Shahid Chamran University of Ahvaz

Conference Research Paper

Use of Corpus Technologies for the Development of Lexical Skills

Albina Fanilevna Mukhamadiarova[1], Lilia Fidarisovna Caserta[2], Mariya Alexandrovna Kulkova[3], & Kirsten Reuter[4]

[1]Kazan Federal University, Kazan, Russia; *liliana_muhamad@mail.ru*

[2]Ferris State University, USA; *casertal@ferris.edu*

[3]Kazan Federal University, Kazan, Russia; *mkulkowa@rambler.ru*

[4]Kazan Federal University, Kazan, Russia; *reuterkirsten@googlemail.com*

Abstract

The introduction of informational technologies in the modern educational process is an essential part of the successful work of the teacher and the guaranty of the achievement of learning outcomes. The text corpus is one of the modern resource tools for the creation of language learning assignments for students. In addition, we present teaching materials for studying the homonymy and polysemy of the words. We offer two different forms of organization of the learning process with emphasis on corpus technology: teacher presenting select assignments to students and students independently working with texts from a text corpus, as a mini-research project.

Keywords: Concordance; Corpus; Corpus' Technology; Development of Lexical Skills; Foreign Language.

1. Introduction

In present times one of the most important and stable tendencies of the development of the international educational process is using modern informational tools and technologies as a part of the learning process (Khakimullina, Ayupova, Zakirova, & Alvarez, 2018; Mukhamadiarova, Kulkova, Caserta, & Reuter, 2020; Nurullina, Muraviyov, Martyanova, & Yarmakeev, 2018). The goal of learning foreign languages is successful communication, both verbally and in writing. As stated by N.D. Galskova and N.I. Gez: Possession of a foreign language vocabulary in terms of semantic accuracy, synonymous wealth, and the appropriateness of its use is an essential prerequisite for the realization of this goal (Galskova & Gez, 2007). «The process of education based on the thematic lexis and the lexis of semantic fields facilitates the quantitative growth of the active vocabulary of students» (Varlamova, Tulusina, Zaripova, & Gataullina, 2016).

The development of corpus' linguistics allows a new and different understanding of language and learning foreign languages because text corpuses give scientists an opportunity to test innovative theories through a multitude of texts. The text corpus brought scientific attention to the compatibility of words, collocation, and chunks (Men, 2015).

Text corpuses allow us to avoid categorical opposition (right vs. wrong) in relation to the German language and give us the ability to project the most likely usage of phrases. In addition, we move from a word to a phrase or a sentence to research collisional (morph syntactical) and collocational (lexical and phraseological) compatibility (Goźdź-Roszkowski & Pontrandolfo, 2017; Hasselg\aard, 2019; Karthick, Shalinie, Umabharathi, & Saroja, 2017; Novakova & Siepmann, 2019).

We agree with scientist J. Sinclair that units of meaning exist in phrases and sentences. Most words exist inside the context of common usage, with the exception of scientific terms and names of animals or plants (Sinclair, 1996).

In methods of teaching a foreign language, special attention must be given to collocations. Scientist J. Hill introduced and developed the term "collocational competence," which is understood as a consistent and typical co-

Journal of Research in Applied Linguistics, 11, Special Issue
Proceedings of the 7th International Conference on Applied Linguistics Issues (ALI 2020),
Saint Petersburg, 13-14 June 2020

Shahid Chamran University of Ahvaz

occurrence of the particular word (Hill, 1999). Corpus' linguistic scientists refer to the term "collocation" in relation to statistically stable linguistic units – a combination of two or more words (Scott & Tribble, 2007).

2. Methods

In this research, we use the following text corpuses: IDS-Mannheim, Wortschatz Universität Leipzig, Digitales Wörterbuch der Deutschen Sprache. Concordance, as a rich illustrative material, was applied in the study and preparation of the exercises. In the process of identifying collocations, we should examine concordance strings and use of extended context functions, since collocations may be separated by other words (Zakirov, Sayapova, & Andryushchenko, 2017).

We also created our own text corpus with the help of the AntConc program. Our corpus consists of 15,678-word combinations and phrases (Mirimanova, 2002). This program is free to access and works with documents formatted as TXT, HTML and XML text files. Our corpus contains 30 files. We chose a collection of texts on several topics based on the academic content of the "Practical course of a second foreign language":

Staatsaufbau, geografische Lage Deutschlands – Government system, Geography of Germany;
Politik – Political system;
Massenmedien – Social media;
Kultur Deutschlands – German culture.

