

БЕЛОРУССКИЙ ГОСУДАРСТВЕННЫЙ УНИВЕРСИТЕТ
ФАКУЛЬТЕТ СОЦИОКУЛЬТУРНЫХ КОММУНИКАЦИЙ

**ЛИНГВИСТИКА, ЛИНГВОДИДАКТИКА,
ЛИНГВОКУЛЬТУРОЛОГИЯ:
АКТУАЛЬНЫЕ ВОПРОСЫ
И ПЕРСПЕКТИВЫ РАЗВИТИЯ**

**Материалы VII Международной
научно-практической конференции**

Минск, 16–17 марта 2023 г.

МИНСК
БГУ
2023

УДК 81(06)+811.1/.8(072)(06)+81:008(06)
ББК 81я431+81.40/59р30-2я431+81.006.3я431
Л59

Научный редактор А. О. Долгова

Редакционная коллегия:

кандидат филологических наук, доцент *О. А. Воробьева* (гл. ред.);
кандидат филологических наук *Н. А. Куркович* (отв. ред.);
кандидат филологических наук, доцент *Е. Л. Левшикова*;
Л. М. Блинкова; *Е. В. Дубровская*; *А. А. Боричевская*;
В. В. Воронович; *Н. В. Ушакова*; *О. С. Гончаренко*

Рецензенты:

кандидат филологических наук, доцент *И. И. Ковалевская*;
кандидат филологических наук, доцент *А. Б. Чиркун*

Лингвистика, лингводидактика, лингвокультурология: актуальные вопросы и перспективы развития : материалы VII Междунар. науч.-практ. конф., Минск, 16–17 марта 2023 г. / Белорус. гос. ун-т ; науч. ред. А. О. Долгова ; редкол.: О. А. Воробьева (гл. ред.) [и др.]. – Минск : БГУ, 2023. – 348 с.
ISBN 978-985-881-462-5.

Рассматриваются современные направления лингвистических исследований, традиции и инновации в обучении иностранным языкам, проблемы языковой картины мира и взаимодействия культур; освещаются вопросы литературоведения, теории перевода и интерпретации текста.

УДК 81(06)+811.1/.8(072)(06)+81:008(06)
ББК 81я431+81.40/59р30-2я431+81.006.3я431

ISBN 978-985-881-462-5

© БГУ, 2023

АЎТАМАТЫЧНАЯ ДЭЛІМАТАЦЫЯ ТЭКСТАВАЙ ІНФАРМАЦЫІ ДЛЯ БЕЛАРУСКАЙ МОВЫ

AUTOMATIC DELIMITATION OF TEXT INFORMATION FOR THE BELARUSIAN LANGUAGE

*Я.С. Зяноўка¹⁾, М.В. Супрунчук²⁾, А.А. Бакуновіч³⁾, М.А. Казлова⁴⁾,
Д.І. Латышэвіч⁵⁾, Ю.С. Гецэвіч⁶⁾*
*Ya. Zianouka¹⁾, M. Suprunchuk²⁾, A.A. Bakunovich³⁾, M.A. Kazlova⁴⁾,
D. Latyshevich⁵⁾, Yu. Hetsevich⁶⁾*

¹⁾АПП НАН Беларусі,
Мінск, Беларусь, *evgeniakacan@gmail.com*

²⁾АПП НАН Беларусі,
Мінск, Беларусь, *suprunchuk@gmail.com*

³⁾АПП НАН Беларусі,
Мінск, Беларусь, *bakunovich.andrei@gmail.com*

⁴⁾АПП НАН Беларусі,
Мінск, Беларусь, *margaryta.kazlova@gmail.com*

⁵⁾АПП НАН Беларусі,
Мінск, Беларусь *david.latyshevich@gmail.com*

⁶⁾АПП НАН Беларусі,
Мінск, Беларусь, *yuras.hetsevich@gmail.com*

¹⁾UIIP of NASB,
Minsk, Belarus, *evgeniakacan@gmail.com*

²⁾UIIP of NASB,
Minsk, Belarus, *suprunchuk@gmail.com*

³⁾UIIP of NASB,
Minsk, Belarus, *bakunovich.andrei@gmail.com*

⁴⁾UIIP of NASB,
Minsk, Belarus, *margaryta.kazlova@gmail.com*

⁵⁾UIIP of NASB,
Minsk, Belarus, *david.latyshevich@gmail.com*

⁶⁾UIIP of NASB,
Minsk, Belarus, *yuras.hetsevich@gmail.com*

Дадзены артыкул прысвечаны аналізу аўтаматычнай сегментацыі тэксту на сінтагмы беларускай літаратурнай мовы. Апісаны распрацаваны прататып інтанацыйнага працэсара, асноўнай задачай якога з'яўляецца вызначэнне мінімальнага сэнсавых адзінак у сказе на аснове знакаў прыпынку і злучнікаў.

