

УДК 811.11
UDC 811.11

Чугаева Татьяна Николаевна
Пермский федеральный исследовательский центр УрО РАН
г. Пермь, Российская Федерация
Tatiana N. Chugaeva
Perm Federal Research Centre UB RAS
Perm, Russian Federation
tatiananch@mail.ru

Байбурова Ольга Васильевна
Пермский государственный национальный исследовательский университет
г. Пермь, Российская Федерация
Olga V. Baiburova
Perm State University
Perm, Russian Federation
olga3079@mail.ru

Вахотин Антон Андреевич
Вятский государственный университет
г. Киров, Российская Федерация
Anton A. Vakhotin
Vyatka State University
Kirov, Russian Federation
samo87@mail.ru

Дмитриева (Мякотникова) Светлана Юрьевна
Пермский национальный исследовательский политехнический университет
г. Пермь, Российская Федерация
Svetlana Y. Dmitrieva (Miakotnikova)
Perm National Research Polytechnic University
Perm, Russian Federation
myakotnikova@yandex.ru

**СОПОСТАВЛЕНИЕ РЕЗУЛЬТАТОВ ЛИНГВОСТАТИСТИЧЕСКОГО
АНАЛИЗА ПЕРЦЕПТИВНЫХ ТИПОВ РУССКОГО И
АНГЛИЙСКОГО СЛОВА (НА МАТЕРИАЛЕ НКРЯ, БНК, АНК)
STATISTIC-LINGUISTIC COMPARATIVE ANALYSIS OF THE
PERCEPTUAL TYPES OF A RUSSIAN AND AN ENGLISH WORD
(BASED ON NCRL, BNC, ANC)**

Аннотация

Корпусные исследования имеют неоспоримые преимущества, однако в соответствии со своими лингвистическими задачами лингвисты применяют различные подходы к изучению обширного материала корпуса. Одни лингвисты описывают содержание корпусов, в то время как психолингвисты объединяют описание корпуса и проведение эксперимента. В настоящей работе используется теоретический систематический подход к описанию частотных страт трёх национальных корпусов - Национального корпуса русского

языка, Британского национального корпуса и Американского национального корпуса. На материале корпусов проводится лингвостатистический анализ для выявления перцептивных типов русского и английского слова (в двух национальных вариантах) с целью построения фонетической (перцептивной) классификации языков, сопоставления перцептивных типов и описания своеобразия звукового строя каждого языка.

В работе были проанализированы частотные страты трёх корпусов; затем в соответствии с системным критерием комплекса перцептивно значимых лингвистических признаков (длина в слогах, ударная гласная, ритмическая структура и др.) отбирались слова, представляющие типы, объединённых одним признаком. Сопоставление полученных данных обнаруживает больше отличительных, чем сходных черт в русском и англоязычных корпусах. Так, в Национальном корпусе русского языка ведущими перцептивными типами оказываются трёхсложные и двусложные слова. В англоязычных корпусах преобладают одно- и двусложные слова. Оба английских корпуса содержат больше слов с ударением на начальные слоги в двухсложном и трёхсложном слове. При анализе русского корпуса не было выявлено существенного преобладания начальнударных двухсложных слов. В англоязычных корпусах самыми частотными являются слова с гласными переднего ряда, а для Национального корпуса русского языка это слова с гласными среднего и переднего ряда. Консонантные коэффициенты слов всех трёх корпусов указывают на меньшую консонантную нагруженность русского слова по сравнению с английским. Сходной чертой для всех трёх корпусов является преобладание слов, начинающихся с согласного звука.

Abstract

Corpus research presents obvious benefits, though linguists approach the material in various ways. For example, corpus linguists approach data in an exploratory way, whereas psycholinguists more often tend to combine corpus data and experimental research. The current work uses the theoretical systemic approach to describe the two frequency strata of the three corpora (Russian National Corpus, British National Corpus and Open American National Corpus) and build the classification of phonetic word types in Russian and English (British and American). The aim of the research is to draw up the phonetic (perceptive) classification of the corresponding languages and to describe the identity of their sound systems based on these types.

The high frequency and the frequency strata of the three corpora have been analyzed to identify the words characterized by the following linguistic features: length in syllables, stressed vowel, rhythmic structure, etc. The data comparison discovered more distinctions than similarities among the three corpora. While the dominant perceptive types in the Russian Corpus were three- and two-syllable words, both English corpora had more one- and two-syllable words. English corpora included more words with the emphasis on the initial syllables in two- and three-syllable words, whereas the Russian Corpus did not have that tendency. In the two English corpora, the words with front vowels turned out most frequent, while in the Russian corpus middle- and front-vowel words had the highest frequency. The consonant coefficients of all the three corpora proved that a Russian word is less consonant-loaded compared to an English word. One similar feature of the three corpora was the predominance of words with word-initial consonant.

Ключевые слова: национальный корпус, частотные страты, лингвистические признаки, фонетические типы, фонетическая (перцептивная) классификация слов.

Keywords: a national corpus, frequency strata, linguistic features, phonetic word types, phonetic (perceptual) classification of words types.

doi: 10.22250/2410-7190_2019_5_3_273_291

1. Введение

В последние десятилетия утвердилась практика проведения разнообразных лингвистических исследований с использованием языковых электронных технологий, которые предлагает корпусная лингвистика.

Появление корпусной лингвистики относят к 60-м гг. XX в. Первые частотно организованные корпуса были созданы на материале немецкого и английского языков, но очень быстро начали появляться корпуса и на базе других языков. Формирование корпусной лингвистики как отдельного раздела языкознания относят к первой половине 90-х гг. XX в.

Сегодня термином *корпус* принято обозначать собрание текстов одного или нескольких языков, связанных между собой определёнными частотными параметрами. С развитием корпусной лингвистики, определение корпуса уточняется и оформляется понятийный аппарат дисциплины. Дж. Синклер, например, описывает корпус как собрание лингвистических текстов естественного происхождения, которые отбираются, чтобы представить вариативность данного языка [Sinclair, 1991, с. 171]. Конкретизируя понятие корпуса, М. Стаббс представляет его как собрание текстов, составленное с определённой целью, обычно для обучения или исследования [Stubbs, 2001, с. 239–240]. Другие исследователи уточняют, что корпус представляет собой собрание письменных и устных высказываний, доступные в электронном виде [Lemnitzer, Zinsmeister, 2006, с. 7]. Российские исследователи описывают корпус текстов как большой, унифицированный, структурированный, размеченный, филологически компетентный массив языковых данных, предназначенный для решения конкретных лингвистических задач [Захаров, Богданова, 2011, с. 7].

