

УДК 81'342.3

ФОНЕТИЧЕСКИЕ ТИПЫ АМЕРИКАНСКОГО СЛОВА (НА МАТЕРИАЛЕ АМЕРИКАНСКОГО НАЦИОНАЛЬНОГО КОРПУСА)

Мякотникова С.Ю.

ГОУ «Пермский научно-исследовательский политехнический университет», Пермь, Россия (614000, г. Пермь, ул. Комсомольский пр., 29), e-mail: myakotnikova@yandex.ru

Корпусная лингвистика используется для анализа и исследования широкого ряда лингвистических явлений. В данном исследовании делается попытка построения классификации фонетических типов американского слова на анализе двух частотных страт Американского национального корпуса (ОАНК). Под фонетическим типом слова понимается группа слов, объединенных тем или иным существенным лингвистическим признаком. Основанием для выделения фонетических типов слова и описания его строевых характеристик может служить набор экспериментально выявленных существенных лингвистических признаков (СЛП). Описание общего облика частотного американского слова по лингвистическим признакам позволяет выявить специфические характеристики ядра американской лексики. Работа выполнена в рамках модели восприятия речи с опорой на лингвистические признаки, предложенной Санкт-Петербургской лингвистической школой.

Ключевые слова: АНК (Американский национальный корпус), существенные лингвистические признаки, фонетические типы слова.

THE PHONETIC TYPES OF AN AMERICAN WORD (ON THE MATERIAL OF THE AMERICAN NATIONAL CORPUS)

Myakotnikova S.Y.

Perm National Scientific Research Polytechnic University, Russian, Perm (614000, Perm, Komsomolsky av., 29), e-mail: myakotnikova@yandex.ru

Corpus linguistics is used to research a wide range of linguistic phenomena. In the current work there has been made an attempt to build the classification of phonetic types of an American word on the analysis of the two the most frequent strata of the American National Corpus (OANC). The set of essential linguistic features lie in the basis of phonetic types of the words. The description of an American word against the linguistic features can help to identify the specific characteristics of the frequent strata of the American lexicon. This work follows the model of speech perception put forward by St. Petersburg linguistic school that states that speech perception is based on the perception of essential linguistic features.

Keywords: American National Corpus, essential linguistic features, phonetic types of the words.

Целью исследования является выстраивание классификации фонетических типов американского слова. Исследование проводилось на материале высокочастотной и частотной страт Американского национального корпуса методом лингвостатистического анализа. Результаты исследования обсуждались на нескольких Международных лингвистических конференциях. На основании результатов были составлены таблицы слов, которые планируется использовать в эксперименте по восприятию изолированного частотного американского слова.

В последнее десятилетие невозможно обойтись без обширной эмпирической базы, которую предлагает корпусная лингвистика для проведения лингвистических исследований. По сравнению с традиционной лингвистикой корпусная имеет ряд очевидных отличий: она занимается описанием языка в том виде, в котором он проявил себя в речи, представленной в виде специально подобранного корпуса текстов. Основное внимание уделяется форме; она

предпочитает количественные методы исследования; и индуктивные методы обработки эмпирического словесного материала.

По определению С.Ю. Богдановой, корпусная лингвистика – это «во-первых, область компьютерной лингвистики, поскольку ее инструментарий связан с машинной обработкой материала. Во-вторых, это деятельность, направленная на создание корпусов текстов, под которыми понимаются репрезентативные собрания текстов в машиночитаемом формате с указанием на автора высказывания, место и время его создания, регистра (жанра), в котором оно употреблено, и др. В-третьих, это деятельность по использованию корпусов текстов для проверки известных лингвистических теорий и созданию новых на основании анализа обширного языкового материала. В-четвертых, это методология, которую можно применить ко многим аспектам языковых исследований» [2, с. 47].

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 др. Национальные корпусы включают не менее 100 млн словоупотреблений, что является залогом возможностей широкомасштабного изучения разноуровневых языковых единиц. Это коллекции устных и письменных текстов различных жанров, стилей, региональных и социальных вариантов, представленных в языке. Весь массив текстов в корпусе систематизирован. Это значит, что в корпусе зафиксировано расположение каждого слова в предложении по отношению к другим словам, а также учитывается частота его использования в данном корпусе [8].

