МУТАХАССИСЛИК УЧУН ИНГЛИЗ ТИЛИ (ESP)

УДК: 372.881.111.1

ATTEMPTS TO CREATE LEARNER CORPUS FOR ESP STUDENTS

Durdona Khikmatullaevna KADIRBEKOVA
Assistant professor (PhD)
Branch of the Russian State University of
Oil and Gas after I. M. Gubkin in Tashkent
Tashkent, Uzbekistan
d_kadirbekova@mail.ru

Abstract

This paper presented a description of processes involved in creating the raw corpus for ESP in the field of oil & gas. English for specific purposes teachers often face dilemma in deciding what lexical units to teach students. In the field of oil and gas, there is no exception on this issue as well. The ESP corpus in the field of oil and gas industry made up of research articles can provide better insights and guide to both oil and gas students and teachers.

Keywords: ESP; corpus; corpus building; vocabulary; oil and gas.

ESP ТАЛАБАЛАРИ УЧУН ЎҚУВЧИ КОРПУСИНИ ЯРАТИШ

Дурдона Хикматуллаевна КАДИРБЕКОВА
Доцент (PhD)
И.М.Губкин номидаги Россия давлат
нефть ва газ университетининг Тошкент
шаҳридаги филиали
Тошкент, Ўзбекистон
d_kadirbekova@mail.ru

Аннотация

ESP (махсус мақсадлар учун инглиз тили) ўқитувчилари талабаларга қандай лексик бирликларни ўргатиш кераклигини аниқлашда қийинчиликларга дуч келадилар. Нефт ва газ соҳасидаги инглиз тили ўқитувчилари фаолиятида ҳам тилни ўқитишда бу каби муаммолар мавжуд. Ушбу мақолада нефт ва газ соҳаси корпусини яратиш жараёни тасвирлаб берилган.

Мазкур корпус соҳага доир илмий-тадқиқий мақолалардан ташкил топган бўлиб, унинг ёрдамида нефт ва газ соҳаси талабалари ва ўқитувчилари ушбу соҳа тилини ўрганишда ва ўқитишда қандай лексик бирликларга аҳамият бериш кераклигини аниқлашлари мумкин.

Калит сўзлар: ESP; корпус; корпус яратиш; лексик бирлик; нефт ва газ.

О СОЗДАНИИ ОБУЧАЮЩЕГО КОРПУСА ДЛЯ ESP СТУДЕНТОВ

Дурдона Хикматуллаевна КАДИРБЕКОВА
Доцент (PhD)
Филиал Российского государственного университета
нефти и газа имени И.М.Губкина в городе Ташкенте
Ташкент, Узбекистан
d_kadirbekova@mail.ru

Аннотация

В статье мы представили описание процессов, связанных с созданием корпуса для ESP (английский язык для специальных целей) в области нефти и газа. Преподаватели английского языка для специальных целей часто сталкиваются с проблемой выбора лексических единиц для обучения учащихся. Область нефти и газа не является исключением. Корпус ESP в области нефтегазовой промышленности, состоящий из научных статей, поможет студентам и преподавателям лучше понимать терминологию этой сферы.

Ключевые слова: английский язык для специальных целей; корпус; создание корпуса; лексические единицы; нефть и газ.

Since the launch of far-reaching higher education reforms, many institutions across Uzbekistan have begun revamping their English for special Purposes (ESP) programs to bring them in line with the needs of local employers. These reforms recognize the correlation between what students learn in class and success in their future professions. Therefore, a main objective of English departments is to ensure that ESP courses impart the key communicative skills that are most useful to graduates in their chosen careers. An important component of an effective ESP course is a textbook or specially prepared lesson material that contains job-related English lessons. However, there often is a mismatch between the content of ESP textbooks or lesson materials and actual workplace language demands.

