

**ФИЛОЛОГИЯ. ТЕОРЕТИЧЕСКАЯ, ПРИКЛАДНАЯ И
СРАВНИТЕЛЬНО-СОПОСТАВИТЕЛЬНАЯ ЛИНГВИСТИКА
PHILOLOGY. THEORETICAL, APPLIED AND COMPARATIVE
LINGUISTICS**

Научная статья
УДК 81'33

Филологические науки

<https://doi.org/10.26907/2658-3321.2023.6.2.282-293>

**ИНСТРУМЕНТЫ ПРИКЛАДНОЙ ЛИНГВИСТИКИ В КОНТРОЛЕ
КАЧЕСТВА ПЕРЕВОДА**

А.М. Шейко

Волгоградский государственный социально-педагогический университет, Волгоград, Россия
Anastasia_sheyko@mail.ru, <https://orcid.org/0000-0002-8669-2292>

Аннотация. В статье предлагается краткий обзор инструментов прикладной лингвистики, используемых в профессиональной переводческой деятельности, а именно облачных сервисов, программ для работы с терминологическими базами и модулей контроля качества письменного перевода (Quality Assurance tools), интегрированных в системы памяти переводов (Translation Memory tools) и САТ-инструменты (Computer Assisted/Aided Translation tool). Перечисленные функциональные характеристики средств контроля качества перевода свидетельствуют о том, что процессу автоматической обработки в большей степени поддаются те составляющие редактуры и постредактуры текста, которые связаны с форматированием и не требуют анализа контекста и прямого участия исполнителя на этапе верификации правильности предлагаемых программой исправлений. Рассматриваемые инструменты автоматизированной обработки текстов представляют собой лингвистические ресурсы и являются ядром прикладной и компьютерной лингвистики, формируя одно из направлений в сфере лингвистических технологий для анализа и обработки естественного языка наряду с машинным переводом, системами распознавания речи, извлечения данных и автоматического реферирования. Несмотря на долгую историю существования и становления прикладной лингвистики как подраздела цифровой филологии (Digital Humanities), ее проблемное поле остается дискуссионным и само по себе становится самостоятельной темой исследований, посвященных понятийному статусу, подходам и методологии данной научной области.

Ключевые слова: качество перевода; автоматизированная проверка качества перевода; САТ-инструменты; программы памяти переводов; лингвистические ресурсы; лингвистические технологии; прикладная лингвистика

Для цитирования: Шейко А.М. Инструменты прикладной лингвистики в контроле качества перевода. *Казанский лингвистический журнал*. 2023;6(2): 282–293. <https://doi.org/10.26907/2658-3321.2023.6.2.282-293>

Original article

Philology studies

<https://doi.org/10.26907/2658-3321.2023.6.2.282-293>

LANGUAGE TECHNOLOGY TOOLS IN TRANSLATION QUALITY ASSURANCE

A.M. Sheiko

Volgograd State Socio-Pedagogical University, Volgograd, Russia
Anastasia_sheyko@mail.ru, <https://orcid.org/0000-0002-8669-2292>

Abstract. The article summarizes a brief description of linguistic tools applied by localization agencies and professional translators: cloud services, evaluation metrics, terminology management, standalone Quality Assurance tools or integrated into Translation Memory databases and Computer Assisted/Aided Translation tools. The functions and features of QA-tools reviewed in the paper imply accurate analysis of the particular context and active involvement of the translator while verifying the quality of a translation. Automated quality assessment tools prove useful eliminating formal (formatting) mistakes considerably reducing the time required for the monotonous repetitive process of proofreading and correction. Automated QA-methods are the main research topic of computational linguistics and language technologies (or natural language processing) along with language modelling, machine translation, natural language understanding and generation, automatic text summarization and data retrieval. The subject, methods, goals and approaches of Computational Linguistics as an interdisciplinary field of Digital Humanities are a matter of controversy and considerable interest, thus encouraging further discussion on the topic.

Keywords: quality assurance; automated quality assurance; computer assisted/ aided translation tools; translation memory tools; language resources; language technology

For citation: Sheiko A.M. Language Technology Tools in Translation Quality Assurance. *Kazan Linguistic Journal*. 2023;6(2): 282–293. (In Russ.). <https://doi.org/10.26907/2658-3321.2023.6.2.282-293>

Масштабное развитие информационных технологий и беспрецедентные темпы цифровизации всех сфер жизнедеятельности человека дали мощный стимул к расширению инструментария прикладных наук, и лингвистика не стала исключением.

