

УДК 811.16

**ЛЕКСИЧЕСКОЕ РАЗНООБРАЗИЕ
КАК ПАРАМЕТР СЛОЖНОСТИ ТЕКСТОВ УЧЕБНИКОВ
АНГЛИЙСКОГО ЯЗЫКА
(НА ПРИМЕРЕ УЧЕБНИКОВ «SPOTLIGHT» И «STARLIGHT»)**

Казачкова М. Б., Галимова Х. Н.

*Одинцовский филиал Московского государственного института
международных отношений (университет), Одинцово, Россия*

E-mail: MBKazachkova@yandex.ru;

Казанский инновационный университет им. В. Г. Тимирязева (ИЭУП),

Казань, Республика Татарстан, Россия

E-mail: galikha@mail.ru

В работе осуществлен сравнительный анализ сложности текстового материала через индекс Лексического разнообразия отечественных учебников «Spotlight» и «Starlight» (8–9 классов). Данные учебники утверждены и рекомендованы Министерством образования и науки РФ и используются в российских школах. Для расчета лексического разнообразия был использован лингвистический анализатор сложности англоязычного текста TextInspector. Под Лексическим разнообразием текста понимается вариативность, богатство, многообразие, диапазон словарного запаса, реализуемого в тексте. Материалом исследования послужил текстовый корпус учебников, объём которого составил 78 текстов и 27215 словоупотреблений. В ходе исследования выявлено: Лексическое разнообразие текстов учебника для 8 класса «Starlight» выше на 10 %, учебника «Spotlight». Индекс Лексического разнообразия текстового материала учебника для 9 класса «Starlight» так же имеет тенденцию к повышению в среднем на 11 % сравнительно с учебником «Spotlight». Показатель сложности текстов Лексическое разнообразие учебника «Starlight» незначительно выше показателя учебника «Spotlight», что говорит о незначительном усложнении учебника.

Ключевые слова: учебный текст, сложность текста, Лексическое разнообразие, текстовый корпус, токен, тайп.

ВВЕДЕНИЕ

Основной целью обучения иностранным языкам является формирование и развитие коммуникативной, речевой и лингвистической компетенций школьников, т. е. обучение практическому овладению иностранным языком. Поэтому, одним из важнейших моментов, который поможет ученикам добиться успеха в изучении языка, является проблема выбора материалов учебного процесса, а именно учебников и учебных пособий.

В настоящее время предлагается широкий спектр отечественных учебно-методических комплексов по иностранному языку. Министерство образования Российской Федерации ведет постоянную работу по обновлению учебников, включенных в федеральный перечень и соответствующих требованиям обновленного ФГОС.

Цель данного исследования – провести сравнительный анализ сложности текстового материала через индекс Лексического разнообразия отечественных учебников для общеобразовательных учреждений (8, 9 классов) под редакцией Ваулиной Ю. Е., Дули Д., Подоляко О. Е. «Spotlight» (Student's books/Английский в фокусе) и «Starlight» (Student's books/Звездный английский) – учебников для общеобразовательных учреждений и школ с углубленным изучением английского языка под редакцией Барановой К. М., Дули Д., Копыловой В. В. и др.

ИЗЛОЖЕНИЕ ОСНОВНОГО МАТЕРИАЛА ИССЛЕДОВАНИЯ

Соотношение понятий «трудность» и «сложность»

Согласно ФГОС, учебный текст – это «особый вид вторичного речевого произведения, обладающий рядом признаков, существенно отличающих его от других видов вторичных текстов. Вторичный учебный текст не заменяет полностью первоисточник, а служит его продолжением и развитием в соответствии с коммуникативной задачей» [9, с. 57].

А. А. Сабина считает учебным «любой текст, полезный при освоении определенной учебной дисциплины» [10].

Мы рассматриваем учебный текст, как «особый вид вторичного текста, так как он является продуктом письменной речевой деятельности автора учебника, созданный на базе первичного текста для передачи знаний учащимся» [2, с. 45]. Соответственно, дидактическая сложность учебных текстов будет отличаться.

