

## **ВЫКАРЫСТАННЕ СІСТЭМ МАШЫННАГА ПЕРАКЛАДУ І СІСТЭМЫ СІНТЭЗУ МАЎЛЕННЯ ДЛЯ ЗАБЕСПЯЧЭННЯ ДАСТУПНАСЦІ ЗАКАНАДАЎЧЫХ ТЭКСТАЎ ДЛЯ ЛЮДЗЕЙ З ІНВАЛІДНАСЦЮ ПА ЗРОКУ**

Ю. С. Гецэвіч<sup>1</sup>, В. В. Варановіч<sup>2</sup>, А. У. Бабкоў<sup>1</sup>

<sup>1</sup>Аб'яднаны інстытут праблем інфарматыкі НАН Беларусі, Мінск;

<sup>2</sup>Беларускі дзяржаўны ўніверсітэт, Мінск

*Падымаецца пытанне даступнасці юрыдычных тэкстаў на дзвюх дзяржаўных мовах для людзей з інваліднасцю па зроку на прыкладзе кодэксаў Рэспублікі Беларусь. Паказваюцца звязаныя з распаўсюджваннем Інтэрнэту змяненні ў адносінах грамадства да людзей, якія слаба бачаць. Звяртаецца ўвага на комплекснасць паняцця «даступнасць заканадаўчых тэкстаў» у білінгвістычным грамадстве; разглядаюцца магчымасці машыннага перакладу на сайце [etalonline.by](http://etalonline.by) і магчымасці выкарыстання партала [corpus.by](http://corpus.by) у мэтаклінгвістычнай апрацоўкі тэкстаў.*

Па інфармацыі Сусветнай арганізацыі аховы здароўя ў свеце налічваецца больш за 40 млн невідзучых і больш за 250 млн людзей з вельмі дрэнным зрокам. Ёсць яшчэ людзі з парушэннямі каляровых адчуванняў, колькасць якіх вар'іруецца ад 4 да 8 % насельніцтва нашай планеты. Толькі ў Беларусі на 1 жніўня 2018 г. пражывала 567,5 тыс. чалавек з інваліднасцю, зарэгістраваных у органах па працы, занятасці і сацыяльнай абароне, што складае амаль 6 % ад агульнай колькасці насельніцтва рэспублікі. Калі дадаць да гэтага 1,9 млн беларусаў, старэйшых за 60 гадоў, большасць з якіх слаба бачыць, то стане відавочна, што дадзеная праблема датычыцца чвэрці жыхароў Беларусі.

Раней людзі, якія слаба бачаць, былі амаль цалкам ізаляваныя ад грамадства. Цяпер грамадства пачынае разумець: людзі з інваліднасцю па зроку існуюць не асобна, а з'яўляюцца яго часткай. Камп'ютар дазваляе ім не замыкацца ў сваім вузкім асяроддзі.

Пра сучасныя тэхналогіяў у віртуальнай прасторы даступна ўсё: і адукацыя, і ўзаемаадносіны, і працаўладкаванне. Чалавецтва прывыкае анлайн дазнавацца пра навіны, будаваць маршруты праезду, працаваць, рабіць пакупкі, кантактаваць і г. д. Аднак, калі інтэрнэт-сайт не створаны з улікам даступнасці, ён можа выключыць частку насельніцтва, якая больш за ўсё мае патрэбу ў ім, – людзей з інваліднасцю па зроку.

Таму 18 кастрычніка 2016 г. Беларусь ратыфікавала міжнародную Канвенцыю аб правах інвалідаў, у артыкуле 9 якой так гаворыцца аб інтэрнэт-даступнасці: «Дзяржавы-ўдзельніцы... прымаюць належныя меры для таго, каб:

развіваць іншыя... формы аказання дапамогі і падтрымкі людзям з інваліднасцю для забеспячэння ім доступу да інфармацыі;

спрыяць доступу людзей з інваліднасцю да новых інфармацыйна-камунікацыйных тэхналогій і сістэм, у тым ліку да Інтэрнэту;

садзейнічаць праектаванню, распрацоўцы, вытворчасці і распаўсюджванню даступных інфармацыйна-камунікацыйных тэхналогій, каб даступнасць такіх тэхналогій і сістэм дасягалася з мінімальнымі выдаткамі» (URL: [http://mintrud.gov.by/gu/new\\_url\\_369854369](http://mintrud.gov.by/gu/new_url_369854369)).