3. Results

Based on the content, learning materials were divided into several groups:

1. Assignments for building lexical/collocational word profiles

Assignments for developing background and linguistic-cultural knowledge
Mini research of materials from the text corpus

Assignments with a focus on the resources from the text corpus were divided into the following groups:

2. Assignments with the use of the concordance
Assignments with the use of statistical frequency and "clouds of words."
Assignments with the use of dictionaries.

The organization of the learning process, with a focus on text corpus technologies, may be divided into two groups:

3. Students working on assignments created by a teacher with the help of the text corpus
Students working independently with the text corpus and mini-research projects on the given topic

For creating assignments, we used the following text corpus technologies:

Concordance for observation of the collocational profile of the word
Keywords or "clouds of words" for specific context and frequency
Dictionary

3.1. Compilation of the lexical profile of the word

Based on a three-step classification, we developed a set of assignments for each word's lexical and collocational profile. The assignments are based on a concordance, the statistical frequency of the text corpus and the vocabulary, and are aimed at the independent work of students with the text corpus, including the implementation of the projects developed by the teacher (Figure 1).
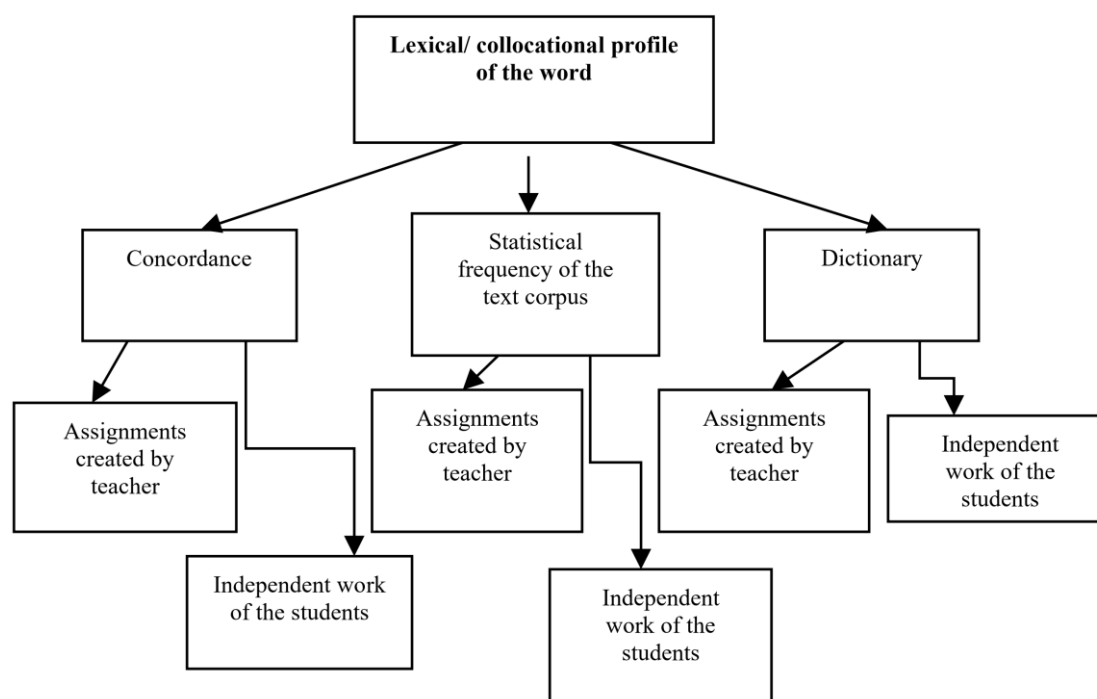
Journal of Research in Applied Linguistics, 11, Special Issue
Proceedings of the 7th International Conference on Applied Linguistics Issues (ALI 2020),
Saint Petersburg, 13-14 June 2020

Shahid Chamran University of Ahvaz

Figure 1. Classification of the assignments for the lexical / collocational profile of the word.

