УДК 81'374.822
DOI: 10.17223/22274200/16/6

Е.Б. Берг, М. Кит

ПОИСКИ РЕШЕНИЯ ПРОБЛЕМ ДВУЯЗЫЧНОЙ ИНТЕРНЕТ-ЛЕКСИКОГРАФИИ В СЛОВАРНОМ ПРОЕКТЕ LexSite

В статье приводятся результаты исследования наиболее популярных двуязычных интернет-словарей. Описываются проблемы их функционирования: отсутствие систематизации переводов, некачественный перевод фразеологизмов, неразличение омоформ, имитация переводов отсутствующих слов, ошибочный грамматический и лексический комментарий, проникновение недостоверной информации в словники. Приведено описание разработанного авторами словарного проекта LexSite, в котором предпринята попытка решить названные проблемы.
Ключевые слова: *интернет-словарь, двуязычная лексикография, словарь LexSite, словари-переводчики, машинный перевод, словник.*

Одним из важнейших инструментов межъязыковой коммуникации являются двуязычные словари. За последние десятилетия словарь как источник информации претерпел существенные изменения. С развитием информационных технологий традиционные печатные словари стали вытесняться словарями электронными, а затем и интернет-словарями [1. С. 188], предлагающими высокую скорость поиска информации и обладающими новыми свойствами [2. С. 2]. Несмотря на многочисленные возможности, открывающиеся перед составителями словарей благодаря новым технологиям [3. С. 143–144], исследователи отмечают, что создание интернет-словарей сталкивается с многочисленными трудностями [4. С. 234], которые, в свою очередь, вызывают системное появление определенных недостатков таких словарей. «Идея двуязычного словаря обманчиво проста: представить все известные смыслы заглавного слова лексическими единицами языка перевода» [5. С. 329]. «К сожалению, выполнение этой задачи в большинстве случаев весьма затруднительно, а временами почти невозможно. Причиной тому являются три качества, присущие естественным языкам… размытость семантического поля, многозначность и отсутствие взаимно-однозначного соответствия между разными лексическими системами» [6. С. 225].

Проблемы, с которыми сталкиваются составители двуязычных и многоязычных интернет-словарей, не зависят от языков[1] и являются общими для большинства словарей. Соответственно, и недостатки у этих словарей оказываются общими. В данной статье остановимся на исследовании словарей, составленных для англо-русской языковой пары.

Для объективной оценки проблем двуязычной интернет-лексикографии и выяснения целесообразности создания нового словаря было проведено исследование четырех интернет-словарей, работающих в англо-русской языковой паре: Мультитран [7] (с 2001 г.), Google Translate [8] (с 2005 г.), ABBYY Lingvo Live [9] (с 2008 г.) и Яндекс.Переводчик [10] (с 2011 г.). Графические интерфейсы этих словарей представлены на рис. 1–4.

Выбор предмета исследования обусловлен степенью популярности данных ресурсов, проявляющейся в их посещаемости (табл. 1)[2], и, как следствие, степенью их влияния на язык и общество.

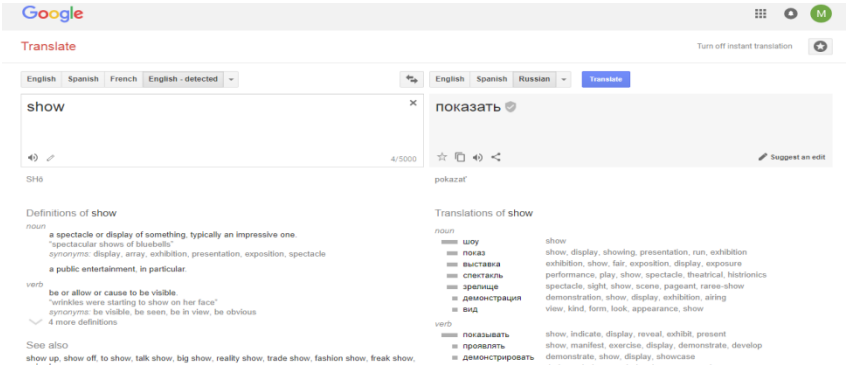

Рис. 1. Графический интерфейс Google Translate

[1] Существенные отличия имеют словари, работающие в паре «живой язык – мертвый язык», например, словарь англо-санскрит. «Особым случаем являются словари языков, носителей которых больше не существует… Такие словари, используемые преимущественно учеными, чаще основываются на дескриптивных объяснениях, чем на эквивалентах языка перевода. Это происходит оттого, что многие слова столь глубоко встроены в культуру языка-источника, что их перевод с помощью значений отдельных слов языка перевода невозможен» [6. С. 214], однако подобные словари не являются предметом рассмотрения данной статьи.

[2] По данным службы SimilarWeb, занимающейся сбором и анализом информации о посещаемости веб-сайтов [11], на июнь 2018 г.

