

Логические функции определения границ и интонационного типа пунктуационных синтагм



О.Г. Сизонов,
соискатель

Описываются правила пунктуационного синтагматического членения и интонационной разметки текста для синтеза русской речи по тексту. Правила учитывают не только типы знаков пунктуации в предложении, но и их ближайшее окружение в тексте. Это позволяет реализовать алгоритм синтеза интонационно-окрашенной речи, избегая монотонности «второго рода», т.е. частой повторяемости одинаковых интонационных конструкций.

Abstract

The rules of punctuation syntagmatic segmentation and intonation marking of a text for Russian text-to-speech synthesis are described. Not only the types of the signs of punctuation in offer, but their nearest surroundings in a text take the rules into account. It is allowed to realize an algorithm of a synthesis intonationally-avoiding monotony «of the second order» i.e. the frequent repeatability of the similar intonation constructions.

Введение

Среди объектов исследования лаборатории синтеза и распознавания речи Объединённого института проблем информатики НАН Беларуси метод мультиволнового синтеза речи по тексту занимает одно из приоритетных направлений как перспективный в плане достижения высокого качества синтезируемого сигнала, в том числе разборчивости и интонационной выразительности. Разработанная в лаборатории компьютерная модель мультиволнового синтеза речи по тексту учитывает специфику лингвистической обработки текстов, фонетическую и просодическую структуру русской речи, особенности артикуляторно-акустических явлений процесса речеобразования [1].

Интонационные характеристики синтезируемой речи, её разборчивость, выразительность и естественность обеспечиваются вычислением необходимых значений час-

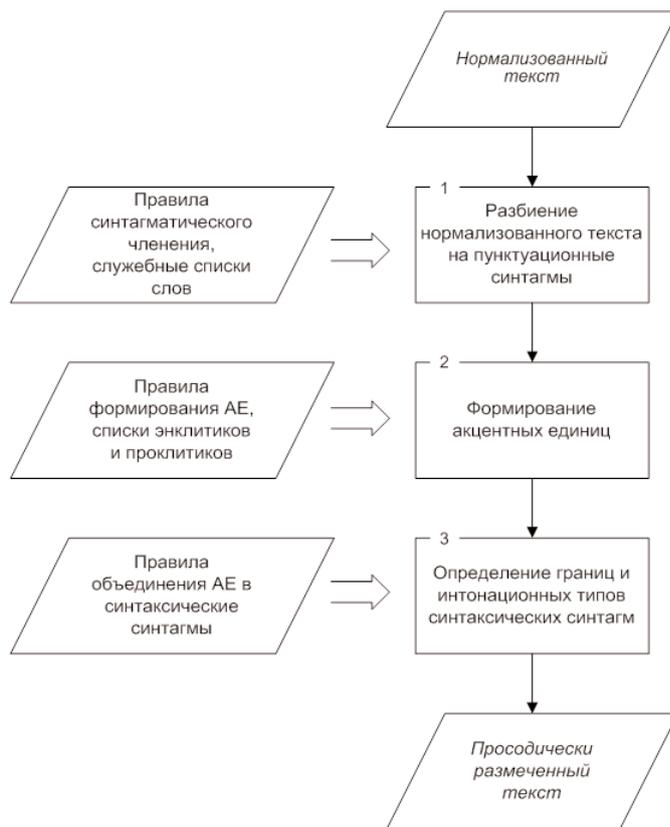


Рис. 1. Структурная схема просодического процессора

тоты основного тона (F_0), длительности звуков и пауз (T), амплитуды звука (A). Эта просодическая информация определяется на основе анализа определённых свойств входного текста, осуществляемого просодическим блоком, который на основе морфологической, синтаксической, пунктуационной информации подаваемого текста обнаруживает в нём минимальные интонационные единицы — синтагмы, определяет их интонационный тип. Маркированный текст является результатом работы, выходными данными просодического процессора.

Последовательность этапов, выполняемых процессором, представлена на рис. 1.

Предметом данной работы является формализация и описание правил просодической разметки, производимой просодическим процессором на первом этапе обработки текста — разбиении поступившего нор-

мализованного текста на пунктуационные синтагмы, включая определение их интонационных типов. Правила описаны в виде логических функций, аргументами которых являются последовательность слов и пунктуационных символов нормализованного текста, служебные списки слов и морфологических признаков.

1. Перечень интонационных типов пунктуационных синтагм

Интонационные типы можно разделить на следующие категории:

- завершённости P ,
- незавершённости C ,
- восклицания E ,
- вопроса Q .

Входящие в эти категории типы обозначаются символом категории и соответствующим индексом.

Все интонационные типы и определяющие их признаки приведены в таблицах 1–4.

