

УДК 811.111

ИЗУЧЕНИЕ ТАМОЖЕННОЙ ТЕРМИНОСИСТЕМЫ АНГЛИЙСКОГО ЯЗЫКА НА БАЗЕ КОРПУСА ТЕКСТОВ

В.О. Гросс¹, М.И. Попова²

Иркутский государственный университет путей сообщения,
664074, Россия, г. Иркутск, ул. Чернышевского, 15.

В статье рассматриваются основные понятия корпусной лингвистики. С целью составления списка основных терминов таможенного дела был построен небольшой специальный корпус текстов, который позволил определить частотность каждого извлеченного термина и создать базу сочетаемости наиболее частотных терминов.

Библиогр. 2 назв.

Ключевые слова: корпусная лингвистика; конкорданс; терминосистема таможенного дела; корпус текстов; частотность термина.

STUDY OF ENGLISH CUSTOMS TERM SYSTEM ON BASIS OF TEXT CORPUS

V. Gross, M. Popova

Irkutsk State Railway Engineering University,
15 Chernyshevsky St., Irkutsk, Russia, 664074.

The article deals with the main notions of corpus linguistics. In order to compile the list of core Customs terms the authors have built a special text corpus. This corpus has allowed identifying frequency of each extracted term and creating the base of the most frequent terms co-occurrence.

Bibliography: 2 refs.

Keywords: corpus linguistics, concordance, Customs term system, text corpus, term frequency.

Целью нашего исследования является изучение терминосистемы таможенного дела в современном английском языке.

Задачи исследования:

- 1) построить корпус текстов по таможенной специальности;
- 2) сформировать конкорданс по специальности «Таможенное дело» с целью использования терминов данной специальности в практике преподавания профессионального английского языка;
- 3) установить основные типы терминов таможенного дела;
- 4) создать базу сочетаемости основных таможенных терминов.

В настоящее время одним из наиболее развивающихся методов лингвистического анализа является корпусная лингвистика, где главная задача – составление моно- и многоязычных корпусов текстов, с помощью которых осуществляется исследование определенных аспектов языка. Широкое применение этого метода объясняется развитием компьютерных технологий. Главным преимуществом корпусов текстов является возможность автоматической обработки языковых данных, что значительно упрощает процесс их обработки и улучшает его качество. Существует уже довольно большой опыт построения корпусов текстов, в том числе специальных (т.е. текстов, принадлежащих к какой-либо предметной области знаний). Однако проблема заключается в том, что доступ к большинству корпусов ограничен.

Корпусом текстов называется совокупность текстов, собранных в соответствии с определенными принципами, размеченных по определенному стандарту. Принято выделять следующие виды корпусов текстов: по форме хранения (в звуковой форме, письменные, смешанные); по языку

¹ Гросс Владислав Олегович, студент 2 курса факультета менеджмента, логистики и таможенного дела, e-mail: grossvlad@mail.ru

Gross Vladislav, a second-year student of Management, Logistics and Customs Affairs Faculty, e-mail: grossvlad@mail.ru

² Попова Мария Иннокентьевна, кандидат филологических наук, доцент кафедры иностранных языков, e-mail: mariainnok@mail.ru

Popova Mariya, Candidate of Philology, Associate Professor of Foreign Languages Department, e-mail: mariainnok@mail.ru

представления текстов (одноязычные и многоязычные); по жанровой принадлежности (литературные, диалектные, разговорные, публицистические, смешанные); по способам доступа (свободно доступные, коммерческие, закрытые); по назначению (исследовательские, иллюстративные); по наличию дополнительной информации (аннотированные и неразмеченные).

Преимуществами работы с такими корпусами являются:

- 1) представление языковых данных в реальном контексте;
- 2) наличие обширной базы данных, достаточной для любого языкового анализа;
- 3) возможность использования созданного корпуса текстов многократно и в разных целях;
- 4) корпус может оказывать огромную помощь при изучении языка, поскольку с помощью

созданного корпуса текстов можно быстро проверить особенности функционирования того или иного иностранного слова или синтаксической структуры.

Конкордансом называется список всех употреблений заданной языковой единицы в контексте (как в ближайшем, так и в расширенном) со ссылками на источник, т.е. конкордансы позволяют увидеть конкретный текст, в котором был найден тот или иной пример употребления языковой единицы, и проанализировать все особенности ее функционирования [1]. Таким образом, конкордансы помогают также решить проблему многозначности термина. Автоматическая выборка заданных языковых единиц из готового корпуса текстов выполняется при помощи специальных компьютерных программ, которые обычно называются *конкордансерами*. Именно конкордансеры выполняют анализ корпуса текстов.

В результате такого анализа формируются алфавитные списки слов с контекстом, и может предоставляться статистическая информация, которая необходима для определения частотности употребления того или иного термина. Отличие словаря от конкорданса заключается в том, что словарь дает описание лексики, которое ориентировано на нормативность, а корпусный вариант (конкорданс) – на исчерпывающее описание.