The importance of English for the oil and gas industry Uzbekistan is country’s unofficial second language, and it plays an important role in commerce and economics. However, oil and gas industry professionals are aware of the status of English as an international language and capitalizing on that trend. The oil and gas industry has a special need for employees who can communicate effectively with the foreigners that visit their work place and use services and technologies they offer, including conferences and other form of meetings. From the office worker to the drilling engineer and environmental engineer in the field, English proficiency is required to find out current news in the sphere or carry out work. In addition, large companies require qualified translations of promotional material and interpretations of business meetings in the field. Preparing students for careers in the oil and gas industry and improving the language skills of current employees requires ESP textbooks or lesson materials

that teachers can use to teach particular language skills, needed vocabulary and forms of communication.

In the development of study materials, ESP teachers should bear in mind specific characteristics of their students. ESP students use English to fulfill their discipline-specific needs. The use of authentic materials is much more widespread in teaching ESP than ELT because authentic materials can show students how ‘real-world’ tasks understood and approached in their subjects. Moreover, studying contemporary research, available literature and on the basis of personal experience, we can identify the key methodological principles and approaches in teaching ESP including exploiting corpus linguistics as a resource for the teaching of specific genres and exploiting authentic materials and tasks.

The reflection of the ICT supported process of instruction requires research continuously and systematically in the whole context as modern information and communication technologies have penetrated the society, including the field of education and brought crucial changes. This situation defines the research topic this paper is dealing with.

Language corpora are extensively used in language technology and linguistic researches. There arose a tremendous interest in building and developing computerized language corpora in recent several decades. The study of e-corpora of various languages offers the students and the researchers an opportunity to work with language data with variety of tools and techniques in terms of computational procedures and programs. There are various types of corpora. In fact, it is a very crucial task of classifying language corpora into different types. However, written corpus, spoken corpus, general corpus, monolingual corpus, bilingual corpus, un-annotated corpus, annotated corpus, parallel and learner corpus are worth to be mentioned.

Consequently, corpus-building skills for various academic disciplines are necessary and crucial for both language teachers and content teachers who should learn how to use corpus-building software and analyze the results. ESP courses based on word lists and concordances become a trend for ESP instructors when teaching ESP courses.

ESP instructors should design curriculum based on acceptable data, such as English textbooks, documents, and information in a specific disciplines to enhance learners’ word power and improve their reading proficiency. Designing content-based courses by building a research articles based and localized corpus is crucial for implementing successful ESP programs. In addition, it is helpful to prepare proper exercises and classrooms activities for listening, speaking, reading, and writing.

Developing ESP course materials, selecting course content, and networking with other colleges are essential elements for the academic English as a foreign language (EFL).

In EFL contexts, English for specific purposes often has great significance for learners, mainly due to its multi-disciplinary nature, and for advanced learners especially important specialized vocabulary (1). For example, vocabulary identified from ESP corpora have served to facilitate learning for disciplines such as nursing, medicine, and tourism (1; 2); however, no corpus yet exists for Oil and gas industry. Creation of an Oil and gas industry corpus will directly assists instructors to identify vocabulary in the field, furthering communication, report writing, and interpretation specialised lexical units.

In the present study, we mainly deal with the building the structure of English Oil and gas industry language un-annotated raw corpus comprising approximately 200 000 words and also try to highlight the issues during the process of building it. This collection of texts would be helpful in the linguistic and non-linguistic studies, cross-linguistic comparisons and, all other areas of language technologies. There are various issues that are associated with the design, development and management of corpus. Such issues vary according to the type and utility of the corpus. In fact, speech corpus development is different from text corpus. Developing a text corpus in ESP is concerned with the issues like the overall size or length of corpus, selection of the type of genres, the number of text and range of writers, data collection, computerizing the data and validation of raw corpus. These are discussed below:

Size or length of corpus is an important factor of consideration. The overall size of ESP corpus in oil and gas is determined as 200 000 words. But before determining the length of the corpus, certain decisions are taken such as – availability of resources, time for data collection and computerizing them. So far as time factor is concerned, the present corpus is expected to be completed within approximately 6 month. The matter of fact is that the length of a corpus is determined not by focusing on the overall length of the corpus, but focusing more on the internal structure of the corpus: the number of genres is to be included in the corpus, the length and number of individual text samples. The expected words would be collected from six main categories: Petroleum geology and geophysics, Reservoir engineering, Pipeline engineering, construction and operation, Mechanical engineering, Chemical and environmental engineering, International oil and gas business

Selection of genres included in the corpus Genres are selected keeping in mind the purpose and utility of a corpus. A large number of written genres are included in the corpus.

