

Федеральное государственное бюджетное образовательное учреждение высшего образования «Тамбовский государственный университет имени Г.Р. Державина»
Педагогический институт
Кафедра лингвистики и гуманитарно-педагогического образования

УТВЕРЖДАЮ:
Директор института



Т. И. Гущина
«22» июня 2023 г.

РАБОЧАЯ ПРОГРАММА

по дисциплине Б1.В.8 Корпусная лингвистика и обучение иностранному языку

Направление подготовки/специальность: 44.03.01 - Педагогическое образование

Профиль/направленность/специализация: Английский язык

Уровень высшего образования: бакалавриат

Квалификация: Бакалавр

год набора: 2020

Тамбов, 2023

Автор программы:

Кандидат педагогических наук, Золотов Питирим Юрьевич

Рабочая программа составлена в соответствии с ФГОС ВО по направлению подготовки 44.03.01 - Педагогическое образование (уровень бакалавриата) (приказ Министерства образования и науки РФ от «22» февраля 2018 г. № 121).

Рабочая программа принята на заседании Кафедры лингвистики и гуманитарно-педагогического образования «29» июня 2022 г. Протокол № 10

Рассмотрена и одобрена на заседании Ученого совета Педагогического института, Протокол от «22» июня 2023 г. № 9.

СОДЕРЖАНИЕ

| | |
|---|----|
| 1. Цели и задачи дисциплины..... | 4 |
| 2. Место дисциплины в структуре ОП Бакалавриата..... | 4 |
| 3. Объем и содержание дисциплины..... | 4 |
| 4. Контроль знаний обучающихся и типовые оценочные средства..... | 12 |
| 5. Методические указания для обучающихся по освоению дисциплины (модуля)..... | 16 |
| 6. Учебно-методическое и информационное обеспечение дисциплины..... | 18 |
| 7. Материально-техническое обеспечение дисциплины, программное обеспечение, профессиональные базы данных и информационные справочные системы..... | 18 |

1. Цели и задачи дисциплины

1.1 Цель дисциплины – формирование компетенций:

ПК-4 Способен использовать иноязычные источники информации, инструменты перевода, произношения на основе информационно-коммуникационных технологий в образовательной и методической деятельности

1.2 Типы задач профессиональной деятельности, к которым готовятся обучающиеся в рамках освоения дисциплины:

- педагогический

1.3 Дисциплина ориентирована на подготовку обучающихся к профессиональной деятельности в сфере: 01 Образование и наука (в сфере начального общего, основного общего, среднего общего образования)

1.4 В результате освоения дисциплины у обучающихся должны быть сформированы:

| Обобщенные трудовые функции / трудовые функции / трудовые или профессиональные действия (при наличии профстандарта) | Код и наименование компетенции ФГОС ВО, необходимой для формирования трудового или профессионального действия | Индикаторы достижения компетенций |
|---|--|---|
| - В/03.6 Педагогическая деятельность по реализации программ основного и среднего общего образования | ПК-4 Способен использовать иноязычные источники информации, инструменты перевода, произношения на основе информационно-коммуникационных технологий в образовательной и методической деятельности | Обращается к иноязычным печатным и аудиовизуальным источникам информации для подбора учебных материалов и их методической адаптации с использованием электронных корпусов и инструментов корпусной лингвистики, включая инструменты перевода и произношения |

1.5 Согласование междисциплинарных связей дисциплин, обеспечивающих освоение компетенций:

ПК-4 Способен использовать иноязычные источники информации, инструменты перевода, произношения на основе информационно-коммуникационных технологий в образовательной и методической деятельности

| № п/п | Наименование дисциплин, определяющих междисциплинарные связи | Форма обучения |
|-------|--|-----------------|
| | | Очная (семестр) |
| | | 3 |
| 1 | Основы перевода | + |

2. Место дисциплины в структуре ОП бакалавриата:

Дисциплина «Корпусная лингвистика и обучение иностранному языку» относится к части, формируемой участниками образовательных отношений, учебного плана ОП по направлению подготовки 44.03.01 - Педагогическое образование.

Дисциплина «Корпусная лингвистика и обучение иностранному языку» изучается в 7 семестре.

3. Объем и содержание дисциплины

3.1.Объем дисциплины: 3 з.е.

Очная: 3 з.е.

| Вид учебной работы | Очная (всего часов) |
|--------------------------------------|------------------------|
| Общая трудоёмкость дисциплины | 108 |
| Контактная работа | 56 |
| Лекции (Лекции) | 16 |
| Лабораторные (Лаб. раб.) | 40 |
| Самостоятельная работа (СР) | 52 |
| Зачет | - |

3.2.Содержание курса:

| № темы | Название раздела/темы | Вид учебной работы, час. | | | Формы текущего контроля |
|-----------|---|-----------------------------|------------------|----|----------------------------|
| | | Лек ции | Лаб · раб. | СР | |
| | | О | О | О | |
| 7 семестр | | | | | |
| 1 | Corpus linguistics as a linguistic discipline | 1 | 4 | 3 | Опрос |
| 2 | Counting, calculating and annotating | 1 | 4 | 5 | Опрос |
| 3 | Looking for lexis | 1 | 4 | 5 | Опрос |
| 4 | Checking collocations and colligations | 1 | 4 | 5 | Контрольная работа |
| 5 | Finding phrases | 2 | 4 | 5 | Опрос |
| 6 | Corpus study of metaphor and metonymy | 2 | 4 | 5 | Контрольная работа |
| 7 | Corpus study of grammar | 2 | 4 | 6 | Опрос |
| 8 | Corpus study of gender aspects of language | 2 | 4 | 6 | Опрос |
| 9 | Corpus study of language change | 2 | 4 | 6 | Опрос |
| 10 | Corpus linguistics in cyberspace | 2 | 4 | 6 | Опрос |

Тема 1. Corpus linguistics as a linguistic discipline (ПК-4)

Лекция.

