdf/machine_translation_for_cross_language_social_media.pdf).
- Castilho, S., Doherty, S., Gaspari, F., Moorkens, J. (2018). Approaches to human and machine translation quality assessment. In J. Moorkens, S. Castilho, F. Gaspari, S. Doherty (Eds), *Translation quality assessment* (pp. 9–38). Cham: Springer.
- Chen, Y., Guo, Zh., Pan, Q. (2021). Analysis on the Characteristics of Travel Vlog Video and Its Impact on Users' Travel Intention. In *Advances in Social Science, Education and Humanities Research. Vol. 554. Proceedings of the 7<sup>th</sup> International Conference on Humanities and Social Science Research (ICHSSR 2021)* (pp. 169–175). Qingdao: Atlantis Press.
- Choi, W., Lee, Y. (2019). Effects of fashion vlogger attributes on product attitude and content sharing. *Fashion and Textiles*, 6 (1), 1–18.
- Crowel, H., Gribben, H., Loo, J. (2014). *Travel content takes off on YouTube*. Retrieved from [https://www.thinkwithgoogle.com/\\_qs/documents/128/travel-content-takes-off-on-youtube\\_articles.pdf](https://www.thinkwithgoogle.com/_qs/documents/128/travel-content-takes-off-on-youtube_articles.pdf).
- Desjardins, R. (2017). *Translation and Social Media*. London: Palgrave Pivot.
- Fernández, A., Matamala, A. (2015). Text-to-speech vs human voiced audio descriptions: A reception study in films dubbed into Catalan. *The Journal of Specialised Translation*, 24, 61–88.
- Frobenius, M. (2011). Beginning a monologue: The opening sequence of video blogs. *Journal of Pragmatics*, 43, 814–827.
- Gambier, Y. (2009). Perception and reception of audiovisual translation: Implications and challenges. In H. Che Omar, H. Haroon, A. Abd Ghani (Eds), *The sustainability of the translation field: The 12<sup>th</sup> international conference on translation* (pp. 40–57). Kuala Lumpur.
- Gambier, Y. (2019). Audiovisual translation and reception. *Slovo.ru: Baltijskij accent*, 1 (10), 52–68.
- Gambier, Y. (2013). The Position of Audiovisual Translation Studies. In C. Millan, F. Bartrina (Eds), *The Routledge Handbook of Translation Studies* (pp. 45–59). London: Routledge.
- Gerber-Morón, O., Soler-Vilageliu, O., Castellà, J. (2020). Effects of screen size on subtitle layout preferences and comprehension across devices. *Hermēneus. Revista de Traducción e Interpretación*, 22, 157–182.
- Grebenev, A. N., Shaiuk, A. Iu. (2018). The structure of a polysemiotic artifact: multimodal analysis of the initial web page of a video blog channel hosted by YouTube. *Filologicheskie nauki. Voprosy teorii i praktiki*, 5-2 (83), 316–322. (In Russian)
- Greco, G. M. (2016). On Accessibility as a human right, with an application to media accessibility. In A. Matamala, P. Orero (Eds), *Researching audio description. New approaches* (pp. 11–33). London: Palgrave MacMillan.
- Han, Ch. (2020). Translation quality assessment: A critical methodological review. *Translator*, 26, 257–273.
- Herring, S. (2013). Discourse in Web 2.0: Familiar, reconfigured, and emergent. In D. Tannen, A. M. Trester (Eds), *Discourse 2.0. Language and New Media* (pp. 1–24). Washington: Georgetown University Press.
- House, J. (2015). *Translation quality assessment: Past and present*. London: Routledge Press.

---

\* Meta is recognized in Russian Federation as an extremist organization.

