Corpus Analysis for Developing Language Competencies in Future Professionals

Análisis de corpus para el desarrollo de competencias lingüísticas en futuros profesionales

Neonilla Barakatova
https://orcid.org/0000-0002-7433-288X
Associate Professor, Department of Philology and Language Communication, Dnipro University of Technology, Dnipro, Ukraine.

Tetiana Oharienko
https://orcid.org/0000-0002-8400-1670
Associate Professor, Department of Social and Humanitarian Disciplines, Donetsk State University of Internal Affairs, Kropyvnytskyi, Ukraine.

Anastasiia Kinashchuk
https://orcid.org/0000-0002-5675-240X
Ph.D., Senior Lecturer, Department of Foreign Languages of the Institute of Economics and Management, National University of Water and Environmental Engineering, Rivne, Ukraine.

Veronika Chekalyuk
https://orcid.org/0000-0003-1223-6646
Ph.D., Kyiv National Taras Shevchenko University, Kyiv, Ukraine.

Daryna Stanko
https://orcid.org/0000-0002-7858-8663
Associate Professor, English Philology Department, Uzhhorod National University, Uzhhorod, Ukraine.

Abstract

Aim. The aim of the study is to demonstrate the effectiveness of using corpus analysis for the development of language competencies of future specialists using specific examples of the Ukrainian and English languages. Methods. The research employs the methods of experimental comparison of training results, questionnaire survey, as well as monitoring and analysis of changes in language skills. Statistical methods were used to process the obtained data. Results. A significant improvement in language competence was observed in the group where corpus analysis was used: the percentage of students who achieved positive results increased from 40% to 70% after the implementation of this method. The difference between pre- and post-training indicators was 30%, which is statistically significant (χ^2 = 27.05, p < 0.001). Conclusions. The study confirmed the effectiveness of using corpus analysis for the development of language competencies of future specialists. The results indicate a significant improvement in the level of language comprehension, oral expression skills, and use of specialized vocabulary in the experimental group (EG) compared to the control group (CG). Prospects. Further research may focus on determining the impact of

corpus analysis on the language learning process in a variety of contexts.

Keywords: communication, educational corpus, higher education, higher education institution (HEI), professional competencies.

Resumen

El objetivo del estudio es demostrar la eficacia del uso del análisis de corpus para el desarrollo de competencias lingüísticas de futuros especialistas utilizando ejemplos específicos de los idiomas ucraniano e inglés. La investigación utiliza métodos de comparación experimental de los resultados de la formación, encuestas por cuestionario, así como seguimiento y análisis de los cambios en las habilidades lingüísticas. Se utilizaron métodos estadísticos para procesar los datos obtenidos. Se observó una mejora significativa en la competencia lingüística en el grupo donde se utilizó el análisis de corpus: el porcentaje de estudiantes que lograron resultados positivos aumentó del 40% al 70% después de la implementación de este método. La diferencia entre los indicadores previos y posteriores al entrenamiento fue del 30%, lo cual es estadísticamente significativo ($\chi^2 = 27{,}05$, $p < 0{,}001$). El estudio confirmó la eficacia del uso del análisis de corpus para el desarrollo de competencias lingüísticas de futuros especialistas. Los resultados indican una mejora significativa en el nivel de comprensión del lenguaje, habilidades de expresión oral y uso de vocabulario especializado en el grupo experimental (GE) en comparación con el grupo control (GC).

Palabras clave: competencias profesionales, comunicación, corpus educativo, educación superior, institución de educación superior (IES).

Introduction

Determined by the key role of language skills in the professional development and social adaptation of specialists in all areas of activity. They are critical to successful communication and career development in a world of globalization and growing international cooperation. The ability to learn not only one's native language, but also foreign languages, in particular English, which is the language of international, scientific, and business communication, is especially relevant today (Lefter et al., 2022).

Corpus are an important tool in acquiring communication competence. Along with traditional grammars and dictionaries, they are mandatory for presenting data as an effective reference system. The traditional methodological principle of visualization under the influence of corpus grammars was replaced by the principle of statistical visualization (Praat, n.d).

Linguo-statistic results are widely presented in the form of histograms, graphs, and word clouds not only in dictionaries, but also in modern textbooks on the English and Ukrainian languages. The technology of creating linguo-methodical materials has undergone significant changes in connection with the transformation of the corpus in the practice of compilation of dictionaries in a professional direction. Corpus analysis tools derived from corpus linguistics can serve as a basis for the creation of linguistic-methodical educational materials for the formation of language skills of future specialists (Cavasso & Taboada, 2021). The article is focused on revealing the role of corpus analysis in building professional communication skills.

Corpus analysis plays a special role in building language competencies of future journalists. It provides students with the opportunity to work with real texts that reflect the variety of speech situations in professional activities. It is important for journalists to understand the peculiarities of speech genres such as news, interviews, analytical materials, and corpus analysis helps to study their stylistics and variability.

