

УДК 028:37.091.64:004:811.161.2

ІНФОРМАЦІЙНА ТЕХНОЛОГІЯ ВИЗНАЧЕННЯ СКЛАДНОСТІ НАВЧАЛЬНИХ ТЕКСТОВИХ ВИДАНЬ

І. В. Огірко¹, З. В. Партико²

¹Українська академія друкарства,
вул. Під Голоском, 19, Львів, 79020, Україна

²Житомирський державний університет ім. І. Франка,
вул. Велика Бердичівська, 40, Житомир, 10002, Україна

В Україні серед затверджених урядом критеріїв оцінювання якості шкільних навчальних видань немає такого показника, як складність (читабельність) тексту. Проте це оцінювання складності тексту (синтаксичної та семантичної) потрібно, як і в країнах Заходу, обов'язково запровадити. Для цього на основі експериментальної системи 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.

Постановка проблеми. 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 — складності (читабельності) тексту, хоча цей критерій є одним із базових і водночас має не тільки формальну, а й автоматизовану процедуру його визначення. На жаль, цей критерій для оцінювання навчальних видань в Україні не застосовують, хоча його використання в країнах Заходу є поширеною процедурою, яка розроблялася й досліджувалася саме для навчальних видань ще з початку минулого століття. До того ж у наш час ця процедура навіть входить до складу текстового процесора Microsoft Word, який застосовують усі, хто використовує операційну систему Windows, а це сотні мільйонів людей.

Тому **мета дослідження** полягає в тому, щоб звернути увагу видавничих працівників та педагогічної громадськості на можливості й необхідності застосування цього критерію для оцінювання навчальних видань в Україні. Відповідно, стаття має оглядовий характер.

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

Методи дослідження. У дослідженні буде використано такі методи:

- традиційний аналіз (для опрацювання літературних даних);

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

Аналіз останніх досліджень і публікацій. У країнах Заходу використання критерію складності (читабельності) для оцінювання якості текстів має понад столітню історію [4]. Значно менше часу досліджують і використовують їх у Росії [5; 8]. В Україні критерій складності практично не застосовують тому, що його застосування має враховувати лінгвістичні особливості саме української мови (використання параметрів і баз даних англійської чи російської мов до україномовних текстів або дасть некоректні результати, або технологічно неможливе). Натомість в Україні під читабельністю найчастіше розуміють графічну придатність літер для читання [9].

Виокремлюють неформалізовані та формалізовані методи визначення складності текстів. Водночас щодо формалізованих методів рівень формалізації доведений до такого ступеня, що ці методи можна автоматизувати. Крім того, на відміну від людини, яка, застосовуючи неформалізовані методи, 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 щонайменше у двох аспектах — синтаксичному і семантичному. Синтаксична складність залежить від середніх довжин слів і речень аналізованого тексту, а семантична — від кількості слів тексту, яких немає в словнику-мінімумі української мови (такий словник для будь-яких мов містить зазвичай приблизно три тисячі слів).

Визначення складності тексту для англійської та російської мов автоматизоване і визначається, наприклад, програмними засобами текстового процесора Microsoft Word, про що вже згадувалось, а також системами, які є в Інтернеті [3; 13].

В Україні не вирішеною раніше частиною заявленої проблеми є відсутність аналогічного програмного забезпечення для промислового (серійного) автоматичного оцінювання складності україномовних текстів. Водночас треба зазначити, що лінгвістичне забезпечення для такого оцінювання свого часу було розроблене й реалізоване як експериментальна система [6].

З огляду на це, 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) певних одиниць мови в інші. Агрегування (наприклад, літер — у слова, слів — у речення, речень — у надфразові єдності (НФС) 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, вимірної в кількості літер; б) середньої довжини речень повідомлення, вимірної в кількості слів, — за принципом: що довшими є слова й речення, то складнішим є текст і навпаки.

Вплив інших показників (наприклад, агрегування речень — у НФС, НФС — 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 має відповідати змозі читачів (школярів) комфортно сприймати тексти з такою довжиною слів і речень.