В настоящее время создано большое количество репрезентативных корпусов текстов языков мира, в том числе национальных корпусов. Это Национальный корпус русского языка, Британский национальный корпус, Американский национальный корпус, Мангеймский корпус немецкого языка, Корпус немецкого языка, Корпус французского языка, Венгерский национальный корпус, Корпус современного китайского языка и другие.

Современные национальные корпуса включают не менее 100 млн. словоупотреблений и представляют собой коллекции устных и письменных текстов различных жанров, стилей, региональных и социальных вариантов языка. Это даёт уникальную возможность изучать разноуровневые языковые единицы [Сысоев, 2010]. Созданные за последние десятилетия на материале различных языков многочисленные частотные словари и корпуса, существенно отличаются в качественном и количественном отношении от первых частотных списков Ф. В. Кэдинга. Они представляют интерес для множества лингвистических дисциплин, поскольку являются достоверным источником сведений о языке благодаря значительному

объёму выборки. Сегодня корпусная лингвистика – это активно развивающееся направление лингвистики, занимающееся разработкой, созданием и использованием текстовых корпусов для решения многих актуальных лингвистических задач.

1.1. Из истории вопроса

Одними из первых к созданию частотных списков приступили немецкие лингвисты: Эдуард Форстеманн (Eduard Förstemann) ещё в 1852 году в своей работе «Numerische Lautverhältnisse im Griechischen, Lateinischen und Deutschen» [Förstemann, 1852] с целью определить частотное соотношение сочетаний звуков в родственных языках, а также Франц Габельсбергер (Franz X. Gabelsberger) с целью выявить частотные характеристики слогов немецкого языка [цит. по : Kaeding, 1897, с. 38].

Первый объёмный частотный словарь немецкого языка был составлен Фридрихом В. Кэдингом [Kaeding, 1897, S. 38]. Словарь показывал частотные соотношения слов, слогов и букв немецкого языка. Этот словарь отличался от предыдущих исследований, прежде всего, обширным объёмом материала для необходимой достоверности частотных значений. Единицами анализа были слог и лексема (словоформа). Общий текст словаря насчитывал 20 млн. слогов или почти 11 миллионов (10970777) лексем.

Кроме того, отдельно рассматривались частотность корней, приставок и суффиксов, отдельных буквосочетаний, что давало представление о характере словообразования в немецком языке. В качестве источников были перечислены более 300 наименований различных печатных текстов; анализировался «язык писателей и образованных людей» XIX века [Meier, 1964]. Следует отметить, что в то время большинство языковедов, впрочем, встретили труд Ф. В. Кэдинга равнодушно или жёсткой критикой, подвергая сомнению достоверность его выводов.

П. Менцерат в своей классической работе «Architectonik des deutschen Wortschatzes» вводит понятие «системной частотности» (systematische Frequenzstatistik), которую он противопоставляет объективной частотности употребления (Gebrauchshäufigkeit), зафиксированной в словаре Ф. В. Кэдинга [Menzerath, 1954, с. 5]. Работа П. Менцерата является «попыткой подвергнуть исследованию не искусственно изолированные факты, но исследовать естественно данное, непосредственно существующее, строя и упорядочивая его по ясным и обозримым критериям» [Menzerath, 1954, с. 2].

В задачу П. Менцерата входило проведение статистического описания фонетических типов немецкого слова, их распределение и, с помощью «системной» частотности слов, выстраивание таким образом классификации словарного состава немецкого языка. Он описывал частотность слова, опосредованную частотностью его признаков. По сути, данная категория описывает частотность формы или типа (класса) слова. Упорядочивание «словарной массы» осуществляется П. Менцератом по трём критериям – количество звуков и слогов и последовательность звуков [Menzerath, 1954, с. 7].

Особый интерес П. Менцерата состоял в том, чтобы определить, какие «теоретические (абстрактные, дедуктивно выводимые типы) существу-

ют в действительности (конкретные, индуктивно выделяемые), то есть являются реализуемыми» [Menzerath, 1954, с. 7].

Таким образом, П. Менцерат обсуждает «разные статистики»; причём, обе они – и объективная, зафиксированная в словаре частота употребления (Gebrauchshäufigkeit), отвергнутая П. Менцератом, и предлагаемая им регулярность, повторяемость, каноничность формы слов (systematische Frequenzstatistik) [Menzerath, 1954, с. 7] представляются нам одинаково значимыми.

В настоящее время тезис о вероятностной организации системы языка и ключевой роли действия вероятностных (частотных) факторов в речевом поведении человека считается очевидным и не нуждается в доказательствах. В начале 50-х годов XX века была отмечена зависимость процессов порождения и восприятия речи от частотности единиц (frequency effect) [Howes, Solomon, 1951]. В речевых механизмах человека установлена определенная иерархическая организация элементов речи в соответствии с частотой их встречаемости; причём от «индекса частоты» слова зависит способ его хранения и поиска в памяти [Фрумкина, Василевич, 1971 ; Залевская, 1990 ; Savin 1963 и др.].

Общепризнанным подходом стал учёт частотных характеристик единиц при исследовании речи [Агибалов, 1995 ; Адмони, 1964 ; Андреев, Зиндер, 1963 ; Ерофеева, 2005 ; Кравец, 1976 ; Топоров, 1959 ; Фрумкина, 1971 ; Углонова, 2004 и др.]. Вероятностные характеристики языковых единиц признаются неотъемлемой частью структуры языка. В подавляющем большинстве современных работ, посвящённых восприятию речи, частотность воспринимаемых лингвистических единиц рассматривается как гиперпризнак, то есть как фактор, оказывающий наибольшее влияние на идентификацию единиц в любых условиях [Штерн, 1992 ; Венцов, Касевич 2003 и др.].

Встречаемость слова в речевой деятельности принято измерять двумя величинами: объективной ($F_{об}$) и субъективной ($F_{суб}$) частотами, которые коррелируют между собой [Фрумкина, Василевич, 1971]. Для установления объективной частоты слова и его элементов (морфем, слогов, фонем) традиционным стало обращение к частотным словарям. Статистические (частотные) словари представляют собой особый тип словарей, в которых лексические единицы рассматриваются вместе с частотами их употребления [Алексеев, 2001 ; Малаховский, 1980], «для определения границ активного словарного состава, давая, таким образом, ясное и чёткое представление «живого словоупотребления» [Засорина, 1977, с. 6].

