

СТАТИСТИЧЕСКИЕ ХАРАКТЕРИСТИКИ СИНТАГМАТИЧЕСКОГО ЧЛЕНЕНИЯ ПРЕДЛОЖЕНИЙ В ПРИЛОЖЕНИИ К СИНТЕЗУ ВЫРАЗИТЕЛЬНОЙ РЕЧИ ПО ТЕКСТУ

Б. М. Лобанов (lobanov@newman.bas-net.by)

Ю. С. Гецевич (mix1122@gmail.com)

Объединенный институт проблем информатики
НАН Беларуси, Минск, Беларусь

В докладе описываются используемый экспериментальный материал, методика и результаты статистической обработки текстового и звукового файлов. Приводятся статистические характеристики особенностей членения предложений на синтагмы, реализуемого в процессе выразительного чтения художественного текста профессиональным диктором (актёром). Описываются пути использования полученных статистических данных в приложении к синтезу выразительной речи по тексту.

Ключевые слова: выразительная речь, синтез речи, синтез выразительной речи, синтагматическое членение.

STATISTICAL CHARACTERISTICS OF SYNTAGMATIC SEGMENTATION OF UTTERANCES FROM THE VIEWPOINT OF EXPRESSIVE TEXT-TO-SPEECH SYNTHESIS

B. M. Lobanov (lobanov@newman.bas-net.by)

Iu. S. Getsevich (mix1122@gmail.com)

Institute of Informatics Problems NAS Belarus, Minsk, Belarus

We describe the results of a statistical study of text segmentation into phrases that occurs during expressive reading of Russian fiction by a professional speaker (actor). The purpose is to find out whether part-of-speech tags could be used to predict breaks between phrases in a sentence. The experimental material was Anton Chekhov's story, *A Hunting Drama*, presented in text (54 thousand words) and sound formats (an audio book with 7 hrs playing time). This material was divided into two parts: the initial segment of the tagged text of the story containing 420 sentences (ca. 6000 words) and the rest of the text (untagged). The untagged part was used for model evaluation. Prosodic phrases were manually tagged by a professional auditor — phonetician who listened to the text. The total number of tagged phrases in the initial 420 sentences was 1516 (of which 710 had pauses no longer than 100 msec and 380 had longer pauses). The average number of phrase breaks in a sentence was 3.6, while the average length of a phrase was 4 words. Pairs consisting of words belonging to 11 different parts of speech or POS-like morphological classes were investigated: adjective, adverb, conjunction, gerund, interjection, parenthetical word, noun, numeral, participle, pronoun, and finite verb. In addition to POS information, the statistical analysis takes account of punctuation marks appearing in the sentence (commas, hyphens, dashes, colons, semicolons and parentheses). Quantitative distributions have also been obtained for phrase breaks occurring in the pairs: "punctuation mark — part of speech", "part of speech — punctuation mark", "space — part of speech", "part of speech — space". Potentials of using this data in expressive text-to-speech synthesis system are considered.

Key words: expressive text-to-speech, speech synthesis, expressive text-to-speech synthesis, syntagmatic segmentation.

Введение

К настоящему времени системы синтеза речи достигли определённого уровня развития и уже используются в ряде практических приложений. Однако комфортность восприятия синтезированной речи в реальных условиях систем массового обслуживания остаётся ещё не вполне удовлетворительной. Мировая тенденция развития речевых технологий указывает на актуальность создания систем синтеза выразительной речи (expressive text-to-speech) [1–3]. Понятие «выразительность речи» сформировалось как междисциплинарное понятие, характеризующее одну из функций устной речи человека [4]. Речевой опыт каждого из нас говорит о том, что, например, два доклада, прочтённые на одну и ту же тему, могут оказать на человека совершенно разный эффект, зависящий от степени выразительности речи диктора.

Одним из главных компонентов звуковой реализации выразительности устной речи является просодика речи и, в частности, при синтезе речи — правильность и качество просодической разметки. Просодическая разметка текста при синтезе речи заключается в членении предложений на синтагмы, в маркировке просодически выделенных слов в синтагме и в установке интонационного типа синтагмы [5]. В данной работе исследуется первый из указанных аспектов просодической разметки. Установка границ синтагм влияет на правильность передачи интонационных характеристик, а также на передачу смыслового содержания текста. Следует отметить, что при членении предложений на синтагмы особенно важно не поставить её границу там, где она может нарушить смысловое восприятие речи, например, между предметом и его признаком.

Подробный обзор и анализ проблем, связанных с локализацией синтагматических границ в естественной речи и возможных подходов к установлению границ синтагм при синтезе речи по тексту дан в работе О. Ф. Кривновой и И. С. Чардина [6]. При этом рассмотрены следующие возможные классы систем синтагматического членения предложений:

1. Системы, которые обходятся анализом структуры текста с помощью обнаруженных эвристик (экспертные системы).
2. Системы, в которых проводится синтаксический анализ с использованием формальных грамматик.
3. Системы, где используется вероятностный анализ текста, основанный на статистической модели, параметры которой получены через обучение по аннотированной тексто-речевой базе данных.