Introducing corpus linguistics. The professor's shoeboxes. The tape-recorder and the computer. What can we get out of corpora? Concordances. Frequency. Criticism from armchair linguists. Types of corpora. Spoken corpora. General corpora. The Bank of English (BoE). The British National Corpus (BNC). The American National Corpus (ANC). Corpus of Contemporary American English (COCA). The International Corpus of English (ICE). Specialised corpora. Parallel and multilingual corpora. Dictionaries as corpora. Text archives as corpora. The web as corpus. Very large and quite small corpora.

Лабораторные работы.

1. Go to <http://corpus.byu.edu/bnc/>. NB: during the session you may be asked to register as a user. Do this – it is free of charge. Make a note of your password. Read the site Introduction and then click on “five minute guided tour” to learn more about the corpus.
2. Click on the word mysterious in the guided tour. It will then occur under “Search string” in the search panel to the left on the screen. Under “Display”, the “Chart” button should be marked. Under “Section”, “Ignore” should be marked. Now press the “Search” button. The results per section of the corpus will be displayed in a bar chart in the righthand part of the screen. The chart shows that mysterious occurs most frequently in Fiction, with 27.4 tokens (instances) per 1 million words, followed by Miscellaneous, Newspapers, Academic and Spoken.
3. Now click on the bar for Fiction in the chart. The first 100 concordance lines with mysterious will be displayed and you will be able to study the contexts in which the word occurs, e.g. what kind of nouns it modifies.
4. Go back to the search panel to the left (it may be partially hidden) and mark “List” instead of “Chart”. Click on “Search” again. Now the total number of occurrences is reported in the right-hand panel. If you click on the word mysterious you get the first 100 concordance lines regardless of subcorpus.
5. Look at this list of five adjectives: big, hot, superfluous, supreme, tragic. Try to guess which is most frequent, second most frequent etc. in the corpus, and then in which subcorpus they occur most frequently. Then search for each of the words in the manner described above and for each of them note down the total frequency and which subcorpus it occurs most frequently in. Don't forget to look at the concordance lines to get an indication of how the words are used.
6. Try searching for some other words that you are curious about.

Задания для самостоятельной работы.

1. What are the main advantages and disadvantages of using corpora for linguistic research?
2. What are a concordance and a concordance line? Why are concordances useful?
3. What are the extra complications with corpora of spoken language?
4. Why is it not always possible to compare data from different corpora?
5. What is the difference between corpora proper and text archives?

Тема 2. Counting, calculating and annotating (ПК-4)

Лекция.

Qualitative and quantitative method. It is common to distinguish two main types of scientific method in linguistics (and in general): qualitative and quantitative. If one uses a qualitative method you make close analyses of, for example, individual texts or grammatical constructions that you know from introspection or get from other sources. Quantitative methods as the name implies, count things and use frequencies and percentages to describe language and to formulate hypotheses and theories. Frequency. Comparing frequencies. Distribution in the corpus. Using percentages and normalising. Representativity. Corpus annotation. Part-of-speech tagging. Parsing. Other types of corpus annotation.

Лабораторные работы.

1. Go to <http://www.american corpus.org/>. NB: during the session you may be asked to register as a user. Do this – it is free of charge. Choose the same password as you chose for BYU-BNC.
2. Find out the frequencies in COCA for the top fifteen noun lemmas in Table 2.3 in the book. You will have to search for one word at a time, using the following method.

3. In the search panel to the left, click “List” and type in the following search string: [time].[nn*]. The square brackets around the search word will retrieve all forms of the lemma (in this case time and times). The second part of the string, [nn*], ensures that only nouns will be retrieved. Note that there are no spaces between the brackets and the period in the middle. Since in this task we are only interested in the relative rank of the nouns, there is no need to calculate tokens per million words: absolute frequencies will do.
4. Go through the procedure in (3) for each of the fifteen top words in Table 2.3 and note down the frequencies. Then order the nouns according to frequency and compare your list based on COCA with the list based on the BNC. Are there large or small differences? What conclusions can you draw from the results?

Задания для самостоятельной работы.

1. What is the difference between qualitative and quantitative methods? How can quantitative methods be combined with qualitative analysis?
2. What is lemmatising? How does it influence frequency counts?
3. Why is it important to normalise frequencies when you compare results from different corpora?
4. What arguments are relevant when we discuss whether a corpus is representative or not?
5. What are the main types of corpus annotation? Why is it important to know the principles behind the tagging of a corpus that you are using?

Тема 3. Looking for lexis (ПК-4)

Лекция.

The role of the lexicon in language. How lexicographers use corpora. The meaning of words. Semantic preference, semantic prosody and evaluation. Collocation is the relation between a word and individual wordforms which co-occur frequently with it. Colligation is the relation between a word and grammatical categories which co-occur frequently with it. Semantic preference is the relation between a word and semantically related words in a lexical field. Semantic prosody is the discourse function of the word: it describes the speaker’s communicative purpose. How words change in frequency over time. How words spread between varieties of English. How authors use words.

Лабораторные работы.

1. Go to <http://corpus.byu.edu/time/>. NB: during the session you may be asked to register as a user. Do this – it is free of charge. Choose the same password as you chose for BYUBNC and COCA.
2. In the search panel, click the “Chart” button and type in the search word regime. Then click “Search”.
3. The results will be displayed in a bar chart showing the frequency distribution per decade. Click in the bar for the 1920s to see the concordance lines. Study the first twenty lines carefully and decide for each instance of regime whether it has negative or positive evaluation or seems to be neutral. If you need to see more context, click on the date to the left of the concordance line. To get back to the concordance from the expanded concordance line, click twice on the return arrow in your web browser. Write down your results.
4. Repeat the procedure for the 2000s.
5. Log in to the Time Corpus. In the search panel, click the “Chart” button. Then type in the search words motor car. The frequency distribution per decade will be given in a bar chart. You can copy and paste the whole chart into a word processing document, or else write down the frequencies per million words. Then repeat the procedure for the following words, one at a time: car, automobile, auto. Report on your findings by making a diagram like Figure 3.6 in the book, or by showing the four bar charts.
6. Investigate at least one other set of words with similar meaning from a field that interests you, for instance sports, politics, fashion or arts, or study at least four single words.

