

УДК 811.512. 122

**КОРПУС КАЗАХСКОГО ЯЗЫКА: ПРОСОДИЧЕСКАЯ РАЗМЕТКА****Базарбаева З.М.***Институт языкознания им. А. Байтурсынова МОН РК, Алматы, e-mail: Zeynepmb@mail.ru*

Целью данной работы является описание просодической разметки публицистического текста для разработки и создания национального корпуса казахского языка. Для достижения цели используются экспериментально-фонетические методы исследования, компьютерная программа. В результате анализа был представлен аннотированный звучащий текст с просодической разметкой, включающий деление на синтагмы с обозначением направления движения основного тона и пауз. Членение на синтагмы определяется смысловым содержанием высказывания, степенью распространенности главных и второстепенных членов. Выявлено, что одно и то же высказывание в зависимости от интенции говорящего может быть произнесено с различным синтагматическим членением. В основе просодической разметки лежит семантико-интонационное деление звучащего текста. С помощью автоматизированной обработки электронных корпусов были выявлены речевые закономерности звучащего текста.

**Ключевые слова:** корпусная лингвистика, просодическая разметка, аннотирование, синтагма, звучащий текст, сегментация, пауза

**CORPUS KAZAKH: PROSODIC MARKING****Bazarbayeva Z.M.***DLitt, Professor of A. Baitursynov Linguistics Institute, Almaty, e-mail: Zeynepmb@mail.ru*

The aim of this paper is to describe prosodic marking journalistic text for the development and establishment of a national body of the Kazakh language. To achieve the objective used experimentally-phonetic research methods, computer program. The analysis was presented annotated sounding text prosodic markings, including the division of Syntagma to indicate the direction of the pitch and pauses. The division Syntagma defined semantic content of speech, the prevalence of primary and secondary members. It was revealed that the same sentence can be pronounced in a different syntagmatic articulation depending on the intentions of the speaker. The basis of prosodic markup is semantic and intonation division sounding text. With the help of the automated processing of electronic housings were identified speech patterns sounding text.

**Keywords:** corpus linguistics, prosodic markup, annotation, syntagma, sounding text, segmentation, pause

В настоящее время в европейском, российском, а также в казахском (последние несколько лет) языкознании все больший интерес вызывают проблемы корпусной лингвистики. Зарождение корпусной лингвистики связано с развитием компьютерных технологий и программного обеспечения, которые дали возможность обработки и разметки больших по объему текстов в электронной форме, аккумулирующих технологии баз данных и информационного поиска. Автоматизированная обработка электронных корпусов дает возможность более точной интерпретации полученных данных.

Авторитет в области корпусной лингвистики В.П. Захаров определяет это направление как «большой, представленный в электронном виде, унифицированный, структурированный, размеченный, филологически компетентный массив языковых данных, предназначенный для решения конкретных лингвистических задач» [2]. Иначе говоря, обработав и систематизировав с помощью компьютерной техники большие объемы текстов, можно статистически обосновывать лингвистические явления, проследить диахронические изменения в речи, найти всю интересующую инфор-

мацию о языке с целью изучения речевых закономерностей для построения функциональной модели языка. Создание, развитие и использование электронных корпусов – это одно из наиболее передовых направлений современной лингвистики.

**Целью** данной работы является описание просодической разметки публицистического текста для разработки и создания национального корпуса казахского языка.

**Материалы и методы исследования**

В работе использовались экспериментально-фонетические методы исследования, а также специально созданная программистами-информатиками компьютерная программа аннотирования звучащего текста. Материалом для анализа послужили отрывки из публицистического текста, начитанные диктором, владеющим орфоэпической нормой казахского языка.

**Результаты исследования и их обсуждение**

Специалисты в области корпусной лингвистики придерживаются определенной лингвистической теории, которая лежит в основе разметки каждого корпуса, поэтому выводы по корпусным данным могут производиться лишь в рамках этой кон-

цепции. Однако в зависимости от структуры языка некоторые теории формализации могут не подходить ко всем исследуемым языкам. В этом случае исследователь находит наиболее приемлемые пути лингвистического аннотирования текстовых данных, учитывающих специфику данного языка.

Выделяют различные виды корпусной лингвистики: письменные и устные (звуковые) тексты, параллельные (тексты нескольких языков), по стилю (публицистические, художественные, официально-деловые, научные, разговорные), по хронологии (синхронические, диахронические) и др. Слишком подробная лингвистическая информация, заключенная в разметке больших по объему корпусов текстов, может быть избыточна, поэтому, как правило, обходятся необходимым в данном исследовании набором меток в зависимости от поставленной цели.