Як приклад подамо один із методів визначення складності, який ми адаптували, — метод Р. Флеша [2; 4], що визначає синтаксичну складність тексту (R) на основі такої функціональної 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 наведено в таблиці.

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, яких немає в словнику-мінімумі читача (школяра), за принципом: що більше слів тексту немає в словнику-мінімумі читача, то складнішим є цей текст і навпаки. Інакше кажучи, слово тексту вважають складним, якщо воно є в словнику-мінімумі читача, і простим (легким), якщо його там немає: що більший цей відсоток, то такий текст семантично складніший.

Таблиця

Оцінка складності тексту залежно від кількісних характеристик його слів і речень

N	L	R	Складність тексту
> 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

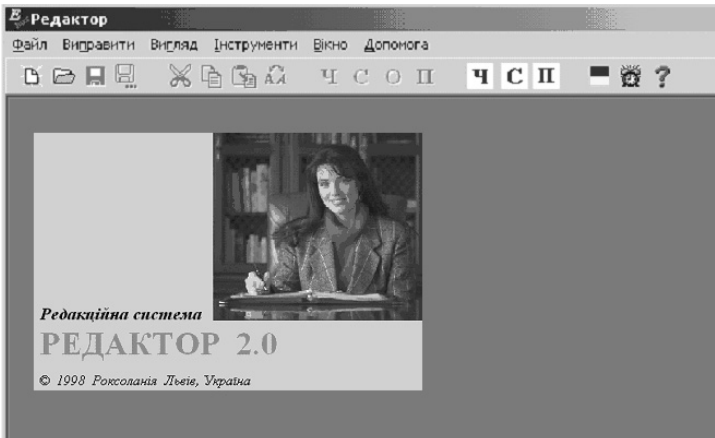
Визначення семантичної складності ґрунтується на тому постулаті, що стандартно 80 % носіїв будь-якої мови мають знати всі слова, наявні в словнику-мінімумі, тобто повинні знати щонайменше три тисячі слів.

Уперше такий метод оцінки семантичної складності запропонували застосовувати І. Дейл та Дж. Челл [1; 4].

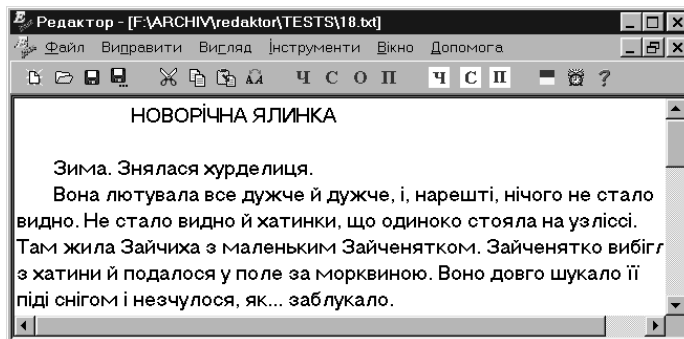
За показником семантичної складності розрізняють такі самі групи текстів, як і за показником синтаксичної: дуже складні; складні; не дуже складні; звичайні; не дуже легкі; легкі; дуже легкі. Кожен із цих рівнів, так само як і для синтаксичної складності, має також своє числове вираження (наприклад, у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

Рівень семантичної, як і синтаксичної, складності тексту має відповідати зможі читачів оптимально сприймати 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 три тисячі слів [11]. Вказані критерії оцінювання складності текстів були реалізовані в експериментальній системі редагування «Редактор» [10]. Приклади її функціонування наведено на рисунках 1–5.



а



б

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

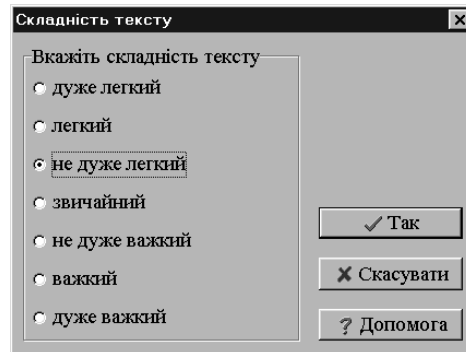


Рис. 2. Діалогове вікно СР «Редактор» для вибору користувачем реципієнтської аудиторії, для якої готувалося повідомлення