Таблица 1

Интонационные типы завершённости

Обозначение	Признак конца синтагмы	Комментарий
P1	[:]	Интонация «двоеточия»
P2	[]]	Интонация «закрывающей скобки»
P3	[:]	Интонация «точки с запятой»
P4	[.] 1-й раз	Интонация «точки» каждой 1-й, 4-й, 7-й и т.д. идущей подряд синтагмы данного типа
P4_1	[.] 2-й раз	Интонация «точки» каждой 2-й, 5-й, 8-й и т.д. идущей подряд синтагмы данного типа
P4_2	[.] 3-й раз	Интонация «точки» каждой 3-й, 6-й, 9-й и т.д. идущей подряд синтагмы данного типа
P5	[...]	Интонация «многоточия»
P6	[.] + абзац	Интонация «абзаца»
P7	Сочинительный союз + [.]	Интонация «точки» при условии, что синтагма начинается с сочинительного союза и ей предшествует синтагма с интонацией незавершённости
P8	Вопросительный союз + [.]	Интонация «точки» при условии, что синтагма начинается с вопросительного союза и ей предшествует синтагма с интонацией незавершённости
P9	Подчинительный союз + [.]	Интонация «точки» при условии, что синтагма начинается с подчинительного союза и ей предшествует синтагма с интонацией незавершённости
P10	Причастие + [.]	Интонация «точки» при условии, что синтагма начинается с причастия и ей предшествует синтагма с интонацией незавершённости
P11	Деепричастие + [.]	Интонация «точки» при условии, что синтагма начинается с деепричастия и ей предшествует синтагма с интонацией незавершённости

Таблица 2

Интонационные типы незавершённости

Обозначение	Признак конца синтагмы	Комментарий
C1	Союз «И»	Интонация «И»
C2	«ИЛИ»	Интонация «ИЛИ»
C3	[.] 1-й раз	Интонация «запятой» для каждой 1-й, 4-й, 7-й и т.д. идущей подряд синтагмы данного типа
C3_1	[.] 2-й раз	Интонация «запятой» каждой 2-й, 5-й, 8-й и т.д. идущей подряд синтагмы данного типа
C3_2	[.] 3-й раз	Интонация «запятой» каждой 3-й, 6-й, 9-й и т.д. идущей подряд синтагмы данного типа
C4	[–]	Интонация «тире»
C5	[(]	Интонация «открывающей скобки»
C6	[, –]	Интонация «запятая — тире»
C7	[.] + сочинительный союз	Интонация «запятой перед сочинительным союзом»
C8	[.] + вопросительный союз	Интонация «запятой перед вопросительным союзом»



C9	[.] + подчинительный союз	Интонация «запятой перед подчинительным союзом»
C10	[.] + причастие	Интонация «запятой перед причастием»
C11	[.] + деепричастие	Интонация «запятой перед деепричастием»

Таблица 3

Интонационные типы восклицания

<i>Обозначение</i>	<i>Признак конца синтагмы</i>	<i>Комментарий</i>
E1	[!] + восклицательное словоодна синтагма	Интонация «восклицания» последней или единственной синтагмы восклицательного предложения, содержащей восклицательное слово
E1_1	[!] + восклицательное слово 1-я синтагма	Интонация «восклицания» каждой нечётной (кроме последней) идущей подряд, содержащей восклицательное слово, синтагмы восклицательного предложения
E1_2	[!] + восклицательное слово 2-я синтагма	Интонация «восклицания» каждой чётной (кроме последней) идущей подряд, содержащей восклицательное слово, синтагмы восклицательного предложения
E2	[!] одна синтагма	Интонация «восклицания» последней или единственной синтагмы восклицательного предложения, не содержащей восклицательного слова
E2_1	[!] 1-я синтагма	Интонация «восклицания» каждой нечётной (кроме последней) идущей подряд и не содержащей восклицательного слова синтагмы восклицательного предложения
E2_2	[!] 2-я синтагма	Интонация «восклицания» каждой чётной (кроме последней) идущей подряд и не содержащей восклицательного слова синтагмы восклицательного предложения

Таблица 4

Интонационные типы вопроса

<i>Обозначение</i>	<i>Признак конца синтагмы</i>	<i>Комментарий</i>
Q1	[?] + вопросительное слово одна синтагма	Интонация «вопроса» последней или единственной синтагмы вопросительного предложения, содержащей вопросительное слово
Q1_1	[?] + вопросительное слово 1-я синтагма	Интонация «вопроса» каждой нечётной (кроме последней) идущей подряд, содержащей вопросительное слово, синтагмы вопросительного предложения
Q1_2	[?] + вопросительное слово 2-я синтагма	Интонация «вопроса» каждой чётной (кроме последней) идущей подряд, содержащей вопросительное слово, синтагмы вопросительного предложения

Q2	[?] Одна синтагма	Интонация «вопроса» последней или единственной синтагмы вопросительного предложения, не содержащей вопросительного слова
Q2_1	[?] 1-я синтагма	Интонация «вопроса» каждой нечётной (кроме последней) идущей подряд и не содержащей вопросительного слова синтагмы вопросительного предложения
Q2_2	[?] 2-я синтагма	Интонация «вопроса» каждой чётной (кроме последней) идущей подряд и не содержащей вопросительного слова синтагмы вопросительного предложения

2. Логические функции блока интонационной разметки пунктуационных синтагм

Основа синтагматического членения текста — выделение по его пунктуационным признакам предложений или их частей, называемых пунктуационными синтагмами [2].

