

Соснин Алексей Владимирович

Национальный исследовательский университет «Высшая школа экономики»,
Россия, 603155, Нижний Новгород, ул. Большая Печерская, 25/12
asosnin@hse.ru

Балакина Юлия Владимировна

Национальный исследовательский университет «Высшая школа экономики»,
Россия, 603155, Нижний Новгород, ул. Большая Печерская, 25/12
logerk3@gmail.com

Кащихин Андрей Николаевич

Национальный исследовательский университет «Высшая школа экономики»,
Россия, 603155, Нижний Новгород, ул. Большая Печерская, 25/12
logerk3@gmail.com

Взаимосвязь экспертных категорий и автоматических метрик, используемых для оценки качества перевода

Для цитирования: Соснин А.В., Балакина Ю.В., Кащихин А.Н. Взаимосвязь экспертных категорий и автоматических метрик, используемых для оценки качества перевода. *Вестник Санкт-Петербургского университета. Язык и литература.* 2022, 19 (1): 125–148. <https://doi.org/10.21638/spbu09.2022.107>

Статья посвящена оценке качества перевода: рассматриваются прикладные и прагматические аспекты оценки качества перевода в условиях стремительного увеличения числа текстов, которые требуется перевести для обеспечения межкультурной коммуникации; суммируется большое количество подходов, каждый из которых при оценке качества перевода имеет свои преимущества и недостатки; анализируется соотношение категорий адекватности и эквивалентности перевода как основных параметров оценки. Заключается, что эквивалентность ориентирована на результат перевода, а адекватность указывает на соответствие процесса перевода данным коммуникативным условиям; выявляется процессуальная сущность первой и дискурсивная — второй. В статье также изучается возможность применения данных критериев для автоматической оценки перевода на основе нескольких актуальных автоматизированных метрик оценки качества перевода Cosine Similarity, Word Error Rate, BLEU, ROUGE, NIST, METEOR; исследуется также потенциальное приложение этих метрик к переводу, осуществляемому человеком. Делается вывод о том, что наиболее эффективной из рассмотренных математических автоматизированных метрик оценки можно признать N-граммную метрику с явно заданным порядком METEOR, при этом подчеркивается, что с успехом ее можно применять лишь для формальных стандартизированных текстов. Что касается перевода художественной литературы, то его оценка по N-граммному соответствию единиц едва ли может быть признана эффективной в силу специфики взаимоотношений в нем формы и смысла, когда даже на уровне предложения обнаруживается масса формальных несоответствий (т. е. используются единицы неэквивалентные), хотя адекватность результата при его оценке экспертом не вызыва-

ет сомнений. Делается общий вывод, что сугубо математические процедуры, которые рассматриваются в статье, не смогут служить полноценными лингвистическими моделями оценки до тех пор, пока они не будут дополнены синтаксическими и семантическими алгоритмами.

Ключевые слова: оценка качества перевода, эквивалентность, машинный перевод, автоматизированные метрики оценки качества перевода, вектор представления слова.

Теоретические подходы к экспертной оценке качества перевода

Одной из наиболее комплексных проблем теории перевода является оценка качества перевода. Ее комплексность заключается в следующем:

- неоднозначность и сложность оценки, в частности из-за большого количества критериев, по которым может быть произведена оценка;
- прикладной характер, поскольку для разных видов перевода одни и те же критерии могут как работать, так и быть неприменимыми;
- отсутствие единообразия критериев: перевод письменный, устный последовательный, устный синхронный, перевод художественного текста требуют применения разных критериев оценки качества.

В настоящем исследовании рассматривается ряд аспектов и подходов к оценке перевода, которые попадают в две широкие категории: экспертные и автоматические. Мы ставили перед собой следующие задачи: выделить среди этих подходов наиболее удачные и эффективные в заданной области; произвести количественное и качественное сравнение подходов из автоматической и экспертной категорий и установить, по возможности, их регулярные соответствия. Была выдвинута гипотеза, что подобные соответствия позволят существенно улучшить критерии оценки перевода в каждой области за счет добавления межкатегориальных параметров оценки, не противоречащих друг другу.

Существует значительное число подходов к оценке качества перевода. Согласно А. А. Аносовой, их общее свойство заключается в том, что дается ряд норм, критериев и требований, которые переводчик должен соблюдать [Аносова 2011: 47]. Одним из наиболее общих критериев качества можно считать норму перевода. Однако, по мнению Ю. М. Скребнева, исследователя, возглавлявшего горьковскую лингвистическую школу, «...число концепций нормы, вероятно, приближается к числу языковедов, специально занимавшихся этим вопросом» [Скребнев 1984: 162]. В. Н. Комиссаров определяет данное понятие как «совокупность автономных, но в то же время связанных требований, которые должны выполняться переводчиком, и действий, которые должны предприниматься переводчиком для выполнения этих требований» [Комиссаров 1990: 145]. Среди требований наиболее часто выделяются следующие: необходимость сопоставлять тексты на исходном языке (ИЯ) и переводящем языке (ПЯ); выявление ошибок, их корректировка; передача содержания текста ИЯ, мыслей автора; как можно более дословная передача смысла отдельных языковых единиц и пр.

Здесь нельзя не заметить некоторую противоречивость выдвигаемых критериев. Как указывает Комиссаров, переводчик каким-то образом должен и буквально

перевести текст на ПЯ (т. е. каждую языковую единицу передать максимально близким, если не полным, эквивалентом), и сохранить весь спектр смыслов, заложенных автором, и, в дополнение к этому, использовать то же самое количество языковых единиц, использованных автором оригинала [Комиссаров 1990: 146–148]. Данное противоречие возникает, во-первых, из-за относительности нормы перевода, а во-вторых, из-за разнообразия видов перевода, о чем говорилось ранее. Действительно, к каждому из видов перевода вышеуказанные критерии и требования применяются в разной мере, и, более того, они не могут быть полностью идентичны даже в пределах одного вида перевода. Проблема оценки перевода встает наиболее остро, поскольку отсутствуют конкретные метрики, по которым можно было бы принять решение — «плохой» ли представлен перевод или «хороший».

Более конкретными, по сравнению с довольно абстрактной нормой перевода, представляются принципы перевода, которые раскрывают отдельные переводческие нормы. Развивая и дополняя классификацию Комиссарова, М. Ю. Илюшкина выделяет следующие принципы перевода:

- норма эквивалентности перевода — насколько тексты ИЯ и ПЯ соотносятся или не соотносятся друг с другом в плане общности заложенного в них содержания;
- жанрово-стилистическая норма перевода — здесь рассматривается соответствие стилистических критериев, представленных в тексте ИЯ;
- прагматическая норма перевода — одна из самых главных норм, заключающаяся в необходимости соответствия прагматической функции текста на ПЯ прагматической функции, которая была в тексте на ИЯ;
- норма адекватности перевода — совокупность представленных выше норм [Илюшкина 2017: 34–36].

Хотя данные принципы и конкретизируют стороны, с которых перевод нужно оценивать, но однозначно с их помощью нельзя сказать, является ли перевод качественным, поскольку каждая из представленных норм не имеет бинарной оценки перевода «хороший/плохой», а имеет только степень соответствия той или иной характеристике текста. Кроме того, нормы отчасти ограничивают друг друга — можно сказать, что всякий перевод должен найти компромисс между всеми этими нормами: не вступать ни с одной из них в противоречие, но в то же время максимально возможно соответствовать каждой из них.

В дополнение к термину «норма перевода» Комиссаров вводит понятие качества перевода, поясняя, что под ним «понимаются результаты процесса перевода, которые определяются степенью его соответствия переводческой норме и характером невольных или сознательных отклонений от этой нормы» [Комиссаров 1999: 115]. Таким образом, каждый из актов перевода должен оцениваться по каждому из классифицируемых аспектов.

Переводческая норма — далеко не единственная норма, по которой перевод может быть оценен. Отечественные и зарубежные исследователи постоянно предлагают развитие идей общей и всеобъемлющей нормы, подразделяя ее на более частные. Например, рассматривая проблемы оценки качества перевода, А. Честерман и Э. Вагнер предлагают собственные нормы, которые, однако, во многом повторяют рассмотренные ранее. К таким нормам относятся:

- приемлемость, которая рассматривается как соответствие перевода стилистическим критериям (acceptability norm);
- эквивалентность, или связь текста перевода с текстом оригинала (relation norm);
- коммуникативная норма, которая определяется как необходимость перевода оказывать то же самое влияние на реципиента, что и оригинальный текст [Chesterman, Wagner 2002].