Корпусные исследования имеют ряд неоспоримых преимуществ: «1) достаточно большой (репрезентативный) объем корпуса гарантирует типичность данных и обеспечивает полноту представления всего спектра языковых явлений; 2) данные разного типа находятся в корпусе в своей естественной контекстной форме, что создает возможность их всестороннего и объективного изучения; 3) однажды созданный и подготовленный массив данных может использоваться многократно, многими исследователями и в различных целях» [3, с. 3].

Корпусы текстов можно использовать для изучения динамики изменений лексических единиц и грамматических конструкций, для исследования особенностей грамматики устной речи, для выявления характеристик определенных типов дискурса и того, до какой степени отдельный текст соответствует моделям дискурса в данном регистре. Отдельно следует отметить возможности употребления корпусов в лингводидактике, а

также в новом наметившемся направлении современной лингвистики «когнитивные технологии в лингвистике» [2].

Выбирая корпус для проведения исследования, необходимо внимательно изучить его описание и определить, все ли задачи могут быть решены с его помощью. Так, например, для исследования американского варианта английского языка можно обратиться к Корпусу Брауна (Brown Corpus), Британскому национальному корпусу (BNC), Американскому национальному корпусу (Open American National Corpus - OANC), Корпусу современного американского английского (Corpus of Contemporary American English - COCA), Корпусу американского английского (Corpus of Historical American English - COHA).

Брауновский корпус, созданный в 1967 г. в Брауновском университете (США) учеными Н. Френсисом (Nelson Francis) и Г. Кучера (Henry Kucera), был первым большим корпусом текстов. Корпус содержал около 500 тысячесловных печатных текстов американского варианта английского языка, принадлежащих 15 наиболее массовым жанрам американской прозы (газетные статьи, рецензии, художественные тексты, научные тексты, тексты религиозного содержания, фантастика, биографии и т.п.) общим объемом около 1 млн слов. По мнению создателей, Брауновский корпус был достаточно представительным для американского варианта английского языка и предназначался для отражения лингвистических особенностей американской печатной прозы [8]. И хотя в свое время корпус интенсивно использовали, несомненно, что в настоящее время он не может применяться в широкомасштабных лингвистических исследованиях. Во-первых, потому, что корпус небольшой по объему, во-вторых, он не содержит образцов устной речи. Британский национальный корпус (BNC), созданный Дж. Личем в 1992 г., содержит 100 млн слов и включает образцы устной речи. Но для исследования американского варианта английского языка этот корпус не подходит из-за значительных лексических и синтаксических различий между двумя вариантами языка. Поэтому в конце XX века возникла острая необходимость создания Американского национального корпуса.

Американский национальный корпус (Open American National Corpus - OANC) [12] разрабатывается с 1990 года и включает образцы письменной и устной речи из разных источников. Он создает наиболее полную картину американского английского 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 завершении работы эта цифра увеличится до 100 миллионов слов. Слова расположены по частоте встречаемости в речи – от наиболее частотных до самых редких. Проект выполняется организацией, которая называется the Linguistic Data Consortium [13]. Этот консорциум, основанный в 1992 году по гранту

агентства Advanced Research Projects Agency (ARPA), включает университеты, компании и государственные исследовательские лаборатории. Организацию возглавляет Пенсильванский университет (The University of Pennsylvania).

В данном исследовании делается попытка построения классификации фонетических типов американского слова на анализе двух частотных страт Американского национального корпуса (ОАНК).

Для определения понятия «фонетический тип» слова целесообразно обратиться к модели восприятия речи с опорой на лингвистические признаки, которая была предложена Санкт-Петербургской лингвистической школой. А.С. Штерн в своей работе «Влияние лингвистических факторов на восприятие речи» [9] доказывает, что восприятие слов зависит от восприятия лингвистических признаков. Под лингвистическими признаками понимаются факторы, влияющие на восприятие речи. Они описывают природу лингвистических явлений разных языковых уровней: звукового, словесного и уровня предложений и текстов и имеют количественную и качественную природу. Среди факторов, влияющих на восприятие слова, называют *частотность слова, длину в слогах, ударную гласную, ритмическую структуру, длину в морфемах, консонантный коэффициент, начальный звук, часть речи*. Лингвистические признаки имеют градации и количественную и качественную природу. Так, например, фактор часть речи имеет градации существительное, прилагательное, глагол. Набор экспериментально выявленных существенных лингвистических признаков (СЛП) может служить основанием для выделения фонетических типов слова и описания его строевых характеристик. Под фонетическим типом слова понимается группа слов, объединенных тем или иным существенным лингвистическим признаком [10]. «Фонетические типы, т.е. группы слов, объединенные тем или иным признаком, характеризуются различными особенностями и вероятностными характеристиками в разных языках» [11, с. 20].