Входные данные блока интонационной разметки пунктуационных синтагм:

$W = \{w_1, w_2, \dots, w_n\}$ — «последовательность слов и пунктуационных символов нормализованного текста, где n — число слов и знаков пунктуации в тексте.

Выходные данные блока интонационной разметки:

S — «последовательность синтагм, формируемых в процессе просодической обработки входного текста;

T — «последовательность интонационных типов синтагм, где T_j — интонационный тип синтагмы S_j .

Каждая функция F определяет, является ли i -й элемент множества W окончанием j -й синтагмы S соответствующего интонационного типа T_j .

2.1. Логические функции определения пунктуационных синтагм с интонацией «незавершённости»

2.1.1. [C1] «устанавливается после слова, за которым идёт союз «И» или «ДА».

¹Пример: *Пьер уже три месяца выбирал карьеру[C1] и ничего не делал[P4].*

$$F_{/C1/} = (W_{i+1} \equiv \langle \text{И} \rangle) \vee (W_{i+1} \equiv \langle \text{ДА} \rangle)$$

2.1.2. [C2] «устанавливается после слова, за которым идёт союз «ИЛИ».

²Пример: *Маленькая княгиня не слыхала[C2] или не хотела слышать его слов[P4].*

$$F_{/C2/} = (W_{i+1} \equiv \langle \text{ИЛИ} \rangle)$$

2.1.3. [C4] «устанавливается после слова, после которого встретился символ «-», отделённый с обеих сторон пробелами.

³Пример: *Этот пресловутый нейтралитет Пруссии[C4] «только западня[P4].*

$$F_{/C4/} = (W_{i+1} \equiv \langle \text{-} \rangle)$$



2.1.4. [C5] «устанавливается после слова, после которого идёт символ «(«, перед которым может быть пробел.

⁴Пример: Богданыч[C5] (Богданычем называли полкового командира[P2]) вас осадил [P4].

$$F_{/C5/} = (W_{i+1} \equiv \langle \langle \rangle \rangle)$$

2.1.5. [C6] «устанавливается после слова, непосредственно после которого встречена последовательность символов «,-», разделённая или нет пробелов, если интонационный тип предыдущей синтагмы отличен от [C6].

⁵Пример: Такая странная антипатия[C6], «думал Пьер[P2], «а прежде он мне даже очень нравился[P4].

$$F_{/C6/} = (W_{i+1} \equiv \langle \langle \rangle \rangle) \wedge (W_{i+2} \equiv \langle \langle - \rangle \rangle) \wedge (T_{j-1} \neq C_6)$$

2.1.6. [C7] «устанавливается после слова, после которого идут символ «,» и сочинительный союз <M1>, разделённые пробелом.

⁶Пример: Графиня хотела хмуриться[C7], но не могла[P7].

$$F_{/C7/} = (W_{i+1} \equiv \langle \langle \rangle \rangle) \wedge (W_{i+2} \in M1)$$

2.1.7. [C8] «устанавливается после слова, после которого идут символ «,» и вопросительно-подчинительный союз <M2>, разделённые пробелом.

⁷Пример: Генерал садился на лошадь[C8], которую подал ему казак[P8].

$$F_{/C8/} = (W_{i+1} \equiv \langle \langle \rangle \rangle) \wedge (W_{i+2} \in M2)$$

2.1.8. [C9] «устанавливается после слова, после которого идут символ «,» и подчинительный союз <M3>, разделённые пробелом.

⁸Пример: Предложение было слишком лестно[C9], чтобы отказаться[P9].

$$F_{/C9/} = (W_{i+1} \equiv \langle \langle \rangle \rangle) \wedge (W_{i+2} \in M3)$$

2.1.9. [C10] «устанавливается после слова, после которого идут символ «,» и причастие, определяемое по списку суффиксов <M4>, разделённые пробелом.

⁹Пример: Вейротер был австрийский генерал[C10], заменивший убитого Шмита[P10].

$$F_{/C10/} = (W_{i+1} \equiv \langle \langle \rangle \rangle) \wedge (W_{i+2} \in M4)$$

2.1.10. [C11] «устанавливается после слова, после которого идут символ «,» и деепричастие, определяемое по списку окончаний <M5>, разделённые пробелом.

¹⁰Пример: Остальная пехота поспешно проходила по мосту[C11], спираясь воронкой у входа[P11].

$$F_{/C11/} = (W_{i+1} \equiv \langle \langle \rangle \rangle) \wedge (W_{i+2} \in M5)$$

2.2. Логические функции определения пунктуационных синтагм с интонацией «завершённость»

2.2.1 [P1] «устанавливается после слова, после которого идёт символ «:», перед которым может быть пробел.