Задания для самостоятельной работы.

1. What are the main advantages of corpus-based lexicography compared to the old way of collecting examples by hand?
2. What does “semantic prosody” mean? Do you know other examples of words which have hidden meanings?
3. What does “the colloquialisation of English” mean? Can you think of any further examples?

4. The chapter gives some examples of how corpus methods can be used in literary studies. Do you think this is a good way to study literature? Are there other aspects of literature that could be investigated by means of corpora?

Тема 4. Checking collocations and colligations (ПК-4)

Лекция.

Two types of collocations. A collocation is a succession of two or more words that must be learnt as an integral whole and not pieced together from its component parts. Collocations in a window. Adjacent collocations. Colligations. It is usually taken to refer to the relation between a node word and grammatical categories such as a preposition or a wh-clause which co-occur frequently with it, or, expressed differently, the grammatical constructions in which a word frequently occurs.

Лабораторные работы.

1. Log in to COCA. In the search panel, click the “List” button, set the “SORT” option on “Frequency” and type in the search word city. Under “POS LIST” (short for part-of speechlist), scroll down to “verb.ALL” and choose this. Now [v*] will appear in the context field. Under CONTEXT, change the window size to -5 +5. Then press “Search”. Study the first thirty verbs. What kind of verbs are they?
2. Repeat the procedure in (1), but set the “SORT” option on “Relevance”. This will give you a list based on mutual information. Compare the verbs you find here with those you found in (1).
3. Investigate a few more words from different word classes in the same manner. Try using different size windows and different spans to the left and to the right and see what changes.

Задания для самостоятельной работы.

1. What are the two main types of collocations mentioned in the chapter and what is the difference between them? Which do you think is most useful for language learners? For linguists?
2. What is meant by a window and how can it be described?
3. How can a word be a collocate of itself?
4. Lists of collocates can be based on absolute frequency or one of a number of statistical measures. What is the difference between the kind of collocates that appear on these lists?
5. What are colligations?

Тема 5. Finding phrases (ПК-4)

Лекция.

Phraseology. Idioms. An idiom is a fixed or semi-fixed expression whose meaning cannot be deduced from its parts. Recurrent phrases. Recurrent phrases are those which were at least two words long and occurred more than once The Phrases in English (PIE) database. This database contains all n-grams (identical strings of words) with a length between one and eight words which occur three times or more in the BNC. A literary application: Dickens’ recurrent long phrases.

Лабораторные работы.

1. Open Google on the Internet. Click on “Advanced search”. Change “10 results” to “30 results”. In the box for “exact phrase” type in: too many * spoil the. Apart from too many cooks spoil the broth, this will retrieve all examples with small modifications like too many chefs spoil the soup and more imaginative variants like Too many clicks spoil the sale and Too many superlatives spoil the telling. Finally, in the “Domain” box, type in “uk” to get examples from British webpages only. Then click on “Search”. Among the first thirty examples, how many were manipulated? How were they manipulated?
2. Go back to the search page and in the search box “without any of these words” add: cook, chef, broth, soup. Then click search again. This will give you only manipulated examples.
3. Try the searches described in (1) and (2) in some other domain, perhaps your home country if it is not the UK. You will find domain abbreviations in Table 10.2 in the book.
4. Investigate some other idioms or proverbs of your own choice. You will see that the extent to which they are manipulated varies considerably.

Задания для самостоятельной работы.

1. What is the difference between idioms like a red herring and recurrent phrases like at the moment? What is their importance for language learning?
2. Idioms like A bird in the hand is worth two in the bush are often used in an abbreviated or manipulated form like It is a case of the 'bird in the hand' (books.google.se/books?isbn=0521589592. . .) and A fish in the hand is worth two on the Net (www.uhh.hawaii.edu/academics/hohonu/writing.php?id=115). What does this imply about how these idioms are stored in the mental lexicon? Can you think of other idioms that are manipulated in this way?
3. 'N-gram' is a purely technical term denoting identical strings of words. PIE extracts n-grams (2-grams up to 8-grams) from the BNC.
4. To what extent are such automatically extracted n-grams interesting from a linguistic point of view?

Тема 6. Corpus study of metaphor and metonymy (ПК-4)

Лекция.

Metaphor. Distinguishing metaphors from non-metaphors. Simile. Conceptual metaphors. Metonymy. The term 'metonymy' is far less well known than 'metaphor' and is not used in everyday language as 'metaphor' is. The semantic process involved is related, but there is one major difference: in typical metaphors, the mapping is from one domain to a different domain, while in metonymy the mapping occurs within the same domain. Starting with the source domain. Starting with the target domain. Starting from manual analysis.

Лабораторные работы.

1. Log in to COCA. Under "Display", click on the "List" button. Type in the search string: life is a * (note the asterisk) and click on "Search".
2. Analyse the first 100 examples given in the right-hand part of the window and try to organise them into semantic groups. How many fit the conceptual metaphors suggested by Lakoff and Johnson? Can you suggest some more conceptual metaphors?
3. Now turn to 'love'. Try searching for love is a *. You will find that many of the words following love is a are adjectives, so you cannot see what metaphor is used. To solve this problem, search for the following string instead: love is a [j*] [n*]. (You can either type in [j*] and [n*] or go down to "POS LIST", scroll down and click on the tags you want one at a time.) This will retrieve all examples like love is a winding way and love is a temporary derangement.
4. Analyse the love examples in the same way as you did the life ones.