Здесь также следует упомянуть термин Л. Н. Латышева «оптимальное переводческое решение». Наполнение этого термина, по сути, повторяет определения, приведенные ранее. По мнению ученого, процесс перевода каждый раз сопряжен с поиском компромисса, который должен удовлетворять нескольким противоречащим друг другу условиям. Переводческое решение отличается от переводческой нормы допущением неоптимальных вариантов перевода, с которыми можно проводить сравнение, в которых можно искать ошибки и на их основе приходиться к более правильному варианту. Другими словами, появляется способ оценки каждого из возможных вариантов перевода — на контрасте с другими [Латышев 2005: 252].

Относительно современная скопос-теория, или теория цели текста, которая пытается предложить теоретическое и практическое решение проблемы оценки качества перевода, характеризуется, по сравнению с предыдущими теориями, отходом от абстрактных категорий и переходом к сугубо прагматическому и прикладному аспектам (см.: [Reiss, Vermeer 1984: 53–54]). Центральное место в переводе — как в процессе, так и результате — отводится скопосу, или цели, которая целиком подчиняет все остальные категории перевода. Перевод осуществляется, адаптируется и интерпретируется исключительно в соответствии с этой целью — вплоть до того, что возможны ситуации, когда тексты на ИЯ и ПЯ не совпадают по впечатлению, которое каждый из них производит на получателя. Скопос задается извне, т. е. либо заказчиком перевода (который может и не иметь переводческой компетенции), либо самим переводчиком после ознакомления с текстом. Как отмечает Э. Прунч, «скопос может быть задан... на основании особой заинтересованности в определении функции перевода. Переводчик определяет скопос на свой страх и риск сообразно своим интересам» [Прунч 2015: 200].

Краеугольным камнем здесь является читательская аудитория, на которую ориентируется перевод. Здесь нельзя говорить о собственно переводе, а, скорее, о некотором виде языкового посредничества, который осуществляется переводчиком. При этом основной целью становится не передача смысла, а апелляция к аудитории (см.: [Евтеев 2017: 265]).

Оценка перевода с помощью данных критериев не может быть объективна, т. к., во-первых, у каждого переводчика (как и заказчика) может быть собственное представление о цели текста на ПЯ, что неизбежно приведет к противоречиям. Во-вторых, перевод, который был сделан на основании только такого критерия, как цель, без учета жанровых, стилистических и других категорий, может полностью лишиться авторского почерка и, в дополнение к этому, исказить или не полностью отразить смысл, заложенный автором (см.: [Портер 2004: 43]).

Рассматривая в качестве примера материал заседаний Европейского парламента, А.-Р. Вуорикоски считает одним из наиболее важных критериев оценки пере-

вода верное и уместное использование терминов, клише, акронимов, собственных имен и ключевых слов. По мнению исследователя, «точная передача специфики определенного жанра предполагает понимание следующих моментов: ожидания аудитории и того, как определенные ключевые слова... соотносятся с интеллектуальной наполненностью и структурой жанра собраний Европейского парламента» [Vuorikoski 2016: 103] (перевод наш, выделено нами. — А. С., Ю. Б., А. К.). Так же как и в рассмотренной ранее скопос-теории, здесь говорится о необходимости учета специфики реципиентов, которые будут воспринимать перевод. Основное же отличие данного подхода — больший упор на конкретные лексические средства, передача которых должны быть произведена с особой тщательностью. При этом другие критерии, например жанрово-стилистический, отходят на второй план.

Адекватность и эквивалентность в экспертной оценке качества перевода

Одним из наиболее значимых подходов к оценке перевода является концепция адекватности и эквивалентности, в которой суммарно учитываются критерии, рассмотренные в предыдущем разделе статьи.

Как у многих терминов переводоведения, у слов «адекватность» и «эквивалентность» существует целый ряд дефиниций, зачастую весьма различных. Так, применительно к уже рассмотренному ранее критерию эквивалентности следует отметить сложность его определения: что же считается эквивалентным? Нередко полагают, что полностью эквивалентных единиц в языках не существует по ряду причин, среди которых структурные, прагматические и, очень часто, семантические несоответствия между лексемами. Тем не менее случаи эквивалентности, близкой к полной, возможны, но наблюдаются они, как правило, в относительно несложных коммуникативных условиях в текстах со сравнительно узким диапазоном функциональных характеристик. Примером того может служить отдельно взятая научная область со своим терминологическим аппаратом, переведенным на разные языки. Таким образом, под эквивалентностью мы будем понимать некоторую предельную величину, к которой стремится хороший перевод.

Ситуация осложняется также тем, что в теории перевода до сих пор не сложилось единого понимания термина «единица перевода», что осложняет интерпретацию и применение на практике определения эквивалентности. Еще в 1959 г. Р. Якобсон начал рассматривать проблемы эквивалентности слов в разных языках. Оказалось, что даже между отдельными словами абсолютной эквивалентности быть не может. Следовательно, нужно искать и использовать другие единицы языка, между которыми можно установить эти отношения. В итоге Р. Якобсон пришел к выводу, что в отношении эквивалентности вступают не отдельные слова/словоформы, а целые предложения или группы предложений, представляющие собой относительно законченное смысловое единство. Сам смысл, заключенный в этих сообщениях, должен быть одинаков, а средства выражения, которые уникальны для каждого из языков и отражают реальность, исходя из народных/культурных особенностей, совершенно не обязательно должны точно проецироваться из одного языка в другой.

Несколько более общую, но схожую точку зрения занимали К. Райс и Г. Фермер. Эквивалентность, по их мнению, может наблюдаться как между отдельными знаками, так и между целыми текстами. Таким образом, нет конкретного элемента для сравнения, поскольку речь идет обо всех языковых единицах. Как и Якобсон, исследователи указали на независимость категории эквивалентности от более мелких или крупных единиц языка: эквивалентность слов еще не означает эквивалентности текстов; верно и обратное, т. е. эквивалентность текстов вовсе не подразумевает эквивалентности всех включенных в них сегментов [Reiss, Vermeer 1984: 128].

Дефиницию эквивалентности можно вывести из определения эквивалентной единицы. Одно из наиболее общих определений этому понятию было дано О. С. Ахмановой в Словаре лингвистических терминов: это «единица речи, совпадающая по функции с другой, способная выполнять ту же функцию, что другая единица речи» [Ахманова 1966: 508]. Здесь представлен один из подходов к данной проблеме — функциональный. Другими словами, две единицы, которые и в тексте на ИЯ и в тексте на ПЯ выполняют в определенном контексте одну и ту же функцию (или, точнее, совокупность функций: коммуникативную, лексическую, прагматическую и т. д.), могут считаться эквивалентными. Важным критерием является именно соотносимость контекстов, так как, будучи расположенными в других местах, эти единицы могут утратить свойство эквивалентности.

Функциональный аспект эквивалентности характеризуется разветвленной структурой. Однако относительно того, что следует считать обязательной функцией, которая бы определяла единицы как эквивалентные, консенсус не достигнут. А. Д. Швейцер полагает, что основная задача переводчика — достижение коммуникативной эквивалентности, которая заключается в необходимости передачи воздействия, которое оказывается текстом на ИЯ на реципиента [Швейцер 1988: 48]. Другие исследователи, в частности В. Н. Комиссаров, выделяют не столько виды эквивалентности, сколько ее уровни, каждый из которых зависит от нескольких факторов: мастерства переводчика, соотносимости или несоотносимости культурных особенностей между народами, характера переводимых текстов, т. е. жанровых и стилистических особенностей, и пр. При этом отмечается, что всякая стратификация условна, так как каждый конкретный случай перевода уникален [Комиссаров 1990: 124].

Обобщая упомянутые и наиболее значимые характеристики эквивалентности, можно вывести следующее ее определение: это характеристика смыслового и интенционального соответствий двух или более языковых единиц, которые в одном и том же контексте выполняют одинаковую функцию.

Изначально, на заре выделения таких характеристик перевода, как адекватность и эквивалентность, они не разделялись вовсе. Например, Дж. Кэтфорд в 1965 г. считал понятия *translation adequacy* («адекватность перевода») и *translation equivalence* («переводческая эквивалентность») взаимозаменяемыми. Однако здесь же им вводятся еще два термина: *formal correspondence* (именно «языковое, или формальное соответствие») и *textual equivalence* («текстуальные, или речевые, эквиваленты»), который конкретизирует более широкое понятие *translation equivalence*. Главное различие между ними заключается в следующем: языковое, или формальное, соответствие характеризует минимальные единицы, которые эквивалентны по отдельности, а текстуальные, или речевые, эквиваленты реализуются только

тогда, когда они погружены в некоторое окружение или контекст [Catford 1965: 103–104].