- Jia, Y., Weiss, R. J., Biadys, F., Macherey, W., Johnson, M., Chen, Z., Wu, Y. (2019). Direct speech-to-speech translation with a sequence-to-sequence model. *INTERSPEECH*. Retrieved from <https://arxiv.org/abs/1904.06037>.
- Johansson, M. (2017). YouTube. In Ch. Hoffmann, W. Bublitz (Eds), *Pragmatics of Social Media* (pp. 173–200). Berlin, Boston: Mouton de Gruyter.
- Kalinin, A. Iu. (2019). On the question of the taxonomy of audiovisual translation in the media. *Teoriia iazyka i mezhkul'turnaia kommunikatsiia*, 4 (35), 104–112. (In Russian)
- Krasnopeyeva, E. (2018). Understanding the Dynamics of User-Generated Translation on YouTube: A Bourdieusian Perspective. *New Voices in Translation Studies*, 18, 38–83.
- Kurz, I. (2001). Conference Interpreting: Quality in the Ears of the User. *Meta*, 2 (46), 394–409.
- Lee, J. E., Watkins, B. (2016). YouTube vloggers' influence on consumer luxury brand perceptions and intentions. *Journal of Business Research*, 12 (69), 5753–5760.
- Lenihan, A. (2014). Investigating language policy in social media: translation practices on Facebook\*. In P. Seargeant, C. Tagg (Eds), *The Language of Social Media* (pp. 208–227). London: Palgrave Macmillan.
- Leve, I. (2019). Logovisualization is a written version of a word in television discourse. *Media Linguistics*, 1 (6), 19–34. (In Russian)
- Lushchikov, V. A., Terskikh, M. V. (2018). Genre-thematic and language features of video blogs. *Vestnik Tambovskogo universiteta*, 4 (14), 57–75. (In Russian)
- Matamala, A. (2019). Voice-over: practice, research and future prospects. In L. Pérez-González (Eds), *The Routledge Handbook of Audiovisual Translation* (pp. 64–81). Milton Park, Regne Unit: Routledge.
- Medvedeva, E. V. (2021). The specifics of the video blog audience and its impact on the process of real communication. *Media Linguistics*, 3 (8), 261–272. (In Russian)
- Mikriukov, V. O., Sarkisova, D. A. (2020). Modern trends and development of travel blogging in Russia. *Mezhdunarodnyi zhurnal gumanitarnykh i estestvennykh nauk*, 1 (8), 96–99. (In Russian)
- Mursekaeva, S. V. (2017). Some features of the video blog as a genre of rhetorical discourse. *Prepodavatel' XXI vek*, 4 (2), 342–347. (In Russian)
- Nord, C., Khoshsaligheh, M., Ameri, S. (2015). Socio-Cultural and Technical Issues in Non-Expert Dubbing: A Case Study. *International Journal of Society, Culture & Language*, 2 (3), 1–16.
- Ortiz-Boix, C., Matamala, A. (2015). Quality assessment of post-edited versus translated wildlife documentary films: a three-level approach. In Sh. O'Brien, M. Simard (Eds) *Proceedings of 4<sup>th</sup> Workshop on Post-Editing Technology and Practice (WPTP4)*. Retrieved from [https://amtaweb.org/wp-content/uploads/2015/10/MTSummitXV\\_WPTP4Proceedings.pdf](https://amtaweb.org/wp-content/uploads/2015/10/MTSummitXV_WPTP4Proceedings.pdf).
- Pérez-González, L. (2019). Fan audiovisual translation. In M. Baker, K. Malmkjær (Eds), *Routledge encyclopedia of translation studies* (pp. 172–177). Routledge.
- Pokazanova, I. V. (2015). Theoretical understanding of the basics of travel blogging. Functional differences between a travel blogger and a travel journalist. *Universum: Filologiya i iskusstvovedenie: elektron. nauch. zhurn.*, 3–4 (17). Retrieved from <http://7universum.com/ru/philology/archive/item/2083>. (In Russian)
- Rezvín, I. I., Rozentsveig, V. Iu. (1964). *Fundamentals of general and machine translation*. Moscow: Vysshiaia shkola Publ. (In Russian)
- Rigual, C. C., Spinolo, N. (2018). Translating and Interpreting orality. *MonTI*, 3, 33–54.
- Rivera-Trigueros, I. (2022). Machine translation systems and quality assessment: a systematic review. *Language Resources & Evaluation* 56, 593–619.
- Romero-Fresco, P. (2018). In support of a wide notion of media accessibility: Access to content and access to creation. *Journal of Audiovisual Translation*, 1 (1), 187–204.
- Shchipsitsina, L. Iu. (2017). *Computer-Mediated Communication: Linguistic Aspect of Analysis*. Moscow: URSS Publ. (In Russian)
- Sosnin, A. V., Balakina, Iu. V., Kashchikhin, A. N. (2022). Relationship between expert categories and automatic metrics used to assess translation quality. *Vestnik of Saint Petersburg University. Language and Literature*, 1 (19), 125–148. (In Russian)
- Tekuteva, I. A. (2016). Genre-thematic classification of video blogging. *Mediasreda*, 11, 107–113. (In Russian)
- Vazquez-Calvo, B., Shafirova, L., Cassany, D. (2019). An overview of multimodal fan translation: Fansubbing, fandubbing, fan translation of games, and scanlation. In M. del Mar Ogea Pozo, F. R. Rodríguez (Eds), *Insights into Audiovisual and Comic Translation. Changing Perspectives on Films, Comics and Video Games* (pp. 191–213). Córdoba: UCO Press.

\* Meta is recognized in Russian Federation as an extremist organization.

- Wattenhofer, M., Wattenhofer, R., Zhu, Z. (2021). The YouTube Social Network. *Proceedings of the International AAAI Conference on Web and Social Media*, 1 (6), 354–361.
- Werner, E. A. (2012). *Rants, reactions, and other rhetorics: Genres of the YouTube vlog*. Dr. Sci. thesis. Chapel Hill.
- Xu, D., Chen, T., Pearce, J., Mohammadi, Z., Pearce, Ph. (2021). Reaching audiences through travel vlogs: The perspective of involvement. *Tourism Management*, 86. Retrieved from <https://www.sciencedirect.com/science/article/abs/pii/S0261517721000455?via%3Dihub>.

Received: March 1, 2023  
Accepted: September 8, 2023

Author's information:

*Andrei Y. Kalinin* — PhD in Philology; [a\\_kalinine@mail.ru](mailto:a_kalinine@mail.ru)