Автоматическая обработка и анализ текстов, машинный перевод, автоматизация контроля качества перевода, лингвистическая экспертиза, распознавание и извлечение информации из текста, автоматический анализ морфологии – вот только некоторые аспекты прикладной лингвистики, междисциплинарное осмысление которых в XXI веке получило «второе дыхание».

Разработка инструментов компьютерного анализа как одно из направлений прикладной лингвистики тесно связана с лингвистическими технологиями, применяемыми для поиска и обработки информации, также эти инструменты называют специальными лингвистическими информационными ресурсами – они представляют собой лингвистические данные и программы, которые непосредственно используются в языковых технологиях (для сравнения, к тематическим лингвистическим информационным ресурсам относятся «любые ресурсы, создаваемые или используемые лингвистами в профессиональной деятельности»).

сти, они выделяются из универсальных структур по тематическому принципу») [1, с. 116].

В этом контексте средства автоматизированной проверки качества перевода, о которых пойдет речь в данной статье, являются «лингвистическими ресурсами», а именно «компьютерными средствами поддержки работы лингвиста» [2, с. 2].

Лингвистические ресурсы являются ярким примером интеграции теоретических знаний исходя из практических потребностей. Так, например, алгоритм работы программ проверки орфографии (spellcheckers) ориентирован на продуктивную лексику, то есть для эффективной работы приложения в его память вводятся лишь наиболее употребительные слова: количество загруженных слов напрямую определяет скорость работы. Таким образом, «теоретический подход о внесении всего объема используемой лексики оказывается в данном случае проигрышным, именно лексика активного словарного запаса помещается в ячейки памяти программы для быстрого анализа, представляя собой наглядный пример оптимизации знаний о языке» [3].

Инструменты проверки качества перевода (Quality Assessment tools, Quality Assurance tools, Quality Assessment automation tools, Translation Quality Assessment tools) широко используются переводчиками, редакторами, корректорами, специалистами по локализации и переводческими агентствами в рамках автоматизированного перевода текстов с использованием компьютерных технологий (иначе именуемых САТ-инструментами или «кошками»). САТ-инструмент (Computer Assisted/Aided Translation tool) представляет собой программное обеспечение для оптимизации переводческого процесса и включает широкий спектр ресурсов, позволяющих создавать базы памяти переводов (Translation Memory (TM)) на основе ранее переведенных сегментов текста (словосочетаний и предложений), пополнять глоссарии, а также осуществлять проверку качества.

Первые ТМ-технологии (translation memory tools, программы памяти переводов) появились на рынке примерно в 1980-х годах: пионерами в данной об-

ласти можно назвать Trados (1984 год) и Star Transit (1986 год). В 1992 году компания IBM Deutschland запустила программу Translation Manager/2 (TM/2), параллельно в том же 1992 году компания GlobalWare анонсировала программу XL8. В 1993 году сотрудничество компаний Telesoft (Испания) и Transsoft (США) положило начало реализации Déjà Vu. При этом модули контроля качества перевода, интегрированные в данные ТМ-инструменты, начали появляться значительно позже. Так, в Star Transit функционал по проверке орфографии, форматирования и единства терминологии появился только в 1998 году, то есть спустя десятилетие, в SDLX (Trados) верификация терминологии и проверка качества были интегрированы в 2003 и 2005 годах. Такая значительная задержка в разработке модулей контроля качества переводов привела к тому, что в данный момент эффективность QA-программ значительно уступает производительности ТМ-инструментов и по многим параметрам не оправдывает ожидания пользователей (наличие ложноположительных ошибок, нераспознанные ошибки, зависание, нестабильность работы, ограниченность сегмента анализа одним предложением, игнорирование контекста и т.п.) [4], [5], [6], [7].

