

ВЕСЦІ

НАЦЫЯНАЛЬНАЙ
АКАДЭМІІ НАВУК БЕЛАРУСІ

Серыя
гуманітарных
навук
№ 2

АСОБНЫ АДЫТАК

2003

УДК 002

Л. І. ПЯТРОВА, Ю. Ф. ШПАКОЎСКИ
ПРАБЛЕМА ІНФАРМАТЫЎНАСЦІ ВЫДАННЯЎ

Беларускі дзяржаўны тэхналагічны ўніверсітэт

(Паступіў у рэдакцыю 28.01.2003)

Галоўнай мэтай любога выдання з'яўляецца перадача змешчанай у ім інфармацыі, якая павінна не толькі дайсці да свайго адрасата (чытача), але і быць адэкватна ім успрынята. Аўтар (а таксама рэдактар) імкнецца да таго, каб чытач атрымаў максімальную колькасць інфармацыі з перадаваемага паведамлення. Але як вызначыць, колькі інфармацыі змяшчаецца ў тэксце і якія фактары ўздзейнічаюць на яе ўспрыманне? Гэтыя пытанні ўяўляюць безумоўную цікавасць для рэдакцыйна-выдавецкіх супрацоўнікаў.

Напрыклад, як падлічыць, колькі інфармацыі заключаецца ў кантэкстах тыпу: «Кожны рукапіс патрабуе рэдагавання», «Тэкст змяшчае інфармацыю аб найноўшых вылічальных машынах», «Усе кіты — млекакормячыя». Пытанне можна паставіць і такім чынам: ці ёсць у вышэйназваных сказах якая-небудзь інфармацыя для пэўнага чытача? Статыстычная тэорыя інфармацыі не зможа даць адказ на гэтае пытанне, таму што не ўлічвае адносін прыёмніка (чытача) да перадаваемага паведамлення, у той час як для аналізу тэкста гэтыя адносін маюць вялікае значэнне (прычым асабліва важнай характарыстыкай з'яўляецца магчымасць прыёмніка да адэкватнага ўспрымання інфармацыі). Для прыкладу возьмем выраз: «У нейкай арганізацыі распрацавана вылічальная машына, якая выконвае 100 млн аперацый за секунду». Відавочна, што для чалавека, малазнаёмага з вылічальнай тэхнікай, гэты выраз не нясе ніякай інфармацыі. Гэты чалавек ведае, што машына лічыць вельмі хутка, і для яго ў дадзены момант розніца паміж хуткасцямі ў 1000 і 100 млн аперацый за секунду не істотна. Для спецыяліста ж гэты выраз у выніку яго асэнсавання азначаў бы вельмі важны факт: у машынах пачалі выкарыстоўваць элементы, заснаваныя на непасрэдным распаўсюджванні электрамагнітнага поля ў асяроддзі [1]. Усё гэта пацвярджае той факт, што атрыманне істотнай інфармацыі магчыма толькі ў тым выпадку, калі прыёмнік (чытач) валодае ўжо некаторым запасам звестак, неабходных для адэкватнага ўспрымання перадаваемага паведамлення.

У сувязі з гэтым уяўляецца мэтазгодным падыход да праблемы вызначэння колькаснай характарыстыкі семантычнай інфармацыі, прапанаваны Ю. А. Шрэйдэрам [1]. Ёсць некаторы спіс назваў аб'ектаў, прэдыкатаў (уласцівасцей) і класаў аб'ектаў. Акрамя таго, паміж некаторымі назвамі аб'ектаў (ці класаў) і назвамі прэдыкатаў вызначаны сувязі наступных тыпаў: «Дадзеная ўласцівасць прымяніма да аб'ектаў дадзенага класа»; «Дадзены аб'ект валодае (не валодае) дадзенай уласцівасцю» і г. д. Спіс тыпаў сувязяў можа папаўняцца. У прыватнасці, акрамя сувязяў паміж парамі назваў можна разглядаць сувязі паміж тройкамі, чацвёркамі назваў і г. д. Перш за ўсё гэты эфект узнікае з-за шматмясцовых прэдыкатаў. Так, напрыклад, прэдыкат «быць суседам» падразумявае два аб'екты. Выказанне «Іван — сусед Пятра» змяшчае два аб'екты і адзін прэдыкат. Адносіны тыпу «Прэдыкат A сапраўдны пры падстаноўцы аб'ектаў X і Y » ці «Прэдыкат A прымянім да аб'ектаў класаў K_1 і K_2 » звязваюць па тры назвы. Прэдыкаты з вялікай колькасцю месцаў прыводзяць да адносін, якія ў сваю чаргу звязваюць вялікія групы назваў. Іншыя прыклады ўтвараюць сувязі тыпу «Усе аб'екты класа A , якія валодаюць уласцівасцю X , уваходзяць у клас B »; «Тыя аб'екты класа A , якія валодаюць уласцівасцю X , уваходзяць у клас B »; «Аб'яднанне (перасячэнне) класаў A і B утвараюць клас C » і да т. п.