The goal of systematic work with a focus on collocational and co-competitive compatibility is to produce strong lexical skills in students. The skills obtained through this work will assist students in developing foreign language communicative competence (Mead & Doecke, 2020). Students are capable of correctly expressing their thoughts using proper grammatical context and phrases. In the process of creating assignments for independent students' research, a dictionary plays an important role. DWDS-Wörterbuch based on the dictionary "Wörterbuch der Deutschen Gegenwartssprache" and in part on "das Große Wörterbuch der Deutschen Sprache". The text corpus continues to grow through the addition of new words and slang expressions. For learning purposes, an electronic educational resource was created; students who learn new words and phrases from the research may use this resource to add this new information to the glossary. First, students use German-language text corpuses (Ramilevna Kayumova, Abel'chaerovna Safina, & Arkadjewna Nefedova, 2019; Valeryevna Erofeeva, Ilgizovna Gimatova, & Vjatcheslavovna Sergeeva, 2019). Secondly, they choose a link to a Wörterbucher. Then they highlight the lexical meaning of the word. The final step is finding examples of word usage in speech and in proverbs if possible. As a result, all the examples are compiled in the word's profile. If there is a problem with the pronunciation of a word, the resource provides the option to hear how the word sounds. Students also can practice the pronunciation of any word. As an example for this learning activity, we can use T. Bartz's "Digitale Sprachressourcen im Deutschunterricht: Korpus-basierte Recherche und Analyse in der, Wörterbuchwerkstatt'. Korpus-basierte, Wörterbuchwerkstatt' im Deutschunterricht" (Bartz, Pölitz, Morik, & Storrer, 2015). The author of the research describes an algorithm for the creation of the Wiki-dictionary by using the German language text corpus DWDS. This research project is based on the students and teachers' cooperative work. We are proposing the creation of a similar dictionary by using the edu.kpfu platform. For example, here, we can demonstrate the lexical profile of the word "Lebkuchen" – "Gingerbread" from a research theme "Feste and Bräuche" – "Holidays." This lexical word profile was created by students (Figure 2). In this image, we can see the gender of the noun, the lexical meaning of the word, an example of usage in speech, an etymological reference, its translation from German to Russian, and typical phrases with the word "Lebkuchen."
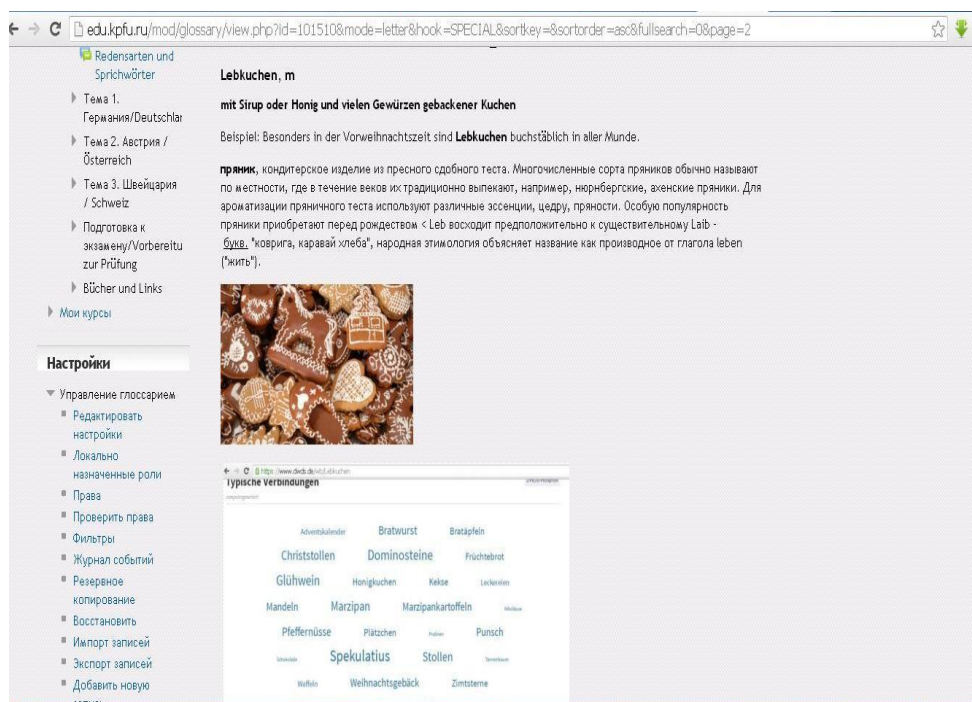
Journal of Research in Applied Linguistics, 11, Special Issue
Proceedings of the 7th International Conference on Applied Linguistics Issues (ALI 2020), Saint Petersburg, 13-14 June 2020

Shahid Chamran University of Ahvaz

Figure 2. Lexical profile of the word Lebkuchen. Notes.