Работа инструментов контроля качества основывается на определенных алгоритмах, позволяющих привести переведенный текст в соответствие с установленными критериями и стандартами всего за несколько минут. Основная задача QA-инструмента – оптимизировать скорость обработки документа, обеспечить гарантированное единство стиля и используемой терминологии. Контроль качества в данном контексте представляет собой систему подсказок, которая поможет избежать языковых нарушений при переводе и оповестит о неправильном написании слова или некорректном использовании соответствия. При подобной проверке качества автоматическое исправление перевода является опциональным – по заданным пользователем параметрам программа выводит отчет с перечнем нарушений, которые необходимо обработать в «ручном» режиме.

Программы для проверки грамматики и правописания обеспечивают соответствие переведенного текста лингвистическим требованиям, то есть выявляют возможные нарушения правописания, пунктуации и грамматических норм, но верификация правильности и исправление таких ошибок требует прямого участия исполнителя перевода, внимательности и безукоризненного владения переводящим языком, а потому не может быть полностью выполнено программой в фоновом режиме. Однако, при установлении несоответствий в форматировании переведенного текста роль САТ-инструментов для помощи переводчику нельзя недооценивать, так как этот однообразный процесс лучше всего поддается автоматизированной коррекции и позволяет многократно повысить скорость обработки текста. К ошибкам форматирования относятся, например, двойной пробел, пропущенная точка в конце предложения и другие нарушения, устранение которых не требует внимательной вычитки текста. Таким образом, субъективное принятие решений и значение «человеческого фактора» в работе программ автоматизированного контроля качества практически сведены к нулю [8].

Итак, современный инструментарий автоматизированного контроля качества предоставляет возможность делать следующие проверки (расположим их по степени важности):

- нарушение порядка, структуры или вложения тэгов / плейсхолдеров (основополагающих элементов организации текста);
- соблюдение терминологии (соответствие глоссарию);
- унификация форматирования;
- сохранение последовательности перевода одинаковых сегментов текста;
- выявление ошибок в цифрах;
- конвертация формата цифр/единиц измерения в языке перевода;
- повторы;

- опечатки;
- расстановка двойных, конечных и парных знаков препинания (в том числе окончание оригинала и перевода на разные знаки пунктуации);
- двойные и лишние пробелы;
- использование неправильного регистра, использование букв разного регистра в оригинале и переводе, постановка двух прописных букв подряд, использование сочетания латиницы и кириллицы;
- полное или частичное совпадение оригинала и перевода (сокращение или увеличение длины сегмента перевода по сравнению с оригиналом);
- распознавание e-mail адресов и ссылок.

На больших объемах текста такую проверку практически невозможно осуществить в ограниченное время без применения автоматизированных инструментов.

Контроль качества представляет собой финальную стадию переводческого процесса и позволяет совместить корректуру и постредатуру документа перед сдачей заказчику. По итогам проверки документа программа выводит отчет, в котором все ошибки сгруппированы по категориям и могут быть исправлены непосредственно в переводе без перехода в стороннее приложение. Процесс формализации и категоризации ошибок по степени важности опционален и настраивается пользователем, в зависимости от текста или клиента. Но не стоит забывать, что при высокой скорости обработки и автоматизации, за решения о внесении корректуры согласно данным QA-инструмента несет ответственность исполнитель, учитывая нормы и правила переводящего языка, контекст, собственные знания и требования заказчика.

Автоматизированный перевод включает широчайший спектр инструментов, которые условно подразделяются на:

- программы для проверки орфографии, пунктуации и грамматики;

— программы для проверки форматирования (отдельные, облачные или встроенные в текстовые редакторы и САТ-инструменты (QA-плагины и QA-модули)).

В настоящее время на рынке переводческих решений представлено большое количество программ контроля качества перевода, различающихся по алгоритму работы, функционалу, доступности в РФ и стоимости.

