



## ARTÍCULOS

UTOPIA Y PRAXIS LATINOAMERICANA. AÑO: 25, n° EXTRA 7, 2020, pp. 185-194  
REVISTA INTERNACIONAL DE FILOSOFÍA Y TEORÍA SOCIAL  
CESA-FCES-UNIVERSIDAD DEL ZULIA. MARACAIBO-VENEZUELA  
ISSN 1316-5216 / ISSN-e: 2477-9555

# Use of Corpus Technologies for the Development of Lexical Skills

*Uso de tecnologías de corpus para el desarrollo de habilidades léxicales*

**Albina Fanilevna MUKHAMADIAROVA**

<https://orcid.org/0000-0001-6918-9209>

[liliana\\_muhamad@mail.ru](mailto:liliana_muhamad@mail.ru)

Kazan Federal University, Kazan, Russia

**Lilia Fidarisovna CASERTA**

<https://orcid.org/0000-0002-1233-9373>

[casertal@ferris.edu](mailto:casertal@ferris.edu)

Ferris State University, USA

**Mariya Alexandrovna KULKOVA**

<https://orcid.org/0000-0002-1787-1406>

[mkulkowa@rambler.ru](mailto:mkulkowa@rambler.ru)

Kazan Federal University, Kazan, Russia

**Kirsten REUTER**

<https://orcid.org/0000-0003-1204-9373>

[reuterkirsten@googlemail.com](mailto:reuterkirsten@googlemail.com)

Kazan Federal University, Kazan, Russia

Este trabajo está depositado en Zenodo:  
DOI: <http://doi.org/10.5281/zenodo.4009657>

### RESUMEN

La introducción de tecnologías de información en el proceso educativo moderno es una parte esencial del trabajo exitoso del maestro y la garantía del logro de los resultados del aprendizaje. El corpus de texto es una de las herramientas de recursos modernas para la creación de tareas de aprendizaje de idiomas para estudiantes. Además, presentamos materiales didácticos para estudiar la homonimia y la polisemia de las palabras. Ofrecemos dos formas diferentes de organización del proceso de aprendizaje con énfasis en la tecnología de corpus: el maestro presenta tareas selectas a estudiantes que trabajan independientemente con textos de un corpus de texto, como un mini proyecto de investigación.

**Palabras clave:** Concordancia, corpus, desarrollo de habilidades léxicas, lengua extranjera, tecnología de corpus.

### 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 independently working students with texts from a text corpus, as a mini-research project.

**Keywords:** Concordance, corpus, corpus technology, development of lexical skills, foreign language.

Recibido: 03-08-2020 • Aceptado: 10-09-2020



## 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 (Mukhamadiarova et al.: 2017, pp.327-334; Khakimullina et al.: 2018, pp.246-253; Nurullina et al.: 2018, pp.461-468). 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 (Гальскова & Гез: 2009). «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 et al.: 2016, pp.273-283).

The development of corpus linguistics allows a new and different understanding of language and learning foreign languages because text corpora 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 (Tamás: 2014; Hausmann: 2003, pp. 309–335).

Text corpora 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 (Гвишиани: 1979; Тер-Минасова: 2007).

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, pp.75–106).

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-occurrence of the particular word (Алексеева: 2011). Corpus’ linguistic scientists refer to the term “collocation” in relation to statistically stable linguistic units – a combination of two or more words (Scott & Tribble: 2006, p.200).

## METHODS

In this research, we use the following text corpora: 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 et al.: 2017, pp. 12-28).

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, pp. 104-115). 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”:

- 1) Staatsaufbau, geografische Lage Deutschlands – Government system, Geography of Germany;
- 2) Politik – Political system;
- 3) Massenmedien – Social media;
- 4) Kultur Deutschlands – German culture.

## RESULTS

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

- 1) assignments for building lexical/collocational word profiles
- 2) assignments for developing background and linguistic-cultural knowledge
- 3) 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:

- 1) assignments with the use of the concordance
- 2) assignments with the use of statistical frequency and “clouds of words.”
- 3) 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:

- 1) students working on assignments created by a teacher with the help of the text corpus
- 2) 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:

- 1) concordance for observation of the collocational profile of the word
- 2) keywords or “clouds of words” for specific context and frequency
- 3) dictionary