Moreover, corpus data analysis helps to improve editing and proofreading skills, which is important for journalists (Bednarek & Carr, 2021).

Corpus Analysis for Developing Language Competencies in Future Professionals. - Eduweb, 2024, abril-junio, v.18, n.2. / 152-166

153

Pedagogical conditions aimed at effective development of students' language competence were created in the course of the experimental work. This goal was achieved through various methods of corpus analysis, which complemented traditional approaches to language learning (Matsera et al., 2023).

1. Being part of database-driven learning, corpus analysis tools rely heavily on the linguistic visibility of correspondence. Traditionally, these tools generate match strings that consist of compilation of corpus texts with the studied lexical units (LU). This approach has proven to be effective in language learning and should be considered a valuable tool. Students with A1-A2 proficiency can perform sorting exercises (Ma et al., 2022).
2. The phenomenon of semantic prosody, which includes subtle features of language use and is confirmed by numerous corpus examples, is widely used in building professional communication skills. Corpus statistics is used to confidently assess the authenticity of speech and to pose complex research tasks to students (Lin & Adolphs, 2023).
3. The corpus tools offer an enhanced context option that provides access to multiple sentences of the source text. Extended context provides important information about the time and circumstances of the creation of the text, the author, and the source of the publication. It also ensures inclusiveness of professional situations considered in the process of training future specialists (Horokhova, 2022).
4. Compilation of a "small" corpus has numerous advantages, in particular, the possibility of independent compilation of professional dictionaries by a team of teachers who are oriented to the future specialists' needs. Large, diverse and representative annotated corpora have been successfully created for languages such as English and Ukrainian. It is, however, important to note that no corpus can effectively serve all purposes (Oleškevičienė et al., 2021). Therefore, it is extremely important for universities to create specialized language corpora adapted to the needs of a particular faculty, department or university. Professionally oriented linguistic databases are created within university projects, for example, a corpus of teachers' speech in classes, a corpus of students' mistakes or a corpus of the studied subject area (Vosiljonov, 2022).

So, the main problem of experimental work is the study of effective methods of building students' language competence. The main focus of the study was the use of corpus analysis as an innovative method to improve English grammar and vocabulary, as well as to improve the level of communication skills. In particular, the research was aimed at determining how the use of corpus analysis can contribute to a better understanding of language structure, differences in the use of words and expressions in context, as well as the quality of students' communication in both Ukrainian and English. This approach made it possible not only to reveal the advantages of corpus analysis in language learning, but also to find out which aspects of language competence can be improved with the help of this method.

The aim of the study is to demonstrate the effectiveness of corpus analysis in the development of language competences of future specialists using the example of the Ukrainian and English languages. This will allow us to show the advantages of corpus analysis in comparison with the traditional method of language learning.

Objectives/questions

1. Comparison of the results of building students' language competencies before and after the application of pedagogical conditions.
2. Conducting a survey among students to assess their receptivity to the use of corpus data in the language learning process.
3. Analysis of changes in students' language skills.

Literature Review

The formation of language competence in pedagogical science is of great importance and continues to attract the researchers' interest f. It is the basis of successful communication when achieving a common goal. The ability to communicate effectively expands opportunities in personal and professional life, opens new horizons for study, work, and communication.

Caratozzolo and Alvarez-Delgado (2021) propose the concept Education 4.0 Framework, which defines an approach to the use of virtual and technological tools to enrich active learning. The authors indicate that the use of such tools helps to increase the efficiency of the educational process and stimulates the active participation of students in their own learning. Such innovations improve awareness among future specialists of the need to develop language skills.

In the later study, Caratozzolo, Rodriguez-Ruiz and Alvarez-Delgado (2022) research the use of natural language processing to assess STEM learning. The authors emphasize the importance of using AI-based tools for automated analysis and assessment of students' educational performance in STEM subjects.

The dissertation of Chua (2020) focuses on a corpus analysis of online discussions to explore the dialogic nature of online communication. The corpus method was used in the work to analyse the structure, means of communication and interaction of the participants of online discussions in order to reveal the features of this form of communication. The research can be useful in the process of developing the methodology of involving corpus methods in the process of language training of future specialists.

The article by Dong and Lu (2020) examine a methodology for developing competence in the use of subject-specific genres through corpus-based genre analysis. Corpus-oriented genre analysis tasks are proposed as an effective method of stimulating students' understanding and use of academic genres within a specific academic subject.

Ferraresi, Aragrande, Barrón-Cedeño, Bernardini and Petrović (2021) explore the competencies and skills required of linguists in the labour market based on an analysis of a corpus of job advertisements. The authors address changes in the demands of linguistics specialists and distinguish the key competencies that are essential for a successful career in this field.

Khaknazarova (2022) studies the role of corpus analysis in learning, focusing on its importance for improving the effectiveness of the learning process. The author describes the methods and approaches to the use of corpora in educational practice and emphasizes the importance of integrating corpus analysis into educational programmes to achieve a higher level of students' language competence.