Отдельные программы контроля качества и лингвистических услуг (standalone autoQA, review and evaluation) [9]:

- | | |
|-------------------------------|------------------------------------|
| 1. QA Distiller Yamagata; | 14. Lingotek; |
| 2. ErrorSpy D.O.G.; | 15. Linguistic ToolBox Lionbridge; |
| 3. Verifika Palex; | 16. LinQ; |
| 4. Xbench Apsic; | 17. Modelfront; |
| 5. AceProof | 18. Ocelot; |
| 6. Bureau Works; | 19. Okapi CheckMate; |
| 7. ContentQuo; | 20. Phoenix ITI; |
| 8. FineQA Tilti Systems; | 21. RIGORA Logrus Global; |
| 9. globalReview Kaleidoscope; | 22. TAUS DQF plugins; |
| 10. Glossa Group; | 23. Televic; |
| 11. InContext QA Lingoport; | 24. TQAuditor; |
| 12. LexiQA; | 25. Vistatec; |
| 13. Lilt; | 26. WebReview Gemino. |

Наиболее известные в РФ QA-программы мы вынесли в начало списка: QA Distiller, ErrorSpy, Xbench и Verifika (русская разработка). Описание функционала, выполняемых типов проверок и других параметров (включая скорость обработки и количество ложноположительных срабатываний) подробно изложено во многих работах [4], [10], [11], [12] [13], а на сайте «Город переводчиков» преимуществам и недостаткам работы программ Xbench и Verifika посвящен отдельный пост с «говорящим» заголовком «Битва титанов» [14].

Не менее внушителен арсенал программ для проверки, автоматического извлечения терминов и управления терминологическими базами (terminology management) [9]:

1. Accolada Uniterm;
2. Acrilinx;
3. Across crossTerm;
4. AnyLexic;
5. Bureau Works;
6. Congree;
7. Copernicus Semantix;
8. Coreon;
9. Ego translating;
10. evoTerm PTS;
11. Flashterm;
12. Fluency Term Server;
13. Fusion Terminology;
14. Glossa Group;
15. Interpretbank;
16. i-Term DANTERM;
17. [i]-Match;
18. LCTerm IAILC;
19. Lexeri Toptranslation;
20. Lexicool Lingo4;
21. Lilt;
22. LogiTerm Terminotix;
23. LookUp D.O.G.;
24. Marfa CAT;
25. MultiLing;
26. Prospector Logrus Global;
27. QTerm memoQ;
28. quickTerm Kaleidoscope;
29. SDL Multiterm;
30. SQUIDDS;
31. SynchroTerm;
32. Termbases.eu;
33. Termbox Lingosail;
34. TERMCAT GesTerm;
35. TermCloud Gemino;
36. Terminator Open Source
Paltform.Terminologue;
37. TermLode Logrus Global;
38. TermWeb Interverbum tech;
39. Termwikipro;
40. Tilde,
41. TippyTerm SysKon;
42. Translate5;
43. WebTerm, TermStar (STAR);
44. Wordbee;
45. WordFinder.

Программные CAT-инструменты хоть и не теряют своей актуальности, но постепенно сдают позиции, уступая место современным облачным технологиям

перевода, позволяющим работать над проектом в режиме реального времени.

Здесь разработчики также предлагают несколько решений:

1. SmartCAT;
2. Lokalise;
3. Phrase Strings;
4. Crowdin;
5. Transifex;
6. POEditor;
7. Smartling;
8. Weglot;
9. Lionbridge;
10. XTM Cloud;
11. Unbabel.

Помимо программ для проверки качества в переводческой среде также используются так называемые метрики (экспертные системы оценки перевода разных типов текстов):

1. Multidimensional Quality Metric (MQM);
2. TAUS Dynamic Quality Framework (DQF);
3. LISA Quality Model;
4. SAE J2450.

В основе работы этих программ лежит индекс Translation Quality Index (TQI), или индекс качества перевода. Он представляет собой количественный показатель качества выполненного перевода в баллах на основе классификации переводческих ошибок, ранжированных по категориям и степени серьезности.

Однако при таком огромном выборе переводческих инструментов, непрерывном развитии ТМ-технологий и инструментов и росте их востребованности особого внимания требует тот факт, что оптимизация переводческого процесса позволяет выполнять большее количество переводов, что в свою очередь приводит к сокращению времени на обработку и редактуру и увеличению количества допущенных ошибок, которые выявляют QA-инструменты, т.е. переводчик оказывается в «замкнутом круге»: технологии, созданные для повышения про-

дуктивности и эффективности в непрофессиональных руках могут иметь прямо противоположный эффект.