А. В. Федоров и Я. И. Рецкер, которые разработали концепцию полноценного, или адекватного, перевода, также не вполне четко разграничивали понятия адекватности и эквивалентности. Согласно исследователям, адекватность оценивается по двум критериям: 1) наиболее полная передача смыслового содержания текста на ПЯ; 2) передача этого содержания с помощью равноценных языковых средств. В понимании исследователей, это те средства, которые выполняют одну и ту же функцию [Рецкер 1974: 168]. Однако формулировка второго критерия адекватности практически дословно повторяет стандартное определение эквивалентности, т. е. при таком подходе эти две характеристики находятся в отношениях «часть — целое».

Развивая идеи Федорова и Комиссарова, нижегородский переводовед В. В. Сдобников утверждает, что качество перевода складывается из двух составляющих: адекватности, которая означает воспроизведение в переводе функции исходного сообщения, и эквивалентности, предполагающей максимальную лингвистическую близость текстов оригинала и перевода [Сдобников 2007: 43].

Комиссаров рассматривает **адекватный** перевод как синоним **хорошего** перевода, то есть такого, который реализует наиболее полную межъязыковую коммуникацию в определенных заранее условиях. Главным же компонентом этой характеристики, как указывает А. М. Финкель, является **цель** перевода, которая должна последовательно учитываться в переводческом процессе [Финкель 2001: 185]. Данное определение тем не менее не вносит практической ясности, поскольку обеспечение межъязыковой коммуникации в полном объеме как критерий качества перевода однозначно оценить нельзя. Ю. Найда отмечает, что само по себе использование адекватности в качестве оценочной категории перевода не имеет смысла, так как основные, по мнению исследователя, характеристики: реакция реципиента перевода и равноценность текстов на ИЯ и ПЯ функциональном аспекте — весьма абстрактны [Найда 1978: 118]. Однако данное утверждение опирается лишь на две из оставшихся характеристик, которые были рассмотрены ранее и по которым можно сделать конкретные выводы о качестве перевода. Главная из них — это, безусловно, переводческая эквивалентность. Продолжая мысль Комиссарова, можно сказать, что эквивалентность, которой маркируется смысловая схожесть или различие единиц языка, тесно соприкасается с адекватностью перевода и входит в это понятие [Комиссаров 1999: 19].

Схожие отличительные особенности понятия «адекватность» выделяются Н. К. Гарбовским. **Адекватный перевод** представляет собой единство прагматического и коммуникативного эффектов, то есть таких же, которые выделяются в исходном тексте. Они могут быть достигнуты даже без наличия эквивалентности на уровнях семантики и синтаксиса [Гарбовский 2004: 301]. С другой стороны, здесь мы говорим о присутствии эквивалентности на уровне всего текста. В дополнение к данному подходу следует упомянуть позицию К. Райса и Г. Вермеера: эквивалентность есть частный случай адекватности — когда в переводе воспроизводится функция текста на ИЯ [Reiss, Vermeer 1984: 215]. Здесь говорится уже не о двух характеристиках качества перевода, а об одной исчерпывающей. Адекватность подразумевает общность не только отдельных знаков (от одного знака до текста), но и культурных особенностей двух текстов. На разных исторических промежут-

ках текст может иметь разное воздействие на народы и их устои, соответствовать им или нет. Исследователями также вносится характеристика **динамичности**, которая присуща переводу [Reiss, Vermeer 1984: 217] (см. также: [Григорьев, Мартыненко 2019: 7–8]).

Для четкого понимания, где заканчивается одна категория и начинается другая, нужно определить, на что ориентирована каждая из них. Эквивалентность как более частная категория описывает отношения между исходными и конечными текстами, т. е. она ориентирована на **результат** перевода — соответствие языковых единиц в разных языках. С другой стороны, адекватность — это категория, присущая переводу скорее как **процессу**. Цель, интенция автора, полнота языковой коммуникации — все это в сумме с эквивалентностью делает, в идеальном случае, перевод **хорошим**, или **правильным**, какой бы абстрактной ни была эта характеристика.

Эквивалентность может быть практически абсолютной, если мы говорим об отдельных языковых знаках или единицах. Однако достичь такого уровня эквивалентности практически невозможно не только для текстов, но даже и для предложений: языки, культура, устои слишком сильно отличаются у разных народов, что неоспоримо влияет на возможность обеспечения эквивалентного перевода, нахождения некоторого **инварианта**, который удовлетворял бы потребности прийти к наиболее полному смысловому единству. Следовательно, сложно говорить и о возможности абсолютно адекватного перевода. Последняя категория подразумевает не столько абсолютное соответствие, сколько некоторый компромисс, на который переводчику приходится идти. В отдельных случаях речь может идти уже о реферативном переводе, пересказе или другом способе передачи мысли автора (имеется в виду перевод-процесс вообще, независимо от того, профессиональный он или нет). Согласно В. В. Мошковичу, причина этого состоит в невозможности передать прагматическую/функциональную наполненность оригинального текста с помощью эквивалентных языковых средств на другом языке [Мошкович 2012: 275]. Отсюда следует, что адекватность носит не идеальный, а оптимальный характер: отклоняются самые неудачные варианты, и остается наилучший из возможных, который совсем не обязательно будет отвечать всем идеальным требованиям.

Суммируя подходы к этим двум категориям, можно заключить, что наиболее конкретными и практически применимыми представляются следующие положения, которые, по сути, являются компиляцией рассмотренных выше:

- **эквивалентность** как категория оценки перевода имеет дело с **единицами перевода** (от слова до текста) в сопоставлении их между исходным и переводящим языками по следующим критериям: лексическому, семантическому и синтаксическому, — а также по функции, выполняемой в данном контексте;
- для достижения **адекватности**, в свою очередь, главным критерием является передача цели и заложенного смысла исходного текста и, в идеальном случае, передача этих элементов с помощью **эквивалентных** единиц.

Таким образом, если эквивалентность отвечает на вопрос о том, соответствует ли конечный текст исходному, то адекватность указывает на то, соответствует ли перевод как процесс данным коммуникативным условиям; это дискурсивная характеристика.

С развитием компьютерных технологий и систем машинного перевода вопрос оценки перевода встает особенно остро: какие критерии и категории могут быть использованы, по каким метрикам может быть оценен перевод? В этой же области рассматривается возможность определения с помощью количественных и статистических средств, насколько конкретный перевод хорош или плох.

Бурно развивающаяся область компьютерной лингвистики предлагает все больше инструментов для достижения этой цели, которые в сочетании друг с другом могут значительно упростить и ускорить сложный процесс перевода: помочь выявить и оценить возможные ошибки, найти узлы текста, от которых можно оттолкнуться при построении системы оценки.

При появлении новых методов возникает вопрос о том, возможно ли применить их к оценке человеческого перевода для ускорения и упрощения процесса. Чтобы это узнать, необходимо рассмотреть автоматические методы оценки перевода и соотнести их с категориями экспертной области, которые, в свою очередь, позволяют оценить перевод человеческий с точки зрения смысла и использованных средств.

Автоматические (машинные) методы оценки качества перевода

Рассмотренные в предыдущем разделе категории, связанные с оценкой качества перевода, можно представить и в машинном эквиваленте, осуществить с помощью компьютерных инструментов. С помощью инструментов автоматической обработки естественного языка можно провести сравнение слов оригинального текста и переводного, выявить трансформации в тексте, которые могли бы улучшить качество перевода, сопоставить тематики¹ текстов. Однако, последние два пункта — выявление **авторских** слов и сравнение текстов по прагматическим компонентам — до сих пор требуют экспертной оценки профессионального переводчика.

В дополнение к этому, при автоматической оценке перевода могут использоваться категории, которые под силу определить и измерить только эксперту: это **полнота** (adequacy)² и **гладкость** (fluency). Последний параметр отвечает за «форму» перевода, и с его помощью оценивается правильность фразы с точки зрения носителя языка (см.: [Митренина 2016: 183]). Параметр гладкости важен при оценке машинного перевода, но не нужен для оценивания профессионального перевода, сделанного человеком.

Способы векторного представления слов

Одним из первых разработанных способов векторной репрезентации лингвистических единиц является представление слова / предложения / целого текста с помощью модели «мешка слов» (Bag-of-Words Model). Суть данного алгоритма заключается в следующем: все слова, которые хотя бы раз встречаются в каком-либо из текстов, собираются вместе и формируется т. н. словарь. Затем для текста создается собственный словарь, и строится вектор, репрезентирующий текст

¹ Т. н. text classification.

² Не следует путать с адекватностью, рассмотренной в предыдущем разделе.

(см.: [Goldberg et al. 2017: 769]). Каждое из чисел в векторе, которое отвечает за то, сколько раз представленное слово встретилось в тексте, образует признак (feature) текста, а тексты, мешки слов которых похожи между собой, будут схожими в лингвистическом и семантическом планах.