¹¹Пример: *Всё только одного желали[P1]: под предводительством государя скорее итти против неприятеля[P4].*

$$F_{/P1/} = (W_{i+1} \equiv «:»)$$

2.2.2. [P2] «устанавливается после слова, если после него идёт символ «)», перед которым может быть пробел, или если непосредственно после слова встречена последовательность символов «,-», разделённая или нет пробелом, и интонационный тип предыдущей синтагмы [C6].

¹²Пример: *Богданыч[C5] (Богданычем называли полкового командира[P2]) вас осадил [P4].*

$$F_{/P2/} = [(W_{i+1} \equiv «)»] \vee [(W_{i+1} \equiv «,-»] \wedge [(W_{i+2} \equiv «-»)^{\wedge} (T_{j-1} \equiv C6)]$$

2.2.3. [P3] «устанавливается после слова, после которого идут символ «;», перед которым может быть пробел.

¹³Пример: *Оттепель и туман продолжались [P3]; за 40 шагов ничего не было видно[P4].*

$$F_{/P3/} = (W_{i+1} \equiv «;»)$$

2.2.4. [P5] «устанавливается после слова, после которого идёт подряд три символа «.» (многоточие), перед которыми может быть пробел.

¹⁴Пример: *Только в Юхнове с Пелагеюшкой сошлись[P5]...*

$$F_{/P5/} = (W_{i+1} \equiv «.»)^{\wedge} (W_{i+2} \equiv «.»)^{\wedge} (W_{i+3} \equiv «.»)$$

2.2.5. [P7] «устанавливается после слова, за которым встречен символ «.»», если предыдущая синтагма имела интонационный тип [C7].

¹⁵Пример: *Графиня хотела хмуриться[C7], но не могла[P7].*

$$F_{/P7/} = (W_{i+1} \equiv «.»)^{\wedge} (T_{j-1} \equiv C7)$$

2.2.6. [P8] «устанавливается после слова, за которым встречен символ «.»», если предыдущая синтагма имела интонационный тип [C8].

¹⁶Пример: *Генерал садился на лошадь[C8], которую подал ему казак[P8].*

$$F_{/P8/} = (W_{i+1} \equiv «.»)^{\wedge} (T_{j-1} \equiv C8)$$

2.2.7. [P9] «устанавливается после слова, за которым встречен символ «.»», если предыдущая синтагма имела интонационный тип [C9].

¹⁷Пример: *Предложение было слишком лестно[C9], чтобы отказаться[P9].*

$$F_{/P9/} = (W_{i+1} \equiv «.»)^{\wedge} (T_{j-1} \equiv C9)$$

2.2.8. [P10] «устанавливается после слова, за которым встречен символ «.»», если предыдущая синтагма имела интонационный тип [C10].

¹⁸Пример: *Вейротер был австрийский генерал[C10], заменивший убитого Шмита[P10].*

$$F_{/P10/} = (W_{i+1} \equiv «.»)^{\wedge} (T_{j-1} \equiv C10)$$



2.2.9. [P11] «устанавливается после слова, за которым встречен символ «.», если предыдущая синтагма имела интонационный тип [C11].

¹⁹Пример: Остальная пехота поспешно проходила по мосту[C11], спираясь воронкой у входа[P11].

$$F_{/P11/} = (W_{i+1} \equiv \langle \cdot \rangle) \wedge (T_{j-1} \equiv C11)$$

2.2.10. [P6] «устанавливается после слова, за которым встречен признак абзаца: последовательность из точки, переноса строки и табуляции.

²⁰Пример: Наташа стала надевать платье [P6].

$$F_{/P6/} = (W_{i+1} \equiv \langle \cdot \rangle) \wedge (W_{i+2} \equiv \langle \downarrow \rangle) \wedge (W_{i+3} \equiv \langle \rightarrow \rangle)$$

2.2.11. [P4] «устанавливается после слова, за которым встречен символ «.», за которым нет признака абзаца, и нет двух идущих подряд символов «.», и интонационный тип предыдущей синтагмы отличен от [C7], [C8], [C9], [C10], [C11], [P4], [P4_1].

²¹Пример: В четверть одиннадцатого наконец сели в кареты и поехали. Но ещё нужно было заехать к Таврическому саду[P4_1]. Перонская была уже готова[P4_2]. Ростовы похвалили её вкус и туалет[P4].

$$F_{/P4/} = (W_{i+1} \equiv \langle \cdot \rangle) \wedge (W_{i+2} \equiv \langle \cdot \rangle) \wedge (W_{i+2} \equiv \langle \downarrow \rangle) \wedge (T_{j-1} \neq C7) \wedge (T_{j-1} \neq C8) \wedge (T_{j-1} \neq C9) \wedge (T_{j-1} \neq C10) \wedge (T_{j-1} \neq C11) \wedge (T_{j-1} \neq P4) \wedge (T_{j-1} \neq P4_1)$$

2.2.12. [P4_1] «устанавливается после слова, за которым встречен символ «.», за которым нет признака абзаца, и нет двух идущих подряд символов «.», и интонационный тип предыдущей синтагмы [P4].