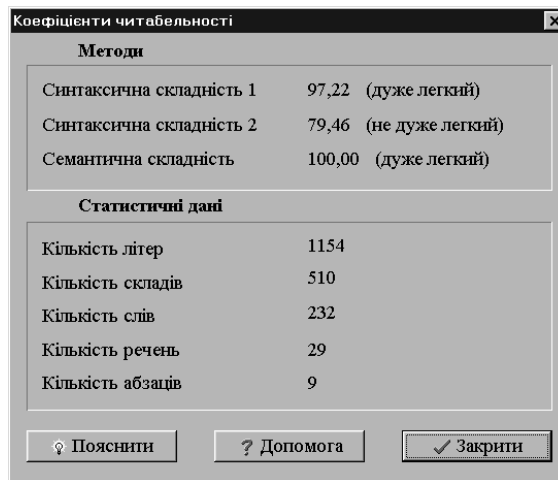


Рис. 3. Вікно визначення синтаксичної складності повідомлення СР «Редактор»

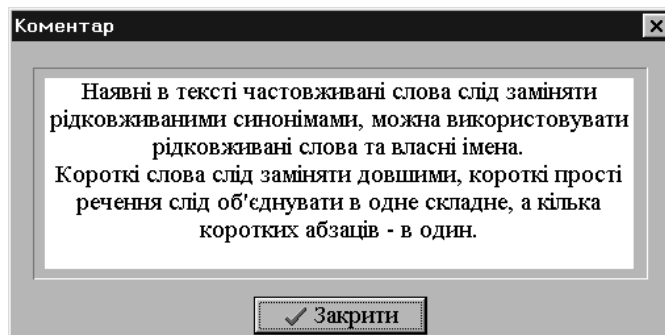


Рис. 4. Пояснення до визначення синтаксичної складності повідомлення СР «Редактор»

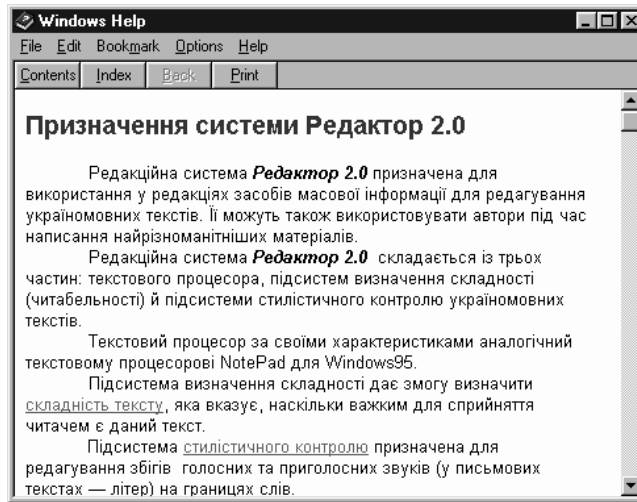


Рис. 5. Приклад допомоги в експериментальній системі редагування «Редактор»

На жаль, в Україні практичного застосування методи визначення складності тексту не отримали, хоча в світі ці методи активно досліджуються й постійно удосконалюються. Причиною відсутності практики застосування цих методів є, за нашими оцінками, людський фактор, а саме — відсутність готовності видавничих працівників і теоретиків видавничої справи на межі ХХ—ХХІ ст. до застосування цифрових технологій з метою об'єктивного визначення характеристик текстів, що публікуються. Співавтором цієї статті вказані методи як класичні були описані в навчальних виданнях, починаючи з 2001 р. (зараз ці видання перевидані втретє)¹.

Висновки. 1. З огляду на світовий досвід в Україні до складу критеріїв оцінювання шкільних та університетських підручників доцільно додати критерій складності (читабельності) у двох його різновидах: синтаксичний та семантичний.

2. Для оцінювання 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.

