

DOI: 10.31548/hspedagog2021.04.048

UDC 371.315:811.111'255

## APPLICATION OF CORPUS TECHNOLOGIES IN TEACHING SPECIALIZED TRANSLATION

STRILETS V. V., PhD in Pedagogy,

Associate Professor of the Department of Foreign Philology and Translation

**National University of Life and Environmental Sciences of Ukraine**

Email: valentynastrilets@nubip.edu.ua

ORCID ID: 0000-0003-0806-0289

**Abstract.** *Corpus technologies (corpora of English and Ukrainian texts and tools for their processing) represent modern specialized discourse and facilitate searching for and comparing different units of translation, which makes them a useful tool for both practicing and trainee translators. The purpose of this article is to determine the role and place of corpus technologies in teaching specialized translation on the example of the oil and gas industry. Comparative and parallel text corpora are characterized. The paper reveals methods of applying mono- and bilingual comparative and parallel corpora and corpus managers for acquiring knowledge about genre-stylistic features of texts; developing skills to distinguish a term and determine its collocation profile and semantic preference; analyze translation techniques; translate collocations, complex noun constructions, verbal phrases, and abbreviations. Examples of relevant exercises and tasks that should be performed at the translation training stage are given. Further research should be aimed at integrating corpus-based tasks into the translation practice stage involving the implementation of a translation project.*

**Key words:** *corpus technologies, comparative corpora, parallel corpora, specialized translation, trainee translators.*

**Introduction.** Specialized discourse presented in texts of various genres is constantly evolving, keeping pace with scientific and technological progress and, accordingly, is supplemented with new concepts that find verbal representation through new terms. Dictionaries, even electronic ones, do not always respond to their emergence in time. These terms are widely represented in arrays of texts – corpora in both source and target language. Monolingual text corpora demonstrate the use of lexical units in context, collocation patterns, syntactic structures, represent background and subject area (extra-linguistic) information, and serve models of texts of certain genres (instructions, patents, company websites, articles, technical advertisements, etc.) representing the textual conventions practicing and trainee translators must be familiar with. In addition to the above features,

bilingual parallel corpora show the solution of translation problems through the use of a variety of translation techniques and transformations. Corpus management tools allow users to quickly find and compare different units of translation. Thus, corpus technologies (corpora of English and Ukrainian texts and corpus managers) are valuable tools for both practicing and trainee translators of specialized discourse. The latter should gain the declarative and procedural knowledge and appropriate skills to use corpus technology to solve translation problems.

### **Analysis of recent studies.**

Corpus technology application in teaching specialized translation has been a research subject in a number of latest publications. S. Frerot [2] and F. Zanettin [8] have outlined the basics of corpus-based methods in the translator training process. S. Sharoff, B. Babych and A. Hartley [6] have proposed a tool (a

kind of corpus manager), which, operating in comparable corpora of English and Russian politics-related texts, finds in the target language equivalents for source language phrases lacking exact dictionary matches. C. I. López-Rodríguez [5] has described the exploitation of online monolingual and multilingual corpora for translating medical texts from English into Spanish through the lens of conventionality and creativity. R. Loock [4] has conducted quantitative and qualitative analysis of comparable electronic corpora consisting of British and American press texts translated from English into French by three different machine translation tools and original French press texts to show the quality of machine translation output and human added value needed during post-editing. E. Symseridou [7] has examined the use of the Web as a Mega Corpus that can be read directly with Google and as a means for constructing corpora automatically with the help of the WebBootCat software. The author's research is based on medical texts in the English-Greek and French-Greek language pairs. J. Braga Riera and N. Maroto [1] have studied how English-Spanish parallel corpora can be used in translation methodology through the analysis of students' translation competences. N. Kübler, A. Mestivier and M. Pecman [3] have presented their empirical study which shows how corpus linguistics techniques can be used in teaching specialized translation, in particular, identifying the most common trainee translators' errors; observing a term's specificities; finding an appropriate support verb (make, do, happen, lead); translating collocations and complex noun phrases; becoming familiar with the genre.

The above analysis demonstrates research findings based on different subject areas, pairs of languages, types of corpora and corpus managers. However, the question of exploiting parallel and comparable corpora in

teaching the oil and gas industry related translation has not been reflected in the literature sources.

**The purpose** of the article is to determine the role and place of corpus technologies in teaching specialized translation on the example of the oil and gas industry.

#### **Methods and materials.**

Research methods include the analysis of corpus-related research findings to specify corpus types, their potential and possible methods of exploiting in teaching specialized translation; the synthesis of teaching techniques in a new academic setting focused on oil and gas subject area to elaborate appropriate tasks and exercises necessary for developing translation skills at the translation training phase. The paper refers to the following types of material: Do-It-Yourself (DIY) comparable corpora of original English and Ukrainian texts compiled from Internet resources (Britannica, oil and gas company websites and journals) and DIY parallel corpora in English and Ukrainian oil and gas related source texts in English and their Ukrainian translations performed by university instructors and Master students involved in translation projects; the same types of texts processed and produced by Master students when undertaking practical training and preparing diploma papers; extracts from monolingual (Ludwig.guru) and bilingual (Reverso Context) corpus-based concordances.