Установлено, что механизм вероятностного прогнозирования обеспечивает «преднастройку организма к действию: ситуация, требующая тех или иных реакций, ещё не возникла, но соответствующие системы уже в состоянии готовности» [Фрумкина, 1971, с. 3]. В таком виде речевая вероятность связана с понятием априорных форм как субъективных категорий ума сопутствующих эффективному восприятию и интерпретации мира человеком, порождаемых опытом [Сегал, 1972, с. 9]. Категории вероятности являются одной из форм представления прошлого опыта и, следовательно, играет ключевую роль при восприятии как объектов действительности, так и человеческой речи.

1.2. Постановка проблемы

Обращение к изучению организации и функционирования перцептивного словаря считается одним из наиболее перспективных направлений перцептивной лингвистики [Венцов, Касевич, 2003 ; Hawkins, 1999]. Ключевая роль слова в процессах восприятия речи предполагает существование своеобразного алфавита полезных признаков, к которым обращается слушающий при восприятии слова [Бондарко, 1981 ; Венцов, Касевич, 2003]. Основной задачей экспериментального изучения восприятия звуковой последовательности остаётся «систематическая квалификация тех особенностей речевого поведения человека, которые позволили бы установить иерархию различных уровней и описать её количественно» [Бондарко, 1981, с. 182].

Ещё А. И. Смирницкий предлагал рассматривать лексику любого языка не как «кучу разнообразных примеров, но как специфическую систему лексических единиц <...> лишь подойдя к лексике как системе можно должным образом выделить в ней существенное и характерное и описать её состав, следуя внутренним связям между его элементами, а не абстрактной классификационной схеме явлений» [Смирницкий, 1998, с. 7]. Он подчёркивал необходимость выяснения фонетических типов слов, характерных для данного языка и построения фонетической классификации слов, которая «безусловно, существенна как для характеристики внешнего облика лексики данного языка, так и для понимания тенденций его развития, так как изменения происходят в конкретных словах» [Смирницкий, 1956, с. 140].

Идеи построения классификации слов в соответствии с принципом «системной частотности» слов (частотности типа) П. Менцерата оказываются созвучными методу выявления фонетических (перцептивных) типов слов конкретного языка на основе комплекса перцептивно значимых лингвистических признаков [Чугаева, 2007, 2009].

Комплекс экспериментально выявленных существенных признаков восприятия русского, английского и немецкого слова, определяющих восприятие [Штерн, 1992 ; Краузе, 2002 ; Чугаева, 2009], может служить системным критерием выявления фонетических типов слов, то есть групп (типов) слов, объединённых тем или иным признаком. Самыми значимыми из них оказались частотность употребления, длина, ударная гласная, акцентно-ритмическая структура и некоторые другие.

Данные характеристики могут быть положены в основу классификации и выделения перцептивных типов того или иного языка, а выявление частотных характеристики данных типов может послужить основанием для определения типичности слов в языке. Поскольку объединение признаков в комплексе признаков не может быть случайным, на этом основании возможно построение классификации перцептивных типов слов конкретного языка, и таким образом, описание звукового облика слова конкретного языка в единстве его существенных, строевых характеристик.

Итак, выявление всего комплекса признаков слова как целостной единицы на материале отдельно взятого языка остаётся актуальной

лингвистической задачей, равно как и описание всей языковой системы с точки зрения своеобразия звукового строя языка в перцептивном аспекте.

Лингвостатистический анализ трёх корпусов (Британского национального корпуса, Американского национального корпуса и Национального корпуса русского языка) был запланирован и проведён для выявления фонетических типов русского и английского (британского и американского вариантов) слова. Сравнение полученных по каждому корпусу данных поможет понять своеобразие перцептивного облика «слова», выявить типологическое сходство британского, американского и русского слова.

2. Эксперимент

В данном исследовании были использованы три национальных корпуса: Британский национальный корпус (British National Corpus), Американский национальный корпус (Open American National Corpus) и Национальный корпус русского языка. Корпуса были организованы на основе принципа объективной частотности, что позволяет осуществить их корректное сопоставление для проведения сравнительного анализа.

Британский национальный корпус (БНК), созданный в 1992 году, содержит 100 млн. слов [British..., 2018]. Материалом является письменная форма речи на основе транскрибированных устных текстов общего характера. Корпус включает все графические и фонетические варианты словоформ с соответствующими индексами частоты. В нём есть имена собственные, дни недели, месяцы, числительные как в цифровой, так и в буквенной форме. В создании корпуса приняли участие ведущие британские издатели словарей Лонгмана (Addison-Wesley Longman) и Чемберса (Larousse Kingfisher Chambers).

Американский национальный корпус (АНК) [Open..., 2018] – это синхронный корпус общего назначения, который включает образцы письменной и устной речи из разных источников. Он снабжён метатекстовой, частеречной разметкой, имеет частичную синтаксическую разметку. Есть также разметка именованных сущностей (Named Entities) – имён собственных для людей, организаций и географических объектов. Слова в корпусе расположены по частоте встречаемости в речи – от наиболее частотных до самых редких. Корпус создаёт наиболее полную картину американского варианта английского языка для дальнейших исследований в области лингвистики, образования, лексикографии и технологии. В настоящее время корпус насчитывает 22 миллиона словоформ американских слов, он постоянно пополняется, и по завершению работы эта цифра увеличится до 100 миллионов слов.

Национальный корпус русского языка (НКРЯ) характеризуется представительностью и сбалансированным составом текстов: корпус содержит по возможности все типы письменных и устных текстов, представленные в данном языке (художественные разных жанров, публицистические, учебные, научные, деловые, разговорные, диалектные и т. п.). Все эти тексты входят в корпус по возможности пропорционально их доле в языке соответствующего периода. На сегодняшний день НКРЯ

содержит 335076 текстов и 364881378 словоупотреблений [Национальный..., 2019].

2.1. Методика исследования

Все исследования проводились по единой методике с опорой на единые теоретические основания – теорию восприятия звучащей речи с опорой на существенные лингвистические признаки Л. Р. Зиндера и А. С. Штерн [Зиндер, Штерн, 1972 ; Штерн, 1992], самыми значимыми из которых были частотность употребления, длина в слогах, ударная гласная, акцентно-ритмическая структура.

На материале английского языка было показано, что комплекс экспериментально выявленных существенных признаков может служить системным критерием для выявления фонетических типов слов – групп слов, объединённых тем или иным признаком [Чугаева 2007, 2009]. Критериями лингвостатистического анализа послужили перцептивно значимые признаки: длина в слогах, ударная гласная, ритмическая структура, консонантный тип, консонантный коэффициент, начальный звук / слог, часть речи [Штерн, 1992 ; Чугаева, 2007]. Процентное соотношение перцептивных типов было получено путём простого статистического подсчёта слов высокочастотных и частотных страт каждого корпуса.