Лингвистическое аннотирование может включать несколько видов разметок: синтаксическую, лексико-семантическую, морфологическую и просодическую [3]. В русском языкознании самым разработанным является морфологическая разметка. С недавних пор в казахском языкознании на основе имеющихся методик аннотирования электронных текстов также начали заниматься морфологической разметкой. Этот вид разметки предусматривает распределение электронных корпусов по частям речи с их грамматическими категориями. На основе правил словообразования большие массивы текстов подвергаются автоматизированной обработке. Аннотирование корпусов, как было сказано выше, осуществляется программными средствами. Иногда из-за отсутствия программных средств некоторые виды разметок осуществляются вручную. Это чаще всего касается просодической разметки, где наряду с автоматизированной обработкой текстов используется и ручная обработка.

Методика просодической разметки устных текстов посредством компьютерных программ наименее разработана в корпусной лингвистике. В русском языке основой просодической разметки является деление на синтагмы и метки ударения. Учитывая фонетическую систему казахского языка просодическая разметка казахских текстов предусматривает синтагматическое деление транскрибированной звучащей речи с обозначением направления движения тона, разметкой синтагм на акцентные единицы (ритмические группы) и пауз хезитации.

Для просодической разметки был выбран озвученный дикторами публицистический текст «Театр». По правилам аудитор-

ского анализа звучащий текст был поделен на синтагмы и акцентные единицы. Каждая синтагма оформлялась соответствующим мелодическим контуром (восходящим, нисходящим, восходяще-нисходящим, нисходяще-восходящим или ровным). В соответствии с задачами исследования озвученные тексты анализировались специальной компьютерной программой, разработанной программистами-информатиками, которая дает возможность фиксировать осциллограмму речевого сигнала, озвученный текст, начало и конец высказывания и синтагмы с обозначением направления движения основного тона синтагмы. Посредством разработанной для просодической разметки компьютерной программы был проанализирован озвученный публицистический текст. В процессе разметки обращалось внимание на процедуру анализа и сопоставления речевых сигналов. При делении текста на синтагмы было важно правильно ставить границу синтагмы, чтобы не нарушалось смысловое восприятие речи. Правила синтагматического членения базируются на грамматическом (пунктуационном, морфологическом и синтаксическом) анализе текста, а также на ситуативном анализе сегментации в звучащей речи. Установка границ синтагм влияет на передачу интонационных характеристик а также на передачу смыслового содержания. Под синтагмой понимается универсальная, характерная для всех языков, смысловая единица, возникающая при сегментации высказывания, текста. Минимальным объемом речевого отрезка, в котором проявляется синтаксическое и интонационное значение является синтагма, несущая большую функциональную нагрузку в тексте [1].

Как показала просодическая разметка, синтагма в зависимости от речевой ситуации и контекста может оформляться словом, словосочетанием и предложением, выражая сложное смысловое единство. Синтагма универсальна и встречается во всех языках, формирование синтагм в потоке речи определяется не жесткими правилами, а обусловлено разнообразными смысловыми связями внутри текста, а также зависит от хода мысли говорящего (диктора), от его видения и понимания ситуации. Например:

Театр өнерінің ең басты құралы – // сөз. (Главное средство театрального искусства – // это слово).

Онсыз әдебиет те, // драматургия да, // қарым-қатынас та болмақ емес. (Без него не может существовать ни литература, // ни драматургия, // ни коммуникация).

Различные синтагмы обладают большей или меньшей смысловой целостностью.

При недостаточности объема синтаксическое значение остается интонационно невыраженным. Например, незавершенность может выражать подчинительные отношения, если подчинительная связь в простом высказывании устанавливается между словами, не образующими синтагму, то она не будет иметь интонационного выражения. В том случае, если слова и словосочетания в простых высказываниях соответствуют синтагме, то они будут выражаться незавершенной интонацией, в частности, в высказываниях с однородными членами. Например:

Оку техникасы фонетика, // грамматика, // орфоэпия, // орфография, // стилистикамен ұштасып жатыр. (Техника чтения связана с грамматикой, // орфоэпией, // орфографией, // стилистикой).