Такі спіс з зададзенымі сувязямі называецца тэзаўрусам. Праз гэта паняцце стала магчымым даць азначэнне колькаснай характарыстыкі семантычнай інфармацыі [1]: колькасцю семантычнай інфармацыі, змешчанай у тэксце T адносна тэзаўруса Q , называецца мера змянення тэзаўруса ў выніку аналізу тэксту T .

Такое азначэнне колькасці інфармацыі зручна для фармальнага апісання некаторых задач, звязаных з эфектыўнасцю чытання рознага роду літаратуры асобамі з пэўнай, адносна аднолькавай ступенню падрыхтаванасці. Дапусцім, ёсць мадэль агульнай для вывучаемай групы асоб сістэмы паняццяў і існуючымі паміж гэтымі паняццямі сувязяў. Тады, калі неабходна вызначыць, якую колькасць інфармацыі можа атрымаць некаторы статыстычны па вызначаных параметрах член дадзенай групы (назавём яго чытач *C*) пры чытанні дадзенага тэксту *T*, патрэбна вызначыць, наколькі змянілася яго сістэма паняццяў і сувязяў, існуючых паміж паняццямі [2].

Аўтары артыкула згодны з працай [2], дзе адзначаецца, што атрыманне такой характарыстыкі вельмі складана, паколькі можна толькі прыблізна ацаніць запас ведаў (а тым больш уяўленняў і ўстановак), якім валодае чалавек да і пасля чытання якога-небудзь тэксту, і метады колькаснага вымярэння будуць настолькі грубымі, што памылка вымярэнняў можа перавысіць велічыню змянення тэзаўруса. Таму распрацаваць метадыку, з дапамогай якой можна было б дастаткова дакладна і адэкватна вызначаць велічыню змянення тэзаўруса, вельмі складана. Самае разумнае ў дадзенай сітуацыі прапанаваць шэраг тэкстаў, з дапамогай якіх тэсціруюць кожнага члена дадзенай групы. Але гэта будзе займаць шмат часу.

Такім чынам, прапануецца выкарыстоўваць ускосныя метады ацэнкі колькаснай інфармацыі, змешчанай у тэксце, з улікам псіхалінгвістычнага аналізу працэсу чытання розных тэкстаў людзьмі з пэўнай ступенню падрыхтаванасці.

У працы [2] вылучаюцца два відавочныя палажэнні, з якімі можна пагадзіцца:

1. Чым лепшыя ўмовы для максімальна магчымага змянення тэзаўруса, тым больш будзе ступень гэтага змянення і, як вынік, тым больш семантычнай інфармацыі змяшчае дадзены тэкст адносна дадзенага тэзаўруса.

2. Неабходнымі, але і недастатковымі ўмовамі таго, каб адбылося найбольшае змяненне тэзаўруса *Q* пры ўздзеянні на яго тэксту *T*, з'яўляюцца:

а) чытач *C* павінен адэкватна зразумець тэкст *T*;

б) узровень падрыхтаванасці чытача *C* не павінен нашмат перавышаць той узровень, на які разлічаны тэкст *T*.

Іншымі словамі, пры чытанні якога-небудзь тэксту найбольшая колькасць інфармацыі можа быць атрымана ў тым выпадку, калі чытач валодае пэўным запасам навыкаў і ўменняў, неабходных для чытання дадзенага тэксту, і ўзровень яго ведаў адпавядае таму ўзроўню, на які разлічаны гэты тэкст (пры тым разумна ўлічваць толькі адэкватна ўспрынятую інфармацыю). Пры аптымальным злучэнні складанасці тэксту і ўзроўню падрыхтаванасці чытача (пры ўсіх іншых роўных умовах) дадзены тэкст будзе змяшчаць максімальную колькасць інфармацыі для дадзенага чытача. Такім чынам, колькасная характарыстыка таго, наколькі складанасць тэксту адпавядае ўзроўню падрыхтаванасці чытача — адна з пэўных ацэнак колькасці інфармацыі, змешчанай у дадзеным тэксце, адносна нейкага тэзаўруса, адпавядаючага ўзроўню ведаў і навыкаў, неабходных для чытання, якімі валодае дадзены чытач.