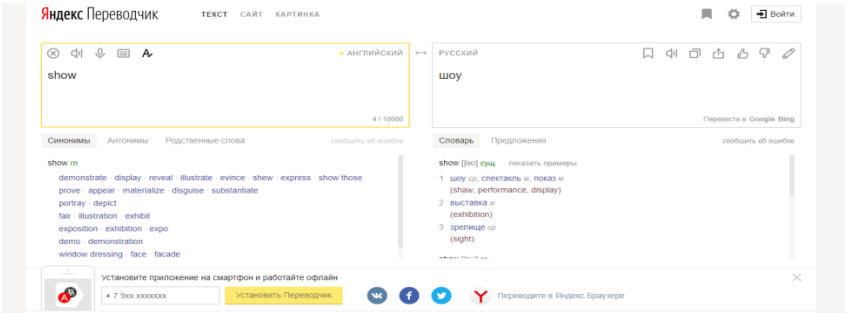

Рис. 2. Графический интерфейс Яндекс.Переводчик

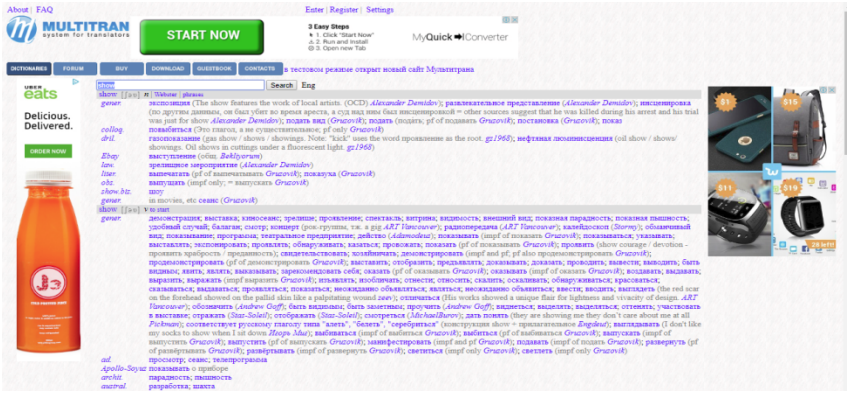

Рис. 3. Графический интерфейс словаря Мультитран

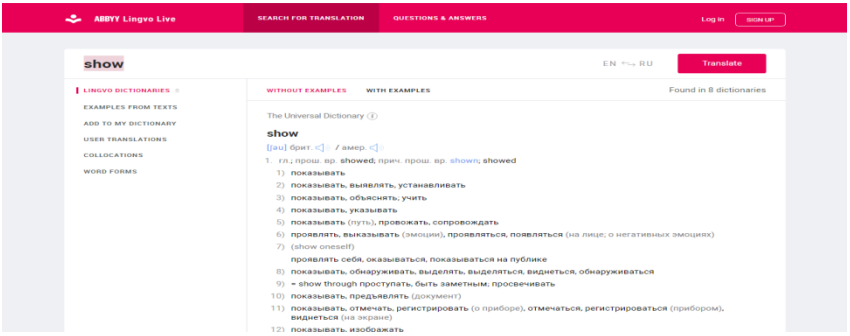

Рис. 4. Графический интерфейс словаря ABBYY Lingvo Live

Таблица 1

Посещаемость сайтов двуязычных (многоязычных) интернет-словарей

Ресурс	Посещений в месяц	Доля российских посетителей, %	Количество российских посетителей в месяц
Словарь-переводчик Google	820 млн	3,8	31 млн 160 тыс.
Словарь-переводчик Яндекс	45 млн	79	35 млн 550 тыс.
Словарь Мультитран	14,7 млн	38,1	5 млн 600 тыс.
Словарь ABBYY Lingvo	3 млн	36,8	1 млн 104 тыс.
Кембриджский словарь	51 млн	менее 1	менее 510 тыс.
Словарь Bab.la	47 млн	менее 1	менее 470 тыс.
Словарь Dict.com	736 тыс.	менее 4	менее 30 тыс.

Как видно из приведенных данных, словарь-переводчик Google ежемесячно обслуживает огромное количество посетителей. И хотя англо-русская языковая пара в нем не единственная и даже не самая востребованная, доля пользователей этой составляющей выражается в десятках миллионов человек. Несмотря на то, что Яндекс тоже предлагает несколько языковых пар, он рассчитан в первую очередь на русскоязычное Интернет-пространство, поэтому доля пользователей англо-русской языковой парой очень велика и так же выражается в десятках миллионов посетителей в месяц. К словарям Lingvo и Мультитран ежемесячно обращаются миллионы пользователей. Следовательно, именно эти четыре словаря, являясь наиболее популярными, оказывают максимальное влияние на представления общества о переводе в англо-русской языковой паре.

Исследуемые словари по механизму функционирования можно разделить на два типа: собственно словари и словари-переводчики. Собственно словари находят переводы, содержащиеся в их словниках, тогда как словари-переводчики комбинируют этот механизм с механизмом использования своих дополнительных алгоритмов, основанных на статистическом анализе. К первому типу относятся Lingvo и Мультитран, ко второму – Google и Яндекс. Большая часть выявленных проблем функционирования интернет-словарей является общей для обоих типов; вместе с тем словарям-переводчикам свойственны еще и специфические проблемы.

Прежде всего, для словарей-переводчиков характерна **скудность результатов**. Например, русское слово *башмак* и Google, и Яндекс пе-

реводят 6 существительными, тогда как существует по меньшей мере 15 научно-технических единиц наряду с тремя общеупотребительными. К слову *таблица*, имеющему не менее 12 значений, Яндекс дает 5 переводов, Google – 6. Из 28 английских значений слова *матрица* Яндекс приводит 5, Google – 7, оставляя за пределами своих словарных статей, например, перевод *stamper*, необходимый переводчику, работающему над текстом по гальванопластике, или *binder*, востребованный в тексте по материаловедению.

В отличие от словарей-переводчиков, собственно словари выдают на запрос значительно большее количество переводов, однако из-за недостаточной систематизации представляемых результатов получить необходимую информацию сложно. В словаре Мультитран первоначальное распределение переводов по тематическим группам исчезло в огромном количестве дополнений, при этом как слова, так и тематические группы нередко повторяются. Словарь Lingvo представляет переводы запрошенного слова последовательно по данным разных словарей, поэтому, как правило, один и тот же перевод показывается многократно.

Словари-переводчики, по определению предлагая выполнение **переводов** не только отдельных слов, но также **словосочетаний и** даже **текстов**, выполняют их крайне **некачественно**. Высмеивание содержания автоматически переведенных текстов стало общим местом в сознании и поведении пользователей интернета, но, как ни парадоксально, низкое качество перевода текстов (или фрагментов текста) не смущает заказчиков переводов и руководителей переводческих организаций, уже повсеместно стремящихся принять на работу не переводчика, а «редактора машинного перевода». При этом предлагаемый «редактору» для исправления машинный перевод – результат деятельности словарей-переводчиков – часто оказывается столь низкого качества, что не может быть отредактирован переводчиком; требуется новый перевод, выполненный с учетом смысла текста, а не заложенных в автоматический переводчик схем, этот смысл игнорирующих. Единственный вариант, когда словарь-переводчик выдает качественный перевод, – наличие готового перевода запрошенного текста в его базе данных. Однако совпадение запрошенного текста с текстами, содержащимися в базе данных, наблюдается нечасто. На практике даже минимальные синтаксические единицы – словосочетания – переводятся с ошибками: словосочетание *громко пища* и Google, и Яндекс переводят

как «loud food» – буквально «громкая еда»; *о пельменном тесте* Google переводит как «about the dumpling test», Яндекс – «about pelmeni test», и то и другое означает «о проверке пельменей».