Для систем синтеза русской речи по тексту, относящихся к первому классу, были предложены достаточно простые правила синтагматического членения предложений, основанные на морфологической информации о словосочетаниях [7, 8]. Прототип системы синтеза русской речи, в которой используется глубокий синтаксический анализ, описан в [9]. Системы, где используется вероятностный анализ текста, разработаны для синтеза речи на ряде европейских языков (см., например, [10]). Первая попытка использования статистических особенностей для синтагматического членения русской речи на небольшом по объёму текстовом и аудио материале была описана в [11]. В данной работе продолжено развитие этого подхода с более детальным рассмотрением различных ситуаций

синтагматического членения. При этом использованы аудиозаписи выразительного чтения художественного текста профессиональным диктором (актёром) и существенно расширенный экспериментальный материал.

1. Экспериментальный материал

В качестве экспериментального материала использована повесть А. П. Чехова «Драма на охоте», представленная в текстовой форме (54 тыс. слов) и в звуковой (аудиокнига с временем звучания — 7 часов в исполнении профессионального диктора А. Балакирева). Экспериментальный материал состоит из двух частей: начального размеченного отрезка повести, состоящего из 420 предложений (около 6000 слов) и остальная неразмеченная её часть. Неразмеченная часть повести использовалась затем для контрольных экспериментов.

Текстовый и звуковой файлы начального отрезка текста предварительно были разбиты на отдельные предложения. Разметка на просодические синтагмы осуществлялась профессиональным аудитором-фонетистом в процессе прослушивания отдельных предложений. Затем эта разметка переносилась на звуковые файлы. Знаком [/] (короткая пауза) аудитором отмечались границы синтагм в тех случаях, когда отсутствовала заметная физическая пауза звука (менее 100 мс), а знаком [//] (долгая пауза) — когда физическая пауза присутствовала (более 100 мс). Наличие границы синтагмы при отсутствии физической звуковой паузы определялась аудитором на основе его представлений о достаточной в интонационном смысле самостоятельности синтагмы.

Общие количественные характеристики используемого размеченного экспериментального материала представлены в таблице 1.

Таблица 1. Количественные характеристики используемого текста

Общ. кол. слов в тексте	Общ. кол. предл. в тексте	Общ. кол. внутр. зн.преп.	Общ. колич. синтагм в тексте	Кол. синтагм с короткой паузой	Кол. синтагм с длинной паузой	Средн. кол. синт. в предл.	Средн. кол. фонетич. слов в синт.
6234	424	485	1516	710	382	3,6	4,1

Ниже приведены 4 примера аудиторской разметки предложений на синтагмы.

- (1) В один из апрельских полудней / тысяча восемьсот восьмидесятого года // в мой кабинет вошел сторож Андрей / и таинственно доложил мне, / что в редакцию явился какой-то господин // и убедительно просит свидания с редактором.
- (2) К пишущим людям / не имею чести / принадлежать, // но, тем не менее, явился к вам / с чисто писательскими целями.

- (3) Был, знаете ли, // судебным следователем в уезде, // прослужил пять с лишком лет, / но ни капитала не нажил, // ни невинности не сохранил...
- (4) Лентяя вдруг ни с того ни с сего / осенила мысль, / что нос моего попугая / очень похож на нос / нашего деревенского лавочника / Ивана Демьяныча, // и с той поры / за попугаем навсегда осталось имя / и отчество // длинноносого лавочника.

Очевидно, что приведённые выше варианты разметки предложений на синтагмы не являются единственно возможными, однако, они объективно отражают те предпочтения, которым следовал данный диктор при их выразительном чтении.

2. Методика и результаты статистической обработки экспериментального материала

Для всех последующих этапов статистической обработки каждое слово анализируемого текста маркируется его грамматическим типом (в нашем исследовании названием части речи). Всего в нашем исследовании используется 11 различных частей речи: *Вводное слово, Глагол, Деепричастие, Междометие, Местоимение, Наречие, Прилагательное, Причастие, Союз, Существительное, Числительное*. В этом списке отсутствуют предлоги и частицы, которые считаются присоединёнными к соседним словам в соответствии с известными правилами образования фонетических слов, а также предикаты, которые мы условно присоединили в одну группу с наречиями. Количественное распределение встречаемости частей речи в анализируемом тексте представлено на рис. 1.