Для начала было важно установить тот объём слов, который будет считаться минимальным и основным набором. Определение границ частотных страт осуществлялось с учётом принципов частотной стратификации словаря, изложенных Л. В. Малаховским [Малаховский, 1980]. Частотные зоны разграничивались на основе логарифмического масштаба: каждая зона охватывает диапазон, равный одному порядку, то есть верхняя граница зоны характеризуется частотой в 10 раз большей, чем нижняя. Таким образом, весь словарь был поделён на пять частотных групп слов, в каждую из которых вошло разное количество словоформ: 1) «очень частые», 2) «частые», 3) «средние», 4) «редкие», 5) «очень редкие» слова. Группа «очень частых» слов или высокочастотная зона (сверхъядро) включает слова с 7-, 6- и 5-значными индексами частоты; «частых» слов или частотная зона (ядро) – с 4-значными; средняя с 3-значными, редкая – 2-значными, очень редкая – 1-значными. Такое разбиение словаря на частотные зоны приблизительно соответствует интуитивному делению слов языка на частотные группы носителями языка [Малаховский, 1968, с. 35].

2.2. Материал исследования

Лингвостатистический анализ был проведён на высокочастотной и частотной стратах корпусов. Из списков слов исключались все имена собственные, аббревиатуры, символы и буквы, все словоформы, классификация которых казалась затруднительной. Объектом статистического анализа стали слова, представляющие знаменательные части речи. В итоге в высокочастотный список британского варианта английского языка вошло 718 словоформ, американского варианта – 106,

русского языка – 644 словоформы. Частотный список представлен 2079, 1787, 7476 словоформами соответственно. Количество отобранных и проанализированных словоформ для каждой страты корпусов представлены в Таблице 1.

Таблица 1. Количество словоформ в двух частотных стратах БНК, АНК и НКРЯ

	БНК	АНК	НКРЯ
Высокочастотная страта	718	106	644
Частотная страта	2079	1787	7476

Очевидное различие количества словоформ в частотных стратах АНК и БНК по сравнению с НКРЯ требует их более скрупулёзного анализа. Одно из объяснений – это лексикализация английского языка [Мельников, 1971, с. 366], проявлением которой является «конверсия». Дело в том, что при совпадении словоформ разных частей речи, подсчитывалась более частотная форма слова, что, несомненно, уменьшило количество слов в частотных стратах английских корпусов.

2.3. Сравнение корпусов по существенным лингвистическим признакам

2.3.1. Длина в слогах

Сравнение БНК, АНК и НКРЯ по признаку «длина в слогах» выявило абсолютное преобладание «короткословной» лексики в частотных слоях АНК и БНК и тенденцию к присутствию «длиннословной» лексики в НКРЯ (61,2% 3–5-сложных слов в обеих стратах корпуса).

Так, высокочастотная страта на 98% в АНК и на 87,8% в БНК представлена одно- и двухсложными словами; а частотная – соответственно на 74,7% и 64,8%. При этом высокочастотная страта НКРЯ только на 65,6% представлена одно- и двухсложными словами, а для частотной страты этот процент ещё ниже – 34,3%.

На долю длинного (3–5-сложного) слова в высокочастотной страте приходится 12,2% в БНК, всего лишь 2% в АНК и 65,6% в НКРЯ. В частотной страте это соотношение составляет 35,2% в БНК; 25,5% в АНК; и больше половины всех словоформ – 62,8% – в НКРЯ (см. табл. 2). Сравнение трёх корпусов по признаку «длина в слогах» обнаруживает абсолютное преобладание «короткословной» лексики (одно- и двусложных слов) в обоих слоях АНК и БНК. Преобладание односложной лексики в первых двух частотных стратах обоих корпусов можно рассматривать как строевую особенность английского слова. Односложное слово является основным перцептивным подтипом слова, в то время как «длинные» слова вытесняются на периферию языка [Чугаева, 2009, с. 95]. В отношении

русского языка отмечается обратная тенденция. Односложные слова занимают всего 4,5% от общего количества слов двух частотных страт НКРЯ; ведущими перцептивными типами русского слова оказываются трёхсложное (34,7%) и двусложное (31,6%) слова. Длинные слова (4-х и более сложные) суммарно также составляют около 30%.

Таблица 2. Распределение слов в высокочастотной и частотной стратах БНК, АНК и НКРЯ по признаку «длина в слогах»

Высокочастотная страта (%)			
	БНК	АНК	НКРЯ
1-сложное слово	54,1	80	13,7
2-сложное слово	33,7	18	51,9
3-сложное слово	9,4	2	23,5
4-сложное слово	2,2	-	7,4
5-сложное слово	0,6	-	3,3
6-сложное слово	-	-	0,2
Частотная страта (%)			
	БНК	АНК	НКРЯ
1-сложное слово	30,3	37,2	3,9
2-сложное слово	34,5	37,3	30,4
3-сложное слово	24,8	16,8	35,3
4-сложное слово	8,1	8,7	19,3
5-сложное слово	2,3	-	8,2
6–7-сложное слово	-	-	2,9

Ряд лингвистов отмечают, что одной из ключевых особенностей развития системы английского языка является склонность к аналитизму. Радикальная типологическая перестройка, происходящая в языке, влияет на все три подсистемы языка. А сосуществование двух типологически разнородных слоев в лексической подсистеме языка: исходной аналитизированной и заимствованной синтетической, находит отражение в полифункциональности частей речи, односложности корня и слова [Ильиш, 1971 ; Мельников, 1971 ; Плоткин, 1989, с. 229].

2.3.2. Ритмическая структура

По признаку «ритмическая структура» двухсложного слова корпуса (табл. 3) сохраняют определённые тенденции: доминирование начальноударных слов в обеих стратах БНК и АНК и незначительное их преобладание в частотной страте НКРЯ.

Таблица 3. Распределение слов в высокочастотной и частотной стратах БНК, АНК и НКРЯ по признаку «ритмическая структура» (двухсложное слово)

Высокочастотная страта (%)			
	БНК	АНК	НКРЯ
Начальноударное	76,6	95	50
Конечноударное	23,4	5	50
Частотная страта (%)			
	БНК	АНК	НКРЯ
Начальноударное	74,7	70	58,6
Конечноударное	25,3	30	41,4

Трёхсложное слово (табл. 4) демонстрирует сходное соотношение между его градациями в высокочастотной и частотной стратах БНК и АНК, однако они не характерны для НКРЯ. Прослеживается тенденция к увеличению количества срединноударных слов в частотной страте БНК и уменьшению их в этой же страте АНК. Большинство трёхсложных слов обеих страт обоих корпусов являются начальноударными или срединноударными. В высокочастотной страте АНК нет слов с основным ударением на последнем слоге. Высокочастотная страта русского корпуса содержит большее количество срединноударных слов, однако данная тенденция не сохраняется в частотной страте НКРЯ, где преобладают начальноударные слова.