Эта проблема распространяется также и на **фразеологизмы**, которые словари-переводчики часто принимают за синтаксическую единицу и переводят пословно или, в других случаях, хотя и рассматривают их как лексическую единицу, переводят неправильно. Например, *бить ключом* Google переводит как «beat the key» (дословно – «бить ключ»). Фразеологизм *all fur coat and no knickers*, означающий «шикарный на вид и пустой внутри» [12] оба словаря-переводчика переводят, во-первых, пословно, а во-вторых, с грамматическими ошибками: «все шубы и без трусов» (Google), «все шубы и без трусиков» (Яндекс). Выражение *утереть нос*, имеющее английский эквивалент *to wipe someone's eye* [13. Т. 3. С. 727]. Google переводит другой идиомой – «to rub one's nose», означающей «ткнуть носом» [14]. Это же выражение Яндекс переводит как «lose nose» («потерять нос»), что не является ни английской идиомой, ни дословным переводом на английский язык.

В связи с тем, что ни один словарь не может содержать абсолютно все слова того или иного языка, корректным ответом на запрос отсутствующего слова в двуязычном интернет-словаре должно являться сообщение об отсутствии слова или о невозможности найти перевод. Однако словари-переводчики никогда не показывают подобное сообщение, а разными способами **имитируют перевод отсутствующих** в их словниках **слов**. Основными способами имитации являются: транслитерация запрошенного слова (причем не всегда корректная), генерирование перевода слова путем перевода морфем запрошенного слова, подбор слова по тематической или графической ассоциации, формирование случайного (иногда стилизованного) слова. Примеры имитации переводов представлены в табл. 2. Строки 1–4 таблицы содержат слова, реально существующие в английском и русском языках; строки 5–8 – несуществующие слова, специально сконструированные для экспериментальной проверки гипотезы об имитации переводов.

Данная проблема имитации переводов словарями-переводчиками является более опасной, чем кажется на первый взгляд, поскольку вследствие высокой степени доверия пользователей к данным словарям она порождает ситуацию фальсификации переводов, выполняемых ежедневно в большом количестве по всему миру.

Таблица 2

Способы имитации переводов слов, отсутствующих в словаре

№	Запрошенное слово	Перевод, выдаваемый словарем-переводчиком	Фактическое значение выданного перевода	Способ имитации перевода
1	голодранец	holodranets (Google)	----------	транслитерация
2	синтепух	synthepus (Google)	----------	частичный морфемный перевод и неправильная транслитерация
3	ряженка	fermented burger (Google)	ферментированный гамбургер	ассоциативная связь и случайное слово
4	попадья	priest (Google)	священник	подбор слова по тематической ассоциации
5	king-of-the-herrings (обыкновенный сельдяной король, Regalecus glesne)	Царь-оф-селедки (Google)	----------	частичный пословный перевод сложного слова с элементами транслитерации
		Кинг-оф-селедки (Яндекс)	----------	
6	сикография	sycography (Google)	----------	морфемный перевод
		ecografia (Яндекс)	----------	формальный подбор ближайшего по буквенному составу слова
7	грандипуль	grandipad (Google)	----------	формирование стилизованной лексемы и частичная транслитерация
		grandeur (Яндекс)	грандиозность	подбор слова с графическим сходством
8	геолостический	geological (Google)	геологический	частичный морфемный перевод с элементами транслитерации
		geologicheskiy (Яндекс)	----------	
9	гигрофарий	hygrophoric (Google)	----------	частичный морфемный перевод
		gyrotary (Яндекс)	----------	формирование случайной лексемы
10	огурел	pickled (Google)	маринованный	подбор слова по тематической ассоциации
		Ogura (Яндекс)	----------	формирование случайной лексемы с графическим сходством

Зачастую пользователи (включая профессиональных переводчиков) мотивируют свое предпочтение исследуемых словарей именно тем, что в них, в отличие от других словарей, всегда можно найти перевод слова, достоверность перевода при этом под сомнение не ставится. И если при переводе с иностранного языка на родной пользователь может квалифицировать предлагаемый перевод как неправильный, то при переводе на иностранный язык ложный перевод, сымитированный словарно-переводческой системой, воспринимается как истинный и, будучи использованным при переводе текста, дает начало процессу переводческой фальсификации.

Общей для словарей-переводчиков и собственно словарей является проблема **неразличения омоформ**. Проявляется она в том, что пользователю, недостаточно хорошо знающему язык оригинала и запрашивающему слова в неосновной форме, при наличии у слова омоформ словарь может выдать перевод не требуемого слова, а другого, совпадающего с запрошенным в отдельных грамматических формах. Например, в качестве переводов русского слова *белила* и Google, и Lingvo дают переводы только существительного *белила*. При запросе слова *дуло* эти же словари дают переводы только существительного *дуло*. В ответ на запрос слова *гостя* Google показывает переводы только для слова *гость*, а Lingvo – только для слова *гостить*.

В некоторых случаях словари показывают омоформы, однако не дают никакого комментария к ситуации представления на один запрос нескольких очевидно разных слов. Так, для русского слова *карьер* Google в качестве основного варианта перевода дает *career* «карьера», а в качестве дополнительных – *quarry* и другие, означающие «карьер». Яндекс показывает переводы и слова *карьер*, и слова *карьера*, не комментируя, в каком отношении друг к другу они находятся. Lingvo дает переводы только для слова *карьер*. Мультитран переводит это слово сначала как *full gallop; career; quarry*, т.е. показывая переводы для слов *карьер* и *карьера* вперемешку, а потом приводит еще 77 вариантов перевода, нередко повторяя один и тот же по несколько раз и не разделяя переводы слов *карьер* и *карьера*. Очевидно, что без пояснений к приведенным переводам разных слов пользователь не может осуществить выбор перевода, актуального для его текста.