The article by Melnyk, Tkachenko and Kalinichenko (2023) deals with the intercorpus analysis of lexico-semantic relations in modern languages. Attention is drawn to the importance of using the corpus approach to study the semantic relations between words in different contexts of communication, which contributes to a better understanding of the language system and the use of language for practical purposes. The research aims to help future specialists to improve their own communication skills by means of corpus linguistics.

The article by Mishchenko (2022) studies the corpus-linguistic approach to the study of English grammar, focusing on the specifics of using corpus data for the analysis of grammatical structures and linguistic regularities. The author highlights the importance of such an approach for improving the quality of teaching and language learning by students.

The work of Romaniuk and Trofimchuk (2021) examines the use of the corpus approach in teaching foreign

languages in HEIs. The authors identifies the advantages of using corpus data to improve students' communication skills and ensure their language competence.

Savchuk (2023) examines the importance of terminology in professional speech in the Ukrainian language. The author emphasizes the importance of corpus analysis for the study and systematization of professional terminology, which contributes to effective communication in the professional sphere.
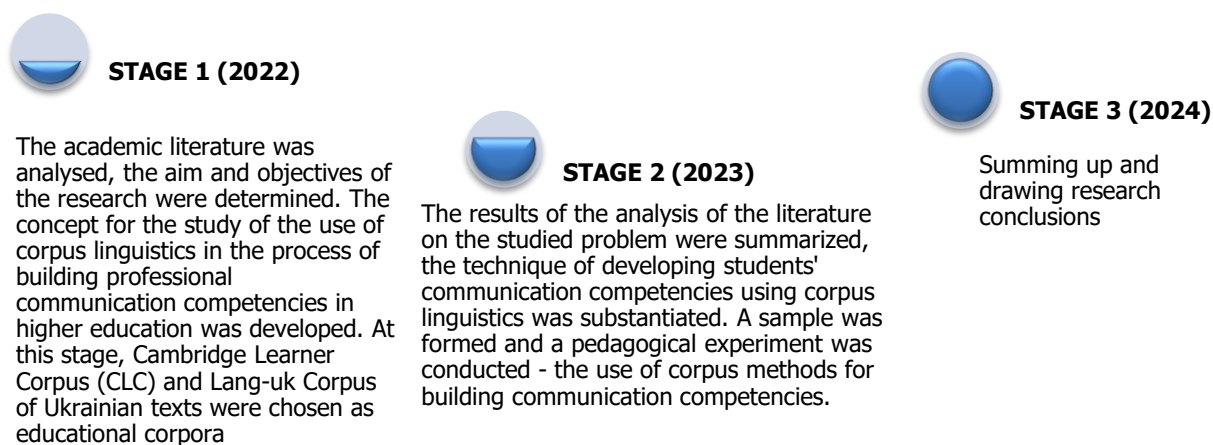
The work of Zhen and Han (2024) examines the issue of representation of national self-identity in mass media. It includes the reflection and expression of identity, values, beliefs and personal characteristics of different social groups or individuals through mass media. According to the researchers, it should include representation of different cultures, ethnic groups, gender identities, social classes, etc. Corpus analysis is an important tool for researching the representation of self-identity in the media, as it allows the analysis of large volumes of texts to identify patterns and trends in the way represented by different social groups and identities.

Understudied issues in the field of corpus linguistics and learning methodology cover various aspects that require more detailed research and attention in the academic community. One such issue is research on the use of corpora to study speech dynamics in online communication. For this purpose, it is necessary to study the changes that occur in the speech behaviour of people in the Internet environment, how these changes affect language structures and ways of expressing thoughts and ideas. Attention should also be paid to the use of corpora for studying the interaction of linguistic means and cultural aspects in communication. This process should include research into the linguistic features that are perceived or detrimental to different cultural groups, and how these features are reflected in the use of linguistic means.

Methodology

Design

The research was conducted in three stages. Figure 1 presents the content of each stage and its duration.

STAGE 1 (2022)

The academic literature was analysed, the aim and objectives of the research were determined. The concept for the study of the use of corpus linguistics in the process of building professional communication competencies in higher education was developed. At this stage, Cambridge Learner Corpus (CLC) and Lang-uk Corpus of Ukrainian texts were chosen as educational corpora

STAGE 2 (2023)

The results of the analysis of the literature on the studied problem were summarized, the technique of developing students' communication competencies using corpus linguistics was substantiated. A sample was formed and a pedagogical experiment was conducted - the use of corpus methods for building communication competencies.

STAGE 3 (2024)

Summing up and drawing research conclusions

Figure 1. Research stages.
Source: developed by the authors of the research.

So, all stages of research and experimental work are defined. This study can be classified as a cross-sectional study: data are collected at the same time and examined at the time they are collected, without follow-up or further analysis.