Таблица 4. Распределение слов в частотных стратах БНК, АНК и НКРЯ по признаку «ритмическая структура» (трёхсложное слово)

Высокочастотная страта (%)			
	БНК	АНК	НКРЯ
Начальноударное	48,2	50	34,3
Срединноударное	49,3	50	47,0
Конечноударное	2,5	0	18,7
Частотная страта (%)			
	БНК	АНК	НКРЯ
Начальноударное	44,7	54	53,1
Срединноударное	51,6	45	28,3
Конечноударное	3,7	1	18,5

Высокочастотная и частотная страты НКРЯ содержат четырёх-, пяти- и шестисложные слова. Ритмическая структура четырёхсложного слова представлена словами с ударением на второй слог (62,5% от всех слов) в

высокочастотной страте и ударением на второй и третий слоги в частотной страте (45,2% и 44,9% соответственно). Пятисложное слово имеет ударение на третьем слоге (71,4% в высокочастотной и 69,3% в частотной страте). Шестисложные слова отсутствуют в высокочастотной страте корпуса, а в частотной страте ударным становится четвёртый слог (69,2%).

Итак, статистический анализ частотного списка словоформ по признаку «ритмическая структура» позволяет сделать следующие выводы. В частотных стратах БНК и АНК преобладают слова с ударением на первом слоге в двухсложных словах и на первый и второй слоги в трёхсложных словах. Анализ НКРЯ не выявил существенного преобладания начальноударных двухсложных слов в обеих стратах. Для слов большей длины отмечается явное стремление ударения к середине слова: преобладание срединноударных слов для трёхсложного слова, примерно равное количество четырёхсложных слов с ударением на второй и третий слоги, доминирование третьего слога как ударного в пятисложных словах и четвёртого слога – в шестисложных словах.

2.3.3. Начальная фонема и консонантный тип

Анализ слов в БНК, АНК и НКРЯ по признаку «начальная фонема» показал значительное преобладание количества слов, начинающихся с согласного в обеих частотных стратах всех трёх корпусов. Наблюдается тенденция увеличения количества слов с начальным гласным при возрастании количества слогов в слове.

2.3.4. Консонантный тип

Данные анализа высокочастотной страты по признаку «консонантный тип» представлены в таблице 5. Самыми представительными типами слогов в БНК оказались CVC, CVCC, CV, CCVC, CCV; АНК представили типы CVC, CV, CVCC, CVCV, VC; НКРЯ: CVCCV, CVCV, CVCVC, CVC, CVCVCV.

Таблица 5. Наиболее распространенные консонантные типы в высокочастотной страте БНК, АНК и НКРЯ (в %)

	БНК	АНК	НКРЯ
CVC	27,4	42,0	7,2
CV	6,3	16,0	-
CVCC	6,9	11,0	-
CVCV	-	6,0	9,9
VC	-	3,0	-
CVCCV	-	-	10,7
CVCVC	-	-	9,7
CVCVCV	-	-	5,5

Согласно исследованиям БНК и АНК, в британском и американском вариантах английского языка ведущим консонантным типом является тип CVC. Система русского языка не выявляет тенденции ведущего типа, консонантные типы слов русского языка располагаются весьма равномерно. Консонантные коэффициенты слов $k=2,34$ (БНК), $k=1,54$ (АНК), $k=1,49$ (НКРЯ) указывают на то, что в частотном русском слове в среднем согласных меньше, чем в частотном британском и американском слове. Разница в консонантных коэффициентах британского и американского высокочастотного слова объясняется тем, что в высокочастотной страте АНК в процентном соотношении больше слов, оканчивающихся на гласный, чем в той же страте БНК – 22%: 6,3%. В. А. Никонов замечает, что отношение количества согласных к количеству гласных в звуковых цепях мира колеблется между 2,0 и 0,5, впрочем, не достигая этих пределов [Никонов, 1966]. Большинство языков мира имеют консонантный коэффициент больше единицы. Это говорит о том, что в звуковой цепочке языка согласных обычно больше, чем гласных [Tambovtsev, 1985].

2.3.4. Ударная гласная

Анализ по признаку «ударная гласная» фонема показал, что в обоих англоязычных корпусах самыми частотными являются слова с гласными переднего ряда /e/ (БНК – 12,8%) или /ɛ/ (АНК – 10%), /i:/ (БНК – 6,5%, АНК – 7,7%) и дифтонгом /ei/ (БНК – 9,2%, АНК – 7,7%). Частотные страты обоих корпусов обнаруживают существование сходных тенденций: они содержат примерно одинаковое в процентном соотношении количество слов с монофтонгами и дифтонгами. Высокочастотные слои БНК и АНК также включают приблизительно равное в процентном соотношении количество слов с монофтонгами и дифтонгами: АНК – 68% и 32%, а БНК – 70,6% и 29,3% соответственно. При этом количество слов с краткой и долгой гласной отличается. АНК содержит 56% словоформ с краткой гласной и 12% с долгой; в БНК – 46,3% слов с краткой гласной и 24,3% с долгой. Анализ частотного списка НКРЯ выявил, что более 75% всех словоформ имеют в качестве ударной гласной одну из трех фонем – [a], [o] и [e]. При этом разница в процентах между этими гласными небольшая – 29,5%, 22,5%, 21,18% соответственно. Наши данные по корпусу согласуются с теми, которые получены ранее другими исследованиями [Уровни языка... 1986, с. 190].

Выявленные статистические различия по указанному признаку имеют большую перцептивную значимость: чем больше градаций имеет какого-либо признак, тем большее дифференциальное значение имеет каждая его градация. Исходя из этого, можно предположить, что в русском языке роль ударной гласной при распознавании может быть несколько ниже, чем в английском.

3. Заключение

Таким образом, в результате предпринятого лингвостатистического анализа частотных страт НКРЯ, БНК и АНК с применением единого комплекса перцептивно значимых лингвистических признаков слова, таких

как «длина в слогах», «ударная гласная», «ритмическая структура», «начальный звук» и консонантный коэффициент, оказалось возможным выявить основные строевые фонетические (перцептивные) типы слова в русском и английском языках (в британском и американском вариантах). Это дало основание, с одной стороны, для построения перцептивной классификации слов русского и английского языков; с другой стороны, их сопоставление оказалось продуктивным с точки зрения описания своеобразия звукового строя данных языков в единстве существенных перцептивных характеристик слова. Такой подход имеет смысл не только в перцептивном аспекте, он является актуальным и с точки зрения общелингвистических проблем, в том числе для характеристики внешнего облика языка.