Носитель английского языка, выполняющий перевод с русского, запросив в словаре Google слово *еду*, в качестве основного варианта перевода получает *food* «еда», а в качестве дополнительных – пять

вариантов перевода на английский язык слова *ехать*. Обратившись с этим же запросом к словарю Яндекс и получив в ответ русское слово *еда* с десятью вариантами его перевода на английский язык и сразу же после него русское слово *ехать* с семью переводами, англоязычный пользователь навряд ли сможет разобраться в предоставленной информации. Lingvo показывает только переводы слова *еда*; Мультитран – сначала переводы слова *еда* (67 вариантов), потом – *ехать* (26) без комментариев относительно их связи с запрошенным словом *еду*. Отсутствие необходимого комментария при представлении омоформ затрудняет пользователю понимание содержания словарной статьи.

Словарные статьи двуязычных интернет-словарей часто содержат избыточный и неупорядоченный по структуре либо **ошибочный грамматический и лексический комментарий**. В частности, в разных словарях нередко встречается неправильное определение части речи. Словарь Мультитран, например, распределив переводы слова *lip* «губа» по частям речи, в группе глаголов указывает существительные *губа, край, выступ, порог* и еще несколько десятков существительных. Восклицание *блин!* по версии этого же словаря относится к существительным. Яндекс дает возможность ознакомиться с синонимами и антонимами к запрошенному слову, каковые отбираются не всегда корректно. Так, в качестве синонимов к слову *рот* представлены, среди прочего, *тор* и *топка*. Синонимом к слову *morning* «утро» Яндекс считает *night* «ночь». Словарь Lingvo на слово *белила* приводит перевод только существительного, а в разделе словарной статьи «Формы слова» на первом месте приводит парадигму спряжения глагола *белить*. Этот же словарь в каждой словарной статье приводит большое количество «Примеров из текстов», зачастую не имеющих отношения к представленному в словаре переводу слова. Вот один из приведенных примеров к существительному *дуло*: «Кричали дрозды, и по соседству в болотах что-то живое жалобно гудело, точно дуло в пустую бутылку. Чехов А.П. Студент». Почти все примеры из текстов к существительному женского рода *полка* фактически содержат существительное мужского рода *полк* в форме родительного падежа: «… последние ряды полка», «Здесь стоит командир полка…», «…прикомандировали к штабу полка…» и т.д. Иногда отмеченная безотносительность примеров усугубляется опечатками, попавшими в их тексты и доводящими содержание словарной статьи до абсурдного. Вот пример из текста для существительного *мята*: «Каждый раз,

когда **мни** ноги касались почвы, я ощущал ужасающий холод, проникавший все глубже. Кларк, Артур Чарльз / Острова в небе». Из приведенного здесь же оригинала следует, что в качестве иллюстрации употребления слова *мята* было использовано притяжательное местоимение *мои*, написанное в русском переводе с опечаткой: «Every time **my** feet touched the ground I could feel the appalling chill striking deeper».

Намерение обогатить словарную статью дополнительными сведениями оборачивается переполнением статьи плохо структурированной и зачастую ложной информацией.

Искажение сути словаря как нормативно-справочного издания (и, соответственно, неспособность выполнения рассмотренными системами справочной, нормативной, систематизирующей и учебной функций), происходящее в результате описанных причин, усугубляется путем активного **привлечения пользователей к составлению словаря**. Словари Google и Яндекс позволяют пользователю «предложить перевод», Мультитран дает возможность добавлять новые переводы и новые словарные статьи. Словарь Lingvo не только приглашает пользователей добавлять свои переводы, но и поощряет их присвоением званий «бронзовый», «серебряный» и «золотой». Привлечение широких масс и поощрение их участия в наполнении словников приводят к появлению не только многочисленных повторов и неправильных переводов, внесенных непрофессиональными «лексикографоманами», но и нецензурных переводов, намеренно добавляемых хулиганами.

Повторы одного и того же варианта перевода, засоряющие лексикографическое описание слова, и наличие неправильных переводов свойственны прежде всего словарям Lingvo и Мультитран. Так, в словаре Lingvo описание слова *proud*, имеющее четырежды (по данным четырех словарей) показанный перевод «гордый», дополнено тремя пользователями все тем же переводом «гордый», причем один из этих пользователей имеет статус «золотой англо-русский». Перевод слова *squirrel* – «белка» – так же показан составителями четыре раза, после чего представлены дополнения трех пользователей: тот же перевод «белка». Еще меньше повезло слову *белка* при переводе на английский язык: кроме того, что в ответ на свой словарный запрос пользователь получает четыре раза показанный перевод «squirrel» (по данным разных словарей), ему предлагается список из семи дополни-

тельных переводов от пользователей (включая имеющих статус «золотого русско-английского» и «бронзового русско-английского»), шесть из которых – «squirrel», а седьмой – «quarrel», являющийся переводом не запрошенного слова *белка*, а слова *ссора*. Аналогичные недостатки имеет описание слова *rise* «подъем; подниматься»: в дополнение к имеющимся словарным переводам пользователями трижды внесен перевод «подъем» и четырежды – «подниматься». Один из пользователей внес перевод «рис», перепутав заглавное слово *rise* со словом *rice*.

В словарь Мультитран «народные лексикографы» вносят целые страницы своих переводов, нередко не являющихся переводами заглавного слова, а представляющих собой их контекстные синонимы, что нарушает описание системных отношений между словами. Например, в качестве одного из переводов слова *beauty* «красота, прелесть» предлагается «гол-красавец». Обоснованность такого перевода мотивируется высказыванием неизвестного автора по поводу гола: «Boy, was it a beauty!». Перевод междометия *блин!* – «why» («почему»). Если следовать такому принципу перевода, то любое русское слово может стать переводом любого английского слова, поскольку существует некоторая вероятность использования его по каким-либо экстралингвистическим причинам в любом значении.