²²Пример: В четверть одиннадцатого наконец сели в кареты и поехали. Но ещё нужно было заехать к Таврическому саду[P4_1]. Перонская была уже готова[P4_2]. Ростовы похвалили её вкус и туалет[P4].

$$F_{/P4_1/} = (W_{i+1} \equiv \langle \cdot \rangle) \wedge (W_{i+2} \neq \langle \cdot \rangle) \wedge (W_{i+2} \neq \langle \downarrow \rangle) \wedge (T_{j-1} \equiv P4)$$

2.2.13. [P4_2] «устанавливается после слова, за которым встречен символ «.», за которым нет признака абзаца, и нет двух идущих подряд символов «.», и интонационный тип предыдущей синтагмы [P4_1].

²³Пример: В четверть одиннадцатого наконец сели в кареты и поехали. Но ещё нужно было заехать к Таврическому саду[P4_1]. Перонская была уже готова[P4_2]. Ростовы похвалили её вкус и туалет[P4].

$$F_{/P4_2/} = (W_{i+1} \equiv \langle \cdot \rangle) \wedge (W_{i+2} \neq \langle \cdot \rangle) \wedge (W_{i+2} \neq \langle \downarrow \rangle) \wedge (T_{j-1} \equiv P4_1)$$

2.2.14. [C3] «устанавливается после слова, за которым встречен символ «,», за которым нет символа «—», и нет слова, определяемого по спискам <M1>, <M2>, <M3>, <M4>, <M5>, и интонационный тип предыдущей синтагмы не [C3] и не [C3_1].

²⁴Пример: Видите[C3], погода мокрая[C3_1], говорил дядюшка[C3_2], отдохнули бы[C3], графинечку бы отвезли в дрожжах[P4].

$$F_{/C3/} = (W_{i+1} \equiv \langle , \rangle) \wedge (W_{i+2} \neq \langle - \rangle) \wedge (W_{i+2} \notin M1) \wedge (W_{i+2} \notin M2) \wedge (W_{i+2} \notin M3) \wedge (W_{i+2} \notin M4) \wedge (W_{i+2} \notin M5) \wedge (T_{j-1} \neq C3) \wedge (T_{j-1} \neq C3_1)$$

2.2.15. [C3_1] «устанавливается после слова, за которым встречен символ «,», за которым нет символа «-», и нет слова, определяемого по спискам <M1>, <M2>, <M3>, <M4>, <M5>, и интонационный тип предыдущей синтагмы [C3].

²⁵Пример: Видите[C3], погода мокрая[C3_1], говорил дядюшка[C3_2], отдохнули бы[C3], графинечку бы отвезли в дрожжах[P4].

$$F/C3_1/ = (W_{i+1} \equiv \langle \rangle) \wedge (W_{i+2} \neq \langle - \rangle) \wedge (W_{i+2} \notin M1) \wedge (W_{i+2} \notin M2) \wedge (W_{i+2} \notin M3) \wedge (W_{i+2} \notin M4) \wedge (W_{i+2} \notin M5) \wedge (T_{j-1} \equiv C3)$$

2.2.16. [C3_2] «устанавливается после слова, за которым встречен символ «,», за которым нет символа «-», и нет слова, определяемого по спискам <M1>, <M2>, <M3>, <M4>, <M5>, и интонационный тип предыдущей синтагмы [C3_1].

²⁶Пример: Видите[C3], погода мокрая[C3_1], говорил дядюшка[C3_2], отдохнули бы[C3], графинечку бы отвезли в дрожжах[P4].

$$F/C3_1/ = (W_{i+1} \equiv \langle \rangle) \wedge (W_{i+2} \neq \langle - \rangle) \wedge (W_{i+2} \notin M1) \wedge (W_{i+2} \notin M2) \wedge (W_{i+2} \notin M3) \wedge (W_{i+2} \notin M4) \wedge (W_{i+2} \notin M5) \wedge (T_{j-1} \equiv C3_1) \quad \notin \quad \wedge$$

2.3. Логические функции определения пунктуационных синтагм с интонацией «вопрос и восклицание»

2.3.1. [Q1] «устанавливается после слова, если после него следует символ «?» и в текущей формируемой синтагме есть вопросительное слово, определяемое по списку вопросительных слов <M6>.

²⁷Пример: И как могла она допустить до этого Курагина[Q1]?

$$F/Q1/ = (W_{i+1} \equiv \langle ? \rangle) \wedge (\exists m, m \in M6 \vee m \in Sj)$$

2.3.2. [Q2] «устанавливается после слова, если после него следует символ «?» и в текущей формируемой синтагме отсутствуют вопросительные слова из списка вопросительных слов <M6>.