Типологическое сравнение звукового облика русского и английского слова выявило больше отличительных черт. Так в НКРЯ ведущими перцептивными типами слова оказались трёхсложные и двусложные слова, в то время как в БНК и АНК преобладали одно- и двусложные слова. Оба английских корпуса содержат больше слов с ударением на начальный слог в двухсложном и трёхсложном слове. При анализе НКРЯ не было выявлено существенного преобладания начальноударных двухсложных слов, но в словах большей длины отмечается стремление ударения к середине слова. В англоязычных корпусах самыми частотными являются слова с гласными переднего ряда /e/, /ɛ/, /i:/, а для НКРЯ – это слова с гласными среднего [a], [o] и переднего ряда [e]. Консонантные коэффициенты слов всех трёх корпусов указывают на меньшую консонантную нагруженность русского слова по сравнению с английским.

Сходной чертой для всех трех корпусов явилось преобладание слов, начинающихся с согласного звука. Однако наблюдалась тенденция увеличения количества слов, начинающихся с гласного, при возрастании количества слогов в слове.

Установление и описание ядра словарного состава языка приобретает ключевую роль в типологических исследованиях, ориентированных на поиск основных тенденций развития языка, а также в прикладных целях для усвоения языка.

Список литературы

1. Агибалов, А. К. Вероятностная организация внутреннего лексикона человека [Текст] : автореф. дис. ... канд. филол. наук 10.02.01 / Агибалов Анатолий Константинович ; Санкт-Петербургский гос. унт. – СПб., 1995. –19 с.
2. Адмони, В. Г. Основы теории грамматики [Текст] / В. Г. Адмони. – М. ; Л. : Изд-во Наука, 1964. –105 с.
3. Алексеев, П. М. Частотные словари: учебное пособие [Текст] / П. М. Алексеев. – СПб. : Изд-во С.-Петерб. ун-та, 2001. –156 с.
4. Андреев, Н. Д. О понятиях речевого акта, речи, речевой деятельности и языка [Текст] / Н. Д. Андреев, Л. Р. Зиндер // Вопросы языкознания. – 1963. – № 3. – С.15–21.
5. Бондарко, Л. В. Фонетическое описание языка и фонологическое описание речи [Текст] / Л. В. Бондарко. – Л. : Изд-во Ленинград. ун-та, 1981. – 200 с.

6. Венцов, А. В. Проблемы восприятия речи [Текст] / А. В. Венцов, В. Б. Касевич. – М. : УРСС Эдиториал, 2003. – 240 с.
7. Ерофеева, Е. В. Вероятностные структуры идиомов: социолингвистический аспект [Текст] / Е. В. Ерофеева. – Пермь : Изд-во Перм. ун-та, 2005. – 320 с.
8. Залевская, А. А. Слово в лексиконе человека: Психолингвистическое исследование [Текст] / А. А. Залевская – Воронеж : Воронеж. ун-т, 1990. – 205 с.
9. Засорина, Л. Н. (ред.). Частотный словарь русского языка [Текст] / Л. Н. Засорина – Москва : Русский язык, 1977. – 936 с.
10. Захаров, В. П. Корпусная лингвистика [Текст] : учебник / В. П. Захаров, С. Ю. Богданова. – Иркутск : ИГЛУ, 2011. – 161 с.
11. Зиндер, Л. Р. Факторы, влияющие на опознание слова [Текст] / Л. Р. Зиндер, А. С. Штерн // Матер. IV Всесоюз. симп. по психолингвистике и теории коммуникации, Москва, 30 мая – 2 июня 1972 г. / АН СССР, Ин-т языкознания. – М., 1972. – С. 100–108.
12. Ильиш, Б. А. Строй современного английского языка. 2-е изд. (на англ. яз.) / Б. А. Ильиш – Л. : Просвещение, 1971. – 215 с.
13. Кравец, А. С. Природа вероятности (философские аспекты) [Текст] / А. С. Кравец. – М. : Мысль. – 1976. – 174 с.
14. Краузе, М. Динамика механизма восприятия слова при различных условиях овладения иностранным языком [Текст] / М. Краузе. – Мюнхен : Verlag Otto Sagner, 2002. – 173 с.
15. Малаховский, Л. В. Принципы частотной стратификации словарного состава языка [Текст] / Л. В. Малаховский // Статистика речи и автоматический анализ текста. – Ч. 2 : Статистика речи. – Л. : Наука, 1980. – С. 99–105.
16. Мельников, Г. П. Детерминанта – ведущая грамматическая тенденция языка [Текст] / Фонетика, Фонология. Грамматика: К 70-летию А. А. Реформатского. – М., 1971. – С. 359–367.
17. Национальный корпус русского языка. [Электронный ресурс]. Режим доступа : <http://www.ruscorgo.ru/corgo-freq.html> (дата обращения 19.05.2018).
18. Никонов, В. А. Фоностатистическое измерение межъязыковых расстояний // Исследования по фонологии : сб. ст. / под ред. С. К. Шаумяна. – М. : Наука, 1966. – С. 285–297.
19. Плоткин, В. Я. Строй английского языка : учеб. пособие [Текст] / В. Я. Плоткин. – М. : Высшая школа, 1989. – 239 с.
20. Сегал, Д. М. Основы фонологической статистики (на материале польского языка) [Текст] / Д. М. Сегал. – М. : Наука, 1972. – 255 с.
21. Смирницкий, А. И. Лексикология английского языка [Текст] / А. И. Смирницкий. – М., 1956. – 260 с.
22. Смирницкий, А. И. Лексикология английского языка [Текст] / А. И. Смирницкий. – М. : Омен, 1998. – 259 с.
23. Сысоев, П. В. Лингвистический корпус в методике обучения иностранным языкам [Электронный ресурс]. – Режим доступа : <http://sun.tsu.ru/mminfo/000349304/09/image/09-099.pdf> (дата обращения 13.04.2019).
24. Топоров, В. Н. О введении вероятности в языкознание [Текст] / В. Н. Топоров // Вопросы языкознания. – 1959. – № 6. – С. 28–35.