Данный метод представляется простым, но не слишком действенным по ряду причин (см.: [Brownlee 2017]):

- размерность: в словаре может быть большое ($> 10\,000$) количество слов; следовательно, каждый текст, даже если он состоит из нескольких предложений, будет представлен вектором большой размерности, что требует серьезного количества ресурсов для хранения;
- разреженность, что напрямую вытекает из предыдущего пункта: текст, состоящий из нескольких предложений, включает в себя не так много уникальных слов; следовательно, в представлении данного текста по словарю большинство чисел в векторе будут нулями, так как вектор строится по словам из словаря;
- игнорирование контекста и порядка слов: числа в векторе представлены хаотично и могут располагаться в любом порядке, что означает игнорирование соседних слов и синтаксического расположения. Тем не менее данная проблема может быть немного сглажена с помощью N-грамм в качестве единиц подсчета (см.: [Goldberg et al. 2017: 769]).

Развитием модели «мешка слов» является метод Word2Vec, предложенный группой ученых под руководством Т. Миколова [Mikolov et al. 2013]. Так же как и в предыдущем случае, при реализации данного алгоритма слово переводится в векторное представление (word embedding) с тем отличием, что в расчет берется контекст представленного слова. При этом размер контекста или окна задается самим исследователем. Модель обучается на корпусе текстов, подбирая для каждого слова вектор, соответствующий его окружению. Таким образом делается предположение, что слова, которые в пространстве полученных векторов находятся близко, имеют и в реальном/литературном мире семантические сходства.

У данного алгоритма существуют два способа применения:

- Continuous Bag-of-Words (CBOW): на вход модели подаются слова из левого и правого контекстов (их количество может быть различно), а на выходе модель должна дать слово, которое стоит «посередине»;
- Skip-gram: обратное по отношению к предыдущему: на входе задано одно слово, а модель должна предсказать левый и правый контексты в определенном окне.

Данный способ получения векторных представлений слов намного эффективнее и точнее модели «мешка слов». Его дальнейшим развитием служат алгоритмы paragraph2vec и doc2vec, которые работают с целыми параграфами и текстами соответственно. Следующим этапом являются глубокие нейронные сети — предмет для отдельного исследования.

Векторы можно сравнивать между собой, вычисляя расстояние между ними. Следовательно, то же самое можно сделать и для векторных представлений текстов — данный способ лежит в основе текстовой классификации (например, по

теме или жанру) и оценке качества машинного перевода, которая происходит по определенным метрикам.

Обзор метрик для оценки качества машинного перевода

Перечислим некоторые метрики (от наиболее простых до комплексных), которые используются для оценки качества машинного перевода, и дадим краткую характеристику каждой из них с указанием их достоинств и недостатков. Далее каждая из них будет проанализирована в ключе рассмотренных ранее категорий из области экспертной оценки перевода — адекватности и эквивалентности.

К наиболее распространенным метрикам относятся следующие:

- 1) Cosine Similarity;
- 2) Word Error Rate (WER);
- 3) BLEU, ROUGE, NIST;
- 4) METEOR.

1. Коэффициент Охай, или косинусный коэффициент (Cosine Similarity), не является метрикой в строгом смысле этого слова. Он необходим для подсчета расстояния между векторами (текстов в данном случае) по следующей формуле (1) [Singhal, 2001: 38]:

$$\text{Similarity}(p, q) = \cos\theta = \frac{p \cdot q}{\|p\| \|q\|} = \frac{\sum_{i=1}^n p_i q_i}{\sqrt{\sum_{i=1}^n p_i^2} \sqrt{\sum_{i=1}^n q_i^2}} \quad (1)$$

Числа из каждого вектора поэлементно перемножаются, а сумма этих произведений делится на норму каждого из векторов. В результате получается косинус угла между векторами в векторном пространстве. По предположению, чем меньше угол, тем больше тексты похожи. Явными достоинствами данного способа являются относительная простота и быстрота вычислений. Однако здесь, как и в ранее рассмотренном «мешке слов», наблюдается проблема с расположением слов и смысловой принадлежности каждого слова, представленного в тексте.

В данном случае можно говорить только об эквивалентности, так как данный способ оценки перевода рассматривает только соответствие единиц в текстах, без учета собственно смыслового критерия³; выдвигается предположение, что если использованы эквивалентные единицы, то и смысл текста на ПЯ будет эквивалентен смыслу текста на ИЯ.

Ни цель, ни смысл, на которые ориентирована адекватность, здесь не рассматриваются.

2. Коэффициент словесных ошибок, Word Error Rate (WER) — это алгоритм, основанный на расстоянии Левенштейна⁴ (т. н. edit distance), который применяется для вычисления необходимых преобразований, чтобы из одного слова получить другое (см.: [Marzal, Vidal 1993]).

³ К тому же не все из возможных эквивалентных категорий (синтаксический уровень, функция единицы языка в данном контексте) учитываются.

⁴ Оригинальная задача была поставлена академиком В. И. Левенштейном [Левенштейн 1965: 845–848].

Кратко опишем, как вычисляется расстояние Левенштейна. Если есть строки *A* и *B*, а также дан словарь конечного количества символов, то задача заключается в нахождении трех чисел (частный случай):

- число необходимых замен одного символа на другой (substitutions)
кот → *код* (замена [*m*] на [*∂*]);
- число необходимых удалений (deletions)
море → *мор* (удаление [*e*]);
- число необходимых вставок (insertions)
мир → *мире* (добавление [*e*]).

WER рассчитывается, исходя из указанных выше метрик, но при этом определенных для каждого слова в строке. Это можно представить в виде формулы (2) или (3):

$$(S + D + I) / N, \quad (2)$$

где *S* — число необходимых замен, *D* — число необходимых удалений, *I* — число необходимых вставок и *N* — количество слов.

Или:

$$(S + D + I) / (S + D + C), \quad (3)$$

где *C* — это количество верных слов, т. е. тех, которые не требуют ни одной из операций.

Данный алгоритм подсчета схожести двух строк берет в расчет только уникальные знаки, что полностью игнорирует их смысл и сочетаемость, а для профессионального перевода эта характеристика не может оставаться в стороне.

Если в контексте этой метрики говорить об адекватности и эквивалентности, то никаких существенных отличий от косинусного коэффициента, описанного в предыдущем пункте, здесь не будет: в расчет берется только эквивалентность единиц текстов на ИЯ и на ПЯ, без привязки к смыслу этих единиц и их сочетаний; тем более не идет речи ни о цели, ни об адекватности в целом.

3. Модель оценки двуязычного соответствия, BiLingual Evaluation Understudy (BLEU), является одной из наиболее популярных метрик, и в ней применяются *N*-граммы (*N*-grams). В наиболее общем смысле, *N*-грамма — это некоторая последовательность символов (от отдельных букв до предложений), идущих друг за другом. В области автоматической обработки естественного языка чаще всего рассматривается последовательность слов — т. н. подпоследовательность (subsequence) данного предложения.

Приведем пример выделения *N*-грамм (от 1 до 3) из предложения «Виктор вышел за продуктами в магазин»:

- юни-граммы: *Виктор*, *вышел*, *за*, *продуктами*, *в*, *магазин*⁵;
- би-граммы: *Виктор вышел*, *вышел за*, *за продуктами*, *продуктами в*, *в магазин*;
- три-граммы: *Виктор вышел за*, *вышел за продуктами*, *за продуктами в*, *продуктами в магазин*.

⁵ Это наиболее общий случай — деление может происходить по фонетическим словам, а предлоги и союзы могут не учитываться.

Алгоритм BLEU, в отличие от ранее рассмотренных метрик, опирается не на количество ошибок (т.е. на пропорцию неправильно переведенного текста), а на то, насколько текст, переведенный машиной, «совпадает» с «идеальным» образцом, который был переведен человеком [Papineni et al. 2002]. В основе метода лежит вычисление количества N-грамм, которые присутствуют в переводе и в оригинале. Чаще всего используется несколько образцовых переводов, каждый из которых был сделан разными людьми. Для каждого из них вычисляется представленная метрика, которая рассчитывается по следующей формуле:

$$BLEU - n = \min \left(1, \frac{\text{output length}}{\text{reference length}} \right) * \sqrt[n]{\prod_{i=1}^n \text{precision}_i}, \quad (4)$$

где:

- precision — описывается как отношение числа совпадающих n-грамм на ИЯ и ПЯ к числу всех «идеальных» N-грамм, которые должны быть (см.: [Olson, Delen 2008: 138]);
- output length — длина текста произведенного перевода;
- reference length — длина «идеального» перевода;
- n — длина N-грамм (задается исследователем).

Первая часть, выбирающая минимум из единицы и отношения переводов идеального и произведенного, необходима для того, чтобы метрика лучше справлялась с очень короткими переводными отрезками, которые максимизируют параметр Precision, не отражая на деле реальной ситуации.