Складанасць вымярэння такой характарыстыкі абумоўлена радам метадычных цяжкасцей. У айчынай літаратуры гэтыя пытанні не абмяркоўваліся, акрамя працы [2], дзе разглядалася прымяненне формул «чытальнасці»*. Таму будзе каштоўным разгледзець метады, распрацаваныя амерыканскімі даследчыкамі.

Даследаванні па чытальнасці пачаліся ў 20-х гадах мінулага стагоддзя. Цікавасць да гэтай праблемы ўзнікла ў даследчыкаў у сувязі з практычнымі патрэбамі педагогікі. Неабходнасць колькасных даследаванняў пры распрацоўцы навуковых асноў для складання вучэбных праграм, шырокае прымяненне эксперымента ў навучанні і патрэба ў дыферэнцыяцыі вучэбнага матэрыялу ў сувязі з увядзеннем абавязковай адукацыі ў ЗША стымулявалі амерыканскіх псіхолагаў і лінгвістаў вырашаць праблему чытальнасці.

У самым шырокім сэнсе пад чытальнасцю разумеюць некаторую характарыстыку друкаванага матэрыялу, якая залежыць ад усіх элементаў дадзенага ўнутры матэрыялу, што ўздзейнічаюць на паспяховасць яго засваення пэўнай групай чытачоў. Мерай такога паспяховага засваення з'яўляецца тое, наколькі сярэдні чытач групы, якая нас зацікавіла, разумее даследуемы матэрыял, у якой меры хуткасць, з якой ён яго чытае, набліжаецца да аптымальнай, і, апошняе, якую цікавасць уяўляе дадзены матэрыял для гэтага сярэдняга чытача.

Разгледзім фактары, якія ўздзейнічаюць на ступень разумення друкаванага матэрыялу.

Большасць аўтараў, якія займаюцца праблемай чытальнасці, разглядаюць працэс разумення друкаванага матэрыялу як нейкае ўзаемадзеянне паміж магчымасцямі пэўнага індыві-

дуума да чытання і разумення прачытанага, з аднаго боку, і складанасцю самога тэксту — з другога.

Магчымы два падыходы да праблемы чытальнасці — якасны і колькасны.

Пры якасным падыходзе з дапамогай масавых апытанняў кампетэнтных асоб спрабуюць вызначыць, чым, па іх меркаванню, абумоўлена чытальнасць друкаванага матэрыялу. Дагэтуль, атрыманыя ў выніку такіх апытанняў, дазваляюць зрабіць агульнае ўяўленне аб фактарах, якія ўздзейнічаюць на поспех засваення друкаванага матэрыялу людзьмі з рознай ступенню падрыхтаванасці.

Пры колькасным падыходзе да вывучэння складанасці тэксту даследчык спачатку выбірае пэўную групу асоб і затым імкнецца даследаваць цяжкасць разумення рознага друкаванага матэрыялу для сярэдняга чытача гэтай групы. Такім чынам, сцверджанне, што тэкст *A* больш складаны, чым тэкст *B*, значыць, што тэкст *A* менш зразумелы выпадковай выбарцы з даследуемай групы асоб, чым тэкст *B*, ці тэкст *A* больш складаны, чым тэкст *B*, калі для адэкватнага разумення тэксту *A* і *B* патрабуецца, каб сярэдні чытач меў больш высокі ўзровень адукацыі.

Пасля высвятлення шляхам апытання ці тэсціравання чытачоў адноснай цяжкасці ўсіх даследуемых тэкстаў эксперыментатар імкнецца вылучыць усе ўласцівасці, якія належаць іх разуменню. Відавочна, што сукупнасць фактараў, якія прадказваюць ступень разумення таго ці іншага тэксту, дае аб'ектыўны метада прагнозу толькі ў тым выпадку, калі можна будзе надзейна вымераць гэтыя фактары.