Задания для самостоятельной работы.

1. What is the difference between metaphor, metonymy and simile?
2. Try to come up with your own examples. Why are metaphors considered to be more powerful than similes?
3. What is the main idea behind conceptual metaphors?
4. What different ways of using corpora in the study of metaphors were introduced in this chapter? Which method do you think seems most viable?
5. It says in the summary that studying metaphors can tell us something about how our minds work. Do you agree? If yes, how? If no, why not?

Тема 7. Corpus study of grammar (ПК-4)

Лекция.

Who and whom. Get-passives. Adjective complementation. Prepositional gerund or directly linked gerund. This topic is devoted to morphology, as in the case of who/whom, but also for grammatical constructions which are distinguished by the use of different lexical items, like be- and get-passives. And there are many other areas of the interface between the lexicon and grammar where individual lexical items exert a considerable influence over which grammatical constructions are chosen, such as adjective complementation (accustomed to) and noun complementation (difficulty (in) finding).

Лабораторные работы.

1. Log in to BYU-BNC. Set the “Display” on “List”. Go to “POS LIST”, scroll down to “pron.ALL” and choose that. [pn*] should now appear in the search box. Type in the word shall and press “Search”. Write down the total figure given at the bottom of the list of phrases on the right. Then repeat for the word will and write down the total figure for that.
2. Log in to COCA (you should be able to stay logged in to BYU-BNC in the background). Repeat the settings described in (i) but in addition, under “Section”, scroll down to 1990–1994 and choose this. The two corpora will be relatively comparable in this limited time period. Search for [pn*] shall and [pn*] will in turn and write down the total figures. Now calculate the percentage of shall in the BNC in the following way. Add the figures for shall and will. Then divide the number of shall by the total and multiply by 100. Note down the percentage.
3. Repeat the procedure in (iii) for COCA. Compare the two percentage figures. If you have done all the operations correctly, there should be a considerable difference between the two corpora in the proportion of shall.
4. Now return to COCA. Keep all the settings except “List”; change this to “Chart”. Search for [pn*] shall again. First study the left-hand part of the chart. In which genres is shall predominantly used in American English? Then look at the right-hand part of the chart. Is there any tendency to a change in frequency over time?
5. Next return to BYU-BNC and repeat the procedure in (4). Can you see any major difference between the corpora in the distribution between the subcorpora?

Задания для самостоятельной работы.

1. The issue of the possible disappearance of the form whom is rather complicated, since it is used in several different constructions. In which of these constructions is it most likely to survive? What other factors are relevant if you want to study the distribution of who and whom?
2. In Table 7.3, two figures are given for COCA: the frequency index (i.e. the proportion of get-passives for each verb) and the absolute frequencies. Why is it important to give the absolute figures as well?
3. Why did Rohdenburg hypothesise that the preposition would be used more often after the plural difficulties than after the singular difficulty in the construction have difficulty/difficulties (in) finding? What did our additional corpus searches find?
4. In the discussion of Tables 7.7 and 7.8, the different usage with Latinate and Germanic words is mentioned. How does the argument go, and can you find any alternative way of interpreting the data?

Тема 8. Corpus study of gender aspects of language (ИК-4)

Лекция.

Language and gender has become a lively field of investigation since the 1970s, inspired by the feminist movement and the feeling that the inequality of men and women in society is mirrored in language and also, conversely, that oppressive societal structures are preserved by language. Some even argue that inequality is created and perpetuated by the language. There are several ways of studying gender role in language. One prominent direction deals with discourse: how men and women interact by means of language in different situations like everyday conversation or business meetings. Who speaks most? Who interrupts most frequently? Who gives more supportive feedback? Who laughs most often? This is normally studied by close analysis of recorded or transcribed speech but is less amenable to automatic computer searches. Another approach is to look at how men and women are referred to in texts, and this is what the next section will be about.

Лабораторные работы.

1. Go to <http://quod.lib.umich.edu/m/micase/>. Click on “General Information about MICASE” at the bottom of the page and read especially “Statistical Overview of Speakers and Speech Events” to get an overview of the contents of this corpus.
2. Click on “Search MICASE” (you may have to go back to the first page to find this link).
3. In the search form, click on “Female” under search attributes and type in sorry in the search box. Click on “Submit search”.
4. Write down the frequency of sorry, which is given at the top of the results page. Note that you immediately get the concordance and can see how the keyword is used.

5. Repeat the procedure in (3) and (4) for the speaker category “Male”.
6. Repeat the procedures in (3)–(5) for pardon and excuse me.
7. The statistical overview of the corpus states that the gender proportions in the corpus are 54 per cent female, 46 per cent male. Compare the frequencies of sorry, pardon and excuse me with those figures. Can you draw any conclusions about male and female language from this? You may want to go back to the concordances to look more closely at how the keywords are used in the discourse situation.

Задания для самостоятельной работы.

1. What kind of information is crucial if you want to use corpora for sociolinguistic studies? What corpora and text archives can be useful for corpus-based gender studies?
2. The chapter describes how some occupational terms have become less sexist over the years. Can you think of other occupational terms which have changed in the same way? Do you think there is a difference in these changes between different genres and between spoken and written English?
3. The results in Table 8.13 are based on search strings which retrieve examples of what men are reported as doing to women in the corpus and vice versa, which perhaps says something about the relation between the sexes. Do you think there would be similar results if the search strings had been he + verb + him and she + verb + her?
4. Have you yourself noticed differences in male and female speech among your friends at school or work?

Тема 9. Corpus study of language change (ПК-4)

Лекция.

In this topic we start out by looking at changes in the use of adverbial likely which seem to be going on right now. We then go back 1,000 years to follow the grammaticalisation of beside(s) through the centuries right up to the present day by means of a number of different corpora. We also saw how the OED online could be used as a corpus, and how a corpus of medieval letters could give insights into the influence of social stratification on language use and language change.

