Igor Ohirko, Zinovii Partyko

СКЛАДНІСТЬ (ЧИТАБЕЛЬНІСТЬ) ЯК ОДНА З ОСНОВНИХ ХАРАКТЕРИСТИК ШКІЛЬНИХ НАВЧАЛЬНИХ ВИДАНЬ

В Україні серед затверджених урядом критеріїв оцінювання якості шкільних навчальних видань відсутній такий показник як складність (читабельність) тексту. Проте такі оцінки складності тексту (синтаксичної та семантичної) потрібно, як і в країнах Заходу, обов'язково запровадити. Для цього на основі експериментальної треба створити промислову комп'ютерну систему визначення складності україномовних текстів. Необхідним є також калібрування таблиці складності до відповідних класів школи та курсів університетів.

Ключові слова: навчальні видання, критерії оцінювання, текст, складність, читабельність, українська мова, комп'ютерна система.

Ihor Ohirko, Zinovii Partyko

COMPLEXITY (READABILITY) AS ONE OF THE MAIN CHARACTERISTICS OF SCHOOL EDUCATIONAL PUBLICATIONS^e

In Ukraine, there are no such indicators as the complexity (readability) of the text among the government-approved criteria for evaluating the quality of school textbooks. However, such assessments of the complexity of the text (syntactic and semantic) must be implemented, as in Western countries. To do this, an industrial computer system for determining the complexity of Ukrainian-language texts should be created on an experimental basis. It is also necessary to calibrate the complexity table to the relevant school classes and university courses.

Key words: *educational editions, evaluation criteria, text, complexity, readability, ukrainian language, computer system.*

Останнім часом в Україні проводиться низка заходів щодо оцінювання якості шкільних навчальних видань [7, с. 19–24]. Проте серед критеріїв оцінювання таких видань немає одного з основних критеріїв — складності (читабельності) тексту, хоча цей критерій є одним із базових і при цьому має не тільки формальну, а й автоматизовану процедуру його визначення. На жаль, цей критерій для оцінки навчальних видань в Україні не застосовують, хоча його використання в країнах Заходу є розповсюдженою процедурою, яка розроблялася й досліджувалася саме для навчальних видань ще з початку минулого століття. Більше того, в наш час ця процедура навіть включена в склад текстового процесора Microsoft Word, яким користуються всі, хто використовує операційну систему Windows, а це сотні мільйонів людей.

З урахуванням сказаного **мета дослідження** полягає в тому, щоб акцентувати увагу видавничих працівників та педагогічної громадськості на можливості й необхідності застосування цього критерію для оцінки навчальних видань в Україні. Відповідно, стаття має оглядовий характер.

Об'єктом дослідження будуть критерії оцінки складності (читабельності) тексту, а **предметом** — фактори, від яких залежить така складність, методика застосування таких критеріїв і особливості застосування критерію складності до україномовних текстів.

У дослідженні буде використано такі **методи**:

- традиційний аналіз (для опрацювання літературних даних);

- метод синтезу (для вироблення рекомендацій щодо використання критерію складності для оцінки якості україномовних шкільних підручників).

У країнах Заходу використання критерію складності (читабельності) для оцінки якості текстів має понад столітню історію [4]. Значно коротшу історію досліджень і такого використання має Росія [5; 8]. В Україні критерій складності практично не використовують через те, що його застосування повинно враховувати лінгвістичні особливості саме української мови (застосування параметрів і баз даних англійської чи російської мов до україномовних текстів дасть або некоректні результати, або є технологічно неможливим). Натомість в Україні під читабельністю найчастіше розуміють графічну придатність літер для читання [9].

Виокремлюють неформалізовані та формалізовані методи визначення складності текстів. При цьому щодо формалізованих методів рівень формалізації доведений до такого ступеня, що ці методи можна автоматизувати. При цьому, на відміну від людини, яка під час застосування неформалізованих методів іноді може визначити складність тексту доволі суб'єктивно (зокрема, залежно від власного досвіду), формалізовані методи, реалізовані в формі комп'ютерних систем, дають об'єктивну оцінку складності, яка не залежить від дослідника.

Вказаний критерій складності можна розглядати як мінімум у двох аспектах — синтаксичному й семантичному. Синтаксична складність залежить від середніх довжин слів і речень аналізованого тексту, а семантична — від кількості слів тексту, які відсутні в словнику-мінімумі української мови (такий словник для будь-яких мов містить, як правило, близько трьох тисяч слів).