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

¹ Психолінгвістичні норми редагування // Партико З. В. Загальне редагування: нормативні основи : навч. посібник. — 3-ге вид., перероб. і допов. — Київ : Видавництво Ліра-К, 2020. — С. 366-370, 374. 2. Мисленнєво опрацювання повідомлення // Партико З. В. Основи редагування: у 2-х кн. / З. В. Партико. — 3-ге вид., перероб. і доп. — Київ : Видавництво Ліра-К, 2020. — Кн. 1 : підручник. — С. 211-213.

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

Список використаних джерел

1. Оборнева И.В. Автоматизация оценки качества восприятия текста // Вестник Московского городского педагогического университета. 2005, №2(5), с. 221–233.
2. Глушань В.М. Компьютерный анализ сложности текстов учебно-методических разработок как средство повышения качества обучения/ Вестник Таганрогского института имени А.П. Чехова, № 2, 2016, с.71-75.
3. Рыбанов А.А. Оценка качества текстов электронных средств обучения// Школьные технологии. 2011. № 6. С. 172-174.
4. Солнышкина С.И., Кисельников А.С. Сложность текста: этапы изучения в отечественном прикладном языкознании. Вестник Томского государственного университета. Филология. 2015. №6 (38), с. 86-99.
5. Кисельников А.С. К проблеме характеристик текста: читабельность, понятность, сложность, трудность//Филологические науки. Вопросы теории и практики, № 11 (53) 2015, часть 2, с.79-84..
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 закладів : посіб. / за заг. ред. О. М. Топузова, Н. Б. Вяткіної. Київ : Педагогічна думка, 2016. 128 с.
8. Мизернов И. Ю., Гращенко Л. А. Анализ методов оценки сложности текста. Новые информационные технологии в автоматизированных системах, 18, 2015. С. 572—581.
9. Читабельність // Літературознавча енциклопедія : у 2 т. / авт.-уклад. Ю. І. Ковалів. — Київ : ВЦ «Академія», 2007. — Т. 2 : М — Я. — С. 577.
10. Партико З. В. Нормативний аспект і автоматизація редагування. Київ : Видавн. центр Інституту журналістики КНУ ім. Тараса Шевченка, 2004. 266 с.
11. Партико З. В. Словник-мінімум української мови. Київ : Видавн. центр Інституту журналістики КНУ ім. Тараса Шевченка, 2004. 80 с.
12. Партико З., Бородчук В., Сорокатиї І. Трипараметричний метод визначення читабельності (складності) україномовних текстів. Палітра друку. 1995. № 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/>. Последний доступ: 05.03.2020. Название с экрана.

REFERENCES

1. Osborneva I.V. (2005). Automation of quality perception assessment of the text // Bulletin of the Moscow City Pedagogical University. №2 (5), p. 221–233. (in Russian)
2. Glushan V.M. (2016). Computer analysis of the complexity of the texts of educational and methodological developments as a means of improving the quality of education / Bulletin of the Taganrog Institute named after A.P. Chekhov, № 2, pp.71-75. (in Russian)