Participants

The lottery method was used to form a sample from the general population, which was carried out in several stages. At the first stage, all elements of the general population were marked. At the second stage, the necessary number of cards was randomly drawn from the deck. These cards were put aside and did not participate in further selection. So, irreversible selection was carried out. The study used a nested sample, that is, several courses were selected from the general population, within which the survey was conducted using a continuous method. The study of the effectiveness of corpus linguistics methods in building the communicative competence in Ukrainian and English languages was conducted at Drahomanov National Pedagogical University (Kyiv). The study involved 190 second- and third-year students of all faculties. Such a sample enables covering the required number of respondents to ensure the reliability of the results. The respondents were divided into two groups — control and experimental. Corpus linguistic tools were used in the experimental group for teaching English.

Instruments

The participants took part in the study through a remote questionnaire survey, which was carried out using Google Forms. The corpus manager Wordsmith 5.0 was used to work with the corpus.

Data collection

1. Experimental comparison of learning outcomes of students who use corpus analysis with the CG that uses traditional learning methods. This method involves dividing students into two groups: experimental and control. The EG studies using corpus analysis to learn language and develop language skills, while the CG uses traditional training methods. After completing the training course, both groups are tested to assess their language knowledge and skills. Test results are compared between groups to determine the effectiveness of corpus analysis in comparison with traditional learning methods.
2. The questionnaire survey among students to assess their receptivity to the use of corpus data in the process of language learning and development of language competencies. The method involves the creation of a questionnaire consisting of questions for students' assessment of their level of interest, knowledge, and experience in using corpus data in the educational process. The Cronbach's alpha coefficient for this questionnaire is 0.77, which is an indicator of high reliability for pedagogical research.
3. Monitoring and analysis of changes in students' language skills after the introduction of corpus analysis into the educational process. This method involves regular monitoring and data collection of students' language skills before and after implementing corpus analysis. The obtained data are analysed in order to identify the impact of corpus analysis on the development of language skills.

Analysis of data

1. The chi-squared test was calculated using the formula:

$$x^2 = (f_1 - f_2)^2 / (f_1 + f_2), \qquad (1)$$

where f_1 i f_2 – frequencies of compared samples.

2. The Cronbach's alpha reliability coefficient indicates the internal consistency of the test items. The Cronbach's alpha is calculated using the formula:

$$\frac{N}{N-1}\left(\frac{\sigma_x^2 - \sum_{i=1}^{N} \sigma_{Y_i}^2}{\sigma_x^2}\right), \qquad (2)$$

where σ_x^2 – total test score variance;

$\sigma_{Y_i}^2$ – i element variance.

3. The Mann-Whitney U test is calculated by using the formula:

$$U = (n_1 \times n_2) + (n_x \times (n_x+1) \,/2) - T_x, \qquad (3)$$

where n_1 – the number of respondents in the EG; n_2 – the number of respondents in the CG; T_x — the larger of the two rank sums; n_x — the number of respondents in the group with a higher rank sum.

Ethical criteria

The research participants were clearly informed about the importance of providing independent and truthful answers to the research questions. The respondents were informed about pedagogical conditions, in particular with the use of corpus linguistics methods for the development of their communicative competence. Ethical requirements regarding integrity, competence, respect for the individual, academic knowledge, and anonymity were observed when working with respondents and conducting questionnaire survey. The respondents' personal data were encrypted to ensure confidentiality. The results of the study are objective and unbiased.

Results

At the beginning and at the end of the study, the success of language competence building was monitored (Table 1).

Table 1.
Comparison of the levels of language competencies of the CG and EG students

Method	Group		Before	After	Difference	χ^2	p-value	U	p-value
Corpus analysis	Experimental group (n = 95)		40%	70%	30%	27.05	<0.001	4425	<0.001
Traditional methods	Control group (n = 95)		40%	55%	15%	8.10	0.004	2975	0.002

Source: developed by the authors of the research

Table 1 shows a comparison of the results of building language competencies of students in groups that studied using corpus analysis and traditional methods. The study was conducted for two groups of students: the EG that used corpus analysis and the CG that studied using traditional methods.

In the experimental group, a significant improvement in language skills was observed: the percentage of students who achieved positive results increased from 40% to 70% after the introduction of corpus analysis. The difference between pre- and post-training results was 30%, which is statistically significant ($\chi^2 = 27.05$, $p < 0.001$).

In the CG, there was also an improvement in language skills, but to a lesser extent than in the EG. The difference between pre- and post-training levels was 15%, which is also statistically significant ($\chi^2 = 8.10$, $p = 0.004$).

The next step was to study the students' receptivity to the use of corpus data in the language learning process. The results of the questionnaire survey are presented in Table 2.