**Results.** The typology of text corpora is diverse as it is based on different criteria. In this paper we focus on comparable and parallel types. The former refer to sets of oil and gas related texts in English and Ukrainian which share common features. The latter include original texts in English / Ukrainian and their translation in Ukrainian / English.

Prior to elaborating the phases of teaching specialized translation we have concretized stages and steps making up

practicing translators' workflow they follow when dealing with specialized discourse and specified instances requiring corpus technology application which are as follows: source text analysis, in particular its genre characteristics; information and terminology search; target text production; editing. In the academic setting the full-scale translation project can be carried out after trainee translators acquire necessary declarative and procedural knowledge and develop appropriate skills, which suggests splitting the academic hours allotted for "Translation and Interpreting Practice" course into two phases: translation training and translation practice. In this article we'll show the application of the above mentioned corpus types in the first phase.

From the "Web as a Corpus" perspective we consider the Internet the primary source for using corpus-based approach when raising trainee translators' awareness of a particular genre (instruction, patent, company website, academic article, technical advertisement, etc.) features; developing their skills to distinguish a subject area term and determine its collocation profile and semantic preference as it contains a wealth of subject area (in our case – oil and gas industry) texts and reference sources which are constantly updated and easily accessed. Typing a particular term in the Google search engine results in an immediate display of matches in the Google search result page where a term appears in bold in natural context. Applying this simple technique trainee translators don't need to use any special corpus processing software as the way of search result presentation is similar to that offered by concordances. For example, when dealing with "cable-tool drilling" term, the first match we can see in the Google search page is a definition provided by Oilfield Glossary: "a method of drilling whereby an impact tool or bit, suspended in the well from a steel cable,

is dropped repeatedly on the bottom of the hole to crush the rock"; the second match is academic.ru translation into Russian which should be translated into Ukrainian as *ударно-канатне буріння*. Having skipped irrelevant (or partly relevant) matches referring to cable-tool drilling in making water wells, we go to the Britannica article entitled "Cable tooling" which demonstrate the term's application in the encyclopedia section related to oil well drilling and then we proceed to studying the web-page from "Oilfield team" – a website of an oilfield workers community, – which, firstly, provides the term's synonym – percussion drilling – and, secondly, its definition, an article on its modern-day application (which is an additional source of subject matter knowledge), and immediate surrounding presented by such word combinations as "repeated strikes for cable-tool drilling", "cable tool drilling rig", "applications for cable-tool drilling" etc. The describes technique based on applying the Internet-based monolingual comparable corpora can be implemented through the following task instructions: 1) Search the Internet for a definition of cable-tool drilling"; 2) Surf the Internet to find synonyms of "cable-tool drilling"; 3) Look through the Google result page matches and highlight those which are relevant / irrelevant to your search; 4) Search the Internet for word combinations comprising "cable tool" etc.

Dealing with collocations – commonly used word combinations – is another difficulty trainee translators frequently encounter. This issue mainly refers to translating specialized texts from Ukrainian into English. When creating the target texts students may have doubts whether a particular collocation, which seems to be semantically acceptable and grammatically correct, sounds natural in English texts belonging to a certain genre. Referring to monolingual resource Ludwig.guru containing extracts from the authentic media (Encyclopedia

Britannica, BBC, The Guardian, The New York Times - Science / Tech, Research Policy, Science Magazine, The Economist, Independent, etc.) can help to solve this translation problem. Thus, “perform drilling operations” has proved to be a commonly used phrase as evidenced by at least 5 text fragments from Forbs, Petroleum Science, Huffington Post, Journal of Loss Prevention in the Process Industries, and Planetary and Space Science, in which it is used both in Active and Passive Voice.

When translating complex noun constructions, verbal phrases, and abbreviations trainee translators can make use of the following corpus-based resources: 1) parallel text extracts provided by Reverso Context resource and 2) the linguistic material provided by DIY bilingual parallel corpora, which we suggest compiling from the English / Ukrainian source texts and corresponding Ukrainian / English target texts created by university instructors and Master students involved in translation projects; the same types of texts processed and produced by Master students when undertaking practical training and preparing diploma papers. For example, typing the query “petroleum production” in the search field of Reverso Context result in a number of English-Russian parallel text fragments containing “petroleum production” in the left-hand English section and “добыча нефти” in the right-hand Russian one. This example also demonstrates the resource’s drawback that is lack of the Ukrainian version, due to which trainee translators have to perform additional translation from Russian into Ukrainian or vice versa.