Список литературы

1. Антопольский А.Б. Лингвистические ресурсы и технологии в России: состояние и перспективы. (Обзор). *Социальные новации и социальные науки*. Москва: ИНИОН РАН. 2021;(2):114–131. URL: https://sns-journal.ru/site/assets/files/1154/sns_2_antopol_skii.pdf. DOI:10.31249/snsn/2021.02.08.
2. Гольдин В.Е. *Введение в электронные лингвистические ресурсы*. URL:http://elibrary.sgu.ru/uch_lit/130.pdf [дата обращения:30.01.2023].
3. Прикладная лингвистика – Анатолий Баранов. URL: <https://www.youtube.com/watch?v=3BzrupwFSu8> [дата обращения: 30.01.2023].
4. Makoushina Ju. Translation Quality Assurance Tools: current state and future approaches. *Translating and the Computer*. 2007;(29). URL: <https://aclanthology.org/2007.tc-1.8.pdf> [дата обращения:30.01.2023]. (In Eng.)
5. Garcia I. Long term memories: Trados and TM turn 20. *Journal of Specialised Translation*. 2005;(4). URL: https://www.jostrans.org/issue04/art_garcia.pdf [дата обращения:30.01.2023]. (In Eng.)
6. From IBM: Translation Manager/2.*Language Industry Monitor*. 1992;(11). URL: <https://aclanthology.org/www.mt-archive.info/90/LIM-1992-11-5.pdf> [дата обращения:30.01.2023]. (In Eng.)
7. XL8 Becomes a Family. *Language Industry Monitor*. 1994;(20). URL: <https://aclanthology.org/www.mt-archive.info/90/LIM-1994-20-4.pdf> [дата обращения:30.01.2023]. (In Eng.)
8. ZehnalováJ. Tradition and Trends in Translation Quality Assessment. *Tradition and Trends in Trans-Language Communication*. 2013;(2). URL: https://www.researchgate.net/publication/294260655_Tradition_and_Trends_in_Translation_Quality_assessment [дата обращения:30.01.2023]. (In Eng.)
9. Translation Quality Assurance Tools: Old and New. URL: <https://www.nimdzi.com/translation-quality-assurance-tools/> [дата обращения: 30.01.2023]. (In Eng.)
10. Полякова Н. В. Сопоставительный анализ современных QA-инструментов для письменного перевода. *Вестник Томского государственного педагогического университета*. 2022;4(222):51–61.
11. Petrova V. Translation Quality Assessment Tools and Processes in Relation to CAT Tools. , *In Proceedings of the Human-Informed Translation and Interpreting Technology Workshop (HiT-IT 2019)*. URL: <https://aclanthology.org/W19-8711.pdf>. DOI:10.26615/issn.2683-0078.2019_011. (In Eng.)
12. Gural S., Chemezov Y. Analysis of Efficiency of Translation Quality Assurance Tools. *Procedia – Social and Behavioral Sciences*. 2014;(154):360–363. URL: https://www.researchgate.net/publication/275244541_Analysis_of_Efficiency_of_Translation_Quality_Assurance_Tools. DOI:10.1016/j.sbspro.2014.10.163 (In Eng.)
13. Daems J., Macken L., Vandepitte S. Quality as the sum of its parts: A two-step approach for the identification of translation problems and translation quality assessment for HT and MT+PE. *Proceedings of the 2nd Workshop on Post-editing Technology and Practice*. 2013. URL: <https://aclanthology.org/2013.mtsummit-wptp.8.pdf>[датаобращения:30.01.2023]. (In Eng.)
14. Битва титанов: Verifika vs. xBench. URL: <http://www.trworkshop.net/forum/viewtopic.php?f=16&t=60871> [дата обращения: 30.01.2023].