Т.Н. Чугаева в своем исследовании «Перцептивный аспект звукового строя английского языка» приводит результаты лингвостатистического анализа фонетического облика английского, немецкого, французского и русского слова, выполненного на материале соответственно Британского национального корпуса, частотного словаря немецкого языка Х. Мейера, словаря французского словаря Ж. Бодо и частотного словаря русского языка С.А. Шарова.

Для выявления фонетических типов американского слова был проведен лингвостатистический анализ частотных страт Американского национального корпуса [12].

Определение частотных страт осуществлялось с учетом принципов частотной стратификации словаря по Л.В. Малаховскому. Словарь был поделен на 5 частотных страт, неодинаковых по количеству включенных слов. Высокочастотная зона (сверхъядро) включает слова с 7-, 6- и 5-значными индексами частоты, частотная (ядро) с 4-значными, средняя с 3-значными, редкая – 2-значными, очень редкая – 1-значными [7].

Высокочастотная группа АНК является самой немногочисленной и насчитывает 187 словоформ; частотная включает 2030 словоформ; средняя исчисляется тысячами, а редкая миллионами словоформ. Анализ проводился по лингвистическим признакам, которые показали свою значимость в многочисленных экспериментах по восприятию английского слова [1; 4-6; 10]. Такими признаками оказались: *ударная гласная, длина в слогах, ритмическая структура, длина в морфемах, начальный звук, часть речи и консонантный коэффициент*. Кроме того, был проведен анализ консонантных типов односложных и двусложных словоформ, которые встретились в высокочастотной и частотной стратах.

Методом исследования был простой статистический подсчет. В каждой страте – высокочастотной и частотной - отбирались все слова, представляющие знаменательные части речи (существительные, прилагательные, глаголы и их формы, наречия) и содержащие градации перечисленных признаков. Полученные списки слов сравнивались между собой. Каждая страта рассматривалась отдельно. Из высокочастотной страты было отобрано и проанализировано 106, а из частотной 1787 словоформ.

Рассмотрим полученные результаты

По критерию *часть речи* в высокочастотной страте доля значимых слов составляет 62,5% от всего количества слов, а служебных – 37,5%. Сверхъядро содержит в 2 раза больше глаголов (41,5%), чем существительных и наречий, на втором месте наречия (20,6%), затем существительные (20%) и меньше всего прилагательных (15%). В сравнении частотная страта сформирована на 93% из значимых слов. На долю служебных слов приходится всего 7%. При этом самыми частотными являются существительные 58% от всего количества значимых слов. На втором месте находятся глаголы – 22%, на третьем прилагательные – 14%, на последнем наречия - 6% от всего количества значимых слов.

Интересно отметить, что если взять первую сотню слов АНК, то 48% в ней составляют служебные слова: числительные, местоимения, союзы, предлоги. При переходе из одной страты в другую доля служебных слов уменьшается в 5,3 раза. В частотной страте в 2,9 раза больше существительных и в 1,9 раза меньше глаголов, количество прилагательных примерно такое же, а наречий в 3,3 раза меньше.

По признаку *длина в слогах* сверхъядро АНК на 98% сформировано из «коротких» слов: односложных (80%) и двусложных слов (18%). Доля трехсложных слов незначительна, всего 2%.

В ядре количество односложных слов уменьшается почти в два раза, а двусложных наоборот возрастает. И тех и других – по 37%. В восемь раз возрастает количество трехсложных слов, их – 16,8%. Появляются и четырехсложные слова – 8,7%, однако полностью отсутствуют пятисложные слова.

Таким образом, мы можем описать сверхъядро, как почти полностью сформированное из «короткого» слова - 98%. В то время как на ядро приходится 74% коротких слов и 25,5% длинных слов. Тем не менее полученные данные говорят об абсолютном преобладании коротких слов в высокочастотной и частотной стратах АНК: 86% «коротких» слов в сравнении с 14% «длинных».