Кроме контекстных синонимов словарные статьи Мультитран переполнены периферийной лексикой – устаревшими, жаргонными, диалектными словами, внесенными пользователями, несмотря на то, что «для включения такой единицы в словарь всегда нужно иметь какое-то дополнительное основание. Таким основанием может служить актуальность слова в современной художественной или общественно-политической литературе, его распространенность в литературе предшествующей эпохи, его включенность в личный тезаурус культурного носителя языка» [15. С. 11]. Анализ периферийных лексем-переводов в словаре Мультитран приводит к выводу об отсутствии оснований включения этих переводов в словарь: уже упоминавшееся существительное *beauty* получило переводы «лепота», «баса», «краля», «красава», «пригожество», «соколена», «красеха», «пригожуня». Слово *handsome* «красивый (о мужчине)» имеет добавленные пользователями переводы «баской», «красовитый», «благолепный», «лепообразный» и другие. Более того, словарная статья словаря окончательно теряет системность изложения и трансформируется в форум народно-

лексикографической общественности, наполняясь, наряду с пользовательскими переводами, комментариями пользователей относительно переводов, внесенных другими, а также лексикографических способностей последних. Например, в перечне вышеприведенных переводов прилагательного *handsome* содержится реплика их «автора»: «now, please stop bothering me with your asinine comments!». Среди переводов слова *quarrel* «ссора» содержится фраза, добавленная одним из пользователей: «Это существительное, а не междометие». Многие пользователи словаря Мультитран в качестве переводов существительного *зло* внесли переводы для наречия *зло*; после каждого из таких переводов (8 раз) следует фраза другого пользователя: «Это наречие, а ни прилагательное, ни существительное» (орфография оригинала).

В словаре-переводчике Google содержатся нецензурные переводы (например, для слов *cock, exhaust, crap* и др.).

Вследствие допуска неограниченного количества лиц к лексикографической практике словарь из нормативно-справочного издания трансформируется в источник бессистемно представленной непроверенной информации, а в некоторых случаях и в площадку для декларирования непрофессионалами их лексикографических воззрений.

К отдельным недостаткам рассмотренных словарей, влияющих на общее качество, можно отнести отсутствие фонетической справки в словаре Мультитран (отсутствие возможности звукового воспроизведения слов), и такие внелексикографические факторы, как **обязательная регистрация** в словаре Lingvo и перегруженность **рекламой** страниц Мультитран.

Результаты исследования позволяют сделать вывод о том, что рассмотренные интернет-словари не являются надлежащими источниками нормативно-справочной информации, снабжая пользователя информацией непредсказуемого качества при отсутствии четкой структуры изложения.

Кроме того, выявление и анализ общих и частных проблем функционирования двуязычных интернет-словарей, а также оценка степени влияния данных словарей на общество привели к выводу о необходимости срочного поиска решений этих проблем. Предлагаемые нами решения воплотились в создании двуязычного интернет-словаря LexSite, функционирующего в англо-русской языковой паре [16].

В основе словника LexSite лежат материалы, накопленные и систематизированные в процессе 20-летней работы компании Language

Interface (США), обеспечивающей лингвистическую поддержку крупных международных проектов в широком диапазоне отраслей, от отчетов и докладов ООН до Международной космической станции. Описание материалов корректировалось по авторитетным лексикографическим источникам: Новому большому англо-русскому словарю в трех томах [13], Большому современному англо-русскому и русско-английскому словарю [17], Большому англо-русскому политехническому словарю [18] и другим англо-русским и русско-английским отраслевым словарям, а также толковым словарям английского языка.

Наряду с общеупотребительной лексикой LexSite содержит термины науки и техники, а также некоторые частотные профессионализмы. Графический интерфейс словаря представлен как на русском, так и на английском языке с возможностью выбора языка в меню (рис. 5, 6).

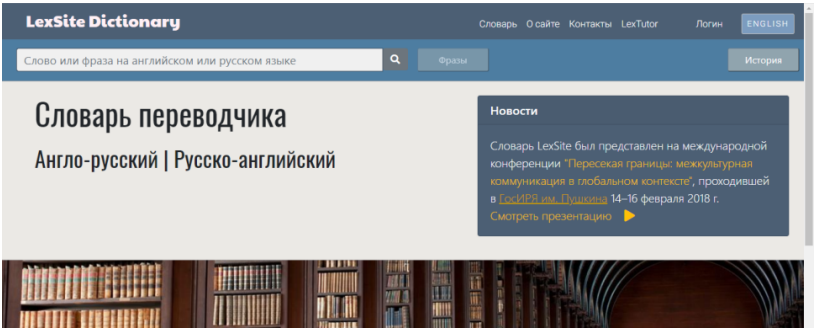

Рис. 5. Графический интерфейс словаря LexSite на русском языке

Рис. 6. Графический интерфейс словаря LexSite на английском языке

Переводы запрошенного пользователем слова размещаются в отдельных столбцах в соответствии с частиречной принадлежностью слова (рис. 7). Такой способ представления переводов позволяет показать большое количество вариантов переводов (включая специальную лексику) в удобной для восприятия форме.

Прилагательные, причастия и порядковые числительные объединены в одну группу («прил. / прич.»); количественные числительные, местоимения, междометия, предлоги и союзы – в группу «прочее» для возможности компактного представления на экране, при этом внутри столбца «прочее» часть речи указывается рядом с каждым словом.

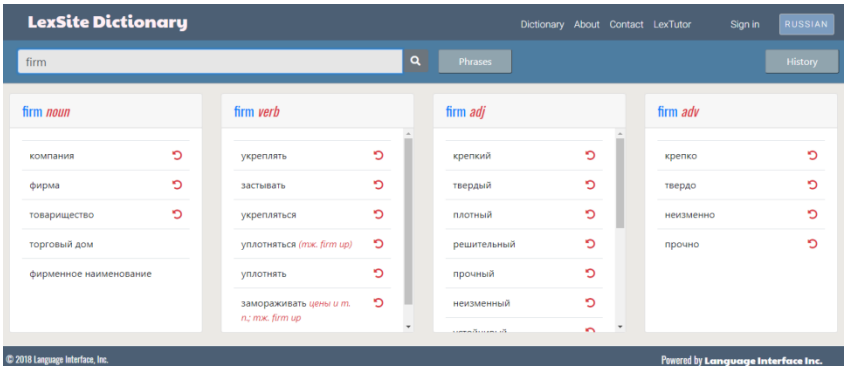

Рис. 7. Демонстрация переводов запрошенного слова

При отсутствии запрошенного слова в словнике словарь не имитирует перевод, а показывает оповещение о невозможности найти слово (рис. 8).