²⁸Пример: Прикажете наших из-под горы кликнуть[Q2]?

$$F/Q2/ = (W_{i+1} \equiv \langle ? \rangle) \wedge (\nexists m, m \in M6 \vee m \in Sj)$$

2.3.3. [Q1_1] «устанавливается в текущей формируемой синтагме, содержащей вопросительное слово из списка <M6> после слова, за которым следует один из символов «-», «(«, «)» или «,» за которым может быть определяемое по множествам <M1>, <M2>, <M3>, <M4>, <M5> слово и при этом интонационный тип следующей синтагмы один из следующих: [Q1], [Q2], [Q1_2], [Q2_2], а интонационный тип предыдущей синтагмы не [Q1_1] и не [Q2_1].

²⁹Пример: А что такое война[Q1_1], что нужно для успеха в военном деле[Q1_2], какие нравы военного общества[Q1]?

$$F/Q1_1/ = [(W_{i+1} \equiv \langle \rangle) \vee W_{i+1} \equiv \langle - \rangle \vee W_{i+1} \equiv \langle (\rangle \vee W_{i+1} \equiv \langle) \rangle] \wedge (W_{i+2} \in M1) \wedge (W_{i+2} \in M2) \wedge (W_{i+2} \in M3) \wedge (W_{i+2} \in M4) \wedge (W_{i+2} \in M5)] \wedge (\exists m, m \in M6 \vee m \in Sj) \wedge [(T_{j+1} \equiv Q1) \vee (T_{j+1} \equiv Q2) \vee (T_{j+1} \equiv Q1_2) \vee (T_{j+1} \equiv Q2_2) \wedge (T_{j+1} \neq Q1_1) \vee (T_{j+1} \neq Q2_1)]$$



2.3.4. [Q1_2] «устанавливается в текущей формируемой синтагме, содержащей вопросительное слово из списка <M6> после слова, за которым следует один из символов «-», «(, «)» или «,» за которым может быть определяемое по множествам <M1>, <M2>, <M3>, <M4>, <M5> слово и при этом интонационный тип следующей синтагмы один из следующих: [Q1], [Q2], [Q1_1], [Q2_1], а интонационный тип предыдущей синтагмы «[Q1_2] или [Q2_2].

³⁰Пример: *А что такое война[Q1_1], что нужно для успеха в военном деле[Q1_2], какие нравы военного общества[Q1]?*

$$F_{/Q1_2/} = [[(W_{i+1} \equiv \langle \langle, \rangle \rangle \vee W_{i+1} \equiv \langle - \rangle \vee W_{i+1} \equiv \langle (\langle W_{i+1} \equiv \langle \rangle \rangle \langle \rangle) \wedge ((W_{i+2} \in M1) \wedge (W_{i+2} \in M2) \wedge (W_{i+2} \in M3) \wedge (W_{i+2} \in M4) \wedge (W_{i+2} \in M5))] \wedge (\exists m, m \in M6 \vee m \in Sj)] \wedge [(T_{j+1} \equiv Q1) \vee (T_{j+1} \equiv Q2) \vee (T_{j+1} \equiv Q1_1) \vee (T_{j+1} \equiv Q2_1) \wedge (T_{j-1} \neq Q1_2) \vee (T_{j+1} \neq Q2_2)]$$

2.3.5. [Q2_1] «устанавливается в текущей формируемой синтагме, не содержащей вопросительного слова из списка <M6> после слова, за которым следует одно из сочетаний символов «и», «или», «да», «-», «(, «)» или «,» за которым может быть определяемое по множествам <M1>, <M2>, <M3>, <M4>, <M5> слово и при этом интонационный тип следующей синтагмы один из следующих: [Q1], [Q2], [Q1_2], [Q2_2], а интонационный тип предыдущей синтагмы не [Q1_1] и не [Q2_1].

³¹Пример: *Любезности это[Q2_1], бабы сказки[Q2_2], или она права[Q2]?*

$$F_{/Q2_1/} = [[(W_{i+1} \equiv \langle \langle, \rangle \rangle \vee W_{i+1} \equiv \langle - \rangle \vee W_{i+1} \equiv \langle (\langle W_{i+1} \equiv \langle \rangle \rangle \langle \rangle) \wedge ((W_{i+2} \in M1) \wedge (W_{i+2} \in M2) \wedge (W_{i+2} \in M3) \wedge (W_{i+2} \in M4) \wedge (W_{i+2} \in M5))] \wedge (\exists m, m \in M6 \vee m \in Sj)] \wedge [(T_{j+1} \equiv Q1) \vee (T_{j+1} \equiv Q2) \vee (T_{j+1} \equiv Q1_2) \vee (T_{j+1} \equiv Q2_2) \wedge (T_{j-1} \neq Q1_1) \vee (T_{j+1} \neq Q2_1)]$$

2.3.6. [Q2_2] «устанавливается в текущей формируемой синтагме, не содержащей вопросительного слова из списка <M6> после слова, за которым следует одно из сочетаний символов «и», «или», «да», «-», «(, «)» или «,» за которым может быть определяемое по множествам <M1>, <M2>, <M3>, <M4>, <M5> слово и при этом интонационный тип следующей синтагмы один из следующих: [Q1], [Q2], [Q1_2], [Q2_2], а интонационный тип предыдущей синтагмы «[Q1_1] или [Q2_1].