По признаку *ударная гласная* из общего количества проанализированных словоформ высокочастотный слой АНК включает 32% слов с дифтонгами в ударной позиции и 68% монофтонгами. Из них с краткой ударной 49%, долгой 12%, ретрофлексной гласной 7%. Сверхъядро АНК сформировано в основном из односложных слов, среди которых чаще всего встречаются слова с краткими гласными переднего ряда [ɛ] (cell), [ɪ] (think) - по 10% словоупотреблений от всего количества слов, [æ] - 8,8% (back), долгой [i:] - 7,7% (people), гласной центрального ряда [ʌ] - 8,8% (much) и дифтонгами [eɪ] - 7,7% (day) и [oʊ] - 8,8% (know) от всего количества словоупотреблений. Также встречаются ретрофлексные [ə] - 3% (work), [ɑə] - 2% (part), [oə] - 1% (more), краткая [ʊ] - 5% (good), долгая [u:] (new) - 4% и дифтонг – [aɪ] (time).

В частотном слое АНК 24% слов – это слова с дифтонгами в ударной позиции и 76% с монофтонгами; из них 42% с краткой гласной, 22% с долгой и 12% с ретрофлексной гласной. В ядре АНК самыми частотными являются краткие гласные [ɛ] - 14,3% (per'cent), [ɪ] - 11% (big), [æ] - 8% (hand) и [ʌ] - 6,4% (re'sult); долгие [i:] - 10% (deal), [ɑ:] - 9% (want) и дифтонг [eɪ] - 10% (take). Довольно часто встречаются долгая [u:] - 3,6% (soon), дифтонги [oʊ] - 4,7% (go) и [aɪ] - 6% (writing); ретрофлексные [ə] - 4,3% (girl), [ɑə] - 2,6% (large), [oə] - 3% (door).

Сверхъядро АНК сформировано в основном из односложных слов, среди которых чаще всего встречаются слова с краткими гласными переднего ряда [ɛ] (cell), [ɪ] (think) - по 10% словоупотреблений от всего количества слов, [æ] - 8,8% (back), долгой [i:] - 7,7% (people), гласной центрального ряда [ʌ] - 8,8% (much) и дифтонгами [eɪ] - 7,7% (day) и [oʊ] - 8,8% (know) от всего количества словоупотреблений. Также встречаются ретрофлексные [ə]

- 3% (work), [aə] - 2% (part), [oə] - 1% (more), краткая [ʊ] - 5% (good), долгая [u:] (new) - 4% и дифтонг – [aɪ] (time).

Несмотря на разное количество слов в высокочастотной и частотной стратах АНК, тенденции употребления определенных фонетических типов сохраняются. Исключениями являются долгая гласная [a:], которая присутствует в ядре, но почти совсем не встречается в высокочастотном слое, и дифтонг [oʊ], который в два раза чаще распространен в высокочастотном слое.

Ритмическая структура высокочастотного двусложного слова может характеризоваться как хореическая (95% слов). Из 19 двусложных слов только одно имеет основное ударение на второй слог, что составляет всего 5% из всех слов. В частотной страте значительно меняется соотношение между хореическими и ямбическими структурами. Среди двусложных слов количество слов с основным ударением на первый слог уменьшается на 20%. А количество слов с основным ударением на второй слог увеличивается на 25%. Соотношение между хореическими и ямбическими словами в частотной страте 70% к 30% соответственно.

В частотной страте резко увеличивается количество трехсложных слов: 283 слова. В сравнении с высокочастотной стратой, где было всего 2 трехсложных слова, одно из которых является начальнударным, а другое среднударным. Соотношение между градациями трехсложного частотного слова следующее: 54% начальнударных слов, 45% срединнударных и только 1% конечнударных слов.

Однако соотношение между градациями трехсложного высокочастотного и частотного слова сохраняется в общих чертах. Практически все «длинные» слова обеих страт являются начальнударными или срединнударными. Среди слов высокочастотной страты нет слов с основным ударением на последний слог, а в частотной страте их всего 3: understand, Japanese, employees.

По признаку *длина в морфемах* высокочастотная страта АНК на 82% состоит из одноморфемного слова, содержит 18% двуморфемных и совсем не включает трехморфемных и четырехморфемных слов. Частотная страта содержит 49% одноморфемных словоформ, 42% двуморфемных, 8% трехморфемных и 1% четырехморфемных словоформ. То есть сверхъядро АНК состоит в основном из одноморфемных слов, ядро же содержит примерно равное количество одно- и двуморфемных слов.

По признаку *начальный звук* исследование показало, что слова в обеих стратах начинаются с согласного звука: сверхъядро в 88,7% и ядро в 86%. Это почти в 7 раз больше, чем с гласного: сверхъядро - 11,3% и ядро - 14% слов.