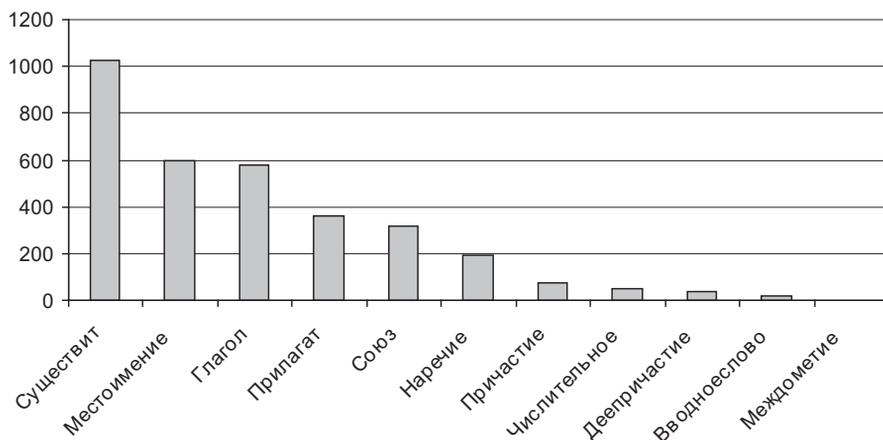
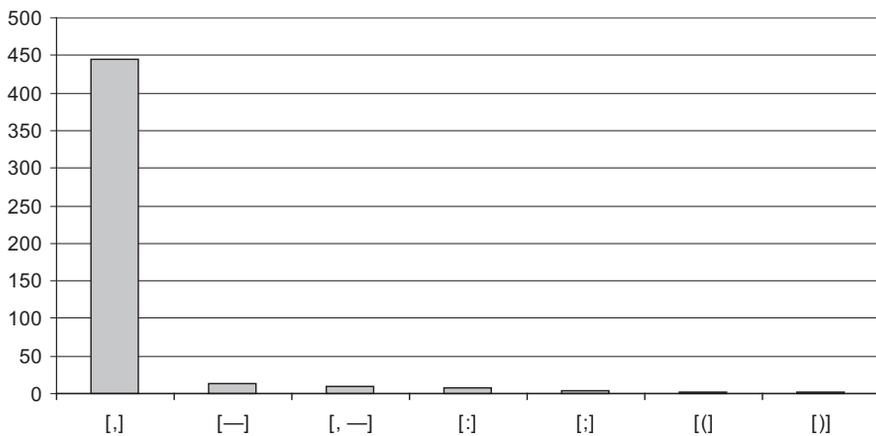
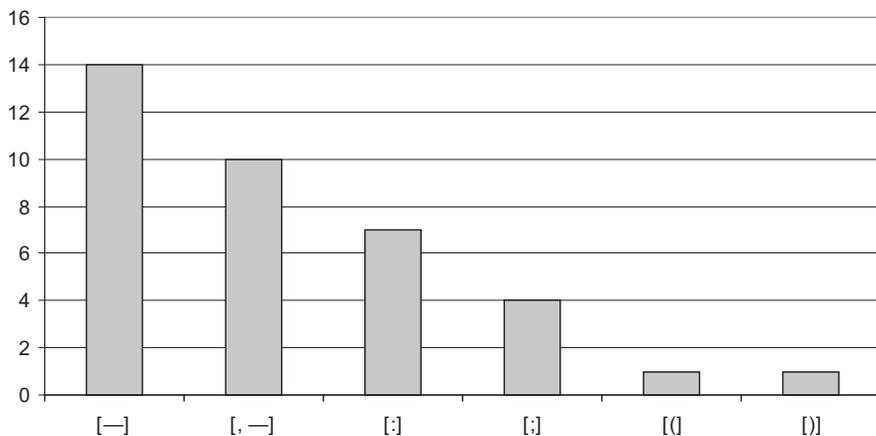


Рис. 1. Количественное распределение частей речи в анализируемом тексте

Кроме информации о частях речи при статистической обработке принимаются во внимание также присутствующие внутри предложений знаки препинания: (,) — запятая, (—) — тире, (,—) — запятая с тире, (:) — двоеточие, (;) — точка с запятой, (()) — открывающая и ()) — закрывающая скобки. Количественное распределение встречаемости знаков препинания в анализируемом тексте представлено на рис. 2а,б.



(А)

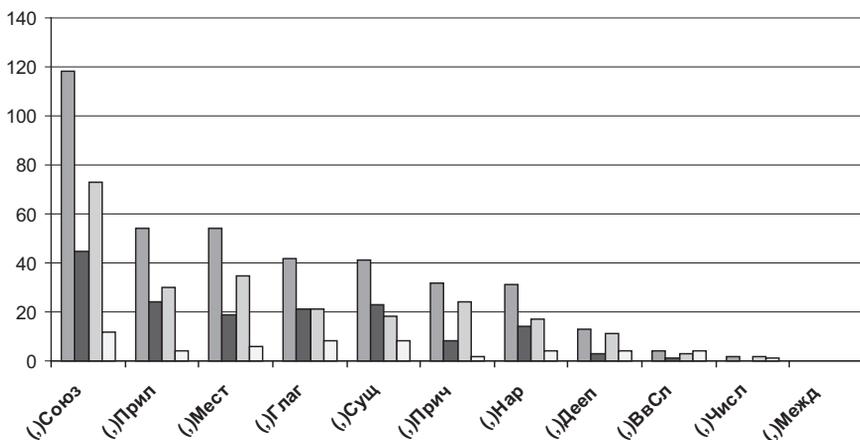


(Б)