Table 2.
Results of the student questionnaire survey at the end of the study

Question	EG (n = 95)	CG (n = 95)	χ²	p-value	U	p-value	Conclusion
The importance of corpus analysis	> 80%	< 50%	25.4	<0.001	4425	<0.001	The EG students value corpus analysis much more
The impact of corpus analysis on language skills	> 70%	< 40%	18.5	<0.001	4025	<0.001	The EG students are more confident in the benefits of corpus analysis for language
Desired implementation of corpus analysis	> 85%	< 60%	22.1	<0.001	4275	<0.001	The EG students are much more willing to use corpus analysis in learning
Agreeing that corpus analysis can help improve the quality of learning	> 90%	< 70%	31.4	<0.001	4500	<0.001	The EG students believe significantly more in the benefits of corpus analysis for better learning

Source: developed by the authors of the research

Table 2 provides the results of the student questionnaire, conducted at the end of the study, regarding the use of corpus analysis methods of acquired communicative competencies in the Ukrainian and English languages. Significant differences in student responses were found in the comparison between the EG and the CG. In the group that used corpus analysis, the vast majority of students considered this method important, compared to less than half of the students in the CG ($\chi^2 = 25.4$, $p < 0.001$).

The majority of the EG students believed that corpus analysis had a positive effect on their language skills, compared to less than 40% in the CG ($\chi^2 = 18.5$, $p < 0.001$). This indicates greater confidence of the EG students in the benefit of corpus analysis for language development. Besides, the vast majority of the EG students agree that corpus analysis can help to improve the quality of learning, compared to less than 70% in the CG ($\chi^2 = 31.4$, $p < 0.001$). This testifies to the greater faith of the EG students in the benefit of corpus analysis to improve the quality of education. The next step was to compare changes in communication skills of the EG and CG students (Table 3).

Table 3.
Results of comparison of changes in communication skills of the EG and CG students

Skills	Group	Before testing	After testing	Difference	χ²	p-value	U	p-value	Significance
Language comprehension	EG (n = 95)	40%	70%	30%	27.05	<0.001	4425	<0.001	High
	CG (n = 95)	40%	55%	15%	8.10	0.004	2975	0.002	Medium
Speaking	EG (n = 95)	45%	75%	30%	25.00	<0.001	4300	<0.001	High
	CG (n = 95)	45%	60%	15%	7.25	0.007	3125	0.004	Medium
Use of professional vocabulary	EG (n = 95)	35%	65%	30%	23.00	<0.001	4225	<0.001	High
	CG (n = 95)	35%	50%	15%	6.00	0.014	3000	0.003	Medium

Source: developed by the authors of the research

Corpus Analysis for Developing Language Competencies in Future Professionals. - Eduweb, 2024, abril-junio, v.18, n.2. / 152-166

159

Table 3 presents the results of a comparison of the change in communication skills between the EG and the CG students before and after testing. The changes in the language comprehension are analysed first. The EG group showed a significant increase in the level of language comprehension after testing, which is confirmed by high significance ($\chi^2 = 27.05$, $p < 0.001$). Compared to the CG, the difference was more noticeable. Similar trends are observed in speaking. The EG group also showed a high level of improvement in speaking after the test, which was confirmed by a high level of significance ($\chi^2 = 25.00$, $p < 0.001$), compared to the CG, where the growth was less noticeable. The EG students showed a significant improvement in the use of professional vocabulary, confirmed by high significance ($\chi^2 = 23.00$, $p < 0.001$), compared to the medium level of growth in the CG.

Compared to the CG, the EG achieved significantly better results through the use of the corpus analysis. The difference in indicators is statistically confirmed by χ^2 values and p-values, which indicate a high statistical significance of the results.

Discussion

According to Drijvers, Grauwin and Trouche (2020), creating a "small" corpus of professional vocabulary has several advantages. According to researchers, teachers can independently adapt the corpus to the needs of future specialists. Although there are large, diverse, and representative annotated corpora for languages such as English and Ukrainian, none of them can effectively meet all needs. Escudero-Mancebo et al. (2022) state that HEIs should develop specialized language corpora adapted to the needs of specific faculties, departments or the university as a whole. Professionally oriented linguistic databases are created under these projects, for example, corpora of classroom discourse, student mistakes, or the language of a certain major.

As Durna and Güneş (2020) state, the creation of a small corpus today is technically possible and justifiable for a university team of specialists thanks to the development of linguistic database management software such as the WordSmith Tools corpus manager. When searching for word combinations or conjugations, semantic prosody indicates the likely use of a word in certain contexts, both positive and negative.

Pérez-Paredes (2020) found that the verb "couse" often accompanies such negative connotation words as cancer, crisis, and delay. This regularity was found in more than 90% of the 250 examined occurrences in the corpus of 1 million word usages and 38 thousand occurrences in the corpus of 120 million word usages.

Furthermore, the Lexical Density Index draws the attention of the future specialist to the essential register features of written and oral speech. For example, according to the latest corpus grammars of the English language, news reports are the most lexically rich, and everyday dialogues are the most lexically sparse, as Zhukovska (2023) and Saddhono et al. (2023) noted. Unlike checked and edited texts of news articles, everyday dialogues between participants take place "live", when the lack of time makes it impossible to edit grammatical means and planning, the correction of what is said takes place in subsequent replicas, and the expressed statement cannot be deleted.