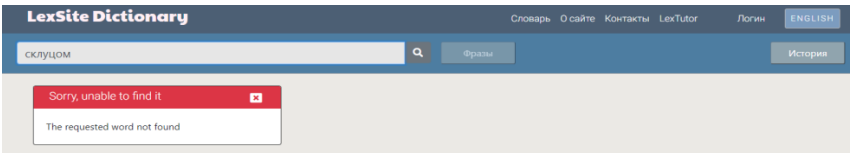

Рис. 8. Оповещение об отсутствии перевода слова

С целью устранения неопределенности при наборе слова с опечатками или орфографическими ошибками словарь показывает окно-оповещение со списком графически сходных слов для выбора необходимого слова (рис. 9).

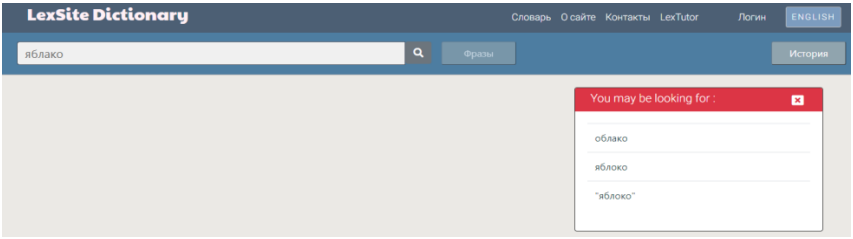

Рис. 9. Устранение неопределенности при опечатках и орфографических ошибках в запрошенном слове

Словарь обеспечивает возможность ознакомиться с переводом словосочетаний, включающих искомое слово, и с употреблением искомого слова в лингвистическом контексте (рис. 10).

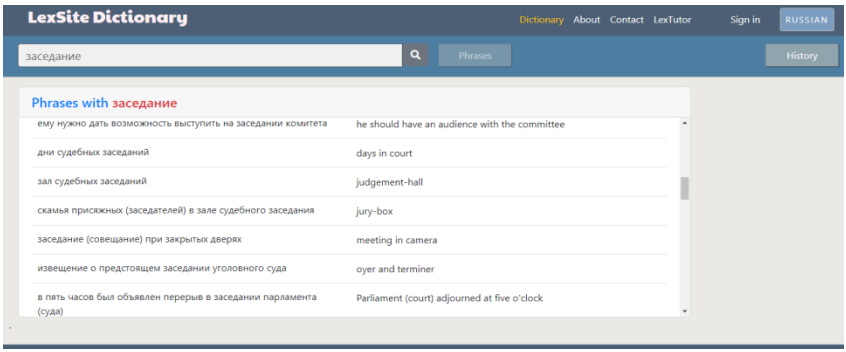

Рис. 10. Перевод словосочетаний с искомым словом и его употребления в лингвистических контекстах

Кроме того, словарь расценивает фразеологические единицы как элементы лексического уровня языковой системы и показывает перевод фразеологических единиц при наличии точных совпадений. Если точных совпадений нет, пользователю будут предложены смысловые эквиваленты искомого фразеологизма. При этом в отличие от автоматического перевода, предлагаемого словарями-переводчиками, описанными выше, всегда сохраняющего вероятность неправильного перевода, словарь LexSite выдает только те фразеологические единицы, которые существуют в языке и содержатся в его словнике (рис. 11).

Программа, обрабатывающая запросы пользователей, учитывает морфологические особенности английского и русского языков и при

наличии омоформ в качестве основного варианта перевода предлагает слово, начальная форма которого совпадает с запрошенным, а во всплывающем окне показывает сообщение о наличии омоформ с возможностью выбора другого слова из всплывающего окна (рис. 12).

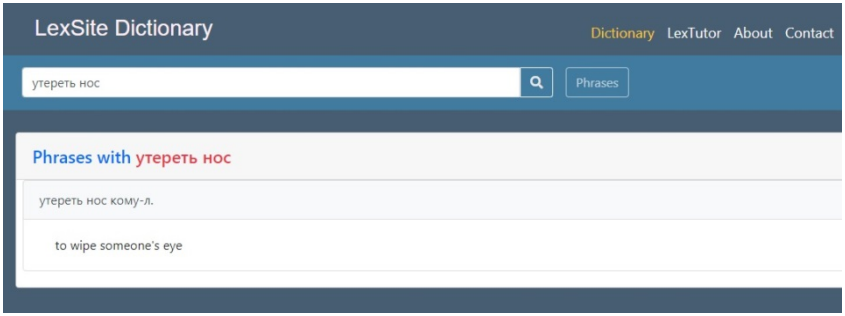

Рис. 11. Фразеологизмы

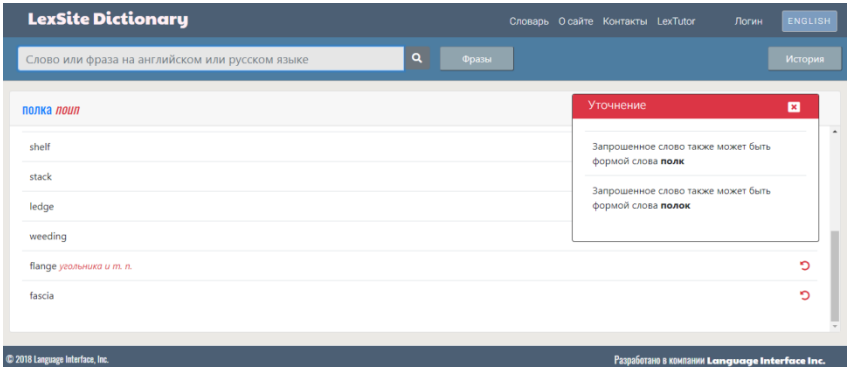

Рис. 12. Демонстрация переводов слова, имеющего омоформы

С целью уточнения лексического значения каждого из полученных пользователем вариантов переводов многозначного слова и выбора оптимального слова для переводимого текста, словарь предлагает обратный перевод каждого однословного варианта перевода без потери результатов поиска: рядом с каждым однословным вариантом перевода находится пиктограмма «красная стрелка», при нажатии на которую появляется всплывающее окно с обратным переводом (рис. 13).

LexSite фиксирует историю словарных запросов пользователя для облегчения повторных запросов, которые могут быть осуществлены путем выбора необходимого слова в «Истории поиска» (рис. 14).