Такім чынам, пры колькасным вывучэнні чытальнасці паўстае праблема высвятлення залежнасці паміж ацэнкай цяжкасці тэксту для пэўнай групы чытачоў, атрыманай у выніку тэсціравання, і аб'ектыўнымі характарыстыкамі гэтага тэксту. Вырашэнне гэтай праблемы можна ўмоўна падзяліць на чатыры асноўныя этапы:

1) пошук аб'ектыўных крытэрыяў для вызначэння складанасці розных тэкстаў для дадзенай групы асоб;

2) выяўленне структурных элементаў, агульных для ўсіх даследуемых тэкстаў, колькасная характарыстыка якіх дапамагала б дакладна адрозніваць складаны матэрыял ад лёгкага;

3) распрацоўка надзейных спосабаў вымярэння колькасных характарыстык такіх элементаў;

4) пошук некаторай функцыі, якая адназначна апісвала б залежнасць паміж паспяховасцю разумення дадзенага матэрыялу выпадковай выбаркай з даследуемай групы асоб і пэўнай камбінацыяй адзначаных вышэй характарыстык.

Вялікую цікавасць уяўляюць першыя два пытанні. Разгледзім, як яны вырашаліся рознымі аўтарамі ў даследаваннях па чытальнасці друкаванага матэрыялу.

Тэсты на разуменне друкаванага матэрыялу. Існуе некалькі метадаў канструявання тэстаў на разуменне друкаванага матэрыялу. Некаторыя з іх прапанаваны Р. М. Фрумкінай [3]:

1. *Метада выбару рэзюмэ.*

Метада заключаецца ў тым, што да тэксту складаецца некалькі рэзюмэ, якія ў рознай ступені адлюстроўваюць яго змест. Адно рэзюмэ заведана з'яўляецца лепшым. Разуменне ацэньваецца па тым, якое рэзюмэ было абрана пасля чытання тэксту як найлепшае.

2. *Метада запавення прабелаў.*

Ставілася задача запавення прабелы ў тэксце. Адзнака выканання задачы вызначалася колькасцю правільна ўгаданых слоў да агульнай колькасці прабелаў.

3. *Пытанні па зместу тэксту.*

Ступень разумення тэксту залежыць ад колькасці правільных адказаў.

Існуюць таксама тэсты амерыканскіх даследчыкаў, якія ў цэлым падобны да пералічаных. Пасля правядзення тэстаў выстаўляецца сярэдняя адзнака ў балах, якая з'яўляецца колькаснай мерай складанасці дадзенага тэксту для вывучаемай групы асоб. Затым вызначаюцца моўныя параметры, якія ўздзейнічаюць на складанасць друкаванага матэрыялу. Па-першае, выбіраюцца тыя з іх, якія прысутнічаюць ва ўсіх даследуемых тэкстах, па-другое, могуць быць надзейна вымераны і, па-трэцяе, дастаткова моцна карэлююць з колькаснай мерай складанасці тэксту, вызначанай адным з указаных вышэй метадаў.

Сярод такіх параметраў, знойдзеных рознымі аўтарамі, можна вылучыць чатыры вялікія групы: а) элементы слоўнікавага складу тэксту; б) фактары, звязаныя са структурай сказаў; в) колькасныя характарыстыкі канцэптальнай насычанасці тэксту; г) цікавасць, з якой чытае разглядаемая група асоб.

Разгледзім больш дэтальна кожны з названых вышэй тыпаў моўных параметраў тэксту.

* Ад англійскага слова «readability».

1. *Слоўнікавы склад* (разнастайнасць слоўніка і яго цяжкасць) найбольш істотна карэлюе з усімі крытэрыямі цяжкасці, выкарыстанымі рознымі аўтарамі.

Аўтары шэрагу работ прыйшлі да аднолькавага выніку, што чым менш розных слоў, тым больш лёгкі матэрыял.

Цяжкасць слоўніка характарызуе разуменне чытачом асобных слоў. У кожнай працы зазначаецца, што лексічная цяжкасць тэксту добра карэлюе са ступенню яго разумення.

Пры вымярэнні гэтай характарыстыкі аўтары зыходзілі з таго, што чым больш прапорцыя незнаёмых ці доўгіх слоў у тэксце, тым цяжэй чытачу ўлавіць яго сэнс, і таму цяжкасць слоўніка вызначалася альбо шляхам супастаўлення слоў тэксту з рознымі частотнымі спісамі, альбо шляхам вымярэння сярэдняй даўжыні слоў, што ўваходзяць у даследуемы адрэзак.