Лабораторные работы.

1. Log in to the Time Corpus. Under “Display”, click on “Chart”. In the search box, type in help to. Then go down to “POS LIST”, choose “verb.ALL” and click on “Search”. Copy the chart into a word processing document or note down the changes in frequency.
2. Repeat (1), but type in just help and choose “verb.INF” under “POS LIST”. Click on “Search” and save the chart or note the figures.
3. Compare and comment on the development of the two constructions, as in help to explain vs. help explain.
4. Try to think of other verbs where there is similar variation and investigate whether there is any ongoing change.

Задания для самостоятельной работы.

1. The introduction to this chapter claims that language changes constantly. Can you think of any further examples of this, in English or any other language? What consequences does this have for synchronic language studies?
2. What does ‘change in apparent time’ mean? Why are results showing change in apparent time more uncertain than results showing change in real time?
3. Some changes originate in informal spoken language (‘change from below’) and other changes are initiated in more formal, written language (‘change from above’). What kind of change is the spread of adverbial likely in the word order will likely in BrE, according to the data presented in this chapter?
4. In what ways do corpus linguists, and especially historical corpus linguists, take spelling variants into account?

Тема 10. Corpus linguistics in cyberspace (ПК-4)

Лекция.

In this topic we have tried to grasp the enormous size of the web and to see how it can be harnessed for linguistic research. We saw the motivations for using it: it is large, it contains many geographical varieties and genres that may be hard to get at elsewhere, it is up to date, it is available and it is free. We also saw the drawbacks: it is messy, it contains lots of dubious material, it is unstable, it can be hard to search with sufficient precision. Nevertheless, more and more linguistic research is carried out on the web.

Лабораторные работы.

1. Go to Google, click on “Advanced search” and type in different from in the
2. search box for “exact phrase”. Make the search and note the frequency figure given.
3. Repeat for different than and different to.
4. Compare your figures with those in Table 10.1. Do the figures indicate that the web has continued to grow? Do the phrases seem to increase at the same rate?
5. Go to Google, click on “Advanced search” and type in realise normalize globalise in the search box for “any of these words”. Under “Domain”, type “uk”.
6. Make the search and note down the frequency.
7. Repeat (4) for the following domains: edu, au, nz, za, hk, in and ph.
8. Repeat (4) and (5), but with the search words spelt realize normalize globalize.
9. Fill in the figures in a table like Table 10.3 in the book. Can you see any patterns in the distribution of these spellings?

Задания для самостоятельной работы.

1. What are the main reasons why linguists use the web as corpus? Can you think of reasons other than those mentioned in this chapter?
2. What are the main problems connected with using the web as corpus?
3. How do interfaces like WebCorp work? What are the advantages and what are the drawbacks?
4. How can regional variation be studied on the web? What are the complications?
5. What kind of corpora can be compiled by means of the web? Do you think that traditional corpus-making will be made obsolete by the web?

4. Контроль знаний обучающихся и типовые оценочные средства

4.1. Распределение баллов:

7 семестр

- текущий контроль – 80 баллов
- контрольные срезы – 2 среза по 10 баллов каждый
- премиальные баллы – 20 баллов

Распределение баллов по заданиям:

| № темы | Название темы / вид учебной работы | Формы текущего контроля / срезы | Мах. кол-во баллов | Методика проведения занятия и оценки |
|--------|---|---------------------------------|--------------------|--|
| 1. | Corpus linguistics as a linguistic discipline | Опрос | 10 | 0-10 в соответствии с процентов правильных ответов. |
| 2. | Counting, calculating and annotating | Опрос | 10 | 0-10 баллов в соответствии с процентом правильных ответов. |
| 3. | Looking for lexis | Опрос | 10 | 0-5 баллов в зависимости от качества ответа. |
| 4. | Checking collocations and colligations | Контрольная работа | 10 | 0-10 баллов в соответствии с процентом правильных ответов. |

| | | | | |
|-----|--|---|-----|--|
| 5. | Finding phrases | Опрос | 10 | 0-10 баллов в соответствии с процентом правильных ответов. |
| 6. | Corpus study of metaphor and metonymy | Контрольная работа(контрольный срез) | 10 | 0-10 баллов в соответствии с процентом выполненных заданий. |
| 7. | Corpus study of grammar | Опрос | 10 | 0-10 в соответствии с процентом правильных ответов. |
| 8. | Corpus study of gender aspects of language | Опрос | 10 | 0-10 баллов в соответствии с процентом правильных ответов. |
| 9. | Corpus study of language change | Опрос | 10 | 0-10 баллов в соответствии с процентом правильных ответов. |
| 10. | Corpus linguistics in cyberspace | Опрос(контрольный срез) | 10 | 0-10 баллов в соответствии с процентом правильных ответов. |
| 11. | Премиальные баллы | | 20 | Дополнительные баллы могут быть зачислены за: - активность на практических занятиях (10 баллов) - написанные научные статьи или участие в научной конференции по тематике дисциплины (10 баллов) - призовые места в конкурсах и олимпиадах по тематике дисциплины (10 баллов) |
| 12. | Индивидуальные задания, с помощью которых можно набрать дополнительные баллы | | 50 | В случае пропусков занятий по уважительным причинам студент имеет право отработать их путем выполнения индивидуального задания, включающего теоретические вопросы, практические задания, лабораторные занятия по всем пропущенным темам, входящим в дисциплинарный модуль. |
| 13. | Итого за семестр | | 100 | |

Итоговая оценка по зачету выставляется в 100-балльной шкале и в традиционной четырехбалльной шкале. Перевод 100-балльной рейтинговой оценки по дисциплине в традиционную четырехбалльную осуществляется следующим образом:

| 100-балльная система | Традиционная система |
|----------------------|----------------------|
| 50 - 100 баллов | Зачтено |
| 0 - 49 баллов | Не зачтено |