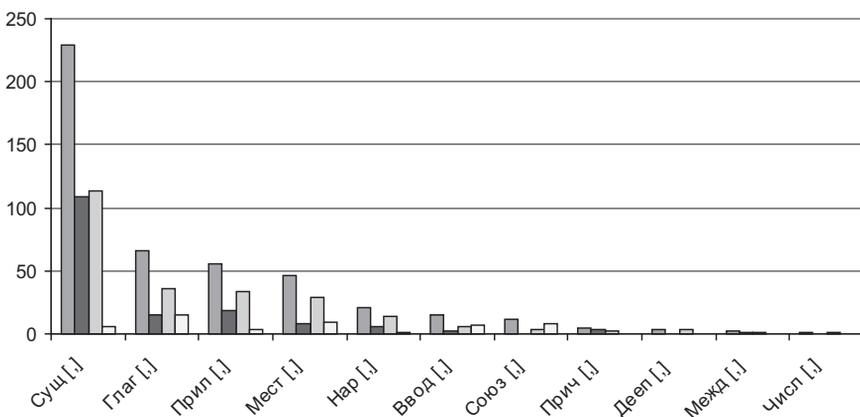
Рис. 2. Количественное распределение встречаемости всех знаков препинания (А) и малочастотных знаков (Б)

Дальнейшее исследование экспериментального материала посвящено подсчёту и анализу количественных распределений встречаемости синтагматической границы (с долгой или с короткой паузой) или её отсутствия для следующих пар:

- «знак препинания — часть речи»,
- «часть речи — знак препинания»,
- «пробел — часть речи»,
- «часть речи — пробел».



(А)

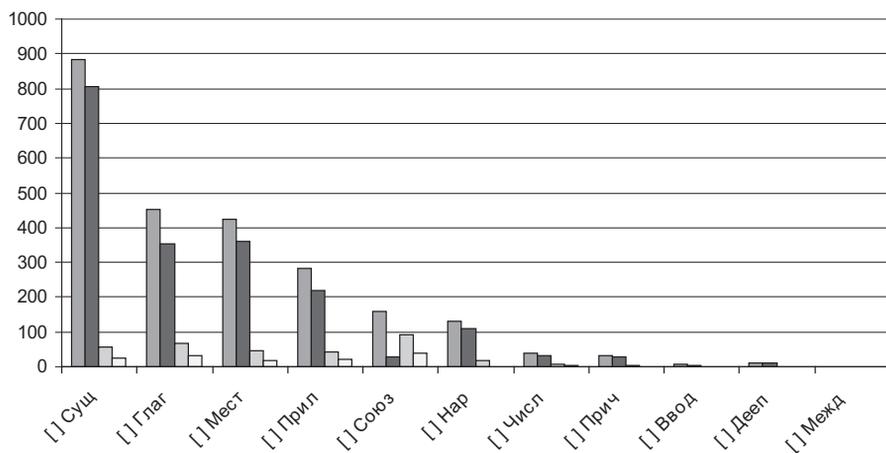


(Б)

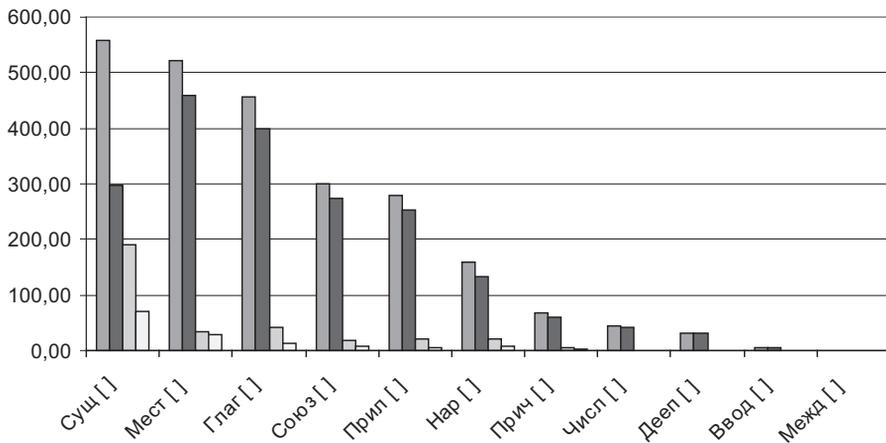
Рис. 3. Количественное распределение встречаемости синтагматической границы в присутствии знака препинания для пар: «запятая — часть речи» (3А), «часть речи — запятая» (3Б) (всего пар — ряд 1, долгая пауза — ряд 2, короткая пауза — ряд 3, нет границы — ряд 4)

В последних двух случаях под пробелом понимается пробел между фонетическими словами в отсутствие между ними знака препинания.