Ключавыя словы: аўтаматычная дэлімітацыя; сінтагма; інтанацыйныя межы; інтанацыйны працэсар; сінтаксіс; пунктуацыя.

This article is devoted to the analysis of automatic textual segmentation into syntagmas of the Belarusian literary language. The developed prototype of an intonation processor is described, the major task of which is to determine the minimum semantic units in a sentence based on punctuation marks and conjunctions.

Keywords: automatic delimitation; syntagma; intonation boundaries; intonation processor; syntax; punctuation

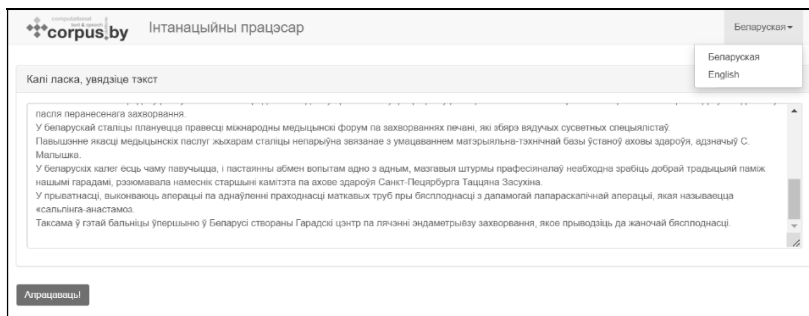
Уводзіны. У сучаснай навуцы пытанне аб зыходнай адзінцы будовы і ўспрымання маўлення не мае адназначнага рашэння менавіта таму, што разглядаецца на аснове розных падыходаў і прынцыпаў. Аднак вядома, што маўленне мае сінтагматычную прыроду і ўяўляе сабой ланцужок не асобных слоў, а адзіных лексічных комплексаў, што фарміруюць сінтагмы. Сінтагма – канкрэтная індывідуальна-аўтарская моўная структура, якая прадстаўляе інтанацыйнае, граматычнае і сэнсвае адзінства некалькіх слоў. Адэкватнае ўспрыманне маўлення забяспечваецца аўтарскай дэлімітацыяй, а правільнае разуменне тэксту дасягаецца дакладным аднаўленнем яго сінтагматычнай структуры і інтанацыі.

Як паказаў агляд наяўных у літаратуры даных, узаемадзеянне фактараў, якія ўплываюць на межы сінтагм, вывучана недастаткова [1]. Выяўлены хутчэй тэндэнцыі, і нават для заўважных канцэпцый няма дастаткова фармалізаванага апісання тэкставых варыяцый, характарыстыкі якіх маглі б служыць ключамі для аўтаматычнай расстаноўкі інтанацыйных межаў і паўз. Няма таксама дастатковых статыстычных даных, якія дазволілі б аддзяляць нарматыўнае (нейтральнае) прачытанне тэксту ад дапушчальнага ці памылковага. Асноўным доказам з’яўляецца сцвярджэнне, што кампаненты інтанацыі звязаны з тымі ці іншымі асаблівасцямі сінтаксічнай структуры сказа [2]. У той жа час узаемаадносінны інтанацыі і сінтаксісу нельга звесці да паралелізму сінтаксічных і інтанацыйных сродкаў. Інтанацыйных мадэляў, дынамічных і меладычных структур, якія маюць фаналагічнае значэнне, заўсёды менш, чым сінтаксічных мадэляў. Адны і тыя ж інтанацыйныя сродкі мовы выкарыстоўваюцца для выражэння розных сінтаксічных значэнняў у розных маўленчых сітуацыях.