References

1. Antopolsky A.B. Language resources and technology in Russia: state and prospects (Review). *Social Novelties and Social Sciences*. Moscow: INION RAN. 2021;(2):114–131. Available from: https://sns-journal.ru/site/assets/files/1154/sns_2_antopol_ski.pdf. DOI:10.31249/snsn/2021.02.08. (In Russ.)
2. Gol'din V.E. *Introduction to digital language resources*. Available from: http://elibrary.sgu.ru/uch_lit/130.pdf [accessed:30.01.2023].
3. Applied Linguistics–Anatoly Baranov. Available from: <https://www.youtube.com/watch?v=3BzrupwFSu8> [accessed: 30.01.2023].
4. Makoushina Ju. Translation Quality Assurance Tools: current state and future approaches. *Translating and the Computer*. 2007;(29). Available from: <https://aclanthology.org/2007.tc-1.8.pdf> [accessed:30.01.2023].
5. Garcia I. Long term memories: Trados and TM turn 20. *Journal of Specialised Translation*. 2005; (4). Available from: https://www.jostrans.org/issue04/art_garcia.pdf [accessed: 30.01.2023].
6. From IBM: Translation Manager/2. *Language Industry Monitor*. 1992;(11). Available from: <https://aclanthology.org/www.mt-archive.info/90/LIM-1992-11-5.pdf> [accessed:30.01.2023].
7. XL8 Becomes a Family. *Language Industry Monitor*. 1994;(20). Available from: <https://aclanthology.org/www.mt-archive.info/90/LIM-1994-20-4.pdf> [accessed:30.01.2023].
8. Zehnalová J. Tradition and Trends in Translation Quality Assessment. *Tradition and Trends in Trans-Language Communication*. 2013;(2). Available from: https://www.researchgate.net/publication/294260655_Tradition_and_Trends_in_Translation_Quality_Assessment [accessed: 30.01.2023].
9. Translation Quality Assurance Tools: Old and New. Available from: <https://www.nimdzi.com/translation-quality-assurance-tools/> [accessed: 30.01.2023]
10. Polyakova N.V. Comparative Analysis of Modern QATools for Written Translation. *Vestnik Tomskogo gosudarstvennogo pedagogicheskogo universiteta*. 2022;4(222):51–61. (In Russ.)
11. Petrova V. Translation Quality Assessment Tools and Processes in Relation to CAT Tools. In *Proceedings of the Human-Informed Translation and Interpreting Technology Workshop (HiT-IT 2019)*. Available from: <https://aclanthology.org/W19-8711.pdf>. DOI:10.26615/issn.2683-0078.2019_011.
12. Gural S., Chemezov Y. Analysis of Efficiency of Translation Quality Assurance Tools. *Procedia – Social and Behavioral Sciences*. 2014;(154):360–363. Available from: https://www.researchgate.net/publication/275244541_Analysis_of_Efficiency_of_Translation_Quality_Assurance_Tools. DOI:10.1016/j.sbspro.2014.10.163.
13. Daems J., Macken L., Vandepitte S. Quality as the sum of its parts: A two-step approach for the identification of translation problems and translation quality assessment for HT and MT+PE. *Proceedings of the 2nd Workshop on Post-editing Technology and Practice*. 2013. Available from: <https://aclanthology.org/2013.mtsummit-wptp.8.pdf> [accessed: 30.01.2023].
14. The Battle of the Titans: Verifika vs. xBench. Available from: <http://www.trworkshop.net/forum/viewtopic.php?f=16&t=60871> [accessed: 30.01.2023].

Автор публикации

Пospelova Софья Николаевна –
кандидат филологических наук, преподаватель
Воронежский государственный университет
Воронеж, Россия
Email: sdankov@yandex.ru
<https://orcid.org/0009-0007-4115-0514>

Раскрытие информации о конфликте

Author of the publication

Pospelova Sofia Nikolaevna –
PhD in Philology, lecturer
Voronezh State University
Voronezh, Russia
Email: sdankov@yandex.ru
<https://orcid.org/0009-0007-4115-0514>

Conflicts of Interest Disclosure

интересов

Автор заявляет об отсутствии конфликта интересов.

Информация о статье

Поступила в редакцию: 1.02.2023

Одобрена после рецензирования: 15.03.2023

Принята к публикации: 5.04.2023

Автор прочитал и одобрил окончательный вариант рукописи.

Информация о рецензировании

«Казанский лингвистический журнал» благодарит анонимного рецензента (рецензентов) за их вклад в рецензирование этой работы.

The author declares that there is no conflict of interest.

Article info

Submitted: 1.02.2023

Approved after peer reviewing: 15.03.2023

Accepted for publication: 5.04.2023

The author has read and approved the final manuscript.

Peer review info

Kazan Linguistic Journal thanks the anonymous reviewer(s) for their contribution to the peer review of this work.