25. Углонова, И. А. Субъективная оценка частоты слова и ее категоризация: онтолингвистический аспект [Текст] : дис. ... канд. филол. наук : 10.02.19 / Углонова Инна Алексеевна ; Пермский государственный университет. – Пермь, 2004. – 194 с.
26. Уровни языка в речевой деятельности [Текст] / отв. ред. Л. В. Бондарко. – Л. : Изд-во ЛГУ, – 1986. – 258 с.
27. Фрумкина, Р. М. Вероятность слова и восприятие речи [Текст] / Р. М. Фрумкина, А. П. Василевич // Вопросы порождения речи и обучения языку. – М., 1967. – С. 17–38.
28. Фрумкина, Р. М. Вероятность элементов текста и речевое поведение [Текст] / Р. М. Фрумкина. – М. : Изд-во Московск. ун-та, 1971. – 168 с.
29. Чугаева, Т. Н. Звуковой строй языка в перцептивном аспекте (экспериментальное исследование на материале английского языка) [Текст] : автореф. дисс. ... докт. филол. наук : 10.02.19 / Чугаева Татьяна Николаевна ; С.-Петербург. ун-т. – СПб, 2009. – 46 с.
30. Чугаева, Т. Н. Перцептивный аспект звукового строя английского языка [Текст] / Т. Н. Чугаева. – Пермь, УрО РАН, 2007. – 246 с.
31. Штерн, А. С. Перцептивный аспект речевой деятельности [Текст] / А. С. Штерн. – СПб. : Изд-во С.-Петербург. ун-та, 1992. – 236 с.
32. Förstemann, E. Numerische Lautverhältnisse im Griechischen, Lateinischen und Deutschen [Text] / E. Förstemann // Zeitschrift für vergleichende Sprachforschung auf dem Gebiete des Deutschen, Griechischen und Lateinischen 1. Bd., 2. H. – 1852. – S. 163–179.
33. Hawkins, S. Looking for invariant correlates of linguistic units: two classical theories of speech perception [Text] / S. Hawkins // The acoustics of speech communication: fundamentals, speech perception theory, and technology / ed. by J. M. Pickett. – Boston ; London ; Toronto : Allin & Bacon, 1999. – P. 198–231.
34. Howes, D. H. Word Frequency, Personal Values, and Visual Duration Thresholds [Text] / D. H. Howes, R. L. Solomon // Psychological Review. – 1951. – Vol. 58. – P. 256–270.
35. Kaeding, F. W.: Häufigkeitswörterbuch der deutschen Sprache.Selbstverlag [Text] / Kaeding F.W // Linguistics and Linguistic Theory. –2009. – 5 (1). – S. 1–26.
36. Lemnitzer, L., Zinsmeister, H. Korpuslinguistik: eine Einführung [Text] / L. Lemnitzer, H. Zinsmeister. – Gunter Narr Verlag, 2006. – 220 S.
37. Meier, H. Deutsche Sprachstatistik [Text] / H. Meier. – Hildesheim, 1964.
38. Menzerath, P. Architektonik des deutschen Wortschatzes. [Text] / P. Menzerath. – Bonn, 1954. – 131 S.
39. Savin, H. B. Wordfrequency effect and errors in the perception of speech [Text] / H. B. Savin // Journal of the Acoustical Society of America. – 1963. – 35(2). – P. 200–206.
40. Sinclair, J. Corpus Concordance Collocation [Text] / J. Sinclair. – OUP, 1991. – 197 p.
41. Stubbs, M. Texts, corpora and problems of interpretation: A response to Widdowson [Text] / M. Stubbs // Applied Linguistics. – 2001. – Vol. 22. – Issue 2. – P. 149–172.
42. Tambovtsev, Yu. A. The Consonantal Coefficient in Selected Languages [Text] / Yu. A. Tambovtsev // Canadian Journal of Linguistics. – 1985. – № 30(2). – P. 179–188.

43. The British National Corpus. – URL : <https://corpus.byu.edu/bnc/> (дата обращения: 15.10.18).
44. The Open American National Corpus. – URL : <http://www.anc.org/> (дата обращения: 15.04.18).

References

1. Agibalov, A. K. (1995). *Veroyatnostnaya organizatsiya vnutrennego leksikona cheloveka* [The probabilistic organisation of an internal lexicon]. Author's abstract of PhD in Philological sci. diss. St.Petersburg : St.Petersburg University
2. Admoni, V. G. (1964). *Osnovy teorii grammatiki* [Grammar theoretic framework]. Leningrad: Publishing House Nauka.
3. Alekseev, P. M. (2001). *Chastotnye slovari: uchebnoe posobie* [Word-Frequency Dictionaries: A coursebook]. St Petersburg : St Petersburg University Press.
4. Andreyev, N. D., Zinder, L. R. (1963). О ponyat'yakh rechevogo akta, rechi, rechevoy deyatel'nosti i yazyka [On the notions of speech-act, speech, speech possibilities and language]. *Voprosy Yazykoznan'iya* [Topics in the Study of Language], 3, 15–21.
5. Bondarko, L. V. (1981). *Foneticheskoe opisanie yazyka i fonologicheskoe opisanie rechi* [Phonetic description of Language and phonological description of speech]. Leningrad : Leningrad University Press.
6. Ventsov, A. V. (2003). *Problemy vospriyatiya rechi* [The issues of speech perception]. Moscow : URSS Editorial.
7. Erofeeva, E. V. (2005). *Veroyatnostnye struktury idiomov: sotsiolingvisticheskiy aspekt* [Probabilistic structure of idioms: Sociolinguistic approach]. Perm : Perm State University Press.
8. Zalevskaya, A. A. (1990). *Slovo v leksikone cheloveka: Psikholingvisticheskoe issledovanie* [A word in a human lexicon: Psycholinguistic study]. Voronezh: Voronezh University.
9. Zazorina, L. N. (1977). *Chastotnyy slovar russkogo yazyka* [Russian language frequency dictionary]. Moscow : Russkiy yazyk Press.
10. Zakharov, V. P. (2011). *Korpusnaya lingvistika: uchebnyk* [Corpus linguistics: A course book]. Irkutsk : IGLU Press.
11. Zinder, L. R., Shtern, A. S. (1972). Faktory, vliyayushchie na opoznanie slova [Factors influencing word recognition]. *Proc. of the 4th All-Soviet-Union Symposium on psycholinguistics and communication theory* (pp. 100–108); Institute for Language Studies, Russian Academy of Sciences. Moscow.
12. Il'ish, B. A. (1971). *Stroy sovremennogo angliyskogo yazyka* [The system of the modern English language]. Leningrad : Prosveshchenie Press.
13. Kravets, A. S. (1976). *Priroda veroyatnosti (filosofskie aspekty)* [The nature of probability: Philosophical aspects]. Moscow : Mysl Press.
14. Krauze, M. (2002). *Dinamika mekhanizma vospriyatiya slova pri razlichnykh usloviyakh ovladeniya inostrannym yazykom* [Dynamics of word perception in various conditions of learning a foreign language]. Munich : Verlag Otto Sagner.
15. Malakhovskiy, L. V. (1980). Printsipy chastotnoy stratifikatsii slovarnogo sostava yazyka [The principles of frequency stratification of the language lexicon]. *Statistika*