З прычыны дамінуючай ролі сінтагмы агульным законам сэнсавага члянэння маўлення з’яўляецца асаблівая, уласцівая кожнай канкрэтнай мове, рытміка-меладычная будова сінтагмы і маркіраванне сінтагматычных межаў. Адным з шляхоў вырашэння пастаўленай праблемы з’яўляецца распрацоўка канкрэтных метадаў і алгарытмаў аналізу і апрацоўкі інтанацыйных асаблівасцей натуральнага маўлення, яго аўтаматычнага сінтагматычнага падзелу і рэалізацыі ўсіх інтанацыйных канструкцый зададзенай мовы, што прывядзе да аўтаматызаванага прайгравання адвольнага тэксту з манерай чытання чалавека, а не штучнай сістэмы [3]. Такім чынам, на падставе дадзенага падыходу супрацоўнікі лабараторыі распазнавання і сінтэзу маўлення АПП НАН Беларусі [4] распрацавалі праграмны прадукт “*Інтанацыйны*

працэсар”[5], які аналізуе сінтаксічную структуру тэкставай інфармацыі і вызначае інтанацыйныя межы сінтагм. Дадзены сэрвіс знаходзіцца ў адкрытым доступе на платформе для апрацоўкі тэкставай і гукавай інфармацыі для розных тэматычных даменаў Corpus.by [6].

Асноўная частка. Прынцып функцыянавання сістэмы прадстаўлены на малюнку 1, згодна з якім уваходны тэкст паслядоўна праходзіць лінгвістычную апрацоўку на аснове ўсіх знакаў прыпынку і злучнікаў, якія сустракаюцца ў тэксце. Сэрвіс просты ў выкарыстанні. Пасля кліка кнопкі “Апрацаваць!” уваходны тэкст разбіваецца на сінтагмы (малюнак 2) у выглядзе спіса слоў з пазнакай іх граматычнага класа і ўказаннем знакаў прыпынку. У хуткім часе плануецца цалкам дапрацаваць сэрвіс для вызначэння ўсіх тыпаў сінтагм, заснаваных на сінтактыка-семантычным падыходзе (вылучэнне фразеалагізмаў і сінтагм, аб’яднаных сэнсавымі адносінамі без абмежавання знакамі прыпынку).



Рыс. 1. Інтэрфейс прататыпа «Інтанацыйны працэсар»

Сэрвіс таксама прадстаўляе наступныя выніковыя даныя:

У акне “Вынік тэкставай апрацоўкі” пералічаны паслядоўны спіс слоў, у якім выводзяцца даныя аб часцінамоўнай прыналежнасці кожнай адзінкі, знойдзенай сістэмай у тэксце: для слова вызначаюцца часціны мовы і іх тэгі, якія з’яўляюцца скарачаным абазначэннем марфалагічных прыкмет слова (малюнак 2). Кожны тэг надаецца слову згодна з асобным слоўнікам, убудаваным у платформу Corpus.by, а менавіта:

– sbm1987 – «Слоўнік беларускай мовы. Арфаграфія. Арфаэпія. Акцэнтуацыя. Словазмяненне / пад рэд. М.В. Бірылы. – Мінск, 1987».

– pou12013 – назоўнікі, згодна з выданнем «Граматычны слоўнік назоўніка / навук. рэд. В.П. Русак. – Мінск : Бел. навука, 2013».

– adjective2013 – прыметнікі, згодна з выданнем «Граматычны слоўнік прыметніка, займенніка, лічэбніка, прыслоўя / навук. рэд. В.П. Русак. – Мінск : Бел. навука, 2013».

– numeral2013 – лічэбнікі, згодна з выданнем «Граматычны слоўнік прыметніка, займенніка, лічэбніка, прыслоўя / навук. рэд. В.П. Русак. – Мінск : Бел. навука, 2013».

– pronoun2013 – займеннікі, згодна з выданнем «Граматычны слоўнік прыметніка, займенніка, лічэбніка, прыслоўя / навук. рэд. В.П. Русак. – Мінск : Бел. навука, 2013».

– verb2013 – дзеясловы, згодна з выданнем «Граматычны слоўнік дзеяслова / навук. рэд. В.П. Русак. – Мінск : Бел. навука, 2013».

– adverb2013 – прыслоўі, згодна з выданнем «Граматычны слоўнік прыметніка, займенніка, лічэбніка, прыслоўя / навук. рэд. В.П. Русак. – Мінск : Бел. навука, 2013».

– zalizniak – «Грамматический словарь русского языка: Словоизменение / А.А. Зализняк. – Москва : Русский язык, 1980. – 880 с.».