Koneva (2020) and Odden, Marin and Rudolph (2021) emphasized that it is necessary to know the rules of querying the corpus and acquire the basic skills of working with such a database in order to use the possibilities of the corpus for both students and teachers. A consistent consideration of the possibilities of corpus linguistics in a linguistic didactic context allows one to convincingly demonstrate its potential for the development of key foreign language competencies. Working on building the capabilities of the corpus allows both the teacher and the student to effectively use the corpus as a large authentic reference system and develop the skills of an autonomous researcher.

Proposing the idea of developing communicative competence, Newman-Griffis, Sivaraman, Perer, Fosler-Lussier and Hochheiser (2021) and Messina, Jones and Poe (2023) see the need for its development in the training of future specialists, as the use of corpora and corpus technologies is a means of supporting improvement existing methods of communication development. The authors rely on the significant potential of corpus learning.

The theoretical significance of this study is the expanded understanding of the effectiveness of corpus analysis as an innovative method in the development of language competencies. The results of the study reveal important aspects where corpus analysis can have the greatest impact on improving the quality of language learning. They can also be used to theoretically rethink the role of corpus analysis in modern education and linguistics.

The practical significance is that the obtained results can be used to develop improved methods of language learning using corpus analysis. They provide teachers and educational institutions with the grounds for implementing this method in the educational process in order to increase the effectiveness of education and improve students' language skills.

Limitations of this study include several factors that may affect its general adaptability and applicability. First, it is a sample limitation, as the study was conducted on a particular group of students at a particular educational institution. The results may be less representative for other contexts and groups of respondents. The second limitation is related to the duration of the study. The time available for data collection and analysis may be limited, which may affect the completeness and representativeness of the results. The time limitation can also make it difficult to assess the duration of the impact of corpus analysis on the development of students' language competencies.

Conclusions

The obtained results emphasize the importance of using corpus analysis in the educational process for building language competencies. The high efficiency of this method confirms its potential in improving language comprehension, speech development, and the use of professional vocabulary. Such results stimulate further research and implementation of corpus analysis in the educational process to improve the quality of education and training of future specialists. *Findings*. The study confirmed the effectiveness of using corpus analysis in the formation of language competencies of future specialists. The results showed a significant improvement in the level of language comprehension, speaking skills, and use of professional vocabulary in the EG compared to the CG. Students who used corpus analysis showed greater interest and willingness to use this method in education. The active influence of corpus analysis on improving the quality of education and the development of language skills indicates the need to include this method in educational practice. The general trend indicates the potential of corpus analysis as an innovative tool in improving the process of language learning and the development of language competencies. *Applications*. This research can be used in the educational field to improve the methods of language learning and build students' language competencies. The results may be useful for teachers and educational institutions seeking to optimize curricula and incorporate innovative teaching methods such as corpus analysis. *Research prospects.* Further research may focus on determining the impact of corpus analysis on language learning in different contexts, such as teaching English as a second language to speakers of other languages, or studying specific groups of speakers, such as linguistic minorities.

Bibliographic References

Bednarek, M., & Carr, G. (2021). Computer-assisted digital text analysis for journalism and communications research: introducing corpus linguistic techniques that do not require programming. *Media International Australia*, *181*(1), 131-151.

Corpus Analysis for Developing Language Competencies in Future Professionals. - Eduweb, 2024, abril-junio, v.18, n.2. / 152-166

161

Caratozzolo, P., & Alvarez-Delgado, A. (2021, October). Education 4.0 framework: enriching active learning with virtual and technological tools. In *Proceedings of the International Conference on Education*. *7*(1), 614-628. TIIKM Publishing. https://tiikmpublishing.com/proceedings/index.php/icedu/article/view/862

Caratozzolo, P., Rodriguez-Ruiz, J., & Alvarez-Delgado, A. (2022, March). Natural language processing for learning assessment in STEM. In *2022 IEEE Global Engineering Education Conference (EDUCON)*. IEEE, Tunisia. https://doi.org/10.1109/EDUCON52537.2022.9766717

Cavasso, L., & Taboada, M. (2021). A groovy delve into the world of online news comments: Unpacking with the Appraisal framework. *Journal of Corpora and Discourse Studies*, *4*, 1-38. https://doi.org/10.18573/jcads.61

Chua, S. M. (2020). *The dialogic nature of online discourse: A corpus analysis of online discussions*. United Kingdom: Open University. Retrieved from https://acortar.link/BMfkeY

Dong, J., & Lu, X. (2020). Promoting discipline-specific genre competence with corpus-based genre analysis activities. *English for Specific Purposes*, *58,* 138-154. https://doi.org/10.1016/j.esp.2020.01.005