Every session of a second language or regional geography course (the German language and geography in our case) should start with a chapter Wort des Tages or Redewendung des Tages – word of the day or phrase of the day – to increase students' interest in learning a foreign language. During the first lesson, a teacher introduced a new word or a proverb, which relates to the theme of the lesson (Elgort, 2018). The usage of the automatically generated text corpus of the German language – Wort des Tages – helps with this activity. This text corpus analyzes a massive number of texts, selects the most often applied words or phrases, and shares this information with students. To confirm the word's lexical meaning and to illustrate the usage of idioms, teachers employ examples from the text corpus DWDS. "Learning an idiomatic language develops a learners' verbal ability to implement communicative intentions in order to express their evaluative opinions" (Konopatskaya, Yarmakeev, Pimenova, & Abdrafikova, 2017). The purpose of creating an algorithm of research through the usage of text corpuses is to teach students ways to independently employ those sources to improve their writing skills in a foreign language (Shtyrlina, 2017). This method helps students to develop language skills, increases motivation to learn a foreign language and enriches the vocabulary of the second language. For future improvement, students may use an online educational resource, like forum Redensarten und Sprichwörter, where learners once a week submit a writing assignment with examples from the text corpus. The students are encouraged to find foreign idioms or proverbs to clarify their meaning, as presented in image 2. Image 3 shows the transformation of the traditional usage of a German proverb. This proverb about winter snowfall is used in a political context; the minister (sie) is compared to Mrs Holle in her ability to cover problem spots with "white flakes" (weiße Flocken) (Figure 3).
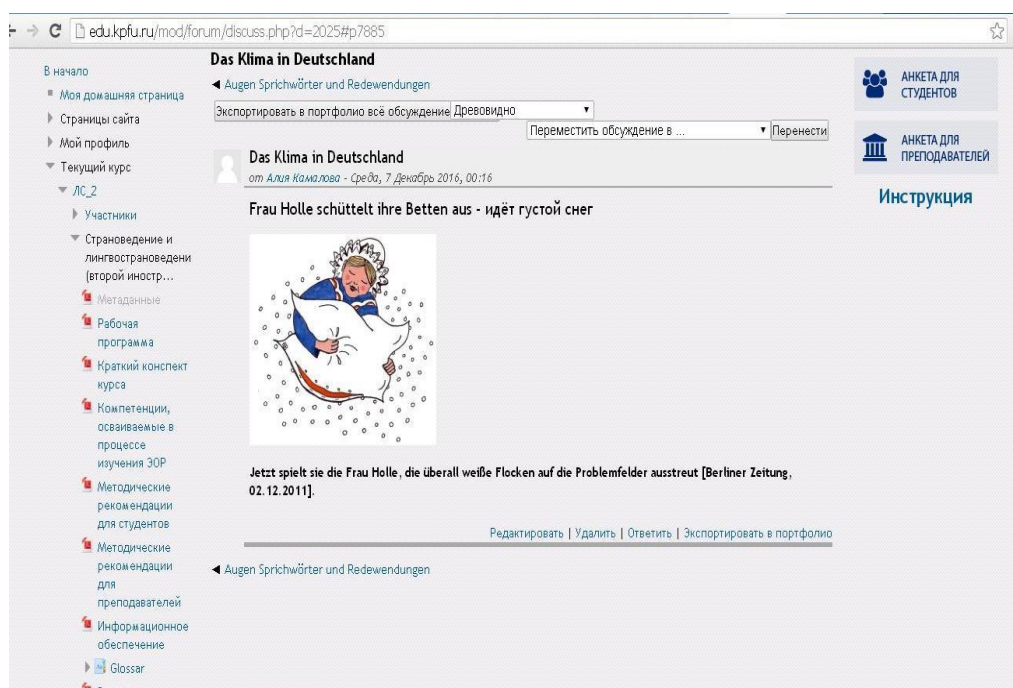
Journal of Research in Applied Linguistics, 11, Special Issue
Proceedings of the 7th International Conference on Applied Linguistics Issues (ALI 2020),
Saint Petersburg, 13-14 June 2020

Shahid Chamran University of Ahvaz

Figure 3. Forum Redensarten und Sprichwörter. Frau Holle schüttelt ihre Betten aus Notes.