На рис.3а представлено количественное распределение встречаемости синтагматической границы для пар: «запятая — часть речи», а на рис. 3б — для пар «часть речи — запятая».



(А)



(Б)

Рис. 4. Количественное распределение встречаемости синтагматической границы в отсутствие знака препинания для пар: «пробел — часть речи» (4а), «часть речи — пробел» (4б) (всего пар — ряд 1, нет границы — ряд 2, долгая пауза — ряд 3, короткая пауза — ряд 4)

Общий вид распределений на рисунках 3а,б позволяет сделать вывод о существенно большей частоте встречаемости короткой или долгой синтагматической границы в присутствии знака препинания в сравнении с частотой отсутствия границы. Из приведенных рисунков видно также, что при выразительном чтении границы синтагм реализуются предпочтительно в виде короткой паузы.

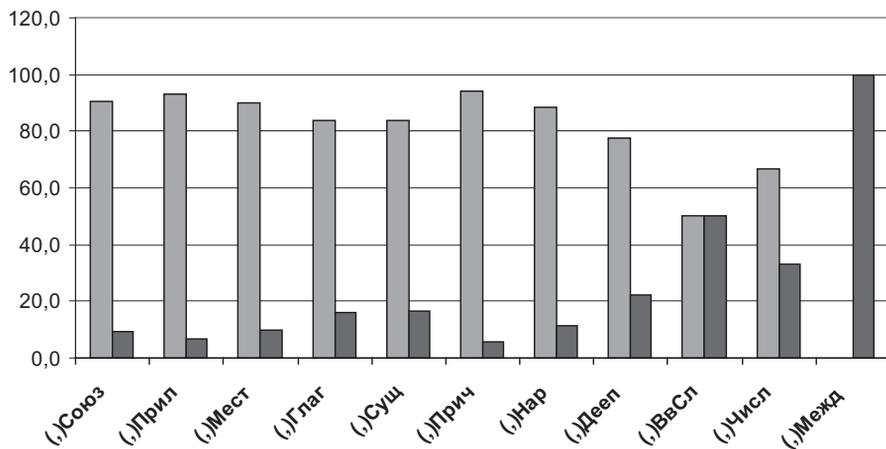
На рис. 4а представлено количественное распределение встречаемости синтагматической границы в отсутствие знака препинания для пар: «пробел — часть речи», а на рис. 4б — для пар «часть речи — пробел».

Общий вид распределений на рисунках 4а,б позволяет сделать вывод, что при отсутствии знака препинания, в отличие от случая рис. 3а,б, частота отсутствия синтагматической границы существенно выше в сравнении с частотой присутствия синтагматической паузы (короткой или долгой). Из приведённых рисунков видно также, что при выразительном чтении, в сравнении с рассмотренным ранее случаем, границы синтагм реализуются (за некоторыми исключениями) в виде долгой паузы.

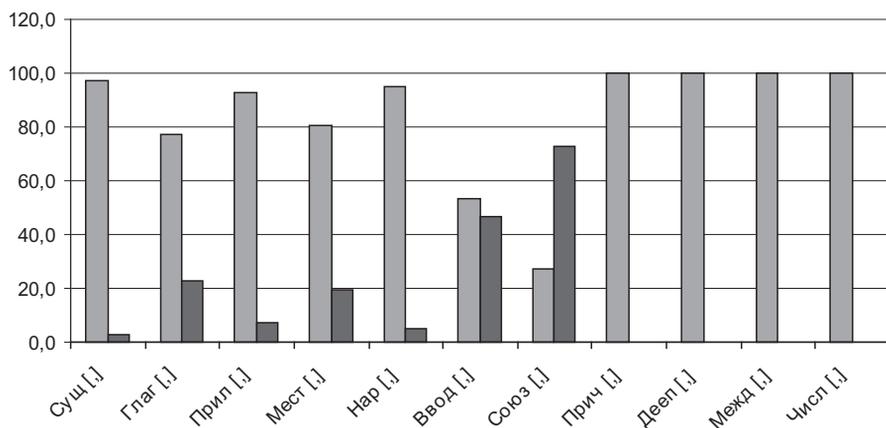
3. Использование статистических данных в приложении к синтезу выразительной речи по тексту

Полученные статистические характеристики могут быть положены в основу вероятностного алгоритма синтагматического членения предложений в системе синтеза речи по тексту. Для этого приведенные на рисунках 3, 4 данные преобразуем в форму распределений статистической вероятности (в %) наступления полной группы событий: «отсутствие — наличие» синтагматической границы. На рис. 5 представлены распределения вероятностей встречаемости синтагматической границы при наличии знака препинания для пар: «запятая — часть речи» (рис. 5а) и «часть речи — запятая» (рис. 5б), а на рис. 6 — распределения вероятностей встречаемости синтагматической границы в отсутствие знака препинания между словами для пар: «пробел — часть речи» (рис. 6а) и «часть речи — пробел» (рис. 6б)