Drijvers, P., Grauwin, S., & Trouche, L. (2020). When bibliometrics met mathematics education research: the case of instrumental orchestration. *Zdm*, *52*, 1455-1469. https://doi.org/10.1007/s11858-020-01169-3

Durna, F., & Güneş, O. (2020). A Corpus Linguistics Investigation into Phrasal Verbs in British Academic Spoken English. *Journal Of Foreign Language Education and Technology*, *5*(1), 204-223. Retrieved from https://www.ceeol.com/search/article-detail?id=887057

Escudero-Mancebo, D., Corrales-Astorgano, M., Cardeñoso-Payo, V., Aguilar, L., González-Ferreras, C., Martínez-Castilla, P., & Flores-Lucas, V. (2022). PRAUTOCAL corpus: a corpus for the study of Down syndrome prosodic aspects. *Language Resources and Evaluation*, *56*(1), 191-224. Retrieved from https://link.springer.com/article/10.1007/s10579-021-09542-8

Ferraresi, A., Aragrande, G., Barrón-Cedeño, A., Bernardini, S., & Milčević Petrović, M. (2021). *Competences, skills and tasks in today's jobs for linguists: Evidence from a corpus of job advertisements*. Zenodo. https://doi.org/10.5281/zenodo.5030879

Horokhova, T. O. (2022). Corpus technologies in the formation of grammatical competence of future language teachers. Theoretical and didactic philology: assets, problems, development prospects. In *Proceedings of the VII International Scientific and Practical Conference: Collection of Scientific Papers*. Grigory Skovoroda University in Pereyaslav, Pereyaslav. Retrieved from http://surl.li/rzrek

Khaknazarova, L. A. (2022). The role of corpus in teaching. *Journal of Innovations in Scientific and Educational Research*, *5*(4), 98-101. Retrieved from https://bestpublication.org/index.php/jaj/article/view/2598

Koneva, M. Z. (2020). Linguistic competence as an important component of foreign language communication of future foreign language teachers. *Scientific Notes*, *31*(70), 28-33. https://doi.org/10.32838/2663-6069/2020.1-3/05

Lefter, I., Baird, A., Stappen, L., & Schuller, B. W. (2022). A cross-corpus speech-based analysis of escalating negative interactions. *Frontiers in Computer Science*, *4*. https://doi.org/10.3389/fcomp.2022.749804

Lin, P., & Adolphs, S. (2023). *Corpus linguistics. In The Routledge Handbook of Applied Linguistics*. London: Routledge. Retrieved from https://acortar.link/aSn2Mn

Ma, Q., Tang, J., & Lin, S. (2022). The development of corpus-based language pedagogy for TESOL teachers: A two-step training approach facilitated by online collaboration. *Computer Assisted Language Learning*, *35*(9), 2731-2760. https://doi.org/10.1080/09588221.2021.1895225

Makhachashvili, R., Bakhtina, A., & Semenist, I. (2021). La función de la inteligencia emocional en la educación digital como el sustrato de la validez de la vida on-line. *Amazonia Investiga*, *10*(45), 20-30. https://doi.org/10.34069/AI/2021.45.09.2

Matsera, O., Tkachuk, T., & Paslavska, I. (2023). Translation as a tool for teaching English: the role of language corpora and translation tools. *Herald of Science and Education*, *10*(16). https://doi.org/10.52058/2786-6165-2023-10(16)-192-202

Melnyk, I. Ye., Tkachenko, L. M., & Kalinichenko, T. M. (2023). Intercorpus analysis of lexical-semantic relations in modern languages. *Transcarpathian Philological Studies*, *30*, 244-249. https://doi.org/10.32782/tps2663-4880/2023.30.45

Messina, C. M., Jones, C. E., & Poe, M. (2023). Prompting reflection: Utilizing corpus linguistic approaches in assessing reflective writing locally. *Written Communication*, *4*0(2), 620-650. https://journals.sagepub.com/doi/abs/10.1177/07410883221149425

Mishchenko, O. V. (2022). *Corpus-linguistic approach to the study of English grammar: special*. Mykolaiv: CHNU named after Petra Mohyly. Retrieved from https://krs.chmnu.edu.ua/jspui/handle/123456789/2181

Newman-Griffis, D., Sivaraman, V., Perer, A., Fosler-Lussier, E., & Hochheiser, H. (2021). TextEssence: A Tool for Interactive Analysis of Semantic Shifts Between Corpora. In *Proceedings of the conference. Association for Computational Linguistics. North American Chapter. Meeting* (Vol. 2021, p. 106). NIH Public Access. Retrieved from https://www.ncbi.nlm.nih.gov/pmc/articles/PMC8212692/

Odden, T. O. B., Marin, A., & Rudolph, J. L. (2021). How has Science Education changed over the last 100 years? An analysis using natural language processing. *Science Education*, *105*(4), 653-680. https://doi.org/10.1002/sce.21623

Oleškevičienė, G. V., Mockienė, L., & Stojković, N. (2021). *Corpus Analysis for Language Studies at The University Level*. Cambridge: Cambridge Scholars Publishing.