В сложном же высказывании, где его компоненты являются отдельными синтагмами, подчинительная связь также будет выражена незавершенной интонацией. Простое нераспространенное высказывание, образующее одну синтагму, оформляется завершенной интонацией. Например, следующее высказывание *Интонация өте маңызды орын алады. (Интонация играет большую роль)* состоит из одной синтагмы, передающей значение законченности и независимости. Оба эти значения передает нисходящая интонация. Но интонационная скрытая внутренняя зависимость друг от друга членов предложения, существующая благодаря подчинительным связям между ними, обнаруживается при их распространности:

Сахна тілінде // интонация өте маңызды орын алады. (В сценической речи // интонация играет большую роль).

Отношения зависимости между членами предложения становятся интонационно выраженными и передаются с помощью восходящей интонации. Восходящий тон, который оформляет неконечные синтагмы, наряду со значением незаконченности передает и эти отношения зависимости. Следовательно, если в объеме синтагмы происходит полное совпадение синтаксических и интонационных значений на основании двух значимых критериев, то интонация в этих конструкциях безвариантна, т.е. может быть использован только один интонационный тип – восходящий или нисходящий. В данном случае под типом понимается движение основного тона, и под вариантом – возможность использования в одной синтагме фонологически противоположного интонационного типа.

В просодической разметке различают грамматические паузы, отделяющие друг

от друга интонационно-оформленные синтагмы и фразы, а также паузы хезитации (неуверенности). Грамматические паузы принимают участие в передаче определенных синтаксических, интонационных и смысловых отношений. Граница акцентных единиц может быть промаркирована переломом тонального контура, сменой просодических характеристик, которые воспринимаются как нарушение плавного течения речи. Синтагма по сравнению с акцентной единицей (ритмической группой) являясь единицей более высокого порядка, шире определяет характеристики предметов и явлений, о которых говорится в речи. Минимальные фонетические и семантико-интонационные единицы, каковыми являются ритмические группы и синтагмы, определяют семантико-интонационное деление речи (текста).

Границы семантико-интонационных, синтаксических единиц синтагм передаются посредством синтагматических ударений. Являясь предельным элементом фразировки, она является результатом синтаксико-стилистического и лексико-фразеологического членения речевого потока. Границы синтагмы в речи достаточно подвижны и часто находят отражение в пунктуации, однако знаки препинания не всегда совпадают с ее границей. Будучи наименьшей знаменательной частью высказывания, синтагма оформляется посредством интонационных средств.

Как показывает просодическая разметка, членение на синтагмы определяется смысловым содержанием высказывания, степенью распространенности главных и второстепенных членов. Одно и то же высказывание в зависимости от интенции говорящего может быть произнесено с различным синтагматическим членением. Рассмотрим следующий пример, озвученный диктором: *Ән өнері ерте заманнан бастау алатыны белгілі. (Известно, что певческое искусство зародилось очень давно)*. Это высказывание может состоять из одной синтагмы, из двух и трех в зависимости от контекста:

Ән өнері ерте заманнан бастау алатыны белгілі.

(Известно, что певческое искусство зародилось очень давно).

Ән өнері // ерте заманнан бастау алатыны белгілі.

(Известно, // что певческое искусство зародилось очень давно).

Ән өнері // ерте заманнан // бастау алатыны белгілі.

(Известно, // что певческое искусство // зародилось очень давно).

Такое соотношение частей высказывания обусловлено определенным осмыслением ситуации говорящим или слушающим. Интонационное оформление синтагмы, в нашем анализе, обеспечивалось мелодическими и темпоральными средствами. Когда несколько ритмических групп объединяются в синтагму, происходит усиление синтагматического ударения за счет ослабления предшествующих ритмических ударений.

#### **Выводы**

Таким образом, просодическая разметка текста показала, что фонетическая целостность синтагмы в речи создается отсутствием паузы внутри синтагмы и общим мелодическим контуром. Синтагмы оформлялись восходящим, нисходящим, восходяще-нисходящим и нисходяще-восходящим

направлением движения основного тона. С помощью автоматизированной обработки электронных корпусов были описаны речевые закономерности звучащего текста. В основе просодической разметки речевого потока лежит семантико-интонационная сегментация. Было выявлено, что просодическое аннотирование звучащего текста определяется паузацией, направлением движения основного тона и переломом мелодического контура на стыках анализируемых сегментов.

#### **Список литературы**

1. Базарбаева З.М. Казахская интонация. – Алматы, 2008. – 281 с.
2. Захаров В.П. Корпусная лингвистика: Учебно-методическое пособие. – СПб., 2005. – 48 с.
3. Плулуня В.А. (ред.). Национальный корпус русского языка: 2003–2005. – М., 2005. – 150 с.