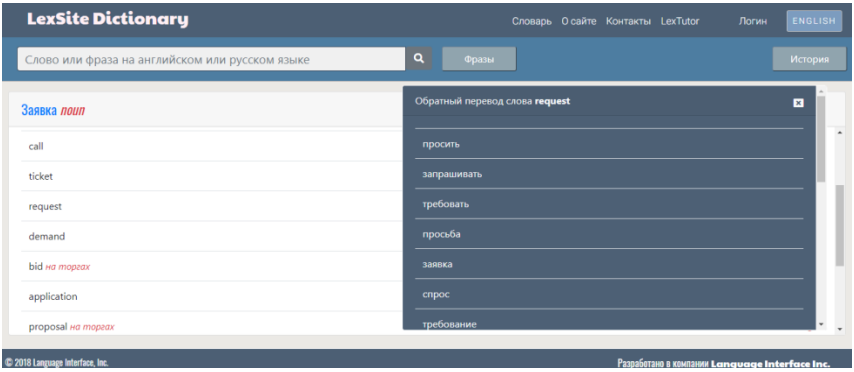

Рис. 13. Демонстрация обратных переводов слова

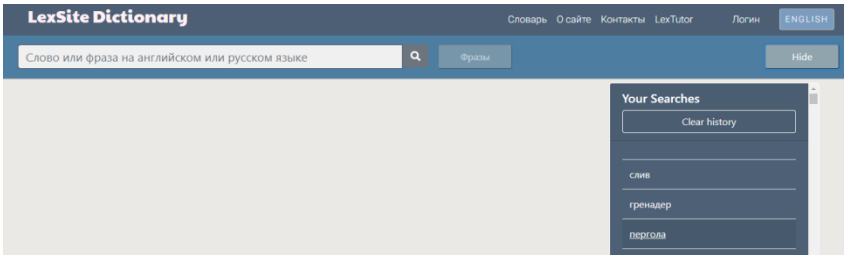

Рис. 14. История поиска

Для сохранения целостности словарной информации, проверенной при формировании словника по авторитетным изданиям словарей, LexSite исключает любую возможность участия пользователей в пополнении словника.

Словарь дает фонетическую справку в виде звукового воспроизведения запрошенных русских и английских слов.

Решение вне-лексикографических проблем двуязычных интернет-словарей в проекте LexSite нашло отражение в следующем. Во-первых, словарь LexSite является общедоступным и не требует регистрации пользователей. Во-вторых, на страницах LexSite отсутствует коммерческая реклама.

Работа над словарным проектом продолжается. Авторы видят свою задачу в обеспечении пользователя достоверной нормативно-справочной информацией при максимально удобном и быстром способе нахождения перевода и такой форме представления переводов, которая способна обеспечить однозначное понимание представленной лексикографической информации.

Литература

1. *Pruvost J.* Colloquium report: Des dictionnaires papier aux dictionnaires electroniques. VIIe Journee des dictionnaires (22 mars 2000) // International Journal of Lexicography. 2000. Vol. 13, is. 3. P. 187–193. DOI: 10.1093/ijl/13.3.187.

2. *Kit M., Berg E.* Online Bilingual Dictionary as a Learning Tool: Today and Tomorrow. URL: https://conference.pixel-online.net/ICT4LL/files/ict4ll/ed0009/FP/3051-ETL1930-FP-ICT4LL9.pdf (дата обращения: 06.06.2018).

3. *De Schryver G.-M.* Lexicographers' Dreams in the Electronic-Dictionary Age // International Journal of Lexicography. 2003. Vol. 16, is. 2. P. 143–199. DOI: 10.1093/ijl/16.2.143.

4. *Zaenen A.* Musings about the Impossible Electronic Dictionary / M.-H. Correard (ed.) // Lexicography and Natural Language Processing. Goteborg, 2002. P. 230–244.

5. *Adamska-Sałaciak A.* Equivalence, Synonymy, and Sameness of Meaning in a Bilingual Dictionary // International Journal of Lexicography. 2013. Vol. 26, is. 3. P. 329–345. DOI: 10.1093/ijl/ect016.

6. *Adamska-Sałaciak A.* Issues in compiling bilingual dictionaries / H. Jackson (ed.) // The Bloomsbury Companion to Lexicography. London, 2013. P. 213–231.

7. *Мультитран*. URL: https://www.Мультитран.ru/ (дата обращения: 06.06.2018).

8. *Google* Translate. URL: https://translate.google.com/ (дата обращения: 06.06.2018).

9. *ABBYY* Lingvo Live. https://www.lingvolive.com/en-us (дата обращения: 06.06.2018).

10. *Яндекс.Переводчик*. URL: https://translate.Яндекс.ru/ (дата обращения: 06.06.2018).

11. URL: https://www.similarweb.com/ (дата обращения: 06.06.2018).

12. *Oxford* Dictionary. URL: https://en.oxforddictionaries.com/definition/fur (дата обращения: 06.06.2018).

13. *Новый* Большой англо-русский словарь : в 3 т. / Ю.Д. Апресян, Э.М. Медникова, А.В. Петрова и др. М. : Рус. яз., 1993. Т. 1. 832 с.; Т. 2. 832 с.; 1994. Т. 3. 832 с.

14. *Словарь* Макмиллан. URL: https://www.macmillandictionary.com/dictionary/british/rub-someone-s-nose-in-something (дата обращения: 06.06.2018).

15. *Апресян Ю.Д.* Лексикографическая концепция Нового большого англо-русского словаря // Новый большой англо-русский словарь : в 3 т. М. : Рус. яз., 1993. Т. 1. С. 6–17.

16. *Словарь* LexSite. URL: www.lexsite-dictionary.com (дата обращения: 06.06.2018).

17. *Мюллер В.К.* Большой современный англо-русский, русско-английский словарь. М.: Цитадель-Трейд, 2009. 1056 с.

18. *Большой* англо-русский политехнический словарь : в 2 т. / М.В. Адамчик. Минск : Харвест, 2004. Т. 1. 784 с.; Т. 2. 784 с.