Визначення складності тексту для англійської та російської мов автоматизовано й визначається, наприклад, програмними засобами текстового процесора Microsoft Word, про що вже згадувалося, а також системами, які представлені в інтернеті [3; 13].

В Україні **невирішеною раніше частиною заявленої проблеми** є відсутність аналогічного програмного забезпечення для промислового (серійного) автоматичного оціню-

вання складності україномовних текстів. При цьому треба зазначити, що лінгвістичне забезпечення для такого оцінювання свого часу було розроблене й реалізоване в якості експериментальної системи [6].

Ураховуючи сказане, нижче подамо опис основних залежностей і методики оцінювання якості україномовних текстів.

Синтаксична складність — це мислительна характеристика тексту, яку визначають на основі ступеня агрегації (об'єднання) одних одиниць мови в інші. Агрегування (наприклад, літер — у слова, слів — у речення, речень — у надфразові єдності (НФЄ) тощо) ґрунтується на такій мислительній операції, як синтезування.

Встановлена експериментальними дослідженнями [2; 4] закономірність констатує, що синтаксична складність повідомлення залежить від двох показників: а) середньої довжини слів повідомлення, виміряної в кількості літер; б) середньої довжини речень повідомлення, виміряної в кількості слів, — за принципом: що довшими є слова й речення, то складнішим є текст і навпаки.

Вплив інших показників (наприклад, агрегування речень — у НФЄ, НФЄ — у блоки) на синтаксичну складність повідомлення настільки малий, що ним можна знехтувати.

За показником синтаксичної складності розрізняють такі рівні складності текстів: дуже складні; складні; не дуже складні; звичайні; не дуже легкі; легкі; дуже легкі. Кожен із цих рівнів має своє числове вираження (наприклад, на шкалі в межах від 0 до 100 балів).

З позиції педагогіки (для текстів шкільних навчальних видань), рівень синтаксичної складності тексту повинен відповідати змозі читачів (школярів) комфортно сприймати тексти з такою довжиною слів і речень.

Як приклад, подамо один із адаптованих нами методів визначення складності (метод Р. Флеша [2; 4]), який визначає синтаксичну складність тексту (R) на основі такої функ-

ціональної залежності:

$$R = 206,835 - 0,846 N - 1,015 L,$$

де N — число складів на 100 слів тексту; L — середня довжина речень, у кількості слів. Значення складності R для англомовних текстів показано в табл. 1¹.

Таблиця 1

*Оцінка складності тексту
залежно від кількісних характеристик його слів і речень*

N	L	R	Складність повідомлення
> 192	> 29	0...30	Дуже важкий
167	25	31...50	Важкий
155	21	31...60	Не дуже важкий
147	17 ¹	61...70	Звичайний
139	14	71...80	Не дуже легкий
131	11	81...90	Легкий
<123	<8	91...100	Дуже легкий

Зрозуміло, що значення складності цієї таблиці для застосування щодо навчальних видань можна відкалібрувати не тільки за вказаними сімома рівнями, а й за класами школи чи курсами університету.

Семантична складність — це мислительна характеристика тексту, яка залежить від кількості слів цього тексту, що відсутні в словнику-мінімумі читача (школяра), за принципом: що більше слів тексту відсутні в словнику-мінімумі читача, то складнішим є цей текст і навпаки. Кажучи іншими словами, слово тексту вважають складним, якщо воно є в словнику-мінімумі читача, й простим (легким), якщо його там нема: що такий відсоток більший, то такий текст є семантично складнішим. Визначення семантичної складності

¹ В усномовному спілкуванні середня довжина речень є меншою і в середньому не повинна перевищувати 12...13 слів.

базується на тому постулаті, що по замовчуванню 80% носіїв будь-якої мови повинні знати всі слова, наявні у словнику-мінімумі, тобто повинні знати як мінімум три тисячі слів. Уперше такий метод оцінки семантичної складності було запропоновано застосовувати І. Дейлом і Дж. Челлом [1; 4].

За показником семантичної складності розрізняють такі самі групи текстів, як і за показником синтаксичної: дуже складні; складні; не дуже складні; звичайні; не дуже легкі; легкі; дуже легкі. Кожен із цих рівнів так само, як і для синтаксичної складності, має своє числове вираження (наприклад, у тих самих межах від 0 до 100 балів).

Рівень семантичної так само, як і синтаксичної, складності тексту повинен відповідати змозі читачів оптимально сприйняти тексти з такою кількістю нових слів.