Акрамя таго, было выяўлена, што вымярэнні семантычнай цяжкасці тэксту добра карэлююць са структурнымі вымярэннямі слоўнікавага складу тэксту. Усё гэта зрабіла ўплыў на адказ ад разгляду семантычнай цяжкасці тэксту дзеля больш практычнага структурнага падыходу.

2. Амаль у кожным даследаванні быў выяўлены ўплыў *структуры сказаў* на цяжкасць разумення друкаванага матэрыялу. Найбольш папулярным метадам ацэнкі структурнай складанасці сказаў было вымярэнне іх даўжыні ў словах. Пры гэтым здаецца відавочным палажэнне, з якога зыходзілі даследчыкі, што чым больш доўгі сказ, тым больш цяжкі тэкст. Іншыя даследчыкі ў якасці характарыстыкі цяжкасці тэксту вымяралі колькасць складаных сказаў, колькасць простых сказаў, а таксама даўжыню сказаў у складах. Усе вымярэнні структуры сказаў дастаткова моцна карэлююць паміж сабой і добра прагназуюць цяжкасць тэксту.

3. Вымярэнне *канцэптуальнай насычанасці тэкстаў* праводзілася ўскосна — праз вызначэнне працэнтных суадносін прыназоўнікавых фраз*. Пры гэтым даследчыкі зыходзілі з таго, што чым больш прапорцыя прыназоўнікавых фраз, тым больш матэрыялу, звязанага з фарміраваннем ідэй і, значыць, тым больш цяжкі тэкст. Колькасць прыназоўнікавых фраз слабей карэлюе з цяжкасцю даследуемага матэрыялу, чым яго слоўнікавы склад ці структура сказаў, і, калі ў формуле ўлічаны гэтыя два фактары, то колькасная характарыстыка канцэптуальнай насычанасці тэксту мала што дадае да агульнага прадказання яго цяжкасці.

4. *Пра колькасную характарыстыку інтарэсу*, які ўзнікае ў сярэдняга чытача даследуемай групы пры чытанні дадзенага тэксту, будзе сказана ніжэй у сувязі з метадыкай вымярэння «індэкса інтарэсу» друкаванага матэрыялу, распрацаванай Р. Флэшам.

Асноўныя даследаванні па чытальнасці. Першыя колькасныя даследаванні па чытальнасці былі даволі недасканалымі. Цяжкасць даследуемых тэкстаў вызначалася з дапамогай суб'ектыўнай ацэнкі эксперыментатараў. Пры апрацоўцы даных быў выкарыстаны вельмі прымітыўны матэматычны апарат; акрамя таго, у якасці фактараў, што прагназуюць цяжкасць, разглядаліся толькі характарыстыкі лексікі тэксту.

Далейшыя даследаванні зводзіліся да пошуку аб'ектыўных характарыстык тэксту, якія добра карэлявалі б з колькаснай ацэнкай вынікаў тэсціравання, і састаўлення ўраўнення рэгрэсіі паміж ацэнкай тэсціравання, з аднаго боку, і ацэнкамі параметраў тэксту — з другога.

Але многія даследчыкі імкнуліся знайсці максімальную колькасць фактараў, якія ўздзейнічаюць на цяжкасць тэксту. Адваротным пунктам сталі працы І. Лорджа [4] і Р. Флэша [5], якія заявілі аб тым, што нягледзячы на існаванне шэрагу фактараў, якія дастаткова моцна ўздзейнічаюць на цяжкасць разумення, карэляцыя паміж імі настолькі вялікая, што толькі некаторыя з іх неабходны для выкарыстання ў якасці надзейных прагназуемых фактараў цяжкасці тэксту.

Лордж уключыў у рэгрэсійную формулу чытальнасці толькі тры фактары — колькасць розных цяжкіх слоў, сярэдняю даўжыню сказа ў словах і колькасць прыназоўнікавых груп. Яшчэ больш спрабавалі формулу чытальнасці Р. Флэш. Акрамя таго, ён даследаваў асноўныя формулы чытальнасці на матэрыяле часопісаў, газет і пераканаўся, што пры выкарыстанні гэтых формул для сярэдняга дарослага чытача былі атрыманы недакладныя, а часам і памылковыя вынікі.

Для тлумачэння гэтага факта Флэш выказаў гіпотэзу, што калі разуменне прачытанага ў слаба падрыхтаванага чытача абумоўлена пераважна ўспрыманням і разуменнем асобных слоў, то разуменне тэксту на больш высокім узроўні падрыхтаванасці залежыць у асноўным ад яго канцэптуальных характарыстык (колькасці думак у тэксце, іх суадносін, складанасці і г. д.).