³²Пример: *Любезности это[Q2_1], бабы сказки[Q2_2], или она права[Q2]?*

$$F_{/Q2_2/} = [[(W_{i+1} \equiv \langle \langle, \rangle \rangle \vee W_{i+1} \equiv \langle - \rangle \vee W_{i+1} \equiv \langle (\langle W_{i+1} \equiv \langle \rangle \rangle \langle \rangle) \wedge ((W_{i+2} \in M1) \wedge (W_{i+2} \in M2) \wedge (W_{i+2} \in M3) \wedge (W_{i+2} \in M4) \wedge (W_{i+2} \in M5))] \wedge (\exists m, m \in M6 \vee m \in Sj)] \wedge [(T_{j+1} \equiv Q1) \vee (T_{j+1} \equiv Q2) \vee (T_{j+1} \equiv Q1_1) \vee (T_{j+1} \equiv Q2_1) \wedge (T_{j-1} \neq Q1_2) \vee (T_{j+1} \neq Q2_2)]$$

2.3.7. [E1] «устанавливается после слова, если после него следует символ «!» и в текущей формируемой синтагме есть восклицательное слово, определяемое по списку восклицательных слов <M7>.

³³Пример: *Ай да Данила Купор[E1]!*

$$F_{/E1/} = (W_{i+1} \equiv \langle ! \rangle) \wedge (\exists m, m \in M7 \vee m \in Sj)$$

2.3.8. [E2] «устанавливается после слова, если после него следует символ «!» и в текущей формируемой синтагме отсутствуют восклицательные слова из списка восклицательных слов <M7>.

³⁴Пример: *Хорошо он себя зарекомендовал в Букарещте[E2]!*

$$F_{/E2/} = (W_{i+1} \equiv \langle ! \rangle) \wedge (\exists m, m \in M7 \vee m \in Sj)$$

2.3.9. [E1_1] «устанавливается в текущей формируемой синтагме, содержащей восклицательное слово из списка <M7> после слова, за которым следует один из символов «(», «)» или «,» за которым может быть определяемое по множествам <M1>, <M2>, <M3>, <M4>, <M5> слово и при этом интонационный тип следующей синтагмы один из следующих: [E1], [E2], [E1_2], [E2_2], а интонационный тип предыдущей синтагмы не [E1_1] и не [E2_1].

³⁵Пример: *Ах[E1_1], я право не думал оскорбить её[E1_2], я так понимаю[E2_1] и высоко ценю эти чувства[E2]!*

$$F_{/E1_1/} = [[(W_{i+1} \equiv \langle \rangle \vee W_{i+1} \equiv \langle - \rangle \vee W_{i+1} \equiv \langle (\langle W_{i+1} \equiv \langle \rangle \rangle) \rangle) \wedge ((W_{i+2} \in M1) \wedge (W_{i+2} \in M2) \wedge (W_{i+2} \in M3) \wedge (W_{i+2} \in M4) \wedge (W_{i+2} \in M5))] \wedge (\exists m, m \in M7 \vee m \in Sj)] \wedge [(T_{j+1} \equiv E1) \vee (T_{j+1} \equiv E2) \vee (T_{j+1} \equiv E1_2) \vee (T_{j+1} \equiv E2_2) \wedge (T_{j-1} \neq E1_1) \vee (T_{j-1} \neq E2_1)]$$

2.3.10 [E1_2] «устанавливается в текущей формируемой синтагме, содержащей восклицательное слово из списка <M7> после слова, за которым следует одно из сочетаний символов «и», «или», «да», «-», «(», «)» или «,» за которым может быть определяемое по множествам <M1>, <M2>, <M3>, <M4>, <M5> слово и при этом интонационный тип следующей синтагмы один из следующих: [E1], [E2], [E1_2], [E2_2], а интонационный тип предыдущей синтагмы «[E1_1] или [E2_1].