Реалізація визначення складності (читабельності) україномовних текстів. Свого часу для визначення складності україномовних текстів критерії їх оцінювання було адаптовано до статистичних особливостей української мови [12], а також було укладено словник-мінімум української мови обсягом у три тисячі слів [11]. Вказані параметри оцінювання складності текстів були реалізовані в експериментальній системі редагування «Редактор» [10]. Приклади її функціонування подано на рис. 1–5.

Висновки. 1. З урахуванням світового досвіду в Україні до складу критеріїв оцінки шкільних та університетських підручників доцільно додати критерій складності (читабельності) у двох його різновидах: синтаксичної та семантичної.

2. Для оцінки рівня семантичної складності доречно використовувати наявний словник-мінімум української мови, укладений на основі шести частотних словників української мови, а також адаптовані до україномовних текстів статистичні характеристики середніх довжин слів і речень.

а

б

Рис. 1. Загальний вигляд експериментальної системи редагування «Редактор»:
а — заставка; б — робоче вікно

Рис. 2. Діалогове вікно СР «Редактор» для вибору користувачем аудиторії, для якої готувалося повідомлення

Рис. 3. Вікно визначення синтаксичної складності повідомлення СР «Редактор»

Рис. 4. Пояснення до визначення синтаксичної складності повідомлення СР «Редактор»

Рис. 5. Приклад допомоги в експериментальній СР «Редактор»

3. Для автоматизованого оцінювання складності (читабельності) україномовних текстів шкільних й університетських навчальних видань доцільно розробити (за аналогією до наявної експериментальної системи «Редактор») промислову систему визначення складності україномовних текстів, яка могла б визначати цю складність для текстів, представлених у різних форматах (txt, doc, epub, html тощо).

4. У межах додаткових педагогічних досліджень необхідно провести калібрування, тобто на шкалах значень складності (читабельності) встановити точне співвіднесення рівнів складності текстів з відповідними рівнями підготовки читачів, які можуть комфортно сприймати тексти такої складності. Калібрування необхідно провести за класами й курсами навчання читачів (школярів і студентів).

СПИСОК ВИКОРИСТАНИХ ДЖЕРЕЛ

1. Dale–Chall readability formula // Wikipedia. URL: https://en.wikipedia.org/wiki/Dale-Chall_readability_formula.

2. Flesch–Kincaid readability tests // Wikipedia. URL: https://en.wikipedia.org/wiki/Flesch-Kincaid_readability_tests.

3. Quick and easy way to test the readability of your work // Readability Test Tool. URL: <https://www.webfx.com/tools/readable/>.

4. Readability // Wikipedia. URL: <https://en.wikipedia.org/wiki/Readability>.

5. Аналіз методів оцінки складності тексту // Аналіз методів оцінки складності тексту. URL: <https://cyberleninka.ru/article/n/analiz-metodov-otsenki-slozhnosti-teksta>.

6. Булик С., Гронська І., Партико З. Редакційна система «Редактор»: контроль складності, стилістики й пунктуації // Друкарський кур'єр. 1998. № 4. С. 12–14.

7. Експертиза шкільних підручників : інструктивно-методичні матеріали для експертизи електронних версій проектів підручників для учнів 8 класу загальноосвітніх навчальних за-

кладів [посібник] / за заг. ред. О. М. Топузова, Н. Б. Вяткіної. К. : Педагогічна думка, 2016. 128 с.

8. Индекс удобочитаемости // Википедия. URL: https://ru.wikipedia.org/wiki/Индекс_удобочитаемости.

9. Легкочитність // Вікіпедія. URL: <https://uk.wikipedia.org/wiki/Легкочитність>.

10. Партико З. В. Нормативний аспект і автоматизація редагування. К.: Видавн. центр Інституту журналістики КНУ ім. Тараса Шевченка, 2004. 266 с.

11. Партико З. В. Словник-мінімум української мови. К. : Видавн. центр Інституту журналістики КНУ ім. Тараса Шевченка, 2004. 80 с.

12. Партико З., Бородчук В., Сорокати І. Трипараметричний метод визначення читабельності (складності) україномовних текстів // Палітра друку. 1995. N 4 С. 54–55.

13. Проверка текста на читабельность для SEO // SEO анализ текста: проверка на уникальность, читабельность, переспам. URL: <https://copywritely.com/ru/readability-checker/>.