Зыходзячы з літаратурных даных і сваіх уласных даследаванняў, Флэш сцвярджаў, што паказчык абстрактнасці слоў, якія ўваходзяць у даследуемы адрэзак, у значнай ступені вызначае ўздзеянне канцэптуальных характарыстык тэксту на яго цяжкасць для сярэдняга дарослага чытача.

* Пад прыназоўнікавай фразай (англ. prepositional phrase) маецца на ўвазе словазлучэнне, якое ўводзіцца прыназоўнікам.

Для атрымання надзейнай ацэнкі ступені абстрактнасці слоў, што ўваходзяць у які-небудзь адрэзак тэксту, Флэш прапанаваў вымяраць колькасць афіксаў у дадзеным адрэзку. Як паказалі яго вымярэнні, існуе моцная карэляцыя ($r = 0,78$) паміж колькасцю абстрактных слоў і колькасцю афіксаў унутры даследуемага адрэзка тэксту.

Аднак у далейшым Флэш замяніў колькасць афіксаў на іншы параметр — колькасць слоў, таму што была выяўлена ненадзейнасць вымярэння колькасці афіксаў. Такую замену можна было зрабіць з-за існавання высокай ступені карэляцыі паміж гэтымі двума параметрамі.

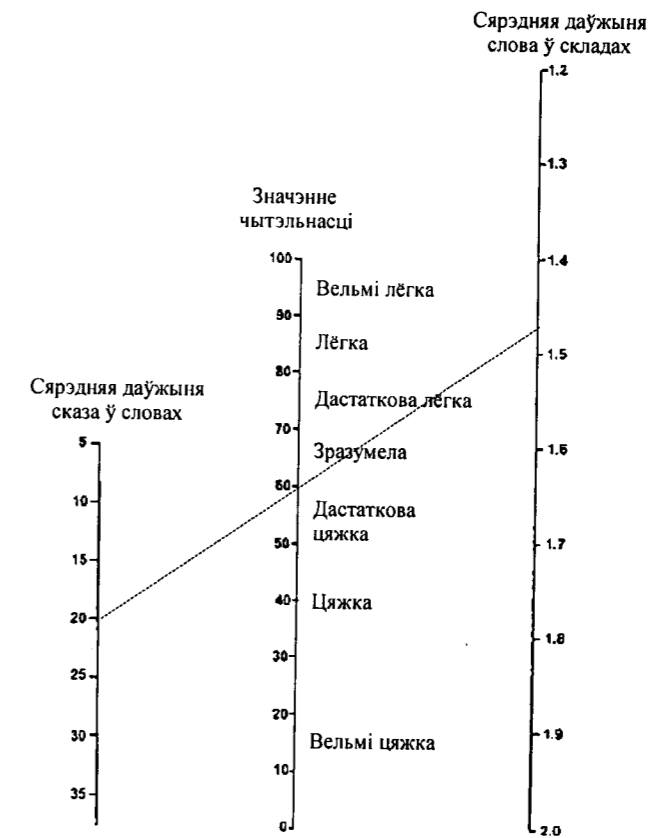
У якасці другога параметра, які прагназаваў цяжкасць тэкстаў для сярэдняга чытача, была ўзята сярэдня даўжыня сказаў даследуемага матэрыялу ў словах, і ўраўненне рэгрэсіі, у якое ўвайшлі два гэтыя параметры, мела наступны выгляд [6]:

$$X_1 = 206,48 - 1,015 \times X_2 - 84,6 \times X_3,$$

дзе X_1 — ацэнка цяжкасці тэксту для сярэдняга дарослага чытача; X_2 — сярэдня даўжыня сказа ў словах; X_3 — сярэдня даўжыня слова ў складах.