3.2. The corpuses' methods of determination of collocations

The teacher-created assignments, which explore collocations, help students to remember correct usage of the phrases and to apply this knowledge in their speech and writing. During a regional geography lesson, students orally present different topics and express their opinions as an essential part of the class communication/discussion. In addition, proverbs and idioms are used in interviews and other modern texts. For this learning goal, we chose collocations most commonly and often used in modern verbal speech and made a list of examples of those collocations. Also, students can find examples of collocations in the text corpus DWDS and independently finish the provided grammar table (Miftakhova, Bochina, & Zhuravleva, 2018).

Table 1. *Collocations "Expressing personal opinion"*

Kollokation zum Thema «Meinung äußern»	Beispiel
eigene, persönliche, subjektive Meinung/ Überzeugung	
kontroverse, konträre, gegenteilige Meinungen	
zu einer Meinung kommen	
sich eine Meinung bilden	
eine Meinung haben/vertreten	
eine Meinung teilen	
die Meinung ändern	
eine Meinung äußern	
die Entscheidung treffen	
vor der Entscheidung stehen	
schwierige, richtige, gut überlegte, endgültige, emotionale Entscheidung	

We will demonstrate the analysis of the students' learning process based on the results of assignment 3. For this assignment, the students are divided into two groups, and each group must learn the frequency of usage of the phrases *Im Vergleich Zu* and *Im Vergleich mit*. The first group must research through the text corpus DWDS, and the second group must use materials from the text courpus IDS-Mannheim.

Journal of Research in Applied Linguistics, 11, Special Issue
Proceedings of the 7th International Conference on Applied Linguistics Issues (ALI 2020), Saint Petersburg, 13-14 June 2020

Shahid Chamran University of Ahvaz

The results from the text corpus DWDS can be seen in table 2.

Table 2. *Comparative analysis of Im Vergleich zu and Im Vergleich mit, text corpus DWDS*

	Im Vergleich Zu	ImVergleichmit
Die Zeit	10146	1635
Tagesspiegel	3400	575
Berliner Zeitung	4463	770
Gesprochene Sprache	28	5

The second group's results are presented in Figure 4.
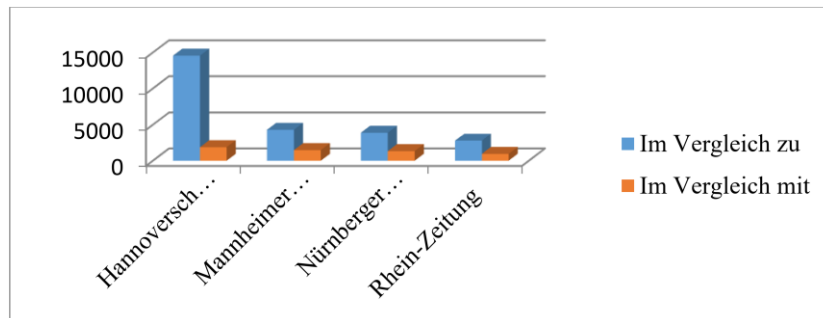
Figure 4. Comparative analysis of *Im Vergleich zu* and *Im Vergleich mit*, text corpus IDS-Mannheim

Students conduct research by systematically selecting, analysing and organizing language facts. Teachers supervise the students' research and provide guidance for the students' independent learning process.

4. Discussion

The goal of assignment 2 is to create a collocational "genealogical tree" for the words das Problem, die Partei, die Hochschule, and begehen.

The purpose of assignment 3 is to identify the frequency of use of the following phrases: Im Vergleich Zu and Im Vergleich mit.