Suggested Solutions for the Bilingual Internet Lexicography Problems in the LexSite Dictionary Project

Voprosy leksikografii – Russian Journal of Lexicography, 2019, 16, pp. 92–112.
DOI: 10.17223/22274200/16/6

Elena B. Berg, Ural State Law University (Yekaterinburg, Russian Federation). E-mail: elenabkct@gmail.com
Mark Kit, Language Interface (Seattle, United States). E-mail: clodpool@gmail.com

Keywords: Internet Dictionary, bilingual lexicography, LexSite dictionary, machine translation systems, machine translation, lexical dataset.

This article discusses identification of problems faced by the contemporary bilingual Internet lexicography and describes the LexSite dictionary the authors developed in the search for solutions. Data used in this research was obtained from four most popular Internet dictionaries users turn to for translations in the English-Russian language pair: Google Translate, Yandex.Translator, ABBYY Lingvo and Multitran. The first two are combined dictionaries-automatic translators while the other two are strictly dictionaries. The authors ran quantitative and qualitative comparison of translations offered by these dictionaries against the meanings of the same lexical units found in English thesauri. They also subjected examples of usage provided by the dictionaries to contextual analysis. They evaluated the completeness of translations given by those dictionaries and the quality of translations of individual lexemes and idioms. To test the translation veracity, the authors queried words that have homoforms and non-existing words made up for these tests followed by the analysis of methods applied to simulate translations for these non-existing words. They also investigated comments of grammatical and lexical nature included in the lexical entries, as well as the impact of users' involvement in adding entries to the dictionaries in terms of presentation and content. This research identified the following major issues faced by bilingual Internet dictionaries: (1) Combined dictionaries provide too few translations while 'pure' dictionaries produce poorly systematized numerous results. (2) Translations of multi-word strings and idioms made by combined dictionaries are of low quality. (3) Combined dictionaries often make up translations of words that they cannot find. (4) Both types of dictionaries often fail to recognize homoforms and give translations of wrong source words. (5) Many entries in these dictionaries come with non-systematic or wrong grammatical and lexical comments. (6) The departure of these dictionaries from a reference source is aggravated due to attempts to involve users in creation of dictionary entries. The outcome of the study suggests that the surveyed online dictionaries are inappropriate sources of the language base information since they produce data of unpredictable quality. The solutions found by the authors have

been implemented in the Internet dictionary LexSite. When unable to find translations for the user query, the dictionary informs the user on its inability to find the word. It translates phrases that include the requested word and its usage in the linguistic context. LexSite shows translations of idioms if exact matches are found. The dictionary recognizes morphological specifics of the languages and, if the requested word has homoforms, provides relevant comments. To disambiguate requested polysemic words, LexSite offers reverse translations while keeping the search results.

References

1. Pruvost, J. (2000) Colloquium report: Des dictionnaires papier aux dictionnaires electroniques. VIIe Journee des dictionnaires (22 mars 2000). *International Journal of Lexicography*. 3 (13). pp 187–193. DOI: 10.1093/ijl/13.3.187

2. Kit, M. & Berg, E. (2016) Online Bilingual Dictionary as a Learning Tool: Today and Tomorrow. *ICT for Language Learning*. Proceedings of the International Conference. Florence. 17–18 November 2016. Padova: [s.n.]. [Online]. Available from: https://conference.pixel-online.net/ICT4LL/files/ict4ll/ed0009/FP/3051-ETL1930-FP-ICT4LL9.pdf. (Accessed: 06.06.2018).

3. De Schryver, G.-M. (2003) Lexicographers' Dreams in the Electronic-Dictionary Age. *International Journal of Lexicography*. 2 (16). pp. 143–199. DOI: 10.1093/ijl/16.2.143

4. Zaenen, A. (2002) Musings about the Impossible Electronic Dictionary. In: Correard, M.-H (ed.) *Lexicography and Natural Language Processing*. Göteborg. pp. 230–244.

5. Adamska-Sałaciak, A. (2013) Equivalence, Synonymy, and Sameness of Meaning in a Bilingual Dictionary. *International Journal of Lexicography*. 3 (26). pp. 329–345. DOI: 10.1093/ijl/ect016

6. Adamska-Sałaciak, A. (2013) Issues in compiling bilingual dictionaries. In: Jackson, H. (ed.) *The Bloomsbury Companion to Lexicography*. London: Bloomsbury. pp. 213–231.

7. *Multitran*. [Online]. Available from: https://www.multitran.ru/. (Accessed: 06.06.2018).

8. *Google Translate*. [Online]. Available from: https://translate.google.com/. (Accessed: 06.06.2018).

9. *ABBYY Lingvo Live*. [Online]. Available from: https://www.lingvolive.com/en-us. (Accessed: 06.06.2018).

10. *Yandeks.Perevodchik*. [Yandex.Translator]. [Online]. Available from: https://translate.yandex.ru/. (Accessed: 06.06.2018).

11. *SimilarWeb*. [Online]. Available from: https://www.similarweb.com/. (Accessed: 06.06.2018).

12. *Oxford Dictionary*. [Online]. Available from: https://en.oxforddictionaries.com/definition/fur. (Accessed: 06.06.2018).

13. Apresyan, Yu.D. et al. (eds) (1993–1994) *Novyy Bol'shoy anglo-russkiy slovar'* [The New Big English-Russian Dictionary]. Vols 1–3. Moscow: Russkiy yazyk.

14. *Makmillan Dictionary*. [Online]. Available from: https://www.macmillandictionary.com/dictionary/british/rub-someone-s-nose-in-something. (Accessed: 06.06.2018).

15. Apresyan, Yu.D. et al. (ed.) (1993) *Novyy bol'shoy anglo-russkiy slovar'* [The New Big English-Russian Dictionary]. Vol. 1. Moscow: Russkiy yazyk. pp. 6–17.

16. *LexSite Dictionary*. [Online]. Available from: www.lexsite-dictionary.com. (Accessed: 06.06.2018).

17. Myuller, V.K. (ed.) (2009) *Bol'shoy sovremennyy anglo-russkiy, russko-angliyskiy slovar'* [Large Modern English-Russian, Russian-English Dictionary]. Moscow: Tsitadel'-Treyd.

18. Adamchik, M.V. (ed.) (2004) *Bol'shoy anglo-russkiy politekhnicheskiy slovar'* [Large English-Russian Polytechnical Dictionary]. Vols 1–2. Minsk: Harvest.