Канвенцыя аб правах інвалідаў – антыдыскрымінацыйны дакумент, які адрозніваецца ад іншых міжнародных дакументаў сваёй інавацыйнасцю і прагрэсіўнасцю. Падпісанне прадугледжвае фактычную рэалізацыю міжнародных абавязацельстваў на ўнутрыдзяржаўным узроўні, уключэнне міжнародна-прававых нормаў у нацыянальную

правую сістэму, дзеля чаго ў Беларусі з 2018 г. распрацоўваецца праект закона «Аб правах інвалідаў і сацыяльнай інтэграцыі».

4 снежня 2017 г. Беларусь у рамках Праграмы развіцця ААН таксама падпісала Інфармацыйную стратэгію Беларусі па паўнапраўным уключэнні (інклюдзіі) людзей з інваліднасцю ў грамадства. А з 1 студзеня 2019 г. уступіла ў сілу пастанова Савета міністраў Рэспублікі Беларусь № 797 ад 23 кастрычніка 2017 г., паводле якой унесены змяненне і дапаўненні ў Палажэнне аб парадку функцыянавання інтэрнэт-сайтаў дзяржаўных органаў і арганізацый (URL: [http://etalonline.by/document/?regnum=c21700797&q\\_id=643794](http://etalonline.by/document/?regnum=c21700797&q_id=643794)). Удакладнена, што інфармацыя на інтэрнэт-сайтах размяшчаецца на рускай і (або) беларускай мовах, а пры неабходнасці таксама на адной або некалькіх замежных мовах. Прымяненне і беларускай, і рускай моў неабходна пры размяшчэнні ўсёй інфармацыі, якая з'яўляецца абавязковай для галоўнай старонкі інтэрнэт-сайта. Устаноўлена, што на інтэрнэт-сайтах неабходна прадугледзець даступнасць інфармацыі ў адпаведнасці з патрабаваннямі, вызначанымі ў тэхнічных нарматыўна-прававых актах. Пры гэтым інтэрнэт-сайт павінен падтрымліваць версію для людзей з інваліднасцю па зроку і быць сумяшчальным з рознымі інтэрнэт-браўзерамі.

Так, з мэтай забеспячэння агульнадаступнасці інфармацыі камітэтам па ахове здароўя Мінгарвыканкама была распрацавана версія інтэрнэт-сайта для людзей з інваліднасцю па зроку. Яна дазваляе ўключаць ці выключаць галасавыя падказкі, якія прагаворваюць назвы раздзелаў інтэрнэт-сайта, кнопак і пунктаў меню, агучваць навіны па-беларуску і па-руску з дапамогай усталяванай сістэмы сінтэзу маўлення, а таксама выбіраць камфортныя для ўспрымання памер шрыфту і каляровую палітру (URL: <https://komzdrav-minsk.gov.by/>).

Для грамадзян Беларусі, маючых праблемы са зрокам, акрамя агульных, існуе дадатковая праблема, звязаная з тым, што ў нашай краіне дзве дзяржаўныя мовы: беларуская і руская. Але ж доўгі час аб даступнасці нарматыўных актаў толькі ўзгадвалася ў кантэксце патрабаванняў да якасці іх моўнага выкладу. Напрыклад, у Законе Рэспублікі Беларусь ад 10 студзеня 2000 г. «Аб нарматыўных прававых актах Рэспублікі Беларусь» у ліку патрабаванняў да мовы дакументаў вызначаны і такія, як яе яснасць, прастата і даступнасць (URL: <http://etalonline.by/?type=text&regnum=h10000361#load>). Забеспячэнне даступнасці і зразумеласці для кожнага грамадзяніна – гэта фактычна канчатковая мэта працэсу ўдасканалення заканадаўства ў любой прававой дзяржаве. Аднак дадзеныя мэты заўсёды будзе заставацца недасягальнай без уліку моўнай сітуацыі ў краіне, без арыентавання на кожнага адрасата прававой інфармацыі (у тым ліку на чалавека з інваліднасцю) як на моўную асобу. Інакш кажучы, якім бы ні быў заканадаўчы тэкст дасканалы апрацаваным з пункту гледжання дакладнасці, лагічнасці, яснасці і г. д., як бы добра не былі забяспечаны ўсе неабходныя ўмовы для азнаямлення з ім, даступнасць тэксту ўсё роўна не будзе поўнай, калі ён створаны і існуе на адной мове ў білінгвістычным грамадстве. Але ў Рэспубліцы Беларусь, нягледзячы на дзяржаўнае двухмоўе, пераважная большасць заканадаўчых дакументаў рэалізавана толькі на адной мове. Так, з 26 кодэксаў Рэспублікі Беларусь, тэксты якіх прадстаўлены на Нацыянальным прававым інтэрнэт-партале pravo.by, 25 апублікаваны толькі па-руску, 1 – толькі па-беларуску.