³⁶Пример: *Ах[E1_1], я право не думал оскорбить её[E1_2], я так понимаю[E2_1] и высоко ценю эти чувства[E2]!*

$$F_{/E1_2/} = [[(W_{i+1} \equiv \langle \rangle \vee W_{i+1} \equiv \langle - \rangle \vee W_{i+1} \equiv \langle (\langle W_{i+1} \equiv \langle \rangle \rangle) \rangle) \wedge ((W_{i+2} \in M1) \wedge (W_{i+2} \in M2) \wedge (W_{i+2} \in M3) \wedge (W_{i+2} \in M4) \wedge (W_{i+2} \in M5))] \wedge (\exists m, m \in M7 \vee m \in Sj)] \wedge [(T_{j+1} \equiv E1) \vee (T_{j+1} \equiv E2) \vee (T_{j+1} \equiv E1_1) \vee (T_{j+1} \equiv E2_1) \wedge (T_{j-1} \neq E1_2) \vee (T_{j-1} \neq E2_2)]$$

2.3.11. [E2_1] «устанавливается в текущей формируемой синтагме, не содержащей восклицательного слова из списка <M7> после слова, за которым следует одно из сочетаний символов «и», «или», «да», «-», «(», «)» или «,» за которым может быть определяемое по множествам <M1>, <M2>, <M3>, <M4>, <M5> слово и при этом интонационный тип следующей синтагмы один из следующих: [E1], [E2], [E1_2], [E2_2], а интонационный тип предыдущей синтагмы не [E1_1] и не [E2_1].

³⁷Пример: *Голубушка[E2_1], мамаша[E2_2], как я вас люблю[E2_1], как мне хорошо[E2]!*

$$F_{/E2_1/} = [[(W_{i+1} \equiv \langle \rangle \vee W_{i+1} \equiv \langle - \rangle \vee W_{i+1} \equiv \langle (\langle W_{i+1} \equiv \langle \rangle \rangle) \rangle) \wedge ((W_{i+2} \in M1) \wedge (W_{i+2} \in M2) \wedge (W_{i+2} \in M3) \wedge (W_{i+2} \in M4) \wedge (W_{i+2} \in M5))] \wedge (\exists m, m \in M7 \vee m \in Sj)] \wedge [(T_{j+1} \equiv E1) \vee (T_{j+1} \equiv E2) \vee (T_{j+1} \equiv E1_2) \vee (T_{j+1} \equiv E2_2) \wedge (T_{j-1} \neq E1_1) \vee (T_{j-1} \neq E2_1)]$$



2.3.12. [E2_2] «устанавливается в текущей формируемой синтагме, не содержащей восклицательного слова из списка <M7> после слова, за которым следует одно из сочетаний символов «и», «или», «да», «—», «(», «)» или «,» за которым может быть определяемое по множествам <M1>, <M2>, <M3>, <M4>, <M5> слово и при этом интонационный тип следующей синтагмы один из следующих: [E1], [E2], [E1_2], [E2_2], а интонационный тип предыдущей синтагмы «[E1_1] или [E2_1].

³⁸Пример: Голубушка[E2_1], мамаша[E2_2], как я вас люблю[E2_1], как мне хорошо[E2]!

$$F_{/E2_2/} = [((W_{i+1} \equiv \langle \rangle \vee W_{i+1} \equiv \langle - \rangle \vee W_{i+1} \equiv \langle (\rangle \vee W_{i+1} \equiv \langle) \rangle) \wedge ((W_{i+2} \in M1) \wedge (W_{i+2} \in M2) \wedge (W_{i+2} \in M3) \wedge (W_{i+2} \in M4) \wedge (W_{i+2} \in M5))) \wedge (\exists m, m \in M7 \vee m \in Sj)] \wedge [(T_{j+1} \equiv E1) \vee (T_{j+1} \equiv E2) \vee (T_{j+1} \equiv E1_1) \vee (T_{j+1} \equiv E2_1) \wedge (T_{j-1} \neq E2_1) \vee (T_{j+1} \neq E2_2)]$$

Заключение

Представленные в данной работе результаты исследований позволяют программно реализовать однопроходный алгоритм блока разбиения на пунктуационные синтагмы, выполняемый просодическим процессором. Приведённые правила интонационного анализа текста позволяют реализовать базовую часть функций просодического процессора. На основании результатов работы этапа разбиения на пунктуационные синтагмы производится членение их на синтаксические синтагмы и формирование акцентных единиц. Алгоритмы и правила такого анализа будут представлены в следующей публикации авторов по теме «Правила просодического анализа текста для синтеза речи».

Литература

1. Лобанов Б.М., Цирульник Л.И. Компьютерный синтез и клонирование речи. Мн.: Белорусская наука, 2008.
2. Алгоритм интонационной разметки повествовательных предложений для синтеза речи по тексту / Л.И. Цирульник, Б.М. Лобанов, О.Г. Сизонов // Труды Международной конференции «Компьютерная лингвистика и интеллектуальные технологии» (Диалог'2008). М.: Наука, 2008. С. 563–568.

Сизонов Олег Геннадьевич —

Окончил факультет информационных технологий Белорусского Национального технического университета. Соискатель учёной степени кандидата технических наук. С 2007 года — младший научный сотрудник лаборатории распознавания и синтеза речи Объединённого института проблем информатики Национальной академии наук Беларуси.

Область научных интересов — методы лингвистической обработки русского текста в синтезе речи по тексту, синтез и обработка речевых и квазиречевых сигналов, применение синтеза речи в системах реабилитации инвалидов по зрению и слуху.