Как отмечает К.С. Добрынина, словарь, анализируя различные употребления лексемы в разных значениях, стремится к нахождению инварианта, конкорданс – к вариативности, и ставит своей первейшей задачей отразить все случаи употребления слова. «Поэтому представление слова в конкордансе ставит во главу угла примеры (контексты словоупотреблений), а в словаре – словарную статью. На различном понимании термина полноты базируется еще одно принципиальное противопоставление словаря и конкорданса: полнота словаря определяется стремлением к исчерпывающему описанию значений, полнота конкорданса определяется исчерпывающим характером описания соответствующего корпуса» [2].

С точки зрения структуры, терминосистема таможенного дела включает не только односложные термины, но и термины-словосочетания. Существующие словари далеко не полно отражают информацию о сочетаемости слов, поэтому исследование сочетаемости лексем на базе корпусов текстов в последнее время является одним из наиболее развивающихся подходов к исследованию языка.

Наше исследование включало следующие этапы:

- 1) построение корпуса текстов по специальности «Таможенное дело»;
- 2) извлечение из корпуса терминов этой предметной области;
- 3) фильтрация терминов с учетом частоты встречаемости;
- 4) анализ наиболее частотных словосочетаний с тем или иным термином.

С целью решения задач нашего исследования был построен небольшой корпус текстов, которые были созданы в течение последних 20 лет, по тематике специальности «Таможенное дело». Для анализа данного корпуса с целью создания конкорданса и базы сочетаемости основных таможенных терминов использовалась программа MonoConcPro.

Созданный корпус текстов имеет следующие характеристики: по форме представляет собой письменные тексты; по объему – небольшой корпус; по языковому покрытию относится к корпусу подязыка таможенного дела, хотя на данном этапе работы недостаточно представлен по объему выборки, однако позволяет сделать некоторые выводы с учетом поставленных перед ним задач; по типу текстов – исследуемые тексты написаны на английском языке и являются первичными текстами; по жанру представляет собой представленные в Интернете научные тексты по основным темам таможенного дела; по варьированию языков является одноязычным корпусом; по доступности корпуса – в перспективе этот корпус предполагается сделать доступным в Интернете для использования в учебных целях.

Для составления нашего списка наиболее употребляемых терминов таможенного дела учитывались два критерия: высокая частотность термина и высокая частотность употребления этого термина с тем или иным словом, поскольку, чем больше представлено словосочетаний с определенным термином, тем более разрешима проблема многозначности этого термина. Так, например, одним из наиболее часто встречающихся терминов в английских текстах таможенной

предметной области является слово *warehousing* (складское хозяйство, складирование, хранение). Корпус текстов позволил проследить его наиболее частотное окружение: *retail warehousing*, *customs warehousing*, *discount warehousing*, *warehousing operations*, *warehousing space*, *warehousing costs* и т.д.

Фрагменты конкорданса на основе составленного корпуса текстов

may be carried out under the customs **[[warehousing]]** arrangements. Item 12 of Annex 72...
transit procedure or the customs **[[warehousing]]** procedure or in a free zone or...
the last day of the month in which the **[[customs]]** debt was incurred. Compensatory ...
which may be carried out under the **[[customs]]** warehousing arrangements. Item 12 of ...
impossible to identify at all times the **[[customs]]** status of these goods. Common storage ...
territory of the Community under the **[[outward processing procedure]]** unless those goods remain...
the declaration placing them under the **[[outward processing procedure]]** and on the basis...
where, prior to being placed under the **[[outward processing procedure]]**, the temporary export...

Результаты сочетаемости наиболее частотных терминов были получены следующим образом: по заданному термину (состоящему из одного слова) определялись все случаи, где он выступает в качестве определения, а также примеры, где функцию определения выполняет сам термин.

В терминоведении принято делить термины по структуре на многокомпонентные термины-словосочетания; термины, образованные путем словосложения; однословные термины. В английской таможенной терминосистеме присутствуют все типы терминов: многокомпонентные – *bonded warehouse* (таможенный склад для хранения нерастаможенных товаров, т.е. не оплаченных пошлиной), *customs harmonization* (таможенная унификация), *double-column tariff* (двухколонный тариф, т.е. максимальная и минимальная ставки таможенного тарифа на импортируемый товар); сложные термины – *warehouse* (*ware* «продукт производства» + *house* «дом»), *drawback* (возврат таможенной пошлины); однословные термины, в структуру которых входит одна основа или основа с аффиксом – *tax*, *tariff*, *waiver*, *commodity*. Количественный анализ отобранных терминов (всего 67 единиц) показывает, что английская таможенная терминосистема имеет четкую тенденцию к использованию терминов-словосочетаний.

Созданный корпус текстов позволил решить ряд задач: систематизировать наиболее часто употребляемые терминологические единицы таможенного дела с целью создания учебного терминологического словаря, определить частотность употребления каждого термина и проследить его окружение, создать базу сочетаемости основных таможенных терминов. В дальнейшем предполагается расширить корпус текстов и сделать его доступным в Интернете, определить сочетаемость каждого термина и провести этимологический анализ таможенных терминов.

Библиографический список

1. Вафин Б.Д. Конкорданс как полноправный продукт лексикографической деятельности (на материале памятников старорусской деловой письменности) [Электронный ресурс]. – URL: http://www.kls.ksu.ru/boduen/bodart_3.php?id=3&num=4000000
2. Добрынина К.С. О методике работы над конкордансами [Электронный ресурс]. – URL: <http://www.gramota.net/materials/2/2012/1/15.html>