Рыс. 2. Вынікі тэкставай апрацоўкі «Інтанацыйным працэсарам»

Таксама праграма выдае дадатковыя даныя лінгвістычнай апрацоўкі. Так, напрыклад, карыстальнік можа азнаёміцца з ўсімі невядомымі словамі ў акне “Невядомыя токены” і са спісам амографу ў акне “Амографы” (малюнак 2). У акне “Апрацоўка знакаў прыпынку” адлюстраваны спіс пунктуацыйных сінтагм з указаннем камп’ютарнай інтанацыйнай мадэлі (у дужках). Паколькі пунктуацыйныя сінтагмы могуць быць рознай даўжыні, пад пунктуацыйнай фразай маецца на ўвазе любая паслядоўнасць словаформаў або фанетычных слоў, раздзеленых знакамі прыпынку. Пры абзначэнні тыпа інтанацыі сінтагмы

ўлічваюцца не толькі знакі прыпынку, але і найбліжэйшы кантэкст у тэксце.

Акно “Апрацоўка злучнікаў” дэманструе разбіцце сказаў на лексічныя сінтагмы згодна са злучнікам як фармальным паказчыкам інтанацыйных межаў. Выкарыстоўваючы прыведзеныя даныя, карыстальнік можа азнаёміцца са спісам знакаў прыпынку, злучнікамі і іх фармальным маркерам, якія сустраляся ў тэксце, з левым і правым кантэкстам (малюнак 3), што апісвае сінтаксічную структуру сказа.



Рыс. 3. Прамежжавыя даныя апрацоўкі знакаў прыпынку і злучнікаў ва ўваходным тэксце

Заклучэнне. Такім чынам, першапачатковым этапам сінтагматычнай дэлімітацыі і інтанацыйнай апрацоўкі ўваходнага тэксту з’яўляецца вызначэнне межаў сінтагм, што магчыма праз выкарыстанне сэрвіса “Інтанацыйны працэсар”. Практыка стварэння аўтаматычных сістэм сінтэзу і распазнавання маўлення паказвае, што найважнейшымі ключамі для вызначэння інтанацыйных межаў пры агучванні тэксту з’яўляюцца сінтактыка-семантычныя адносіны ўнутры сказа. Аднак для аўтаматычнай дэлімітацыі тэксту неабходны аўтаматычны сінтаксічны аналіз, які на бягучы момант цалкам не рэалізаваны. У той жа час распрацоўка канкрэтных прыкладанняў для розных моў сведчыць аб тым, што для расстаноўкі інтанацыйных межаў аднаўленне поўнай сінтаксічнай структуры сказа неабавязковае. Прызнана, што арганізацыя прасадыхных складнікаў прасцейшая, чым звязаныя з імі сінтаксічныя залежнасці. Стварэнне прататыпнай сістэмы “Інтанацыйны працэсар” для падзелу тэкставай інфармацыі на мінімальныя сэнсавыя адрэзкі дазваляе дадзены факт. Сэрвіс ажыццяўляе пошук сінтагм праз знакі

прыпынку і злучнікі. На наступным этапе адбудзецца аналіз алгарытмаў на аснове тэсціравання існуючых лінгваакустычных рэсурсаў, якія пакрываюць усю разнастайнасць сінтаксічных канструкцый беларускай літаратурнай мовы. Гэта дазволіць выявіць недакладнасці ў прадстаўленым прататыпе, якія будуць выпраўлены ў найбліжэйшы час.