Pérez-Paredes, P. (2020). *Corpus linguistics for education: A guide for research*. London: Routledge. https://doi.org/10.4324/9780429243615

Praat. (n.d.). *Praat: Doing phonetics by computer*. Retrieved from http://www.praat.org/

Romaniuk, S., & Trofimchuk, V. (2021). Corpus approach in teaching foreign languages in higher education institutions. *Innovation in Education*, *1*(13), 192-199. https://doi.org/10.35619/iiu.v1i13.332

Saddhono, K., Rohmadi, M., Setiawan, B., Suhita, R., Rakhmawati, A., Hastuti, S., & Islahuddin, I. (2023). Corpus linguistics use in vocabulary teaching principle and technique application: A study of Indonesian language for foreign speakers. *International Journal of Society, Culture & Language, 11*(1), 231-245. Retrieved from https://www.ijscl.com/article_697566.html

Savchuk, N. (2023). Terminology as an important component of professional speech in the Ukrainian language. *Science and Perspectives*, *7*(26). https://doi.org/10.52058/2695-1592-2023-7(26)-113-126

Vosiljonov, A. (2022). Basic theoretical principles of corpus linguistics. *Academicia Globe*, *3*(02), 173-175.

Zhen, H., & Han, B. (2024). Original Research Article The role of iconography in shaping Chinese national identity: Analyzing its representation in visual media and political propaganda. *Journal of Autonomous Intelligence*, *7*(3). https://doi.org/10.32629/jai.v7i3.1516

Zhukovska, V. V. (2023). Corpus technologies and genre-analytical approach in teaching English for academic purposes. *Discourse of professional and creative communication: linguistic, cultural, cognitive, translation and methodical aspects*, 126-128. Retrieved from http://eprints.zu.edu.ua/38987/1/Zhukovska.pdf

Corpus Analysis for Developing Language Competencies in Future Professionals. - Eduweb, 2024, abril-junio, v.18, n.2. / 152-166

163

Questionnaire for assessing the level of students' interest, knowledge, and experience in using corpus data in the educational process

1. How do you rate your level of knowledge about corpus analysis?
- Very low
- Low
- Medium
- High
- Very high

2. Have you had experience using corpus data for educational purposes before?
- Yes
- No

3. Do you consider corpus analysis an important tool in language learning?
- Yes
- No

4. Do you have skills in working with corpus tools (e.g., search interfaces, filters, etc.)?
- Yes
- No

5. To what extent do you consider yourself interested in studying language and its structure using corpus analysis?
- Very interested
- Interested
- Neutral
- Not very interested
- Not interested

6. Do you have experience using Internet resources with corpus data (for example, websites with corpus texts)?
- Yes
- No

7. Do you know how to effectively use corpus data to learn language and improve language skills?
- Yes
- No

8. Do you think corpus analysis can help improve your language skills?
- Yes
- No

9. Do you have experience using corpus data analysis software?
- Yes
- No

10. How often do you use corpus data in your teaching or research?

- Every day
- Several times a week
- Several times a month
- Rarely
- Never

11. What specific aspects of language would you like to study using corpus analysis? (e.g. vocabulary, syntax, stylistics, etc.)

12. How do you rate the availability of corpus analysis resources for your language learning?
- Freely available
- Available
- Neutral
- Not available
- Not available at all

13. Have you used corpus data in your previous learning or research?
- Yes
- No

14. How do you rate the difficulty of corpus analysis for your level of knowledge?
- Very difficult
- Difficult
- Medium
- Easy
- Very easy

15. Are you confident in your ability to analyse and interpret corpus data?
- Yes
- No

16. How often do you look for additional information or resources on corpus analysis to support your learning?
- Every day
- Several times a week
- Several times a month
- Rarely
- Never

17. How desirable do you consider the introduction of corpus analysis into the educational process of your educational institution?
- Highly desirable
- Preferable
- Neutral
- Less desirable
- Not desirable

18. How do you rate your readiness to use corpus analysis in learning and research?
- Ready
- Partially ready
- Not ready

Corpus Analysis for Developing Language Competencies in Future Professionals. - Eduweb, 2024, abril-junio, v.18, n.2. / 152-166

165

19. What advantages do you see in using corpus analysis compared to traditional language teaching methods?
20. What disadvantages do you see in the use of corpus analysis in language learning?
21. Would you like additional training in corpus analysis to improve your skills?
22. How do you rate the level of support and availability of corpus resources in your educational institution?
23. Do you agree that corpus analysis can help to improve the quality of your teaching and the development of language competencies?
24. Are you ready to accept an additional task to study corpus analysis during your year of study?
25. Would you like to be able to share your own findings and conclusions from corpus analysis with other students or researchers?