3. Rybanov A.A. (2011). Assessment of the quality of texts of electronic teaching aids // School technologies. No. 6. S. 172-174. (in Russian)
4. Solnyshkina S.I., Kiselnikov A.S. (2015). The complexity of the text: the stages of study in the domestic applied linguistics. Bulletin of Tomsk State University. Philology. No. 6 (38), p. 86-99. (in Russian)
5. Kiselnikov A.S. (2015). On the problem of text characteristics: readability, comprehensibility, complexity, difficulty // Philological Sciences. Questions of theory and practice, No. 11 (53) part 2, pp. 79-84 . (in Russian)
6. Bulyk, S., Hronska, I., & Partyko, Z. (1998). Redaktsiina systema «Redaktor»: kontrol skladnosti, stylistyky y punktuatsii: Drukarskyi kur'ier, 4, 12–14 (in Ukrainian).
7. Ekspertyza shkilnykh pidruchnykiv: instruktivno-metodychni materialy dlia ekspertyzy elektronnykh versii proektiv pidruchnykiv dlia uchniv 8 klasu zahalnoosvitnikh navchalnykh zakladiv (2016) / za zah. red. O. M. Topuzova, N. B. Viatkinoi. Kyiv : Pedahohichna dumka (in Ukrainian).
8. Mizernov I. Yu., Grashchenko L. A. (2015). Analysis of methods for assessing the complexity of the text. New information technologies in automated systems, 18, P. 572-581. (in Russian)
9. Readability (2007). // Literary Encyclopedia: in 2 volumes / ed. YI Kovalev. - Kyiv: VC "Academy" - Vol. 2: M - Ya. - P. 577. (in Ukrainian).
10. Partyko, Z. V. (2004). Normatyvnyi aspekt i avtomatyzatsiia redahuvannia. Kyiv : Vydavn. tsentr Instytutu zhurnalistyky KNU im. Tarasa Shevchenka (in Ukrainian).
11. Partyko, Z. V. (2004). Slovnyk-minimum ukrainskoi movy. Kyiv : Vydavn. tsentr Instytutu zhurnalistyky KNU im. Tarasa Shevchenka (in Ukrainian).
12. Partyko, Z., Borodchuk, V., & Sorokatyi, I. (1995). Tryparametrychnyi metod vyznachennia chytabelnosti (skladnosti) ukrainomovnykh tekstiv: Palitra druku, 4, 54–55 (in Ukrainian).
13. Proverka teksta na chitabel'nost' dlja SEO. SEO analiz teksta: proverka na unikal'nost', chitabel'nost', peresпам. Retrieved from <https://copywritely.com/ru/readability-checker/>. Poslednij dostup: 05.03.2020. Nazvanie s jekrana (in Russian).

DOI 10.32403/2411-9210-2020-2-44-151-159

INFORMATION TECHNOLOGY OF DETERMINING THE COMPLEXITY OF EDUCATIONAL TEXT EDITIONS

I. Ohirko ¹, Z. Partyko ²

¹*Ukrainian Academy of Printing,
19, Pid Holoskom St., Lviv, 79020, Ukraine*

²*Zhytomyr Ivan Franko State University,
40, Velyka Berdychivska St., Zhytomyr, 10002, Ukraine
ogirko@gmail.com*

Among the criteria approved by the Government of Ukraine for assessing the quality of school publications, there is no such indicator as the complexity

(readability) of the text. If we consider the text of the publication as a set of letters (for oral text — a set of sounds), then complexity is understood as a complex whose properties depend on the number of letters (sounds) in the set, their specifics (semantics) and their combination (syntax). Since such assessments of the complexity of the text have been used in Western countries for almost hundred years, it is also advisable to introduce them in Ukraine.

Syntactic complexity is understood as the dependence of this property of texts from two indicators: first, from the average length of words in the text (in the number of letters), and, secondly, from the average length of sentences (in the number of words). Semantic complexity is understood as the number of words in the text that are absent in the minimum dictionary of any language (for our case, such a dictionary should be the minimum dictionary of the Ukrainian language). Such a minimum dictionary of the Ukrainian language with a volume of three thousand words (the volume of a minimum dictionary of three thousand words is its standard volume for any language) was compiled by one of the authors of this article and published in 2004. The compilation of this dictionary was performed by a specially developed program based on six frequency dictionaries of the Ukrainian language.

The formulas for calculating the complexity of Ukrainian-language texts were translated on the basis of formulas for English-language texts, adjusted for statistical differences in Ukrainian-language texts, in particular the average length of words.

To test the functioning of methods for determining the complexity of Ukrainian-language texts, an experimental system “Editor” has been developed. Testing the system on a number of specially selected texts (texts were a priori calibrated for complexity) gave positive results, i.e. indicated approximately the same complexity that was determined a priori. Nowadays, such systems for the analysis of English and Russian texts function as part of a number of word processors (such as Microsoft Word) and as stand-alone systems (on Internet sites).

Thus, the results of experimental use of the “Editor” system give grounds to assert the need to create an industrial system for determining the complexity of the texts of Ukrainian-language publications. During the development of such a system, it will be necessary to carry out additional calibration of the table of complexity of texts of publications to the appropriate classes of schools and university courses (nowadays a table containing seven levels of complexity is typical).

Keywords: *educational editions, evaluation criteria, text, complexity, readability, Ukrainian language, computer system.*

Стаття надійшла до редакції 12.11.2020

Received 12.11.2020