

В. Е. Войнилова

Тверской государственный университет, магистрант

Научный руководитель: к.ф.н. Е. П. Денисова

СОСТАВЛЕНИЕ ОТРАСЛЕВОГО ГЛОССАРИЯ: ОТ ТЕОРИИ К ПРАКТИКЕ

Несмотря на многие современные реалии, внутри мирового сообщества продолжается межкультурное взаимодействие, обмен профессиональной информацией, в процессе которого существенную роль играет использование терминологии. Однако порой профессиональная коммуникация может сопровождаться коммуникативными сбоями, вызванными расхождением в трактовке терминов на зарубежных и отечественных предприятиях.

На минимизацию коммуникативных сбоев в профессиональном дискурсе и обеспечение адекватной интерпретации текста направлено создание и использование в практике тематического глоссария – словаря специализированных терминов, относящихся к той или иной отрасли знаний, включающего в себя их толкование, комментарии по употреблению, примеры использования, а также перевод.

Следовательно, составление отраслевого глоссария, а также выявление наиболее существенных характеристик языковых единиц, позволяющих включить их в такой глоссарий, является важным шагом при подготовке переводчика к работе. Перед ним также может стать вопрос, имеет ли смысл в целях облегчения своей работы обратиться к современным техническим возможностям, например, программам-экстракторам и системам автоматизированного перевода.

Вместе с тем стоит отметить необходимость разграничивать понятие «термин» с лингвистической точки зрения и практической точки зрения, используемой при работе над составлением глоссария.

Термины – это специальные слова, ограниченные своим особым назначением; слова, стремящиеся быть однозначными как точное выражение понятий и название вещей. Такие единицы необходимы в науке, технике, политике и дипломатии. Учёные, определяя понятие «термин», соотносят его с понятием специальной области знаний: термин – это «языковой знак (слово / словосочетание), соотнесённый со специальным понятием, явлением или предметом» [Реформатский 1986: 165].

Б.Н. Головин выдвигает на первый план функции термина, т. е. его «профессиональное значение, выражающее и формирующее профессиональное понятие и применяемое в процессе (и для) познания и освоения некоторого

круга объектов и отношений между ними – под углом зрения определённой профессии» [Головин 1981: 5].

В программах-экстракторах и системах автоматизации и оптимизации перевода термин представляет собой скорее слово, частотное для того или иного текста и подходящее под определенные формальные критерии (максимальная длина в словах, минимальная частотность и минимальная длина слова (критерии, используемые в программе Phrase)). Такие программы представляют собой системы для автоматического выделения терминов и в зависимости от используемой технологии делятся на три группы:

1. лингвистические системы, работа которых основана на лингвистических данных. Они, в свою очередь, делятся на два вида: системы, использующие информацию, касающуюся конкретного термина, и системы, использующие общеязыковую информацию, выделяющие, например, базовые языковые структуры – именные группы;

2. статистические системы, использующие сугубо статистический подход и считающиеся традиционными для извлечения терминов в силу того, что они не зависят от выбора языка. Такие системы работают за счет выделения двух или более лексических единиц, частота совместной встречаемости которых больше некоторого заданного уровня. Основным недостатком статистических систем является неспособность извлекать низкочастотные термины;

3. гибридные системы, сочетающие в себе статистический метод с использованием некоторой лингвистической информации [Bourigault, Condamines 1999: 3].

Алгоритм работы специалиста над составлением глоссария может быть следующим: переводчик загружает необходимый документ в выбранную программу-экстрактор для автоматического поиска терминов, затем оценивает выдачу. Нами в качестве материала для исследования был использован документ «Leitfaden Umweltmanagement nach ISO 14001» («Руководство по системе экологического менеджмента и аудита по стандарту 14001»), разработанный немецкой фирмой GUTcert, специализирующейся на организации и проведении экологического аудита [Leitfaden URL]. Данный документ общим объемом в 54 страницы посвящен правилам и нормативам проведения экологического аудита.