БІБЛІАГРАФІЧНЫЯ СПАСЫЛКІ

1. Зяноўка, Я.С. Перспектывы развіцця аўтаматычнай апрацоўкі вуснага маўлення / Я.С. Зяноўка, М.В. Супрунчук, Ю.С. Гецэвіч // Лінгвістыка, лінгводидактыка, лінгвокультуралогія: актуальныя пытанні і перспектывы развіцця / Беларускі гос. ун-т ; редкол.: Н. А. Курковіч (гл. ред.) і др. – Мінск : БГУ, 2022. – С. 252-257.
2. Зяноўка, Я.С. Прынцыпы аўтаматычнага вызначэння інтанацыйных партрэтаў беларускага маўлення / Я. С. Зяноўка // Лінгвістыка, лінгводидактыка, лінгвокультуралогія: актуальныя пытанні і перспектывы развіцця : матэрыялы IV Міжнароднага навучна-практ. канф., Мінск, 19–20 сакавіка 2020 г. / редкол.: О.Г. Прохоренка (отв. ред.) [і др.]. – Мінск : БГУ, 2020. – С. 472-477.
3. Zianouka, Y. Automatic Generation of Intonation Marks and Prosodic Segmentation for Belarusian NooJ Module / Y. Zianouka, Y. Hetsevich, D. Latyshevich, Z. Dzenisiuk // 15th International Conference, NooJ 2021 / France ; ed. Magali Bigey, Annabel Richeton, Max Silberstein, Izabella Thomas. – Besançon : Springer, Cham, 2022. – P. 231-242.
4. Лабараторыя распазнавання і сінтэзу маўлення [Электронны рэсурс]. – 2022. Рэжым доступу: <http://ssrlab.by/>. – Дата доступу: 11.07.2017.
5. Інтанацыйны працэсар [Электронны рэсурс]. – 2023. Рэжым доступу : <https://corpus.by/IntonationalProcessor/?lang=be>. – Дата доступу : 29.01.2023.
6. Платформа для апрацоўкі тэкставай і гукавай інфармацыі для розных тэматычных даменаў corpus.by // [Электронны рэсурс]. – 2022. Рэжым доступу : <http://corpus.by/>. – Дата доступу : 12.01.2023.

АВТОМАТИЧЕСКОЕ ПОСТРОЕНИЕ ЧАСТОТНОГО СЛОВАРЯ

AUTOMATIC CONSTRUCTION OF FREQUENCY DICTIONARY

Ю.Ю. Красовская

J.J. Krasowskaja

Белорусский государственный университет,
Минск, Беларусь, hisradaar@gmail.com
Belarusian State University,
Minsk, Belarus, hisradaar@gmail.com

Частотный словарь включает в себя слова или словоформы какого-либо текста или корпуса текстов и частоту их встречаемости в исследуемом корпусе текстов. Составление частотного словаря посредством программного обеспечения позволяет автоматизировать процесс. Автоматизация с применением лемматизации позволит упростить долгий про-

<i>Трегубова Ю.А.</i> Национально-культурная составляющая иноязычного художественного текста: вопросы перевода	252
<i>Троцинская-Степушина Т.Е.</i> Феномен Чехова	256
РАЗДЕЛ 4. АКТУАЛЬНЫЕ НАПРАВЛЕНИЯ ПРИКЛАДНОЙ ЛИНГВИСТИКИ	262
<i>Бабаян М.А.</i> Мем как феномен современной интернет-коммуникации	262
<i>Глазкова С.Н.</i> Дериват в языковом сознании школьника как гештальт	267
<i>Жданович М.А.</i> Применение облака слов в специальном переводе	274
<i>Зяноўка Я.С., Супрунчук М.В., Бакуновіч А.А., Казлова М.А., Латышэвіч Д.І., Гецэвіч Ю.С.</i> Аўтаматычная дэлімітацыя тэкставай інфармацыі для беларускай мовы.....	278
<i>Красовская Ю.Ю.</i> Автоматическое построение частотного словаря... ..	283
<i>Лучкина Н.В., Мирзоева С.А., Проценко И.Ю.</i> Особенности межличностного общения в интернет-пространстве	288
<i>Рыжковіч Г.Ч.</i> Фразеалагізмы са словам <i>время</i> ў назвах акаўнтаў сацыяльнай сеткі Instagram	294
<i>Садовская Е.Ю.</i> О токсичности в межпоколенческом дискурсе	297
<i>Селезнева В.С.</i> Метаязыковой инструментарий описания экспертной системы	301
РАЗДЕЛ 5. ПЕРВЫЕ ШАГИ В НАУКЕ	305
<i>Богущ В.А., Буйновская А.Э.</i> Фразеологизмы о мужчинах и женщинах в английском языке: лингвокультурологический анализ	305
<i>Гончаренко О.С., Данильчик У.А.</i> О стратегии использования Jolly Phonics в обучении детей чтению	308
<i>Кривошеева Д.А., Гончаренко О.С.</i> Метод полного физического реагирования (TPR) в обучении детей	310
<i>Дубровская Е.В., Ломонос М.А.</i> Проблемы применения тестирования при контроле усвоения иностранного языка	314