Исходя из данных, представленных на рис. 5, можно сделать следующие основные выводы. Вне зависимости от типа части речи, следующей за знаком препинания (рис. 5а), синтагматическая граница (с короткой либо долгой паузой) присутствует в подавляющем большинстве случаев. Исключение составляют пары с междометиями, вводными словами и числительными, где вероятность отсутствия синтагматической границы весьма высокая. Это хорошо согласуется с нашими априорными представлениями. Похожие выводы можно сделать и исходя из анализа данных рис. 5б. Вероятность отсутствия синтагматической границы весьма высокая для пар «союз — (,)» и «вводное слово — (,)». Отметим, что полученное 100% присутствие синтагматической границы для пар с причастиями, деепричастиями, вводными словами и междометиями хотя и правдоподобно, но объясняется, скорее всего, их низкой частотностью в исследуемом материале.



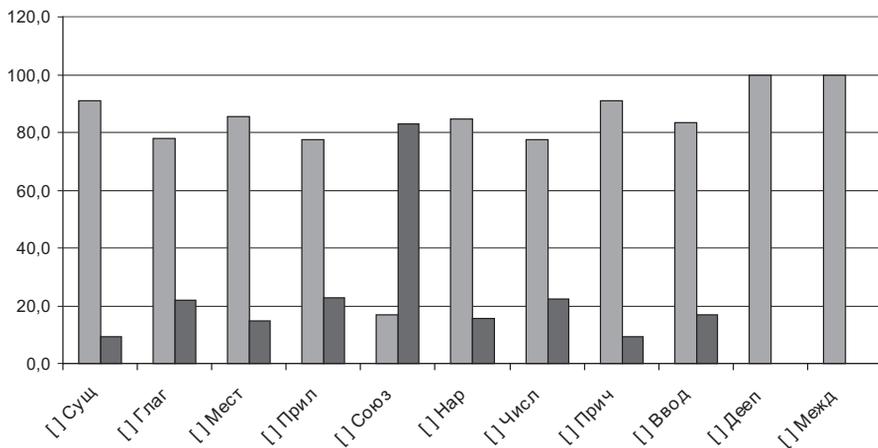
(А)



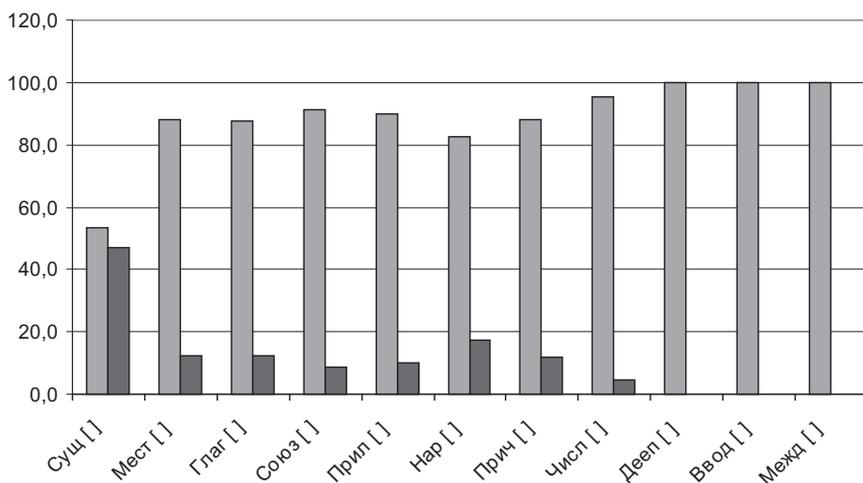
(Б)

Рис. 5. Процентное соотношение встречаемости синтагматической границы при наличии знака препинания для пар: «запятая — часть речи» (а), «часть речи — запятая» (б) (есть граница с паузой короткой или долгой — ряд 1, граница отсутствует — ряд 2)

Исходя из данных, представленных на рис. 6, можно сделать следующие основные выводы. Вне зависимости от типа части речи, следующей за пробелом (рис. 6а), вероятность наличия синтагматической границы (с короткой либо долгой паузой) в большинстве случаев значительно более низкая, чем её присутствие. Исключение составляет пара с союзами (как правило, союзы «И», «ИЛИ»), для которой вероятность присутствия синтагматической границы весьма высокая. Это хорошо согласуется с нашими априорными представлениями. Похожие выводы можно сделать и исходя из анализа данных рис. 6б.