Одной из основных характеристик слова при восприятии является соотношение гласных и согласных в слоге, или *консонантный коэффициент*, который определяется по формуле $k=C/V$ (отношение количества согласных к количеству гласных в слове). Для слов высокочастотной страты АНК этот коэффициент составил $k=1,54$. Для сравнения возьмем данные, которые приводит В.А. Никонов: для английского слова $k=1,52$; для немецкого $k=1,7 - 1,8$; для русского $k=1,38$. Таким образом, частотное американское слово оказалось средне консонантно нагруженным. Анализ выявил 13 самых распространенных консонантных типов. Для высокочастотной страты самыми встречаемыми оказались: CVC – 42% (got) от общего количества словоупотреблений; CV – 16%, CVCC – 11%, CVCV – 6%, VC – 3% от общего количества слов. Среди частотной страты самые распространенные типы – это CVC – 47%, CVCC – 24,4%, CCVC – 8,6%, CCVCC – 3,6%, CCV – 2,8%, VC – 2% от общего количества словоупотреблений.

Таким образом, частотное американское слово на 86% «короткое» слово, в котором ударной чаще всего является гласная переднего ряда [ε], [i], [i:] или дифтонг [ei]. По ритмической структуре оно начально- или срединноударное. Частотное американское слово средне консонантно нагруженное, с доминирующим типом слога CVC.

Описание общего облика частотного американского слова по лингвистическим признакам позволяет выявить специфические характеристики ядра американской лексики. «Сопоставление данных, полученных по одной методике на материале разных языков (английском, русском, немецком, французском), является интересным в плане их типологического сходства и с точки зрения их звукового строя» [10, с. 55].

Изучение представленности фонетических типов американского слова в разных частотных слоях устойчивого ядра языка представляется важным как при моделировании процессов восприятия английской речи, так и при обучении языку.

Список литературы

1. Байбурова О.В. Механизмы восприятия разносложных типов английского слова : дис. ... к. филол. наук. - Пермь, 2008. - С. 223.
2. Богданова С.Ю. Возможности корпусной методологии в решении лингвистических задач // Вестник Иркутского государственного лингвистического университета. – 2012. - № 2 (18). - С. 47-53.
3. Захаров В.П. Корпусная лингвистика : учебник / В.П. Захаров, С.Ю. Богданова. – Иркутск : ИГЛУ, 2011. – 161 с.

4. Кокорина Т.В., Чугаева Т.Н. Влияние контекста на восприятие английского слова // Ученые записки ПГТУ. - Пермь, 2002. - С. 61-67.
5. Конарева О.Ю., Чугаева Т.Н. Фонетическая интерференция при восприятии английского односложного слова // Проблемы социо- и психолингвистики... / отв. ред. Е.В. Ерофеева. - Пермь, 2002. - Вып. 3. - С. 83-67.
6. Ощепкова О.В. Влияние морфемной структуры на восприятие «длинного» английского слова в родном и неродном языках // Проблемы социо- и психолингвистики... / отв. ред. Е.В. Ерофеева. - Пермь, 2004. - Вып. 5. - С. 58-62.
7. Малаховский Л.В. Принципы частотной стратификации словарного состава языка // Статистика речи и автоматический анализ текста. - Л., 1980. - С. 99-105.
8. Сысоев П.В. Лингвистический корпус в методике обучения иностранным языкам. – URL: <http://www.lib.tsu.ru/mminfo/000349304/09/image/09-099.pdf>.
9. Штерн А.С. Влияние лингвистических факторов на восприятие речи : автореф. дис. ... канд. филол. наук. — Л. : ЛГУ, 1981. - С. 24.
10. Чугаева Т.Н. Перцептивный аспект звукового строя английского языка. – Пермь : УрО РАН, 2007. – 246 с.
11. Чугаева Т.Н. Перцептивный аспект звукового строя английского языка : дис. ... д-ра филол. наук. - СПб., 2009. - С. 346.
12. American National Corpus. - URL: <http://www.americannationalcorpus.org/>.
13. About LDC | Linguistic Data Consortium. - URL: <http://www.ldc.upenn.edu/About>.

Рецензенты:

Чугаева Т.Н., д.фил.н., зав. кафедрой иностранных языков и философии Пермского научного центра Уральского отделения Российской академии наук, г. Пермь.

Байкова О.В., д.фил.н., зав. кафедрой лингвистики и перевода Вятского государственного гуманитарного университета, г. Киров.