Defining number of text and range. After selecting the genres, next task is to determine how many the numbers of texts and the range of writers to be included in the corpus. There are a huge number of texts available in the languages, but we are very selective in determining the number of texts. In the selection of material, we are very much selective. Regarding the text selection we consider the time factors so that we can include texts that are most up to date (last 3 years).

Data collection is a crucial task of building a corpus. There are various ways to collect written texts for ESP corpus such as buying printed texts, use of library (with necessary permission), photocopying and scanning the texts etc. In this context, the issue of copyright is well maintained.

Computerizing data. After data collection, we prepare for entering those data in an electronic format. It is a very laborious process. The composer has the most important task of entering the metadata. He has to give certain information about the text, for example, genre of the text, type of the text (report, article, letter, instruction, etc.), the name of the text, the name of author and editor, name of publisher, date of publication, place of publication, the page numbers of the texts etc.

Validating the raw corpus. The process of validating the whole raw corpus starts just after the completion of entering the computerized data. It is done by the experts, who possesses linguistic knowledge of English. Sometimes, the data compiler validates the data himself. But the cross-validation of the data is best deserved.

In this paper, we have presented a description of processes involved in creating the raw corpus for ESP in the field of oil & gas. Corpus is being regarded as a multi-dimensional in nature. Corpus opens up new avenues in the field of language technology, communication, exchange of information, translation, language education and linguistic activities etc. In the future, it should be our great responsibility to create bigger corpora, consisting of billions of words, for ESP. Besides, steps are to be taken in annotating the raw corpus which would result in building morphological analyzer, spell checking tool, concordancer, machine translation, speech recognition etc. in the language of oil and gas industry.

To study about various naturally occurring phenomenons on natural language text, a well structured text corpus is very much essential. The quality and structure of a corpus can directly influence on performance of various Natural Language Processing (NPL) applications. Language technology development works in different languages have been started at various levels, and research and development works started demanding a structured and well covered ESP Corpus. Here we presented various issues related to building an ESP text corpus. We reviewed our experience with constructing one such corpus including about 200 000 words of research articles in the

field oil and gas in English language. It will provide a significant effort by serving as an important research tool for teaching ESP and NLP researchers.

The creation of oil and gas corpus in ESP is essential as it will reveal the specialized vocabulary that must be instilled in the students. The corpus created will not only help students but also facilitates the lecturers and instructors in the process of teaching and learning. The teaching and learning of English for oil and gas can be enhanced with the existence of the oil and gas corpus. Oil and gas students may be exposed to the vocabulary important in their field in a more comprehensive way. With the creation of this corpus, English for oil and Gas material developers would have an idea and guidelines on the vocabulary needed to be taken into consideration when developing a material or an course book. On the other hand, the production of the oil and gas dictionaries can be developed focusing on the frequently used words in the field oil and gas.

REFERENCES

1. Abedi Z. Mobaraki M. (2014) The Effect of Grammatical Collocation Instruction on Understanding ESP Texts for Undergraduate Computer Engineering Students/ *Iran Journal of Language Teaching and Research*, Vol. 5, No. 3, pp. 631–641.
2. Özdemir N.Ö. (2013) Using Corpus Data to Teach Collocations in Medical English/ Journal of Second Language Teaching and Research. Volume 3. Issue 1, pp. 37–52.
3. Kwary D.A. (2018) A Corpus and a Concordancer of Academic Journal Articles / Data in Brief 16, pp. 94–100.
4. Marinov S. Training ESP Students in Corpus Use – Challenges of Using Corpus-based Exercises with Students of Non-philological Studies/ Teaching English with Technology. 13(4), 49–76.