4.1. Assignments for exploration of homonymy and polysemy of the words

The science of text corpus linguistics can be beneficial to the process of studying foreign words with more than one meaning. The teacher chooses various text with a target word, so students must understand the meaning of the word based on different contexts (Sakaeva, 2018). The words in italicized print functions as a help word because it hints to a correct understanding of the target word, given in bold print. For example,

1. a) Heike Weber und Roland Au fanden alle Gerichte sehr schmackhaft [Die Zeit, 11.03.2015 (online)].

b) Sawtschenko war in Einem international kritisierten Mordprozess von einem russischen Gericht zu 22 Jahren Lagerhaft verurteilt Worden Die Zeit, 28.11.2016 (online)

2. a) Liebe Leser, Weihnachten ist nicht bloß das Fest der Geschenke und der gebratenen Enten [Die Zeit, 20.12.2000, Nr. 52].

Zum ersten Gang gab es gebratene Enten, kalten Schinken, Forellen, blauen Hecht, Taubenpasteten, Fleischtorten und gedämpftes Schweinefleisch [Die Zeit, 11.05.2006, Nr. 20].

b) Die Zeitungsente war von einer Truppe in die Welt gesetzt worden, die sich für Obamas Klimapläne ausspricht – eine Pressemittelung mit einem gefälschten Logo reichte, um die Medien für einige Stunden zu narren.

Journal of Research in Applied Linguistics, 11, Special Issue
Proceedings of the 7th International Conference on Applied Linguistics Issues (ALI 2020), Saint Petersburg, 13-14 June 2020

Shahid Chamran University of Ahvaz

4.2. Concordance assignments for words similar in spelling but different in meaning

Students will have a list of words that have similar spellings but different meanings from the text corpus DWDS. Those words are very difficult to learn for foreign students who are studying the German language. We use the web source Deutsche Welle to select words.

The teacher selects examples from the text corpus and gives examples to students, so they may work independently to find the meanings of homophones and the difference in function of these words in the foreign language. The examples look like this:

4. Und was deprimiert stärker: keine Arbeit zu haben oder trotz großen politischen Engagements dauerhaft nur Almosen zu empfangen?

Aber ist sie nicht deprimiert über den Zustand der Partei, über das Postengeschacher, über interne Kämpfe, die extern niemanden interessieren?

Bundeskanzler Schröder bezeichnete das anhaltende Umfragetief der SPD als deprimierend.

5. Conclusion

The present research demonstrates how text corpuses and their technology can be helpful in learning lexical skills. Analysis of existing materials and methods of teaching foreign languages allows us to make the following conclusion – the text corpuses have extensive data and great potential as a learning tool for the German language. Text corpuses are very helpful in learning lexical meanings of words and phrases. Text corpuses may be used as research material for creating assignments covering the lexical and collocational profiles of the words. The dictionary, concordance and statistic frequency of the text corpuses are highly productive teaching instruments. One of the advantages of corpus linguistics, as a science, is active language learning. Unlike traditional passive processes for language learning, the use of corpus linguistics encourages students to study through research, guided discovery, and exploration of foreign languages. In this process, the students' research is an important link to developing their knowledge of the language.

References

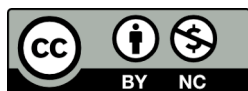Bartz, T., Pölitz, C., Morik, K., & Storrer, A. (2015). Using data mining and the clarin infrastructure to extend corpus-based linguistic research. *Selected Papers from the CLARIN 2014 Conference, October 24-25, 2014, Soesterberg, The Netherlands*, (116), 1–13. Linköping University Electronic Press.

Elgort, I. (2018). Technology-mediated second language vocabulary development: A review of trends in research methodology. *Calico Journal*, *35*(1), 1–29.

Galskova, N. D., & Gez, N. I. (2007). Theory of teaching foreign languages. *Linguodidactics and Methods (4th Ed.). M: Publishing Center «Academy.*

Goźdź-Roszkowski, S., & Pontrandolfo, G. (2017). *Phraseology in Legal and Institutional Settings: A Corpus-based Interdisciplinary Perspective*. Routledge.

Hasselg\aard, H. (2019). *Phraseological teddy bears: Frequent lexical bundles in academic writing by Norwegian learners and native speakers of English.*

Hill, J. (1999). Collocational competence. *Readings in Methodology*, *162*.

Karthick, S., Shalinie, S. M., Umabharathi, S., & Saroja, S. K. (2017). Parallel hierarchical subspace clustering for segmenting large text corpuses. *2017 International Conference on Trends in Electronics and Informatics (ICEI)*, 1–6. IEEE.