### 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).

Intertítulos o apartados

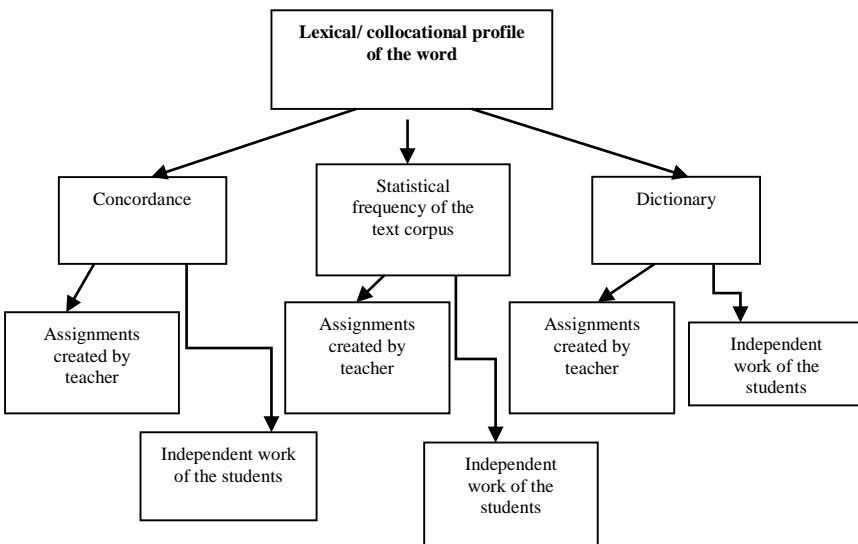


Figure 1. Classification of the assignments for the lexical / collocational profile of the word.

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. 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: 2017). 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 und 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."

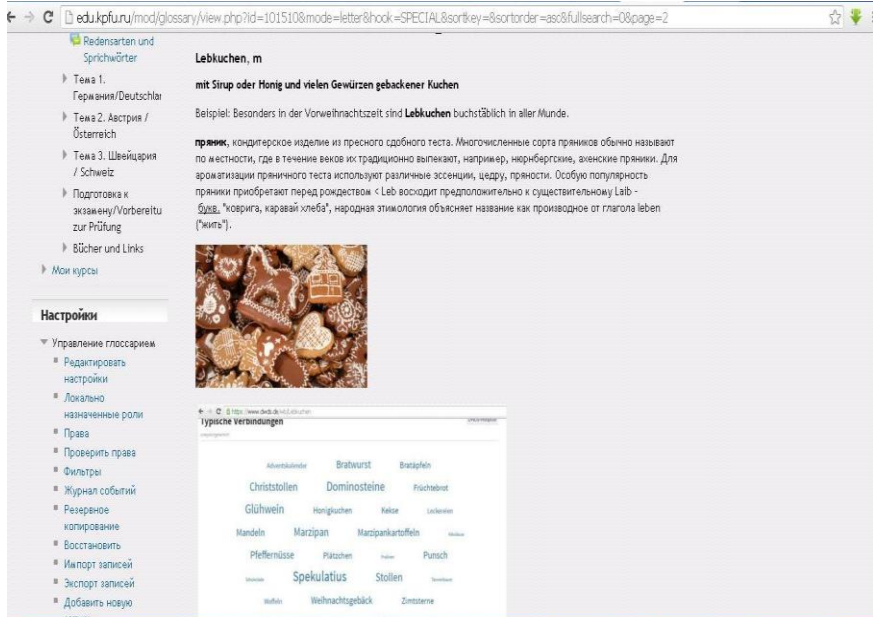


Figure 2. Lexical profile of the word Lebkuchen. Notes.

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, pp. 1-29). 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 et al.: 2017, pp.1783-1788). The purpose of creating an algorithm of research through the usage of text corpora is to teach students ways to independently employ those sources to improve their writing skills in a foreign language (Shtyrina: 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).

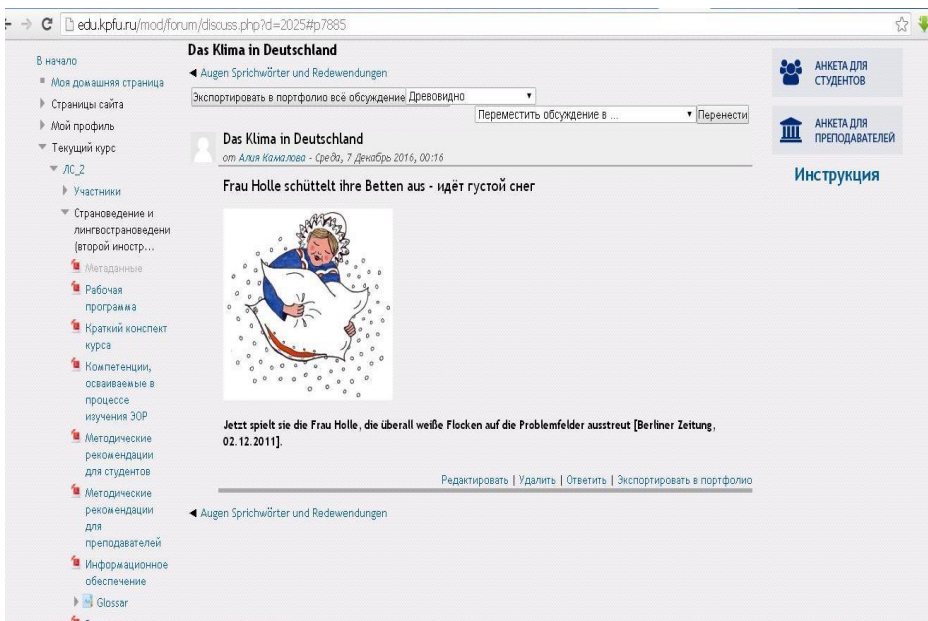


Figure 3. Forum Redensarten und Sprichwörter. Frau Holle schüttelt ihre Betten aus Notes.