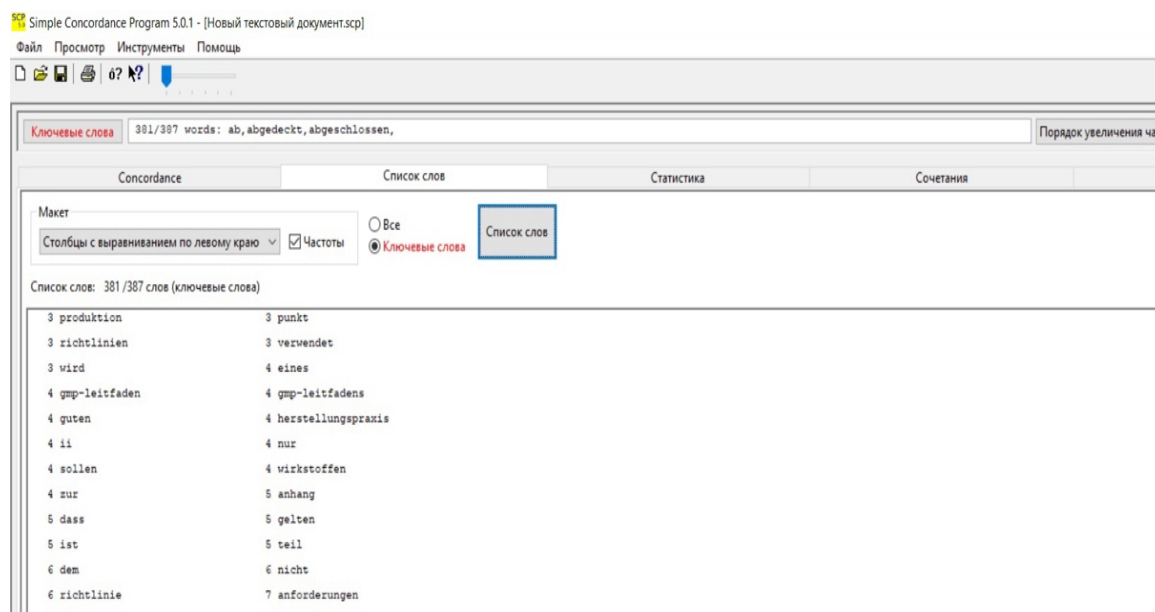
В рамках исследования мы использовали следующие программы-экстракторы: Term Extraction (разработчик – компания FiveFilters.org), Simple concordance program 5.0.1 и TerMine (разработчик – компания NaCTem) и систему «Phrase».

Приведем далее некоторые результаты, полученные нами благодаря использованию указанных экстракторов:

1. Term Extraction – программное обеспечение, позволяющее извлекать термины с помощью веб-сервиса. На вход подается текст или ссылка на сайт, из которого извлекаются термины. На выходе пользователь получает списки терминов в виде таблицы: в первом столбце приводятся наиболее употребительные термины, во втором столбце – количество употреблений данного термина в тексте, в третьем столбце – количество слов в терминологическом словосочетании. Всего программа выделила около 100 предположительных терминов:

getroffen werden müssen	1	3
Nachweisdokumente erzeugt werden	1	3
Größe einer Organisation	1	3
Beschreibungen mit Verlinkungen	1	3
das UMS führen	1	3
der Öffentlichkeit	1	3
Kenntnis aller Mitarbeitenden	1	3
Analyse von Risiken	1	3
ihrer Beeinflussbarkeit durch	1	3
jedes Managementsystem torpedieren	1	3

2. Simple concordance program – программа, позволяющая создавать списки слов и искать в текстовых файлах слова и фразы. В программе можно настроить поиск по частотности, результатом ее использования стал список наиболее частотных слов в анализируемом документе:



3. TerMine – онлайн-сервис, позволяющий извлекать термины из текста или с Интернет-страницы. Результат ее использования – таблица, состоящая из трех столбцов: порядковый номер термина, сам термин, а также его терминологичность – способность выступать в качестве термина для данной предметной области, определяемая частотностью. Всего программа выделила около 200, приведем первые 10 из них:

Rank	Term	Score
1	leitfaden umweltmanagement	48.543285
2	die organisation	19.570162
4	des ums	18.975882
5	die organisation muss	18.106995
6	der umweltleistung	14.867231
7	die organisation muss dokumentierte	14.464286
8	organisation muss	13.453125
9	verbesserung der	12.847059
10	verbesserung der umweltleistung	12.46573