Для алгоритма BLEU характерна та же проблема, что и для ранее рассмотренных метрик: синтаксические позиции элементов не учитываются в оценке, хотя и в меньшей степени, так как используются N-граммы, а не отдельные слова. Таким образом, если рассматривать уровень предложения, может сложиться ситуация, когда смысл предложения в переводе будет диаметрально противоположным тому, который содержится в идеальном переводе, но BLEU будет достаточно высоким — вплоть до 1. Например:

- идеальный перевод: *police killed the gunman*;
- произведенный перевод 1: *police kill the gunman*;
- произведенный перевод 2: *the gunman kill police*.

Слово *killed* в идеальном переводе не соответствует *kill* в предложенных переводах, т.е. N-граммы с этим словом учитываться не будут. Это всего лишь один из способов оценки; помимо него также применяется оценка на тексте, который прошел предобработку. Можно было бы сделать лемматизацию или стемминг текста и оценивать качество уже на приведенных к общему виду формах.

Предложенные переводы имеют одинаковое количество схожих с идеальным текстом N-грамм: одну би-грамму и три юни-граммы. В зависимости от сложности вычислений, размера входного текста и желания исследователя размер окна выбора N-грамм может быть увеличен.

Основным достоинством данного метода по сравнению с рассмотренными ранее является учет соседствующих слов, т.е. вполне возможным оказывается нахождение устойчивых выражений, фразеологизмов и авторских слов. К категориям

ошибок, которые могут быть обнаружены с помощью данной метрики, можно отнести ошибки в словах и неправильное расположение единиц. Однако чтобы обнаружить данные ошибки, необходимы модификации алгоритма.

Главный недостаток, который делает использование данной метрики для оценки человеческого перевода практически невозможным, заключается в ее реализации: если в предложениях (идеальном и произведенном) нет ни одной одинаковой N-граммы, то значение данной метрики будет равно 0. В предыдущих же разделах отмечалось, что далеко не всегда можно сделать такой перевод, который бы использовал лексические и синтаксические эквиваленты. Смысл текстов на ИЯ и ПЯ может быть практически идентичен, но средства, с помощью которых эта идентичность была достигнута, могут существенно отличаться. Указанный недостаток привел к появлению модификаций данной метрики: NIST, ROUGE (с вариациями).

4. N-граммная метрика с явно заданным порядком, Metric for Evaluation of Translation with Explicit ORdering (METEOR), является одной из наиболее современных метрик оценки качества машинного перевода (ее последняя версия датирована 2014 г.). Алгоритм основывается на двух часто используемых метриках оценки качества в машинном обучении: precision («точность») и recall («полнота»). Последовательность реализации следующая:

- в качестве единицы оценки выступает предложение, которое разбивается на N-граммы;
- строится соответствие (alignment) между элементами из идеального перевода и произведенного (каждой N-грамме⁶ из произведенного перевода должен соответствовать 0 или 1 N-грамма из идеального перевода);
- вычисляется точность (precision) для N-грамм по следующей формуле:

$$P = m / w_t, \quad (5)$$

где m — число N-грамм, которые присутствуют как в идеальном переводе, так и в произведенном; w_t — общее число N-грамм в произведенном переводе;

- вычисляется полнота (recall) по следующей формуле:

$$R = m / w_r, \quad (6)$$

где m — то же, что и у точности; w_r — количество N-грамм в идеальном переводе;

- вычисляется гармоническое среднее точности и полноты по следующей формуле⁷:

$$F_{mean} = (\alpha * \beta * P * R) / (\alpha * R + \beta * P); \quad (7)$$

- изначально учитываются только юни-граммы, которые будут давать довольно высокие результаты для P и R, что совсем не обязательно означает хорошее построение перевода. Чтобы этого избежать, используется включение

⁶ Возможно использование лемматизации и стемминга для приведения языковых единиц к единообразному виду.

⁷ В котором каждая из метрик может иметь собственный вес — этот гиперпараметр настраивается вручную. В формуле это α и β , которые чаще всего находятся в следующем диапазоне: (0, 100).

штрафа (penalty) за неправильно подобранные N-граммы. Оценка штрафа рассчитывается по следующей формуле:

$$p = 0,5 * (c / u_m) ^ 3, \quad (8)$$

где c — несколько юни-грамм, которые соответствуют друг другу в произведенном и идеальном текстах (чем больше соответствий в текстах, тем меньше будет количество таких отрезков⁸, а следовательно, и значение c); u_m — количество юни-грамм, которым нашлось соответствие;

- конечное значение *METEOR* вычисляется следующим образом:

$$M = F_{mean} * (1 - p). \quad (9)$$

Несомненным плюсом данного способа оценки качества является возможность работы с большим количеством языков — необходимы лишь параллельные корпуса текстов. Как и у рассмотренных ранее N-граммных методов оценки, у данной метрики возникает та же проблема: не всегда перевод может быть произведен с помощью однозначно соответствующих лексических средств [Banerjee, Lavie 2005]. Для решения этой проблемы в следующих версиях данной метрики были введены дополнительные параметры:

- сравнивается не только точное соответствие лексических единиц и N-грамм, но и дополнительно их вариации — стеммы (основы слова), леммы (нормальные/словарные формы), синонимы⁹, перифразы. Если лексическая единица в гипотетическом переводе совпадает хотя бы с одной из данных вариаций, то значение соответствий тоже увеличивается;
- должна существовать таблица перифраз (paraphrase table) для определенного языка. Данные таблицы, изначально составлявшейся вручную, могут с помощью современных методов извлекаться автоматически [Denkowski, Lavie 2014].

Благодаря этим улучшениям решается одна из главных проблем, описанных в предыдущих метриках: необходимость точного соответствия единиц.

На практике данная метрика показывает лучший результат, чем алгоритм BLEU: на уровне корпуса соответствие с человеческой оценкой у *METEOR* равно 0,964, тогда как у BLEU — 0,817 [Banerjee, Lavie 2005].

С позиций экспертных категорий метрику *METEOR* выгодно отличает не такое пристальное внимание к абсолютно точному соответствию единиц — перевод может быть произведен и не полностью эквивалентными единицами. Кроме того, метрика учитывает эквивалентность на уровне перифраз, что приближает ее к человеческой оценке. О приоритете цели и смысла здесь, как и в других рассмотренных автоматических метриках, речи не идет, но метрикой учитываются новые уровни эквивалентности, позволяющие точнее отделять «плохой» перевод от «хорошего». Как и в случае с BLEU, для выделения ошибок алгоритму необходимы надстройки; спектр возможных ошибок расширяется за счет нахождения не только неточных соответствий, но и перифраз — алгоритм будет выделять мень-

⁸ Для идеального соответствия значение c будет равно единице.

⁹ Для этого должен быть подключен словарь синонимов, например, WordNet для английского языка.

шее количество ошибок¹⁰, так как будут учитываться стеммы, леммы, перифразы. Следовательно, редактору или самому переводчику будет легче сосредоточиться на оставшихся недочетах.

Названные выше метрики уже достаточно долгое время используются для автоматической оценки машинного перевода. Однако их практически невозможно использовать для оценки человеческого перевода по следующей причине: большинство из современных метрик — N-граммные. Перевод, например художественный, редко может быть оценен именно по N-граммному соответствию единиц в силу своей специфики, где даже на уровне предложения зачастую наблюдаются несоответствия, т. е. используются единицы неэквивалентные. Другие же виды текстов, более формальные и «сухие»: научные статьи, техническая документация — могут быть переведены с «N-граммной точностью»; следовательно, этот перевод может быть оценен с помощью названных метрик.

Полнота и гладкость как вспомогательные категории автоматизированной оценки перевода

На современном этапе развития компьютерной лингвистики не всякий категориальный параметр оценки перевода может быть реализован автоматически. К таковым, в частности, относятся полнота и гладкость перевода.

Первая из них имеет дело с собственно смыслом переведенного текста: насколько полно сообщение на ИЯ было передано средствами ПЯ, как точно был воссоздан смысл. При оценивании перевода по этому параметру во внимание не берется то, какими именно средствами было передано изначальное сообщение: главное, чтобы сообщение на ПЯ было тем же по содержанию, что и на ИЯ.

За оценку средств языка, с помощью которых был произведен перевод, отвечает второй параметр — гладкость. Здесь речь идет именно о единицах языка всех уровней: насколько представленный перевод отвечает законам языка переводящего безотносительно смысла, который в переводе был заложен.

Строго говоря, два данных параметра используются как вспомогательные при автоматической оценке машинного перевода. Когда автоматический метод не может вынести однозначного вердикта насчет качества перевода, решающей становится оценка эксперта.