Разгледзім больш дэтальна сэнс велічыні X_1 . Яе колькаснае значэнне змяняецца ад 0 да 100. Ацэнка 100 адпавядае прадказанню таго, што чалавек з мінімальным узроўнем адукацыі зможа адказаць на тры чвэрці пытанняў любога надзейнага тэксту ці, што тое самае, дадзены тэкст даступны любому чалавеку, які толькі навучыўся чытаць. Ацэнка 0 адпавядае матэрыялам, якія даступны толькі для вузкага кола спецыялістаў з высокай падрыхтоўкай (узровень навуковых прафесійных часопісаў). Астатнія ацэнкі цяжкасці вар'іруюць паміж гэтымі дзвюма, прычым для кожнай ацэнкі Флэша ўказан узровень адукацыі асоб, неабходны, па-першае, для элементарнага разумення сэнсу, і, па-другое, для лёгкага чытання тэкстаў з такім індэксам цяжкасці (табліца і намагама) [6]. Напрыклад, калі пры падстаноўцы ў формулу Флэша неабходных характарыстык тэксту ($X_2 = 20$, $X_3 = 1,5$) атрымаем $X_1 = 60$, то можна з дастатковай ступенню ўпэўненасці сказаць:

- 1) для разумення дадзенага тэксту неабходна адукацыя ў аб'ёме 8 класаў сярэдняй школы;
- 2) гэты тэкст будзе лёгкім для людзей з вышэйшай адукацыяй;
- 3) тэкст напісан на ўзроўні матэрыялаў, звычайна публікуемых у пэўным часопісе;
- 4) для сярэдняга чытача дадзены тэкст з'яўляецца стандартным па цяжкасці.



Намагама для вызначэння чытальнасці друкаванага матэрыялу

Значэнне	Узровень адукацыі
90—100	5-ы клас
80—90	6-ы клас
70—80	7-ы клас
60—70	8-ы і 9-ы класы
50—60	10—12-ы класы (сярэдня школа)
30—50	каледж
0—30	выпускнік каледжа

Зыходзячы з вышэйсказанага, можна зрабіць вывад, што формула Р. Флэша і шмат іншых формул амерыканскіх даследчыкаў па чытальнасці друкаваных матэрыялаў прымянімы толькі для англійскай мовы (і менавіта толькі для аднаго віду выданняў — вучэбнага), таму што ўраўненне рэгрэсіі складалася ў залежнасці ад колькасных рэзультатаў тэсціравання і параметраў тэксту, і зразумела, што ў нашых умовах гэтыя лічбы будуць зусім іншыя. Але сама метадыка вызначэння цяжкасці тэксту можа быць выкарыстана ў далейшых даследаваннях па дадзенаму пытанню.

Трэба адзначыць, што Р. Флэш прапанаваў таксама формулу для прагназавання цікавасці, якую будзе выклікаць дадзены матэрыял у дарослага чытача з сярэднім узроўнем падрыхтоўкі.

Пры канструяванні гэтай формулы Флэш зыходзіў з таго агульнапрызнанага факта, што інтарэс, які ўзнікае пры чытанні якога-небудзь матэрыялу, дапамагае недастаткова падрыхтаванаму чытачу пераадолець моўныя цяжкасці, матывацыйныя фактары павышаюць яго ўвагу і сканцэнтраванасць, і ён у стане засвоіць змест матэрыялу, які пры адсутнасці гэтага інтарэсу быў бы для яго вельмі цяжкім.

Ураўненне рэгрэсіі, якое ўстанаўлівае залежнасць ступені інтарэсу, узнікаючага пры чытанні пэўнага адрэзка тэксту, ад моўных параметраў гэтага адрэзка, мае наступны выгляд [2]:

$$X_1 = 0,3635 \times X_2 + 0,314 \times X_3,$$

дзе X_1 — «індэкс інтарэсу» Флэша; X_2 — колькасць слоў наступных тыпаў:

- усе асабовыя займеннікі,
 - усе назоўнікі і займеннікі, род якіх можа быць вызначаны без кантэксту альбо праз функцыі дэнатата, альбо пры наяўнасці ў слове марфемы, якая ўказвае на род (напрыклад, бацька, актрыса),
 - англійскія словы «reople» (народ, людзі) і «folk» (людзі);
- X_3 — колькасць сказаў наступнага тыпу:
- сказы, якія ўваходзяць у простую мову,
 - сказы, якія змяшчаюць пытанні, каманды, патрабаванні, а таксама ўсе іншыя сказы, непасрэдна звернутыя да чытача,
 - воклічы,
 - граматычна няпоўныя сказы ці фрагменты сказаў, сэнс якіх можна зразумець толькі з кантэксту.

Згодна з гіпотэзай Флэша, абодва гэтыя фактары з'яўляюцца хаця і не прамым, але адэкватным вымярэннем таго, наколькі даследуемы матэрыял напісаны ў гутарковым стылі, і, такім чынам, цікавы сярэдняму чытачу.