## 2 The corpora's 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 et al.: 2018, pp.1118-1121).

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	

**Table 1.** Collocations “Expressing personal opinion”

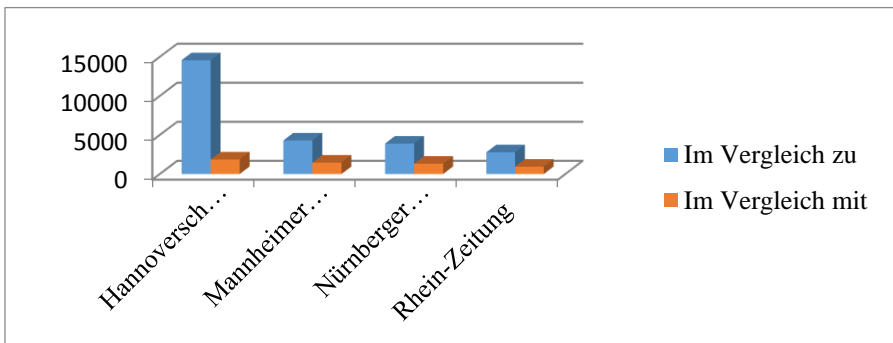
We will demonstrate the analysis of the students’ learning process based on the results of assignment 3. For this assignment, the students are divided into two groups, and each group must learn the frequency of usage of the phrases Im Vergleich zu and Im Vergleich mit. The first group must research through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

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

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

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

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, analyzing and organizing language facts. Teachers supervise the students' research and provide guidance for the students' independent learning process.

## **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*.

3 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, pp.108-115). 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, Fleischorten und gedämpftes Schweinefleisch [Die Zeit, 11.05.2006, Nr. 20].

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

4 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:

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

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

3. Bundeskanzler Schröder bezeichnete das anhaltende Umfragetief der SPD als deprimierend.

## **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.

## ACKNOWLEDGEMENTS

The work is performed according to the Russian Government Program of Competitive Growth of Kazan Federal University.

## BIBLIOGRAPHY

- BARTZ, T (2017). "Digitale Sprachressourcen im Deutschunterricht: Korpus-basierte Recherche und Analyse in der "Wörterbuchwerkstatt" Korpus-basierte ,Wörterbuchwerkstatt' im Deutschunterricht", Skiba, Wolf-Dirk & Lombardi, Alessandra (Hrsg.) (in Vorbereitung). Korpora im Sprachunterricht. Tagungsband zur gleichnamigen Sektion der XV. Internationalen Tagung der DeutschlehrerInnen, Bozen, Bozen-Bolzano University Press.
- ELGORT, I (2018). "Technology-mediated second language vocabulary development: A review of trends in research methodology". *calico journal*, 35(1), pp. 1-29.
- HAUSMANN, FJ (2003). "Was sind eigentlich Kollokationen?" In: Steyer, Kathrin (ed.): *Wortverbindungen – mehr oder weniger fest*. Berlin, de Gruyter, pp. 309–335.
- KHAKIMULLINA, RR, AYUPOVA, RA & ZAKIROVA, LR (2018). "Maria Luisa Ortiz Alvarez, "Internet technologies in teaching foreign languages", *Revista Publicando*", 5(17), pp. 246-253.
- KONOPATSKAYA, EA, YARMAKEEV, IE & PIMENOVA, TS (2017). "Teaching Idiomatic English in ESP Class", *Quid-Investigacion Ciencia y Tecnologia*, vol. 1, Special Issue, pp. 1783-1788.
- MEAD, P & DOECKE, B (2020). "Pedagogy. In *Oxford Research Encyclopedia of Literature*".
- MIFTAKHOVA, AN, BOCHINA, TG & ZHURAVLEVA, YA (2018). "Gender interpretation of Russian lexeme баба/baba in internet discourse", *Herald NAMSCA* 3, pp.1118-1121.
- MIRIMANOVA, MS (2002). "Tolerance as a problem of education", *Personality development*, 2, pp. 104-115.
- MUKHAMADIAROVA, AF, KULKOVA, MA & FIRSOVA, EV (2017). "Application of Corpus Technologies in Teaching German Vocabulary", *Astra Salvensis*, Supplement, 1, pp. 327-334.
- NURULLINA, GM, MURAVIYOV, AF, MARTYANOVA, AA & YARMAKEEV, IE (2018). "Project technology in the development of communicative competence in schoolchildren: Extracurricular classes of Russian language", *Cypriot Journal of Educational Sciences*, 13(4), pp. 461-468.
- SAKAEVA, L (2018). "Translation features of author neologisms on the example of Modern English prose". *Revista San Gregorio*, (23), pp.108-115.
- SCOTT, M & TRIBBLE, C (2006). "Textual Patterns: key words and corpus analysis in language education: *Studies in Corpus Linguistics*", Amsterdam/Philadelphia: John Benjamins, p.200.