- rechi i avtomaticheskiiy analiz teksta. Ch. 2: Statistika rechi* [Speech statistics and automatic text analysis. Part 2. Speech statistics]. Leningrad : Nauka Press.
16. Melnikov, G. P. (1971). Determinanta – vedushchaya grammaticheskaya tendentsiya yazyka [Determinanta as the leading grammatical tendency of the language]. *Fonetika. Fonologiya. Grammatika: K 70-letiyu A. A. Reformatskogo* [Phonetics. Phonology. Grammar: To the 70th anniversary of A. A. Reformatyky]. Moscow.
 17. Natsional'nyy korpus russkogo yazyka [Russian National Corpus]. Retrieved June 6, 2019 from <<http://www.ruscorpora.ru/>>.
 18. Nikonov, V. A. (1966). Fonostatisticheskoe izmerenie mezh"yazykovykh rasstoyaniy [Phonostatistical description of interlinguistic differences]. In S. K. Shaumyan (Ed.), *Issledovaniya po fonologii* [Studies in Phonology] (pp. 285–297). Moscow : Nauka Press.
 19. Plotkin, V. Ya. (1989). Stroy angliyskogo yazyka: ucheb. posobie [The system of English: A coursebook]. Moscow : Vysshaya Shkola Press.
 20. Segal, D. M. (1972). *Osnovy fonologicheskoy statistiki (na materiale polskogo yazyka)* [The foundation of the phonological statistics (Based on Polish)]. Moscow : Nauka Press.
 21. Smirnit'skiy, A. I. (1956). *Leksikologiya angliyskogo yazyka* [English Lexicology]. Moscow : Foreign Languages Publishing House.
 22. Smirnit'skiy, A. I. (1998). *Leksikologiya angliyskogo yazyka* [English Lexicology]. Moscow : Omen Press.
 23. Sysoyev, P. V. (2010) Lingvisticheskiiy korpus v metodike obucheniya inostrannym yazykam [Linguistic corpus in teaching foreign languages]. *Yazyk i kultura* [Language and Culture], 1, 99–111.
 24. Toporov, V. N. (1959). O vvedenii veroyatnosti v yazykoznanie [On introducing probability into the Study of languages]. *Voprosy Yazykoznaneniya* [Topics in the Study of Language], 6, 28–35.
 25. Uglanova, I. A. (2004). *Sub"ektivnaya otsenka chastoty slova i ee kategorizatsiya: ontolingvisticheskiiy aspekt* [Subjective evaluation of a word frequency and its categorisation: Ontolingvistic aspect]. PhD in Philological sci. diss. Perm : Perm State University.
 26. Urovni yazyka v rechevoy deyatelnosti [Language levels in speech] (1986). L. V. Bondarko (Ed.). Leningrad : Leningrad State University Press.
 27. Frumkina, R. M., Vasilevich, A. P. (1967). Veroyatnost slova i vospriyatie rechi [The probability of a word and speech perception]. In A. A. Leontyev, T. V. Ryabova (Eds.), *Voprosy porozhdeniya rechi i obucheniya yazyku* [The issues of speech production and language teaching] (pp. 17–38). Moscow : Moscow State University Press.
 28. Frumkina, P. M. (1971). *Veroyatnost elementov teksta i rechevoe povedenie* [The probability of the text elements and speech]. Moscow : Moscow State University Press.
 29. Chugaeva, T. N. (2009) *Zvukovoy stroy yazyka v pertseptivnom aspekte (eksperimentalnoe issledovanie na materiale angliyskogo yazyka)* [The perceptual aspect of the sound system of a language (An experimental study of English)]. Author's abstract of Doctoral in Philological sci. diss. St Petersburg : St Petersburg State University.
 30. Chugaeva, T. N. (2007). *Pertseptivnyy aspekt zvukovogo stroya angliyskogo yazyka* [Perceptual aspect of the sound system of the English language]. Author's abstract of Doctor in Philological sci. diss. Perm : UB RAS Press.

31. Shtern, A. S. (1992). *Pertseptivnyy aspekt rechevoy deyatel'nosti* [Perceptual aspect of speech activity: An experimental study]. St Petersburg : St Petersburg State University Press.
32. Förstemann, E. (1852). Numerische Lautverhältnisse im Griechischen, Lateinischen und Deutschen. *Zeitschrift für vergleichende Sprachforschung auf dem Gebiete des Deutschen, Griechischen und Lateinischen* 1. Bd., 2. H., 163–179.
33. Hawkins, S. (1999). Looking for invariant correlates of linguistic units: two classical theories of speech perception. In J. M. Pickett (Ed.), *The acoustics of speech communication: fundamentals, speech perception theory, and technology* (pp. 198–231). Boston ; London ; Toronto : Allin & Bacon.
34. Howes, D. H., Solomon, R. L. (1951). Word frequency, personal values, and visual duration thresholds. *Psychological Review*, 58, 256–270.
35. Kaeding, F. W. (2009). Häufigkeitswörterbuch der deutschen Sprache. Selbstverlag. *Linguistics and Linguistic Theory*, 5 (1), 1–26.
36. Lemnitzer, L., Zinsmeister, H. (2006). *Korpuslinguistik: Eine Einführung*. Gunter Narr Verlag.
37. Meier, H. (1964). *Deutsche Sprachstatistik*. Hildesheim.
38. Menzerath, P. (1954). *Architektonik des deutschen Wortschatzes*. Bonn.
39. Savin, H. B. (1963). Wordfrequency effect and errors in the perception of speech. *Journal of the Acoustical Society of America*, 35 (2), 200–206.
40. Sinclair, J. *Corpus Concordance Collocation*. OUP, 1991.
41. Stubbs, M. (2001). Texts, corpora and problems of interpretation: A response to Widdowson. *Applied Linguistics*, 22 (2), 149–172.
42. Tambovtsev, Yu. A. (1985). The Consonantal Coefficient in Selected Languages. *Canadian Journal of Linguistics*, 30 (2), 179–188.
43. The British National Corpus. – Retrieved: October 15, 2018 from <<https://corpus.byu.edu/bnc/>>.
44. The Open American National Corpus. – Retrieved: April 15, 2018 from <<http://www.anc.org/>>.