(А)



(Б)

Рис. 6. Процентное соотношение встречаемости синтагматической границы в отсутствие знака препинания для пар: «пробел — часть речи» (а) и «часть речи — пробел» (б) (нет границы — ряд 1, есть граница с паузой короткой или долгой — ряд 2)

Опираясь на представленный выше комплекс статистических характеристик, решающие правила вероятностного синтагматического членения при синтезе речи в строгом математическом смысле должны основываться на формуле Байеса:

$$P(A | B_{ij}) = \frac{P(B_{ij} | A) \cdot P(A)}{P(B_{ij})}$$

Здесь:

A — интересующее нас событие: появление синтагматической границы между парами частей речи в тексте (паузы в речи)

B_{ij} — текущее событие в тексте: одна из множества возможных ij -пар частей речи в промежутке между словами;

$P(A)$ — априорная вероятность гипотезы A (как часто в среднем появляется граница между любыми словами). Определяется исходя из подсчитанного среднего числа слов в синтагме (см. таблицу 1);

$P(B_{ij} | A) = P_b(ij)$ — вероятность наступления события B_{ij} при истинности гипотезы A . Найденное заранее распределение вероятностей пар частей речи B_{ij} , между которыми присутствует граница синтагмы (пауза);

$P(B_{ij})$ — вероятность наступления события B_{ij} . Найденное заранее распределение вероятностей пар частей речи B_{ij} вне зависимости от того присутствует либо отсутствует между ними граница синтагмы (пауза).

$P(A | B_{ij}) = P_a(ij)$ — вероятность гипотезы A при наступлении события B_{ij} (апостериорная вероятность). Искомое распределение вероятностей присутствия границы синтагмы (паузы) при условии, что в данной позиции текста находится пара частей речи — B_{ij} ;

Рассмотрим далее в качестве иллюстрации упрощённый алгоритм. Процедура вероятностного синтагматического членения предложений с использованием данных, приведенных на рис. 5, 6, может быть построена следующим образом. На первом этапе определяются наиболее вероятные границы синтагм на стыке фонетических слов, разделённых знаком препинания (назовём их «пунктуационными» синтагмами). Для этого определяются среднее процентное соотношение встречаемости синтагматической границы при наличии знака препинания для пар: «запятая — часть речи» (а), «часть речи — запятая» (б), исходя из данных, приведенных на рисунках 5а и 5б. По некоторому заранее выбранному порогу принимаются решения и проставляются границы найденных таким образом пунктуационных синтагм. Далее все слова внутри пунктуационных синтагм считаются разделёнными только пробелами. К каждой из полученных таким образом последовательностей слов применяется описанная выше процедура, но только с использованием данных, приведенных на рисунках 6а и 6б. В результате каждая пунктуационная синтагма разбивается в свою очередь на последовательность внутренних синтагм (назовём их «синтаксическими» синтагмами).

Наибольший интерес, с нашей точки зрения, представляет проверка работоспособности изложенных правил при сегментации многословных предложений без единого знака препинания, которые часто встречаются в текстах. Рассмотрим пример такого рода предложения, не входившего в обучающую выборку (предлоги присоединены знаком «Ъ»):

- (1) *«Но молодая жена упорно продолжала отстирывать белую въкровавых пятнах рубаху мужа посиневшими отъхолода руками въжелезном тазике съледяной водой.»*

На рис. 7 для этого предложения графически представлены нормированные значения статистической вероятности наличия синтагматической границы между словами, рассчитанными в соответствии с данными, приведенными на рисунках 6а и 6б, а также возможные положения границ синтагм при двух порогах принятия решений — 0,45 и 0,25.

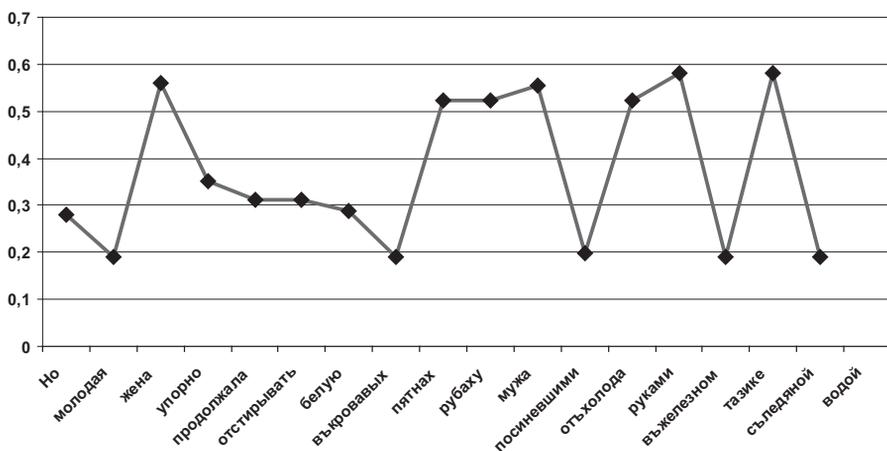


Рис. 7. Возможные положения границ синтагм при двух порогах принятия решений — 0,45 и 0,25 для предложения «Но молодая жена упорно продолжала отстирывать белую в кровавых пятнах рубаху мужа посиневшими от холода руками в железном тазике с ледяной водой»

Ниже на 3-х примерах показаны возможные границы синтагм для максимального объёма синтагм (1) — при пороге 0,45 и для минимального объёма: (2) — при пороге 0,45 и (3) — при пороге 0,25.