Параметры полноты и гладкости, как правило, используются совместно, чтобы объективно оценить перевод с разных сторон. Продемонстрируем эффективность такого синергетического подхода на конкретном примере, предварительно обговорив процедуру оценивания:

- тексты на ИЯ и ПЯ предлагаются нескольким экспертам (обычно четырем или более);
- каждому переводу ставится две оценки по шкале от 1 (минимальная оценка: не отвечает требованиям категории) до 5/7/10 (максимальная оценка: перевод полностью удовлетворяет категории).
- для каждого из переводов оценки по соответствующим категориям суммируются и от суммы берется среднее значение для каждой из категорий;

¹⁰ По сравнению с WER, BLEU.

- итоговая оценка формируется как среднее из значений полноты и гладкости (см.: [Snover et al. 2009]).

Пример алгоритма оценки в действии (перевод с русского языка на английский):

Предложение на ИЯ: *Она пойдет в магазин.*

Возможные переводы:

1. *She will go shopping.*
2. *He does not go to the store.*
3. *She goes there.*
4. *She travel for there.*

Оценки, выставленные экспертами, можно представить таблицей (см. табл. 1), где для примера будет использована шкала от 1 до 7.

Таблица 1. Экспертные оценки перевода по параметрам полноты и гладкости

| Категория | Эксперт 1 | | Эксперт 2 | | Эксперт 3 | | Эксперт 4 | |
|---------------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|
| | Полнота | Гладкость | Полнота | Гладкость | Полнота | Гладкость | Полнота | Гладкость |
| Предложение 1 | 7 | 7 | 7 | 7 | 7 | 7 | 7 | 7 |
| Предложение 2 | 2 | 6 | 1 | 6 | 2 | 7 | 1 | 7 |
| Предложение 3 | 3 | 6 | 3 | 6 | 4 | 6 | 1 | 6 |
| Предложение 4 | 3 | 3 | 3 | 2 | 4 | 2 | 2 | 1 |

Средние значения по категориям представлены в таблице 2.

Таблица 2. Средние оценки перевода по параметрам полноты и гладкости

| Категория | Полнота | Гладкость |
|---------------|---------|-----------|
| Предложение 1 | 7 | 7 |
| Предложение 2 | 1,5 | 6,5 |
| Предложение 3 | 2,75 | 6 |
| Предложение 4 | 2,75 | 2 |

Итоговые оценки качества перевода экспертами сведены в таблицу 3.

Таблица 3. Итоговые оценки качества перевода

| Вариант перевода | Оценка |
|------------------|--------|
| Предложение 1 | 7 |
| Предложение 2 | 4 |
| Предложение 3 | 4,375 |
| Предложение 4 | 2,375 |

На основе итоговых оценок представляется возможным ранжировать предложенные варианты перевода от лучшего к худшему. Последовательность будет следующей: 1, 3, 2, 4.

Предложение под номером 2, которое по параметру гладкости получило высокую оценку (6,5), совсем не передает смысл, заложенный в предложении на ИЯ и, соответственно, получает низкую оценку полноты (1,5). Благодаря взаимодополнению методик оценивания мы получаем возможность, во-первых, рассмотреть перевод с нескольких сторон и, во-вторых, получить конкретное значение оценки, которое необходимо для сравнения. Наконец, это позволяет заключить, какие из аспектов — сами языковые средства или их построение и использование — некорректно переданы в переводе.

Однако при использовании данного подхода к оценке перевода нельзя забывать о человеческом факторе, поскольку в нем задействованы живые люди — эксперты, каждый из которых имеет собственную компетенцию и навыки, которые могут входить в противоречие с компетенциями других оценивающих¹¹. К тому же в изначальном своем виде, где оценка происходит на основе присвоения некоторого рейтинга, данный способ оценивания перевода не дает представлений о том, где именно произошли ошибки при переводе. Об этом можно узнать только от самих экспертов, причем составление отчета по каждому из переводов может отнять большое количество времени.

Исходя из приведенного анализа категорий полноты и гладкости, представляется возможным провести их аналогию с категориями адекватности и эквивалентности из области экспертной оценки перевода. Полнота перевода, рассмотренная в данном пункте, при сопоставлении может соотноситься с адекватностью: насколько полно было передано сообщение, были ли переданы прагматические установки текста на ИЯ, заложена ли в тексте на ПЯ та же интенция. Гладкость перевода возможно соотнести с эквивалентностью: насколько тексты на ИЯ и на ПЯ похожи в использованных языковых средствах, насколько (не)правильно построен перевод с точки зрения языка. Для последнего сравнения необходима оговорка: имеется в виду тот случай, когда текст на ИЯ построен по всем правилам языка; следовательно, если перевод сделан эквивалентными средствами и по эквивалентным для ПЯ правилам, то это означает, что эксперт, квалифицирующий такой перевод по категории гладкости, поставил бы высокую оценку.

Приводя сравнение параметров оценки, авторы не ставили своей целью выявить полное категориальное соответствие, т. к. оно не представляется возможным по причине взаимопроникновения категорий адекватности и эквивалентности. Полнота перевода может считаться скорее категорией, которая включает в себя как адекватность, так и эквивалентность. Сопоставление полноты с адекватностью было предложено ранее. Полноту, рассматривающую не только смысл сообщения, но и то, как оно было передано, т. е. его жанровую и стилистическую принадлежности, возможно соотнести с эквивалентностью, которой на разных уровнях языка оценивается соответствие языковых единиц ИЯ и ПЯ (не только на лексическом и синтаксическом уровнях, но и с точки зрения стилистической отнесенности).

¹¹ Этот недостаток может быть компенсирован с помощью привлечения большего числа экспертов, что сгладит неверные оценки. Однако это, очевидно, увеличит время, необходимое для обработки результатов.

Наиболее логичным, исходя из данного соответствия, выглядит однозначное отделение смысла сообщения, его прагматических характеристик — полноты перевода — и того, как текст построен на ПЯ, — гладкости перевода — со следующим дополнением: смысл должен учитываться в первую очередь. Если значение высказывания не было передано в переводе, то нет смысла и говорить о том, насколько перевод соответствует правилам ПЯ, ведь перевод, в первую очередь, является средством передачи сообщения из одной системы знаков в другую (см.: [Сдобников, Петрова 2007: 110]). Таким образом, ценность гладкости перевода как категории его оценки неизбежно ставится под сомнение из-за названной причины. Ее, скорее, следует рассматривать в разрезе того, как переводчик или переводящая система — если говорить о машинном переводе — умеет работать с переводящим языком.

Выводы

Оценка качества перевода как один из прикладных аспектов переводоведения требует комплексного рассмотрения. С лингвистических позиций наиболее предпочтительной и логичной концепцией оценки представляется квалификация перевода на основе адекватности и эквивалентности.

Эквивалентность как более частная категория описывает отношения между исходными и конечными текстами, т. е. она ориентирована на результат перевода — соответствие языковых единиц в разных языках. Эквивалентность может быть практически абсолютной, если мы говорим об отдельных языковых знаках или единицах. Однако подобную ситуацию сложно представить не только для текстов, но и для предложений: языки, культура, ценности слишком сильно отличаются у разных народов, что, бесспорно, влияет на возможность достижения абсолютно эквивалентного перевода. С другой стороны, адекватность — это категория, присущая переводу скорее как процессу. Она носит не идеальный, а оптимальный характер: отклоняются самые неудачные варианты, и остается наилучший из возможных, который совсем не обязательно будет отвечать всем идеальным требованиям.

Таким образом, если эквивалентность отвечает на вопрос о том, соответствует ли конечный текст исходному, то адекватность указывает, соответствует ли перевод как процесс данным коммуникативным условиям; это дискурсивная характеристика.

Критерии эквивалентности и адекватности перевода, подробно разработанные в теоретическом переводоведении, находят свое применение в области статистической и автоматической оценки качества перевода. Нами были проанализированы автоматические метрики и их связь с экспертными категориями. Наиболее эффективной из них можно считать N-граммную метрику с явно заданным порядком МЕТЕОР, однако с успехом ее можно применять лишь для формальных стандартизированных текстов. Из экспертных категорий в автоматических метриках, рассмотренных в данной работе, находит отражение только эквивалентность. Адекватность — более комплексная категория — не репрезентируется в данных метриках, что делает невозможным использование автоматических метрик для оценки человеческого перевода без существенных дополнений и без включения в процесс оценки перевода эксперта.

Наконец, сугубо математические процедуры, которые были рассмотрены, не могут служить полноценными лингвистическими метамоделями оценки до тех пор, пока они не будут дополнены синтаксическими и семантическими алгоритмами.