The above mentioned resources are also useful for analyzing translations techniques and transformations applied by both professional and high performance trainee translators. The appropriate tasks can be as follows: 1) Look through the source and target texts and find the instances of addition /

omission / transposition / substitution. Explain the reasons for their application. 2) Look through the source and target texts, identify lexical transformations and comment on their application.

**Conclusions.** The study conducted in this paper has shown possible applications of comparable and parallel corpora in teaching English-Ukrainian oil and gas area related translation at the translation training phase. Further research should be aimed at integrating corpus-based tasks into the translation practice stage involving the implementation of a translation project.

### References

1. Braga Riera, J., Maroto, N. (2019). Using bilingual parallel corpora in translation methodology: an analysis of students’ translation competences in the UCMA-MUST corpus (English-Spanish). *Quaderns de Filologia: Estudis Lingüístics XXIV*: 39 – 58. doi: 10.7203/QF.24.16298.
2. Frerot, S. (2016.). Corpora and corpus technology for translation purposes in professional and academic environments: major achievements and new perspectives. *Cadernos de Tradução*, 36 (1), 36 – 61. URL: [http://www.scielo.br/scielo.php?script=sci\\_arttext&pid=S2175-79682016000500036](http://www.scielo.br/scielo.php?script=sci_arttext&pid=S2175-79682016000500036) (access date: 25.10.202)
3. Kübler, N., Mestivier, A., Pecman, M. (2018). Teaching specialised translation through corpus linguistics: translation quality assessment and methodology evaluation and enhancement by experimental approach. *Meta*, 63(3), 807–825. URL: <https://doi.org/10.7202/1060174ar>.
4. Loock, R. (2020). No more rage against the machine: how the corpus-based identification of machine-translationese can lead to student empowerment. *The Journal of Specialized Translation*, 34. URL: [https://www.jostrans.org/issue34/art\\_loock.php](https://www.jostrans.org/issue34/art_loock.php).

5. López-Rodríguez, C. I. (2016). Using corpora in scientific and technical translation training: resources to identify conventionality and promote creativity. *Cadernos de Tradução*, 36 (1), 88 – 120. URL: <http://dx.doi.org/10.5007/2175-7968.2016v36nesp1p88>.

6. Sharoff, S., Babych, B., Hartley, A. (2006). Using Comparable Corpora to Solve Problems Difficult for Human Translators. *ACL 2006, 21st International Conference on Computational Linguistics and 44th Annual Meeting of the Association for Computational Linguistics, Proceedings of the Conference*, Sydney, Australia, 17-

21 July 2006. URL: [https://www.researchgate.net/publication/220875185\\_Using\\_Comparable\\_Corpora\\_to\\_Solve\\_Problems\\_Difficult\\_for\\_Human\\_Translators](https://www.researchgate.net/publication/220875185_Using_Comparable_Corpora_to_Solve_Problems_Difficult_for_Human_Translators).

7. Symseridou, E. (2018). The web as a corpus and for building corpora in the teaching of specialised translation: the example of texts in healthcare. *FITISPOS International Journal*. 5 (1), 60 – 82. URL: <https://doi.org/10.37536/FITISPos-IJ.2018.5.1.160>.

8. Zanettin, F. (2002). Corpora in translation practice. URL: [https://www.academia.edu/30887110/Corpora\\_in\\_translation\\_practice](https://www.academia.edu/30887110/Corpora_in_translation_practice).

## ЗАСТОСУВАННЯ КОРПУСНИХ ТЕХНОЛОГІЙ У НАВЧАННІ ГАЛУЗЕВОГО ПЕРЕКЛАДУ

Стрілець В. В.

**Анотація.** Корпусні технології (корпуси англо- й україномовних текстів та інструменти для їх опрацювання) репрезентують сучасний фаховий дискурс та дозволяють швидко знаходити й зіставляти різні одиниці перекладу, що робить їх цінними засобами як для письмового перекладача-практика, так і майбутнього перекладача. Мета цієї статті – визначити роль і місце корпусних технологій у навчанні письмового галузевого перекладу на прикладі нафтогазової галузі. Охарактеризовано зіставні й паралельні корпуси текстів. Розкрито способи застосування моно- й білінгвальних зіставних і паралельних корпусів та корпус-менеджерів для набуття знань про жанрово-стилістичні особливості текстів; розвиток умінь виокремлювати термін і з'ясувати його колокаційний профіль та семантичну преференцію; аналізувати застосовані автором перекладацькі прийоми; перекладати колокації, складні іменникові конструкції, дієслівні словосполучення, абрєвіатури. Наведено приклади відповідних вправ і завдань, які доцільно виконувати у циклі занять тренування в перекладі. Окреслено перспективу подальших наукових пошуків, які слід спрямувати на інтеграцію завдань із застосуванням корпусних технологій у цикл занять із практики в перекладі, які передбачають виконання перекладацького проекту.

**Ключові слова:** корпусні технології, зіставні корпуси, паралельні корпуси, галузевий переклад, майбутні перекладачі.