Адзін з асноўных фактараў, якія стрымліваюць практычнае забеспячэнне двухмоўя ў прававой сферы Рэспублікі Беларусь, – нявырашанасць праблемы якаснай і хуткай лінгвістычнай апрацоўкі тэкстаў вялікага памеру, што сведчыць аб актуальнасці якаснага машыннага перакладу.



Атрымаць неафіцыйную беларускамоўную версію рускамоўнага нарматыўнага акта або неафіцыйную рускамоўную версію беларускамоўнага нарматыўнага акта сёння магчыма, калі звярнуцца на сайт Нацыянальнага цэнтра прававой інфармацыі Рэспублікі Беларусь (URL: <http://etalonline.by/>) і націснуць кнопку «Машыны пераклад». Пераклад азначаўляецца за некалькі хвілін, а яго якасць дазваляе ў цэлым зразумець дакумент, але звяртае на сябе ўвагу сэнсавая нераўнацэннасць лексічных адзінак або сінтаксічных канструкцый зыходнай рускай мовы і беларускай мовы, напрыклад (руская лексема – прапанаваная машынным перакладчыкам лексема – правільны аналаг): *отношения – прапанаваная машынным перакладчыкам лексема – правільны аналаг: **отношения – дачыненні – адносіны; настоящий – сапраўдны – дадзены; управлением, действующим на кіраванні на працы – упраўленнямі на працы; законодательством, действующим на кіраванні на працы – упраўленнямі на працы; территория, – заканадаўствам, якое ёе тэрыторыя, – заканадаўствам, якое дзее на яе тэрыторыі, – заканадаўствам, якое дзее на яе тэрыторыі; родственные, – рабочнікамі, існуючымі ў блізкай родстве, – рабочнікамі, якія складаюцца ў блізкай сваяцтве, – рабочнікамі, якія знаходзяцца ў блізкай сваяцтве; ответственных за соблюдение законности – адказных за прытрымліванне законнасці – адказных за захаванне законнасці.***

Менавіта таму пастрадаванне тэкстаў лінгвістамі – гэта неабходная праца, якая можа быць аптымізавана пры дапамозе праграмага забеспячэння (URL: <http://corpus.by/>), распрацаванага супрацоўнікамі лабараторыі распазнавання і сінтэзу маўлення Аб'яднанага інстытута праблем інфарматыкі (АПП) НАН Беларусі. У прыватнасці, пры перакладзе тэкстаў кодэксаў Рэспублікі Беларусь карыснымі з'яўляюцца алгарытмы сістэмы праверкі правапісу, статыстычнай і дэведачнай інфармацыі аб сімвалах, статыстыкі ўжывання адвольных сімвальных паслядоўнасцей у электронным тэксце, пошуку і выпраўлення памылак у напісанні літар «у» і «ў», распазнавання і вылучэння ў тэксце амографу.

Машыны перакладчык можа быць удасканалены, калі ён будзе мець магчымасць пастаянна напрацоўваць свае навыкі ў адной і той жа прадметнай галіне, грунтоўчыся на правілах, якія выбіраюцца з зададзеных крыніц інфармацыі, а з удасканаленнем машыннага перакладчыка пачне вырашацца і праблема забеспячэння даступнасці заканадаўства для носьбітаў розных моў (у тым ліку сярод людзей з інваліднасцю па зроку) – адна з галоўных задач, што вырашаюцца ў кожнай дзяржаве ў мэтах інфармацыйнай бяспекі [1].

У лабараторыі распазнавання і сінтэзу маўлення АПП НАН Беларусі існуе шэраг распрацаваных сінтэзатараў беларускага і рускага маўлення, сярод якіх стацыянарны [2], мабільны і анлайн-сінтэзатар [3], што дазваляе людзям з інваліднасцю па зроку праз аўтаматычнае агучванне тэксту сінтэзатарам праслухоўваць патрэбны ім кантэнт на розных прыстасаваннях.

Дадзеныя распрацоўкі таксама выкарыстоўваюцца для працы з юрыдычным даменам беларускай лексікі пры перакладзе з рускай мовы на беларускую і далейшым агучванні кодэксаў Рэспублікі Беларусь, у тым ліку Жыллёвага, Банкаўскага, Ляснога, Кодэкса аб зямлі, Кодэкса аб шлюбе і сям'і, Кодэкса аб адміністрацыйных правапарушэннях.