Словари и справочные материалы

- Ахманова 1966 — Ахманова О. С. Эквивалент. *Словарь лингвистических терминов*. М.: Советская энциклопедия, 1966. С. 508.
- Латышев 2005 — Латышев Л. К. *Технология перевода: учеб. пособие для студентов лингвистич. вузов и фак.* М.: Академия, 2005.

Литература

- Аносова 2011 — Аносова А. А. Критерии оценки качества перевода поликодового текста. *Научно-технические ведомости СПбГПУ. Гуманитарные и общественные науки*. 2011, (3): 46–49.
- Гарбовский 2004 — Гарбовский Н. К. *Теория перевода*. М.: Изд-во Моск. гос. ун-та, 2004.
- Григорьев, Мартыненко 2019 — Григорьев Ю. Д., Мартыненко Г. Я. Комбинаторные варианты рифм в сонетах Рильке: квантитативно-типологический подход. *Вестник НГУ. Серия: Лингвистика и межкультурная коммуникация*. 2019, (1): 5–20.
- Евтеев 2017 — Евтеев С. В. Перевод: эквивалентно — насколько возможно, и адекватно — насколько нужно. *Вестник Брянского государственного университета*. 2017, (1): 262–267.
- Илюшкина 2017 — Илюшкина М. Ю. *Теория перевода: основные понятия и проблемы*. М.: Флинта, 2017.
- Комиссаров 1990 — Комиссаров В. Н. *Теория перевода (лингвистические аспекты)*. М.: Высшая школа, 1990.
- Комиссаров 1999 — Комиссаров В. Н. *Общая теория перевода. Проблемы переводоведения в освещении зарубежных ученых*. М.: ЧеРо; Юрайт, 1999.
- Левенштейн 1965 — Левенштейн В. И. Двоичные коды с исправлением выпадений, вставок и замещений символов. *Доклады АН СССР*. 1965, 163 (4): 845–848.
- Митренина 2016 — Митренина О. В. Машинный перевод. В кн.: *Прикладная и компьютерная лингвистика*. Николаев И. С., Митренина О. В., Ландо Т. М. (ред.). М.: URSS, 2016. С. 156–189.
- Мошкович 2012 — Мошкович В. В. Взаимосвязь и взаимовлияние адекватности и эквивалентности. *Вестник ЮУрГПУ*. 2012, (10): 270–279.
- Найда 1978 — Найда Ю. К науке переводить. *Вопросы теории перевода в зарубежной лингвистике*. М.: Междунар. отношения, 1978. С. 114–137.
- Портер 2004 — Портер Л. Г. Количественные критерии адекватности поэтического перевода. *Мир перевода*. 2004, 1 (11): 39–54.
- Прунч 2015 — Прунч Э. *Пути развития западного переводоведения. От языковой асимметрии к политической*. М.: Р. Валент, 2015.
- Рецкер 1974 — Рецкер Я. И. *Перевод и переводческая практика*. М.: Междунар. отношения, 1974.
- Сдобников 2007 — Сдобников В. В. *Методические основы подготовки переводчиков: нижегородский опыт*. Н. Новгород: Изд-во Нижегород. гос. лингвистич. ун-та им. Н. А. Добролюбова, 2007.
- Сдобников, Петрова 2007 — Сдобников В. В., Петрова О. В. *Теория перевода*. М.: АСТ: Восток-Запад, 2007.
- Скробнев 1984 — Скробнев Ю. М. *Основы стилистики английского языка*. Киев: Вища школа, 1984.
- Финкель 2001 — Финкель А. М. О точности стихотворного перевода. *Московский лингвистический журнал*. 2001, 5 (2): 127–186.
- Швейцер 1988 — Швейцер А. Д. *Теория перевода: Статус, проблемы, аспекты*. М.: Наука, 1988.
- Banerjee, Lavie 2005 — Banerjee S., Lavie A. METEOR: An Automatic Metric for MT Evaluation with Improved Correlation with Human Judgments. In: *Proceedings of Workshop on Intrinsic and Extrinsic Evaluation Measures for MT and/or Summarization at the 43rd Annual Meeting of the Association of Computational Linguistics (ACL 2005)*. Ann Arbor, 2005. P. 65–72.

- Brownlee 2017 — Brownlee J. A Gentle Introduction to the Bag-of-Words Model. In: *Deep Learning for Natural Language Processing*. <https://machinelearningmastery.com/gentle-introduction-bag-words-model/> (accessed: 03.02.2020).
- Catford 1965 — Catford J.C. *A Linguistic Theory of Translation: An Essay in Applied Linguistics*. Oxford: Oxford University Press, 1965.
- Chesterman, Wagner 2002 — Chesterman A., Wagner E. *Can Theory Help Translators? A Dialog between the Ivory Tower and the Wordface*. Manchester: St. Jerome Publ., 2002.
- Goldberg et al. 2017 — Goldberg Y., Levy O., Søgaard A. A Strong Baseline for Learning Cross-Lingual Word Embeddings from Sentence Alignments. In: *Proceedings of the 15th Conference of the European Chapter of the Association for Computational Linguistics*. Vol. 1. 2017. P. 765–774.
- Denkowski, Lavie 2014 — Denkowski M., Lavie A. Meteor Universal: Language Specific Translation Evaluation for Any Target Language. In: *Proceedings of the Ninth Workshop on Statistical Machine Translation*. Baltimore, 2014. P. 376–380.
- Marzal, Vidal 1993 — Marzal A., Vidal E. Computation of Normalized Edit Distance and Applications. *IEEE Trans. Pattern Analysis and Machine Intelligence*. 1993, 15 (9): 926–932.
- Mikolov et al. 2013 — Mikolov T., Chen K., Corrado G., Dean J. Efficient Estimation of Word Representations in Vector Space. In: *Proceedings of the International Conference on Learning Representations*. 2013. P. 1–12.
- Singhal 2001 — Singhal A. Modern Information Retrieval: A Brief Overview. *IEEE Data Eng. Bulletin*. 2001, (24): 35–43.
- Olson, Delen 2008 — Olson D. L., Delen D. *Advanced Data Mining Techniques*. Berlin; Heidelberg: Springer-Verlag Berlin Heidelberg, 2008.
- Papineni et al. 2002 — Papineni K., Roukos S., Ward T., Zhu W.-J. BLEU: a Method for Automatic Evaluation of Machine translation. *Proceedings of the 40th Annual Meeting of the Association for Computational Linguistics (ACL)*. Philadelphia, 2002. P. 311–318.
- Reiss, Vermeer 1984 — Reiss K., Vermeer H. J. *Grundlegung einer allgemeinen Translationstheorie*. Tübingen: De Gruyter, 1984.
- Snover et al. 2009 — Snover M., Madnani N., Dorr B., Schwarz R. Fluency, Adequacy, or HTER? Exploring Different Human Judgments with a Tunable MT Metric. In: *Proceedings of the Fourth Workshop on Statistical Machine Translation Athens, Greece, March 30–31, 2009*. Athens, 2009. P. 259–268.
- Vuorikoski 2016 — Vuorikoski A.-R. Importance of being accurate. Труды Санкт-Петербургского государственного университета культуры. Культура и деловой иностранный язык. Т. 214: м-лы III междунар. науч.-практ. конф. 12–13 марта 2015 г., Санкт-Петербург. СПб, 2016. P. 101–109.

Статья поступила в редакцию 25 октября 2020 г.
Статья рекомендована к печати 29 ноября 2021 г.

Alexey V. Sosnin

HSE University,
25/12, Bolshaia Pecherskaia ul., Nizhnii Novgorod, 603155, Russia
asosnin@hse.ru

Julia V. Balakina

HSE University,
25/12, Bolshaia Pecherskaia ul., Nizhnii Novgorod, 603155, Russia
jbalakina@hse.ru

Andrey N. Kashikhin

HSE University,
25/12, Bolshaia Pecherskaia ul., Nizhnii Novgorod, 603155, Russia
logerk3@gmail.com

Interdependence of expert categories and automated metrics applied to evaluate translation quality

For citation: Sosnin A. V., Balakina Ju. V., Kashikhin A. N. Interdependence of expert categories and automated metrics applied to evaluate translation quality. *Vestnik of Saint Petersburg University. Language and Literature*. 2022, 19 (1): 125–148. <https://doi.org/10.21638/spbu09.2022.107> (In Russian)

The article evaluates the quality of translation; we consider the applied and pragmatic aspects of such evaluation in the conditions of the current rapid increase in the number of texts to be translated. The article summarizes a plethora of assessment principles, each having its merits and drawbacks, and examines the correlation between the categories of adequacy and equivalence as the basic assessment parameters. We conclude that equivalence is oriented towards the result of translation, whereas adequacy indicates whether translation as a process corresponds to the communicative situation given. Adequacy is thus viewed as a discursive feature. The article also looks into the possibility of applying these assessment criteria to automated evaluation of translation quality, and compares the specific computer metrics or, algorithms: *Cosine Similarity*, *Word Error Rate*, *BLEU*, *ROUGE*, *NIST*, *METEOR*. We analyze the potential for applying these metrics to man-made rather than machine translated. The n-gram metric for evaluation of translation with explicit ordering *Meteor* is the most efficient algorithm; however, it can only be applied advantageously to formal standardized texts. The article points out that literary translation can hardly be evaluated on the basis of n-gram correspondences due to its characteristic feature: ample use of non-equivalent units even at the sentence level. The purely mathematical procedures overviewed in the article cannot serve as linguistic meta-models of assessment unless supplemented with syntactic and semantic algorithms.