4.2 Типовые оценочные средства текущего контроля

Контрольная работа

Тема 4. Checking collocations and colligations

1. Go to <http://corpus.byu.edu/bnc/>. NB: during the session you may be asked to register as a user. Do this – it is free of charge. Make a note of your password. Read the site Introduction and then click on “five minute guided tour” to learn more about the corpus.
2. Click on the word mysterious in the guided tour. It will then occur under “Search string” in the search panel to the left on the screen. Under “Display”, the “Chart” button should be marked. Under “Section”, “Ignore” should be marked. Now press the “Search” button. The results per section of the corpus will be displayed in a bar chart in the righthand part of the screen. The chart shows that mysterious occurs most frequently in Fiction, with 27.4 tokens (instances) per 1 million words, followed by Miscellaneous, Newspapers, Academic and Spoken.
3. Now click on the bar for Fiction in the chart. The first 100 concordance lines with mysterious will be displayed and you will be able to study the contexts in which the word occurs, e.g. what kind of nouns it modifies.

4. Go back to the search panel to the left (it may be partially hidden) and mark “List” instead of “Chart”. Click on “Search” again. Now the total number of occurrences is reported in the right-hand panel. If you click on the word mysterious you get the first 100 concordance lines regardless of subcorpus.
5. Look at this list of five adjectives: big, hot, superfluous, supreme, tragic. Try to guess which is most frequent, second most frequent etc. in the corpus, and then in which subcorpus they occur most frequently. Then search for each of the words in the manner described above and for each of them note down the total frequency and which subcorpus it occurs most frequently in. Don’t forget to look at the concordance lines to get an indication of how the words are used.
6. Try searching for some other words that you are curious about.

Тема 6. Corpus study of metaphor and metonymy

1. Give a definition of collocation.
2. Give a definition of colligation.
3. Explain the difference between tokens and wordforms.
4. Give a definition of semantic prosody.
5. Explain the difference between lemmas and lexemes.

Опрос

Тема 1. Corpus linguistics as a linguistic discipline

1. What are the main advantages and disadvantages of using corpora for linguistic research?
2. What are a concordance and a concordance line? Why are concordances useful?
3. What are the extra complications with corpora of spoken language?
4. Why is it not always possible to compare data from different corpora?
5. What is the difference between corpora proper and text archives?

Тема 2. Counting, calculating and annotating

1. What is the difference between qualitative and quantitative methods? How can quantitative methods be combined with qualitative analysis?
2. What is lemmatising? How does it influence frequency counts?
3. Why is it important to normalise frequencies when you compare results from different corpora?
4. What arguments are relevant when we discuss whether a corpus is representative or not?
5. What are the main types of corpus annotation? Why is it important to know the principles behind the tagging of a corpus that you are using?

Тема 3. Looking for lexis

1. What are the main advantages of corpus-based lexicography compared to the old way of collecting examples by hand?
2. What does “semantic prosody” mean? Do you know other examples of words which have hidden meanings?
3. What does “the colloquialisation of English” mean? Can you think of any further examples?
4. The chapter gives some examples of how corpus methods can be used in literary studies. Do you think this is a good way to study literature?

Тема 5. Finding phrases

1. What is the difference between qualitative and quantitative methods? How can quantitative methods be combined with qualitative analysis?
2. What is lemmatising? How does it influence frequency counts?
3. Why is it important to normalise frequencies when you compare results from different corpora?
4. What arguments are relevant when we discuss whether a corpus is representative or not?

5. What are the main types of corpus annotation? Why is it important to know the principles behind the tagging of a corpus that you are using?

Тема 7. Corpus study of grammar

1. The issue of the possible disappearance of the form *whom* is rather complicated, since it is used in several different constructions. In which of these constructions is it most likely to survive? What other factors are relevant if you want to study the distribution of *who* and *whom*?
2. In Table 7.3, two figures are given for COCA: the frequency index (i.e. the proportion of get-passives for each verb) and the absolute frequencies. Why is it important to give the absolute figures as well?
3. Why did Rohdenburg hypothesise that the preposition *with* would be used more often after the plural difficulties than after the singular difficulty in the construction *have difficulty/difficulties (in) finding*? What did our additional corpus searches find?
4. In the discussion of Tables 7.7 and 7.8, the different usage with Latinate and Germanic words is mentioned. How does the argument go, and can you find any alternative way of interpreting the data?

Тема 8. Corpus study of gender aspects of language

1. What kind of information is crucial if you want to use corpora for sociolinguistic studies? What corpora and text archives can be useful for corpus-based gender studies?
2. The chapter describes how some occupational terms have become less sexist over the years. Can you think of other occupational terms which have changed in the same way? Do you think there is a difference in these changes between different genres and between spoken and written English?
3. The results in Table 8.13 are based on search strings which retrieve examples of what men are reported as doing to women in the corpus and vice versa, which perhaps says something about the relation between the sexes. Do you think there would be similar results if the search strings had been *he + verb + him* and *she + verb + her*?
4. Have you yourself noticed differences in male and female speech among your friends at school or work?

Тема 9. Corpus study of language change

1. The introduction to this chapter claims that language changes constantly. Can you think of any further examples of this, in English or any other language? What consequences does this have for synchronic language studies?
2. What does 'change in apparent time' mean? Why are results showing change in apparent time more uncertain than results showing change in real time?
3. Some changes originate in informal spoken language ('change from below') and other changes are initiated in more formal, written language ('change from above'). What kind of change is the spread of adverbial likely in the word order *will likely* in BrE, according to the data presented in this chapter?
4. In what ways do corpus linguists, and especially historical corpus linguists, take spelling variants into account?