У прыватнасці, падчас працы над Банкаўскім кодэксам (колькасць слоў у рускамоўным тэксце – 43 507) найбольшую цяжкасць выклікаў пераклад спецыяльных *прудэнцыяльных, бенефіцыйар – бенефіцыйар, факторинг – факторынг, отзыв (лицензии) – адкліканне (ліцэнзіі)*. Таксама цяжка паддаваліся алгарытмаванню ў сістэмах машыннага перакладу наступныя асаблівасці беларускай мовы (у параўнанні з рускай): – канчаткі назоўнікаў мужчынскага роду ў родным склоне: *фактару* (у значэнні 'чыннік') і *фактара* (у значэнні 'фізічная ці юрыдычная асоба');

– адрозненні ў граматычным афармленні канструкцый з ступенямі параўнання прыметнікаў і прыслоўяў: *цена не ніжже минимальной – цана, не ніжэйшая за мінімальную;*

– дзееспрыметнікі, дзееспрыметныя звароты: *нормы банкоўскага заканадаўства, содэржацца в міжнародных договорах – нормы банкаўскага заканадаўства, якія змяшчаюцца ў міжнародных дагаворах;*

– пераважнае выкарыстанне меснага склону пасля прыназоўніка «па» (у рускай мове – давальны склон): *правовыя акты по бухгалтёрскому учету – прававыя акты па бухгалтёрскім уліку.*

Пасля перакладу ўсё тэкст Банкаўскага кодэкса быў агучаны мадэрнізаваным анлайн-сінтэзатарам BarysBelHigh, які дае магчымасць атрымліваць больш чыстае гучанне за кошт выдалення часткі басовага спектра.

Далей лабараторыяй распазнавання і сінтэзу маўлення плануецца прапаноўваць для тэсціравання і наступнай практычнай эксплуатацыі сістэмы сінтэзу і распазнавання маўлення па тэксце, сістэмы машынага перакладу тэкстаў юрыдычнага і медыцынскага дамена, удзельнічаць у развіцці і нападзенні дзяржаўных інфармацыйна-прававых рэсурсаў. Для гэтага яшчэ 7 жніўня 2018 г. АПП НАН Беларусі і Нацыянальным цэнтрам прававой інфармацыі Рэспублікі Беларусь быў заключаны дагавор аб супрацоўніцтве ў галіне навуковых распрацовак па развіцці і ўдасканаленні інфармацыйных тэхналогій і сістэм, нацыянальных інфармацыйных рэсурсаў і іх рацыянальным выкарыстанні, у тым ліку па выпрацоўцы карпусоў беларускай, рускай і англійскай лексікі з дапамогай партала corpus.by.

Такім чынам, ва ўмовах новага інфармацыйнага асяроддзя чалавечай цывілізацыі – інфасферы, якая цягне за сабой радыкальныя сацыяльныя змены і істотным чынам датычыцца жыцця практычна кожнага грамадзяніна, – у Беларусі робяцца вельмі сур’ёзныя і патрэбныя крокі ў бок дапамогі людзям з інваліднасцю па зроку. Значную ролю пры гэтым адыгрывае забеспячэнне даступнасці інфармацыі для безбар’ернай навігацыі па заканадаўстве Рэспублікі Беларусь усіх катэгорый моўных носьбітаў, нягледзячы на іх асаблівасці.

## Спіс літаратуры

1. Гецэвіч, Ю. С. Выкарыстанне сістэм машынага перакладу і сінтэзу маўлення для забеспячэння даступнасці заканадаўчых тэкстаў на розных мовах у Рэспубліцы Беларусь / Ю. С. Гецэвіч, А. А. Кірлун // Інфармацыйныя тэхналогіі і право (Правовая інфарматызацыя – 2018) : матэрыялы VI Міждунар. навуч.-практ. конф., Мінск, 17 мая 2018 г. / Нац. цэнтр прававой інформ. Респ. Беларусь ; под общ. ред. Е. И. Коваленко. – Мінск, 2018. – С. 123–128.
2. Гецевич, Ю. С. Система синтеза белорусской речи по тексту / Ю. С. Гецевич, В. М. Лобанов // Речевые технологии. – 2010. – № 1. – С. 91–100.
3. Гецэвіч, Ю. С. Распрацоўка сінтэзатара беларускага і рускага маўленняў па тэксце для мабільных і інтэрнэт-платформаў / Ю. С. Гецэвіч, Д. А. Пакладок, Д. В. Брэж // Развитие информатизации и государственной системы научно-технической информации (РИНТИ–2012) : докл. XI Міждунар. конф., Мінск, 15 нояб. 2012 г. – Мінск : ОИПИ НАН Беларусі, 2012. – С. 254–259.