Keywords: evaluation of translation quality, adequacy, equivalence, machine translation, automated metrics for evaluation of translation, word embedding.

References

- Аносова 2011 — Anosova A. A. Criteria of Polycode Text Translation Quality. *Nauchno-tehnicheskie vedomosti SPbGPU. Gumanitarnye i obshchestvennye nauki*. 2011, (3): 46–49. (In Russian)
- Гарбовский 2004 — Garbovskii N. K. *Theory of Translation*. Moscow: Izdatel'stvo Moskovskogo gosudarstvennogo universiteta Publ., 2004. (In Russian)
- Григорьев, Мартыненко 2019 — Grigoriev Yu. D., Martynenko G. Ya. Combinatorics of Rhyming Variants in Sonnets by R. M. Rilke: A Quantitative and Typological Approach. *Vestnik Novosibirskogo gosudarstvennogo universiteta. Seriya: Lingvistika i mezhkul'turnaia kommunikatsiia*. 2019, (1): 5–20. (In Russian)

- Евтеев 2017 — Evteev S. V. Translation: Equivalency as Possible, Adequacy as Needed. *Vestnik Brianskogo gosudarstvennogo universiteta*. 2017, (1): 262–267. (In Russian)
- Илюшкина 2017 — Ilyushkina M. Yu. *Theory of Translation: Basic Concepts and Issues*. Moscow: Flinta Publ., 2017. (In Russian)
- Комиссаров 1990 — Komissarov V. N. *Theory of Translation (Linguistic Aspects)*. Moscow: Vyshaia Shkola Publ., 1990. (In Russian)
- Комиссаров 1999 — Komissarov V. N. *General Theory of Translation. Issues of Translation Studies as Viewed by Foreign Scholars*. Moscow: CheRo Publ.; Iurait Publ., 1999. (In Russian)
- Левенштейн 1965 — Levenshtein V. I. Binary Codes with the Correction of Symbol Omissions, Insertions, and Substitutions. *Doklady Akademii nauk SSSR*. 1965, 163 (4): 845–848. (In Russian)
- Митренина 2016 — Mitrenina O. V. Machine Translation. In: *Prikladnaia i komp'uternaia lingvistika*. Moscow: URSS Publ., 2016. P. 156–189. (In Russian)
- Мошкович 2012 — Moshkovich V. V. Interconnection and Mutual Influence Between Adequacy and Equivalence. *Vestnik Iuzhno-Uralskogo gosudarstvennogo gumanitarno-pedagogicheskogo universiteta*, 2012, (10): 270–279. (In Russian)
- Найда 1978 — Naida E. Toward a Science of Translating. In: *Issues of Translation Theory in Foreign Linguistics*. Moscow: Mezhdunarodnye otnosheniia Publ., 1978. (In Russian)
- Портер 2004 — Porter L. G. Quantitative Criteria of Translation Adequacy. *The World of Translation*. 2004, 1 (11): 39–54. (In Russian)
- Прунч 2015 — Prunč E. *Ways of development of Western translation studies. From linguistic asymmetry to political*. Moscow: R. Valent Publ., 2015. (In Russian)
- Рецкер 1974 — Retsker Ya. I. *Translation and Practice of Translation*. Moscow: Mezhdunarodnye otnosheniia Publ., 1974. (In Russian)
- Сдобников 2007 — Sdobnikov V. V. *Methodological Basis for Teaching Translation: Nizhny Novgorod Experience*. Nizhnii Novgorod: Izdatel'stvo Nizhegorodskogo gosudarstvennogo lingvisticheskogo universiteta Publ., 2007. (In Russian)
- Сдобников, Петрова 2007 — Sdobnikov V. V., Petrova O. V. *Theory of Translation*. Moscow: AST Publ., 2007. (In Russian)
- Скребнев 1984 — Skrebnev Yu. M. *Fundamentals of the Stylistics of the English Language*. Kiev: Visha shkola, 1984. (In Russian)
- Финкель 2001 — Finkel' A. M. Towards the Precision of Verse Translation. *Moskovskii lingvisticheskii zhurnal*. 2001, 5 (2): 127–186. (In Russian)
- Швейцер 1988 — Shveitser A. D. *Theory of Translation: Status, Issues, Aspects*. Moscow: Nauka Publ., 1974. (In Russian)
- Banerjee, Lavie 2005 — Banerjee S., Lavie A. METEOR: An Automatic Metric for MT Evaluation with Improved Correlation with Human Judgments. In: *Proceedings of Workshop on Intrinsic and Extrinsic Evaluation Measures for MT and/or Summarization at the 43rd Annual Meeting of the Association of Computational Linguistics (ACL 2005)*. Ann Arbor, 2005. P. 65–72.
- Brownlee 2017 — Brownlee J. A Gentle Introduction to the Bag-of-Words Model. In: *Deep Learning for Natural Language Processing*. <https://machinelearningmastery.com/gentle-introduction-bag-words-model/> (accessed: 03.02.2020).
- Catford 1965 — Catford J. C. *A Linguistic Theory of Translation: An Essay in Applied Linguistics*. Oxford: Oxford University Press, 1965.
- Chesterman, Wagner 2002 — Chesterman A., Wagner E. *Can Theory Help Translators? A Dialog between the Ivory Tower and the Wordface*. Manchester: St. Jerome Publ., 2002.
- Goldberg et al. 2017 — Goldberg Y., Levy O., Søgaard A. A Strong Baseline for Learning Cross-Lingual Word Embeddings from Sentence Alignments. In: *Proceedings of the 15th Conference of the European Chapter of the Association for Computational Linguistics*. Vol. 1. Valencia, 2017. P. 765–774.
- Denkowski, Lavie 2014 — Denkowski M., Lavie A. Meteor Universal: Language Specific Translation Evaluation for Any Target Language. In: *Proceedings of the Ninth Workshop on Statistical Machine Translation*. Baltimore, 2014. P. 376–380.
- Marzal, Vidal 1993 — Marzal A., Vidal E. Computation of Normalized Edit Distance and Applications. *IEEE Trans. Pattern Analysis and Machine Intelligence*. 1993, 15 (9): 926–932.

- Mikolov et al. 2013 — Mikolov T., Chen K., Corrado G., Dean J. Efficient Estimation of Word Representations in Vector Space. *Proceedings of the International Conference on Learning Representations*. 2013. P. 1–12.
- Singhal 2001 — Singhal A. Modern Information Retrieval: A Brief Overview. *IEEE Data Eng. Bulletin*. 2001, (24): 35–43.
- Olson, Delen 2008 — Olson D. L., Delen D. *Advanced Data Mining Techniques*. Berlin; Heidelberg: Springer-Verlag Berlin Heidelberg, 2008.
- Papineni et al. 2002 — Papineni K., Roukos S., Ward T., Zhu W.-J. BLEU: a Method for Automatic Evaluation of Machine translation. In: *Proceedings of the 40th Annual Meeting of the Association for Computational Linguistics (ACL)*. Philadelphia, 2002. P. 311–318.
- Reiss, Vermeer 1984 — Reiss K., Vermeer H. J. *Grundlegung einer allgemeinen Translationstheorie*. Tübingen: De Gruyter, 1984. 245 S.
- Snover et al. 2009 — Snover M., Madnani N., Dorr B., Schwarz R. Fluency, Adequacy, or HTER? Exploring Different Human Judgments with a Tunable MT Metric. In: *Proceedings of the Fourth Workshop on Statistical Machine Translation Athens, Greece, March 30–31, 2009*. Athens, 2009. P. 259–268.
- Vuorikoski 2016 — Vuorikoski A-R. Importance of being accurate. *Trudy Sankt-Peterburgskogo gosudarstvennogo instituta kul'tury. Kul'tura i delovoi inostrannyi iazyk*. Vol. 214: Materialy III mezhdunarodnoi nauchno-prakticheskoi konferentsii, 12–13 marta 2015, Sankt-Peterburg. St Petersburg, 2016. P. 101–109.

Received: October 25, 2020

Accepted: November 29, 2021