- (1) *Но молодая жена // упорно продолжала отстирывать белую в кровавых пятнах рубаху мужа // посиневшими от холода руками // в железном тазике // с ледяной водой.*
- (2) *Но молодая жена // упорно продолжала отстирывать белую в кровавых пятнах // рубаху // мужа // посиневшими от холода // руками // в железном тазике // с ледяной водой.*
- (3) *Но // молодая жена // упорно // продолжала // отстирывать // белую // в кровавых пятнах // рубаху // мужа // посиневшими от холода // руками // в железном тазике // с ледяной водой.*

Приведенные варианты синтагматического членения, на наш взгляд, вполне допустимы. Как показано в [12], для перцептивной сегментации на синтагмы весьма характерна индивидуальная вариабельность стратегий, проявляющаяся в выделении синтагм существенно разного объёма — от одного фонетического слова до пяти-семи.

Заключение

В результате проведённого исследования получены объективные статистические данные особенностей синтагматического членения в процессе выразительного чтения текста профессиональным диктором. Намечены пути использования статистического подхода для реализации алгоритмов синтагматического членения в синтезаторе русской речи по тексту. Авторы ясно осознают, что полученных статистических данных во многих случаях недостаточно для безошибочного членения предложений на синтагмы. Однако, ввиду сравнительной простоты полученных правил, их использование, например, в приложениях с ограниченными вычислительными ресурсами, может оказаться целесообразным. Предполагается дальнейшее развитие данного подхода с использованием более детальной информации о морфологической структуре слов и синтаксической структуре предложений. Предварительные результаты синтеза речи по тексту с использованием статистического алгоритма синтагматического членения будут продемонстрированы во время доклада.

Авторы выражают свою признательность Белорусскому фонду фундаментальных исследований за поддержку данной работы.

References

1. *Golovin I. B.* 1983. Fundamentals of Speech Culture [Osnovy Kul'tury Rechi].
2. *Lobanov B. M.* 2008. Computational Synthesis and Speech Cloning [Komp'yuternyi Sintez i Klonirovanie Rechi].
3. *Hiromichi Kawanami et al.* Designing Speech Database with Prosodic Variety for Expressive TTS system, available at: <http://gandalf.aksis.uib.no/irec2002/pdf/337.pdf>
4. *Hongwu Yang, Shuang Li, Lianhong Cai.* Toward Synthesizing Expressive Mandarin Speech, available at: <http://www.w3.org/2005/08/SSML/Papers/Tsinghua.pdf>
5. *Iomdin L. L., Lobanov B. M., Getsevich Iu. S.* 2004. The Talking ETAP. Using the ETAP Parser in Russian Speech Synthesis [Govoriashchii "ETAP": Opyt Ispol'zovaniia Sintaksicheskogo Analizatora Sistemy ETAP v Russkom Rechevom Sinteze]. This compendium : 874–877.
6. *Khomitsevich O. G., Solomennik M. V.* 2010. Automatic Pausing in the System of Russian Text-To-Speech Synthesis [Avtomaticheskaia Rasstanovka Pauz v Sisteme Sinteza Russkoi Rechi po Tekstu]. *Komp'yuternaia Lingvistika*

- i Intellektual'nye Tekhnologii: Trudy Mezhdunarodnoi Konferentsii "Dialog 2010" (Computational Linguistics and Intelligent Technologies: Proceedings of the International Conference "Dialog 2010") : 531–537.
7. *Krivnova O. F., Chardin I. S.* 1999. Pausing in Natural and Synthesized Speech [Pauzirovanie v Estestvennoi I Sintezirovannoi Rechi]. Teoriia I Praktika Reshevykh Issledovani (ARSO-99). Materialy Konferentsii, (Proc. of Conference "Theory and Practice of Linguistic Researches"), available at: <http://www.russian.slavica.org/article9348.html>
 8. *Lobanov B. M.* 2008. The Algorithm of Text Segmentation into Syntactic Syntagmas for Speech Synthesis [Algoritm Segmentatsii Teksta na Sintaksicheskie Sintagmy dla Sinteza Rechi]. Komp'iuternaia Lingvistika i Intellektual'nye Tekhnologii: Trudy Mezhdunarodnoi Konferentsii "Dialog 2008" (Computational Linguistics and Intelligent Technologies: Proceedings of the International Conference "Dialog 2008") : 323–329.
 9. *Lobanov B., Tsirulnik L.* 2006. Statistical Study of Speaker's Peculiarities of Utterances into Phrases Segmentation. Speech Prosody: Proceedings of the 3-rd International Conference, 2 : 557–560.
 10. *Pitrelli J. F.* et al. 2006. The IBM Expressive Text-to-speech Synthesis System for American English. Audio, Speech, and Language Processing, IEEE Transactions on V., 14 (4) : 1099–1108.
 11. *Ventsov A. V., Kasevich V. B., Slepokurova N. A.* 1993. Perceptive Segmentation of a Sounding Text [Pertseptivnaia Segmentatsiia Zvuchashchego Teksta]. Problemy Fonetiki : 242–273.