Khakimullina, R., Ayupova, R., Zakirova, L., & Alvarez, M. L. O. (2018). Internet Technologies In Teaching Foreign Languages. *National Academy of Managerial Staff of Culture and Arts Herald*, (3).

Journal of Research in Applied Linguistics, 11, Special Issue
Proceedings of the 7th International Conference on Applied Linguistics Issues (ALI 2020),
Saint Petersburg, 13-14 June 2020

Shahid Chamran University of Ahvaz

Konopatskaya, E. A., Yarmakeev, I. E., Pimenova, T. S., & Abdrafikova, A. R. (2017). Teaching idiomatic English in ESP class. *QUID: Investigación, Ciencia y Tecnología*, (1), 1783–1788.

Mead, P., & Doecke, B. (2020). Pedagogy. In *Oxford Research Encyclopedia of Literature*.

Men, H. (2015). *Vocabulary increase and collocation learning: A corpus-based cross-sectional study of Chinese EFL learners* (PhD Thesis). Birmingham City University.

Miftakhova, A., Bochina, T., & Zhuravleva, Y. (2018). Gender Interpretation Of Russian Lexeme БАБА/ВАВА IN INTERNET DISCOURSE. *National Academy of Managerial Staff of Culture and Arts Herald*, (3).

Mirimanova, M. S. (2002). Tolerance as education problem. *Development of the Personality*, (2), 282–303.

Mukhamadiarova, A. F., Kulkova, M. A., Caserta, L. F., & Reuter, K. (2020). Use of Corpus Technologies for the Development of Lexical Skills/Uso de tecnologias de corpus para el desarrollo de habilidades lexicales. *Utopía y Praxis Latinoamericana*, *25*(S7), 185–195.

Novakova, I., & Siepmann, D. (2019). *Phraseology and Style in Subgenres of the Novel: A Synthesis of Corpus and Literary Perspectives*. Springer Nature.

Nurullina, G. M., Muraviyov, A. F., Martyanova, A. A., & Yarmakeev, I. E. (2018). Project Technology in the Development of Communicative Competence in Schoolchildren: Extracurricular Classes of Russian Language. *Cypriot Journal of Educational Sciences*, *13*(4), 461–468.

Ramilevna Kayumova, A., Abel'chaerovna Safina, R., & Arkadjewna Nefedova, L. (2019). Issues and Trends in Phraseodidactics. *Research in Applied Linguistics*, *10*(Proceedings of the 6th International Conference on Applied Linguistics Issues (ALI 2019) July 19-20, 2019, Saint Petersburg, Russia), 694–701.

Sakaeva, L. (2018). Translation features of author neologisms on the example of Modern English prose. *Revista San Gregorio*, (23), 108–115.

Scott, M., & Tribble, C. (2007). Textual patterns: Key words and corpus analysis in language education. *TESL-EJ*, *11*(2).

Shtyrlina, E. G. (2017). Concept as a linguistic guideline in teaching Russian as a foreign language. *Modern Journal of Language Teaching Methods*, *7*(12), 90–96.

Sinclair, J. (1996). The search for units of meaning. *The Search for Units of Meaning*, 1000–1032.

Valeryevna Erofeeva, I., Ilgizovna Gimatova, L., & Vjatcheslavovna Sergeeva, E. (2019). Functioning of the Lexeme 'Intellect' in Russian and German-Speaking Discourses of the Mid-XXth Century. *Research in Applied Linguistics*, *10*(Proceedings of the 6th International Conference on Applied Linguistics Issues (ALI 2019) July 19-20, 2019, Saint Petersburg, Russia), 748–757.

Varlamova, E. V., Tulusina, E. A., Zaripova, Z. M., & Gataullina, V. L. (2016). Lexical, semantic and culturological approaches to the teaching of a second language. *Analele Universității Din Craiova. Seria Științe Filologice. Lingvistică*.

Zakirov, A., Sayapova, A., & Andryushchenko, O. (2017). The Incident In Forming Adultery Motif In The Artistic Structure Of The Novel "Anna Karenina" By Leo Tolstoy. *Revista Publicando*, *4*(13 (2)), 894–903.