Тема 10. Corpus linguistics in cyberspace

1. What are the main reasons why linguists use the web as corpus? Can you think of reasons other than those mentioned in this chapter?
2. What are the main problems connected with using the web as corpus?
3. How do interfaces like WebCorp work? What are the advantages and what are the drawbacks?
4. How can regional variation be studied on the web? What are the complications?
5. What kind of corpora can be compiled by means of the web? Do you think that traditional corpus-making will be made obsolete by the web?

4.3 Промежуточная аттестация по дисциплине проводится в форме зачета

1. Corpus linguistics as a discipline.
2. Types of corpora.
3. Counting, calculating and annotating in corpus linguistics.
4. Collocations and colligations in corpus linguistics.
5. Corpus linguistics in cyberspace.

Типовые задания для зачета (ПК-4)

1. Log in to BYU-BNC. Set the “Display” on “List”. Go to “POS LIST”, scroll down to “pron.ALL” and choose that. [pn*] should now appear in the search box. Type in the word shall and press “Search”. Write down the total figure given at the bottom of the list of phrases on the right. Then repeat for the word will and write down the total figure for that.
2. Log in to COCA (you should be able to stay logged in to BYU-BNC in the background). Repeat the settings described in (i) but in addition, under “Section”, scroll down to 1990–1994 and choose this. The two corpora will be relatively comparable in this limited time period. Search for [pn*] shall and [pn*] will in turn and write down the total figures. Now calculate the percentage of shall in the BNC in the following way. Add the figures for shall and will. Then divide the number of shall by the total and multiply by 100. Note down the percentage.
3. Repeat the procedure in (iii) for COCA. Compare the two percentage figures. If you have done all the operations correctly, there should be a considerable difference between the two corpora in the proportion of shall.
4. Now return to COCA. Keep all the settings except “List”; change this to “Chart”. Search for [pn*] shall again. First study the left-hand part of the chart. In which genres is shall predominantly used in American English? Then look at the right-hand part of the chart. Is there any tendency to a change in frequency over time?
5. Next return to BYU-BNC and repeat the procedure in (4). Can you see any major difference between the corpora in the distribution between the subcorpora?

4.4. Шкала оценивания промежуточной аттестации

| Оценка | Компетенции | Дескрипторы (уровни) – основные признаки освоения (показатели достижения результата) |
|---------------------------------|-------------|--|
| «зачтено» (50 - 100 баллов) | ПК-4 | Демонстрирует адекватный уровень знания корпусной лингвистики: анализирует и интерпретирует данные, прослеживает междисциплинарные связи с теоретической лингвистикой и методикой преподавания иностранного языка, но не всегда уверенно. Ответ построен в целом логично, материал излагается в общем понятно, иногда не хватает аргументации. Имеются неточности в употреблении терминов. |
| «не зачтено» (0 - 49 баллов) | ПК-4 | Демонстрирует незнание корпусной лингвистики: не может анализировать и интерпретировать данные. Ответ – несвязный. Не знает терминов. |

5. Методические указания для обучающихся по освоению дисциплины (модуля)

5.1 Методические указания по организации самостоятельной работы обучающихся:

Приступая к изучению дисциплины, в первую очередь обучающимся необходимо ознакомиться содержанием рабочей программы дисциплины (РПД), которая определяет содержание, объем, а также порядок изучения и преподавания учебной дисциплины, ее раздела, части.

Для самостоятельной работы важное значение имеют разделы «Объем и содержание дисциплины», «Учебно-методическое и информационное обеспечение дисциплины» и «Материально-техническое обеспечение дисциплины, программное обеспечение, профессиональные базы данных и информационные справочные системы».

В разделе «Объем и содержание дисциплины» указываются все разделы и темы изучаемой дисциплины, а также виды занятий и планируемый объем в академических часах.

В разделе «Учебно-методическое и информационное обеспечение дисциплины» указана рекомендуемая основная и дополнительная литература.

В разделе «Материально-техническое обеспечение дисциплины, программное обеспечение, профессиональные базы данных и информационные справочные системы» содержится перечень профессиональных баз данных и информационных справочных систем, необходимых для освоения дисциплины.

5.2 Рекомендации обучающимся по работе с теоретическими материалами по дисциплине

При изучении и проработке теоретического материала необходимо:

- просмотреть еще раз презентацию лекции в системе MOODLe, повторить законспектированный на лекционном занятии материал и дополнить его с учетом рекомендованной дополнительной литературы;
- при самостоятельном изучении теоретической темы сделать конспект, используя рекомендованные в РПД источники, профессиональные базы данных и информационные справочные системы;
- ответить на вопросы для самостоятельной работы, по теме представленные в пункте 3.2 РПД.
- при подготовке к текущему контролю использовать материалы фонда оценочных средств (ФОС).

5.3 Рекомендации по работе с научной и учебной литературой

Работа с основной и дополнительной литературой является главной формой самостоятельной работы и необходима при подготовке к устному опросу на семинарских занятиях, к дебатам, тестированию, экзамену. Она включает проработку лекционного материала и рекомендованных источников и литературы по тематике лекций.

Конспект лекции должен содержать реферативную запись основных вопросов лекции, в том числе с опорой на размещенные в системе MOODLe презентации, основных источников и литературы по темам, выводы по каждому вопросу. Конспект может быть выполнен в рамках распечатки выдачи презентаций лекций или в отдельной тетради по предмету. Он должен быть аккуратным, хорошо читаемым, не содержать не относящуюся к теме информацию или рисунки.

Конспекты научной литературы при самостоятельной подготовке к занятиям должны содержать ответы на каждый поставленный в теме вопрос, иметь ссылку на источник информации с обязательным указанием автора, названия и года издания используемой научной литературы. Конспект может быть опорным (содержать лишь основные ключевые позиции), но при этом позволяющим дать полный ответ по вопросу, может быть подробным. Объем конспекта определяется самим студентом.