4. Система «Phrase» – облачная система управления переводами, а также инструмент автоматизированного перевода, имеющий опцию выделения терминов в переводе. Настройки процесса поиска терминов помогают исключить из поиска служебные слова, например, артикли и предлоги, и тем самым сделать выборку терминов более эффективной. Далее представлен фрагмент списка предположительных терминов, созданный этой программой:

de-de
Abbildung
aber
aber auch
Abfälle
abgeleitet
Ablauf
Abläufe
Abteilungen
Abwasser
achten
Akademie
Aktivitäten
Aktualität
aktuelle
aktuellen
Akzeptanz
alle
allen
aller
also
Analyse

Возможность включения многих приведенных выше терминов в тематический глоссарий по экологическому аудиту вызывает сомнения, поскольку результаты, полученные с помощью программ-экстракторов, являются неточными и неполными.

В этом случае переводчику придется выделять термины в документе самостоятельно, что может занять гораздо больше времени, но даст необходимый вдумчивый подход специалиста, учитывающего не только частотность использования того или иного слова, но и наличие контекста, сложности упомянутых категорий, важность их адекватной передачи в рамках осуществления письменного или устного перевода.

Тем не менее, среди отобранных программами-экстракторами единиц нами были выделены слова, которые были вынесены в списки терминов всеми четырьмя программами, их число составило около 130 единиц.

Выбранные как автоматически, там и вручную единицы нуждаются в проверке, насколько они соответствуют параметрам, предъявляемым к терминам в лингвистическом смысле посредством поиска и анализа частотности их использования в научных статьях или отраслевых журналах (при условии, что он хочет создать именно терминологический глоссарий, а не просто список частотных в данном тексте единиц).

Все термины можно поделить на 3 группы:

- отраслевые – слова или сочетания слов, обозначающие специальные понятия, употребляемые в определенной отрасли науки, техники, искусства;
- корпоративные – слова или сочетания слов, используемые внутри отдельно взятой компании, могут быть как неологизмами, так и общепотребимыми словами с переосмысленным значением;

- авторские – неологизмы, созданные автором, или общеупотребимые слова, которые в определенном контексте приобретают новые значения, вложенные в них автором.

Проверить является ли единица полноправным термином, можно, проанализировав ее употребление в профессиональном дискурсе, например, в материалах, созданных на языке перевода экспертами для других людей, в энциклопедиях, в проектной документации. Для этого нами с помощью программы «Архивариус», предназначенной для создания, ведения (пополнения) и управления текстовыми базами данных, был проработан корпус из 30 документов на немецком языке (в общей сложности 200 страниц). Этот анализ помог существенно сократить корпус единиц, поскольку мы выносили в свой глоссарий только те слова, которые встречались в разных источниках более трех раз. Подтвержденные термины были переведены на русский язык и еще раз проверены, теперь уже в русскоязычном корпусе статей и журналов, состоявшем из 30 документов (в общей сложности 150 страниц). В итоге корпус единиц, которые можно назвать терминами с лингвистической точки зрения и внести в глоссарий, составил 30 единиц.

На основании проведенного анализа можно сделать следующий вывод: программы-экстракторы не способны качественно и надежно выделять термины из загруженного в них текста, т. к. их работа основана только на сравнении частотности использования слов, встречающихся в тексте, с частотностью слов в языке, поэтому переводчику придется прорабатывать интересующий его документ самостоятельно.

ЛИТЕРАТУРА

Головин Б. Н. Типы терминосистем и основания их различия // Термин и слово. Горький: Горьк. гос. ун-т им. Н. И. Лобачевского, 1981. С. 3–10.

Реформатский А. А. Современные проблемы русской терминологии. М.: Наука, 1986. С. 165–198.

Cabré M. T., Estopà R., Vivaldi J. Recent advances in computational terminology. Amsterdam/Philadelphia: John Benjamins Publishing Company, 2001. 88 p.

Bourigault D., Condamines A. Terminology & Artificial Intelligence. Toulouse: Université de Toulouse, 1999. 246 p.

Leitfaden Umweltmanagement nach ISO 14001. URL: https://www.gut-cert.de/files/content/download/Publikationen/Leitfaden/Leitfaden_UMS_v01.pdf. (дата обращения 11.11.2022).