SHTYRLINA, EG (2017). "Concept as a linguistic guideline in teaching Russian as a foreign language", *Modern Journal of Language Teaching Methods*, 7(12), pp. 88-94.

SINCLAIR, J (1996). "The search for units of meaning", *Textus*, 9 (1), pp. 75–106.

TAMÁS, K (2014). "Kaffee oder Tee? Textkorpusbasierte Kollokationsforschung und ihre Realisierung in der Lernerlexikographie", Bassola, Péter et al. (Hrsg.): *Zugänge zum Text*. Frankfurt: Peter Lang., pp. 217-245.

VARLAMOVA, EV, TULUSINA, EA, ZARIPOVA, ZM & GATAULLINA, VL (2016). "Lexical, semantic and culturological approaches to the teaching of a second language", *Analele Universitatii din Craiova - Seria Stiinte Filologice, Lingvistica*, 38(1-2), pp. 273-283.

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(1), pp. 12-28.

АЛЕКСЕЕВА, ЛБ (2011). "Методика формирования коллокационной компетенции у студентов неязыковых факультетов в процессе обучения английской научной речи: Дис... канд. пед. наук", РГПУ им. А.И. Герцена, СПб. 210 (In Russian).

ГАЛЬСКОВА, НД & ГЕЗ, НИ (2009). "Теория и методика обучения иностранным языкам: лингводидактика и методика: Учеб. пособие для студ. лингв. ун-тов и ф-тов ин.яз. высш. пед. учеб. заведений", 6-е изд., стерео-типн., М., Изд. центр «Академия», 336 с (In Russian).

ГВИШИАНИ, НБ (1979). "Полифункциональные слова в языке и речи: Учеб. пособие", М., Высш. Школа, 200 с (In Russian).

ТЕР-МИНАСОВА, СГ (2007). "Словосочетание в научно-лингвистическом и дидактическом аспектах", 3-е изд., М., Изд-во ЛКИ, 143 с (In Russian).

## **BIODATA**

**A.F MUKHAMADIAROVA:** Born in 1991. Candidate of Philology. In 2017 she graduated from the Institute of Philology and Intercultural Communication of KFU, specialization: Pedagogical education. Qualification: Master. In 2019, she defended her thesis on the topic "Comparative analysis of phraseological and paremiological units with the coloronim component (based on German, Russian and Tatar languages)". Senior Lecturer, Department of Theory and Practice of Teaching Foreign Languages, IFMK KFU. Research interests: phraseology, corpus linguistics, methods of teaching foreign languages.

**L.F CASERTA:** Born in 1973. In 1996 she graduated from the Faculty of Russian Philology of Kazan State Pedagogical University. Direction (specialty) Russian language, literature, world art culture. The theme of the thesis: "Problems of literary and artistic criticism on the pages of the magazine "World of Art". Senior Lecturer, Department of English, Literature and World Languages, College of Arts and Sciences, V. Ferris University. Research interests: linguistics and methods of teaching foreign languages.

**M.A KULKOVA:** Born in 1980. Doctor of Philology. In 2002 she graduated from Kazan Federal University (formerly Kazan State Pedagogical University), specialty "Philology", qualification "Teacher of German and English". In 2011, she defended her thesis for the degree of Doctor of Philological Sciences on the topic "Cognitive and semantic space of folk signs." Professor, Department of Theory and Practice of Teaching Foreign Languages, IFMK KFU. Research interests: pragmalinguistics, paremiology, cognitive linguistics, corpus linguistics, German, Russian.

**K REUTER:** Born in 1990. In 2018, she graduated from the Institute of German Studies, Contemporary Literature and Linguistics, University of Otto von Guericke in Magdeburg. The theme of the master's thesis: "The depiction of slowness in literature, in art and in music based on the novel Detection of slowness" (Die Darstellung der Langsamkeit in Literatur, Kunst und Musik auf Grundlage des Romans "Die Entdeckung der Langsamkeit" von Sten Nadolny). Lecturer, Department of Theory and Practice of Teaching Foreign Languages, IFMK KFU. Research interests: modern German literature, German media words, linguistics.