В процессе работы с основной и дополнительной литературой студент может:

- делать записи по ходу чтения в виде простого или развернутого плана (создавать перечень основных вопросов, рассмотренных в источнике);
- составлять тезисы (цитирование наиболее важных мест статьи или монографии, короткое изложение основных мыслей автора);
- готовить аннотации (краткое обобщение основных вопросов работы);
- создавать конспекты (развернутые тезисы).

5.4. Рекомендации по подготовке к отдельным заданиям текущего контроля

Собеседование предполагает организацию беседы преподавателя со студентами по вопросам практического занятия с целью более обстоятельного выявления их знаний по определенному разделу, теме, проблеме и т.п. Все члены группы могут участвовать в обсуждении, добавлять информацию, дискутировать, задавать вопросы и т.д.

Устный опрос может применяться в различных формах: фронтальный, индивидуальный, комбинированный. Основные качества устного ответа подлежащего оценке:

- правильность ответа по содержанию;
- полнота и глубина ответа;
- сознательность ответа;
- логика изложения материала;
- рациональность использованных приемов и способов решения поставленной учебной задачи;

- своевременность и эффективность использования наглядных пособий и технических средств при ответе;
- использование дополнительного материала;
- рациональность использования времени, отведенного на задание.

Устный опрос может сопровождаться презентацией, которая подготавливается по одному из вопросов практического занятия. При выступлении с презентацией необходимо обращать внимание на такие моменты как:

- содержание презентации: актуальность темы, полнота ее раскрытия, смысловое содержание, соответствие заявленной темы содержанию, соответствие методическим требованиям (цели, ссылки на ресурсы, соответствие содержания и литературы), практическая направленность, соответствие содержания заявленной форме, адекватность использования технических средств учебным задачам, последовательность и логичность презентуемого материала;
- оформление презентации: объем (оптимальное количество), дизайн (читаемость, наличие и соответствие графики и анимации, звуковое оформление, структурирование информации, соответствие заявленным требованиям), оригинальность оформления, эстетика, использование возможности программной среды, соответствие стандартам оформления;
- личностные качества: ораторские способности, соблюдение регламента, эмоциональность, умение ответить на вопросы, систематизированные, глубокие и полные знания по всем разделам программы;
- содержание выступления: логичность изложения материала, раскрытие темы, доступность изложения, эффективность применения средств ИКТ, способы и условия достижения результативности и эффективности для выполнения задач своей профессиональной или учебной деятельности, доказательность принимаемых решений, умение аргументировать свои заключения, выводы.

6. Учебно-методическое и информационное обеспечение дисциплины

6.1 Основная литература:

1. Баранов А.Н. Введение в прикладную лингвистику : [учеб. пособие]. - 4-е изд., испр. и доп.. - М.: Книжный дом "ЛИБРОКОМ", 2013. - 367 с.
2. Гвишиани Н.Б. Практикум по корпусной лингвистике. - М.: Высшая школа, 2008. - 191 с.

6.2 Дополнительная литература:

1. Кристалл Д. Кембриджская энциклопедия английского языка. - 2-е изд.. - Кембридж: Изд-во Кембриджского ун-та, 2005. - 499 с.

6.3 Иные источники:

1. British National Corpus - <http://www.natcorp.ox.ac.uk>
2. American National Corpus - <http://www.anc.org/>
3. Michigan Corpus of Academic Spoken English (MICASE) - <http://quod.lib.umich.edu/m/micase/>
4. Phrases in English (PIE) - <http://pie.usna.edu/>
5. WebCorp Live - <http://www.webcorp.org.uk/live/>

7. Материально-техническое обеспечение дисциплины, программное обеспечение, профессиональные базы данных и информационные справочные системы

Для проведения занятий по дисциплине необходимо следующее материально-техническое обеспечение: учебные аудитории для проведения занятий лекционного и семинарского типа, групповых и индивидуальных консультаций, текущего контроля и промежуточной аттестации, помещения для самостоятельной работы.

Учебные аудитории и помещения для самостоятельной работы укомплектованы специализированной мебелью и техническими средствами обучения, служащими для представления учебной информации большой аудитории.

Помещения для самостоятельной работы укомплектованы компьютерной техникой с возможностью подключения к сети "Интернет" и обеспечением доступа в электронную информационно-образовательную среду Университета.

Для проведения занятий лекционного типа используются наборы демонстрационного оборудования, обеспечивающие тематические иллюстрации (проектор, ноутбук, экран/ интерактивная доска).

Лицензионное и свободно распространяемое программное обеспечение:

Microsoft Windows 10

LibreOffice

Операционная система "Альт Образование"

Microsoft Office Профессиональный плюс 2007

Профессиональные базы данных и информационные справочные системы:

1. Научная электронная библиотека eLIBRARY.ru. – URL: <https://elibrary.ru>
2. Университетская библиотека онлайн: электронно-библиотечная система. – URL: <https://biblioclub.ru>
3. Президентская библиотека имени Б.Н. Ельцина. – URL: <https://www.prilib.ru>
4. Российская национальная библиотека. – URL: <http://nlr.ru>
5. Российская государственная библиотека. – URL: <https://www.rsl.ru>
6. Консультант студента. Гуманитарные науки: электронно-библиотечная система. – URL: <https://www.studentlibrary.ru>
7. Электронный каталог Фундаментальной библиотеки ТГУ. – URL: <http://biblio.tsutmb.ru/elektronnyij-katalog>
8. Электронная библиотека РФФИ. – URL: <https://www.rfbr.ru/rffi/ru/library>
9. Научная электронная библиотека Российской академии естествознания. – URL: <https://www.monographies.ru>

Электронная информационно-образовательная среда

https://auth.tsutmb.ru/authorize?response_type=code&client_id=moodle&state=xyz

Взаимодействие преподавателя и студента в процессе обучения осуществляется посредством мультимедийных, гипертекстовых, сетевых, телекоммуникационных технологий, используемых в электронной информационно-образовательной среде университета.