Пры гэтым неабходна памятаць, што індэкс інтарэсу Флэша характарызуе толькі тое, наколькі інтарэс, які ўзнікае пры чытанні, змяншае цяжкасць дадзенага адрэзка, і гэты індэкс нельга разглядаць як вымярэнне таго, наколькі наогул цікавы той ці іншы тэкст.

Трэба адзначыць, што формулы чытальнасці Флэша атрымалі шырокае распаўсюджанне ў самых розных галінах чалавечай дзейнасці, усюды, дзе паўстае пытанне аб дыферэнцыяцыі друкаванага матэрыялу ў залежнасці ад падрыхтаванасці яго патэнцыяльнага чытача.

Пры практычным прымяненні формул чытальнасці трэба мець на ўвазе, што гэтыя формулы прызначаны не для дакладнага вымярэння цяжкасці друкаванага матэрыялу, а для прагнозу гэтай цяжкасці.

Аднак нават вельмі прыблізнае аб'ектыўнае вымярэнне цяжкасці друкаванага матэрыялу дасць магчымасць вырашаць розныя задачы. Пасля таго, як будзе распрацавана аб'ектыўная метадыка для колькаснага вымярэння цяжкасці тэксту на рускай, беларускай мовах, можна будзе вырашаць наступныя задачы:

1. Ёсць тэксты T_1, T_2, \dots, T_n і тэзаўрус Q . У якім з дадзеных тэкстаў змяшчаецца найбольшая колькасць інфармацыі адносна тэзаўруса Q ці, што тое самае, пры аналізе якога з гэтых тэкстаў тэзаўрус Q зменіцца ў найбольшай ступені?

Прыкладам практычнага выкарыстання гэтых формул можа служыць самае шырокае іх прымяненне ў педагагіцы. Яны былі выкарыстаны для вызначэння цяжкасці падручнікаў і дадатковай літаратуры для школьнікаў таго ці іншага класа, дапаможнікаў і лекцыйных матэрыялаў для студэнтаў, вучэбнай літаратуры для асоб, якія займаюцца самападрыхтоўкай. Па меркаванню большасці спецыялістаў, вывады, зробленыя аўтарамі работ па колькаснаму вымярэнню цяжкасці вучэбнага матэрыялу, значна дапамаглі складальнікам падручнікаў і настаўнікам вырашыць важную праблему адпаведнасці складанасці кнігі і ўзроўню вучня.

2. Ёсць тэкст T і тэзаўрусы Q_1, Q_2, \dots, Q_n . Адносна якога з дадзеных тэзаўрусаў у тэксце T змяшчаецца найбольшая колькасць інфармацыі ці, іншымі словамі, які з тэзаўрусаў, якія нас цікавяць, зменіцца ў найбольшай ступені пры аналізе тэксту T ?

Такім чынам, распрацоўка метадыкі дазволіць вырашыць шэраг пытанняў, якія, безумоўна, будуць мець цікавасць не толькі для рэдакцыйна-выдавецкіх супрацоўнікаў, настаўнікаў, вучоных, але і для ўсіх тых, хто звязаны з пытаннямі перадачы інфармацыі.

Літаратура

- Шрейдер Ю. А. О количественных характеристиках семантической информации // НТИ. 1963. № 10. С. 33—48.
- Мацковскій М. С. Применение формул читабельности для получения некоторых количественных характеристик семантической информации // НТИ. 1969. Сер. 2. № 6. С. 3—7.
- Фрумкина Р. М. Понимание текста в условиях ограниченного знания словаря // НТИ. 1965. № 4. С. 44—48.
- Loge I. Predicting readability. «Teacher's College Record». 1944. P. 45, 404—419.
- Flesch R. Estimating the comprehension difficulty of magazine articles. «J. Gener. Psychol.». 1943. P. 28, 63—80.
- www.mang.canterbury.ac.nz/courseinfo/AcademicWriting/Flesch.htm.

L. I. PYATROVA, YU. F. SHPAKOWSKY

PROBLEM OF INFORMATION CAPACITY AND COMPREHENSION IN PUBLICATIONS

Summary

Some methods to evaluate text semantic load have been tried. Comprehension tests for printed material have been proposed. Main linguistic parameters reflecting complexity and readability of the text (vocabulary, sentence composition, conceptual saturation, attractive topic) are discussed. Similar experiments of US colleagues are compared.