Сложность – это базовая характеристика любого академического текста, которая понимается как «сложность его чтения и восприятия в соответствии с возрастными характеристиками читателя и рассматривается как критическая переменная, которая зависит от различных факторов, например, таких как жанр и степень изученности темы» [5, с. 22].

На данный момент существуют различные подходы к пониманию сложности текстовой единицы.

До сих пор имеет место смешивание понятий «сложности» и «трудности». В ряде работ сложность трактуется как характеристика текста, зависящая от внутренних параметров текста, влияющих на понимание темы, скорости чтения, уровня интереса к прочитанному [1; 7; 14]. Трудность, в отличие от сложности, определяется количеством учебного труда, затрачиваемого на усвоение учебного материала [15]. Зачастую уровень трудности учебного материала не всегда совпадает с его сложностью. Так, к примеру, И. Я. Лернер соотносит трудность текста с субъективной характеристикой учащегося, т.е. уровнем его подготовленности учащегося и когнитивными способностями. Трудность текста он ставит в зависимость от его сложности [7], а по отношению к задаче считает, что «трудность характеризует возможность субъекта преодолеть объективную сложность задачи» [7, с. 34], что означает: один и тот же материал субъективно может восприниматься по-разному разными учениками.

Параметры сложности текста

Я. А. Микк выводит сложность текста на основе четырех основных параметров: информативность текста, сложность предложений, ясность текстовой структуры и абстрактность. В самом общем представлении сложность текста определяется морфологическими, лексическими, синтаксическими характеристиками текста. К морфологическим характеристикам относятся средняя длина слова (среднее количество морфем в слове). Лексические параметры сложности текста – это частотность, абстрактность, этноспецифичность, неоднозначность, уникальность, референциальная специфичность и др. Синтаксическая сложность соотносится к

комплексностью модели, порядком слов в предложении, длиной предложений, количеством слов в абзацах, расстоянием между зависимыми словами и др. Текст также может характеризоваться информационной (когнитивной) сложностью и абстрактностью [12].

На сегодняшний день европейскими образовательными стандартами для анализа сложности текстов используются 3 базовых фактора: количественные параметры, качественные параметры и уровень подготовленности читателя [16].

Количественные характеристики рассчитываются с помощью различных компьютерных программ и текстовых профайлеров (например, Rulingva¹ – программа автоматического анализа для русскоязычных текстов, CoMetrix² и TextInspector³ – текстовые анализаторы для англоязычных текстов, включающие следующие показатели: среднее количество слогов в слове, общее количество слов и слогов в тексте, количество предложений, среднее количество слов в предложении, общее количество определённых частей речи и индекс удобочитабельности по Флеш-Кинкейду⁴ [16; 17].

Качественные характеристики включают такие параметры сложности, как структура текста, визуальная поддержка, ясность и точность языка, синтаксическая сложность, нарративность, конкретность и абстрактность слов, когезия и когеренция [5; 6].

Расчет индекса Лексического разнообразия (ЛР/ТТР)

Индекс Лексического разнообразия (ЛР) – это количественная характеристика текста при оценке его сложности [11].

Впервые термин «лексическое разнообразие» был предложен Скоттом Джарвисом, американским лингвистом, который утверждал, что ЛР следует рассматривать как обобщающий термин, аналогичный экологическому разнообразию. Согласно модели Джарвиса, ЛР включает изменчивость, объем, равномерность, редкость, дисперсию и несоответствие [8].

На наш взгляд, ЛР текста есть «вариативность, богатство, многообразие, диапазон словарного запаса, реализуемого в тексте» [4] и является одним из ключевых параметров, которые используются для оценки разнообразия и качества текстового материала.

Индекс ЛР может быть рассчитан различными методами, одним из которых является ТТР (Type-Token Ratio), который позволяет определить отношение числа уникальных слов к общему числу слов в тексте.

Формула, по которой рассчитывается ТТР (ЛР):

¹ «Интеллектуальные технологии управления текстами» на базе Учебного корпуса русского языка (<https://rulingva.kpfu.ru>).

² CoMetrix. <https://www.cometrix.com.au/>

³ TextInspector/ <https://textinspector.com/>

⁴ Рассчитывается по формуле $206.835 - (1.015 \times ASL) - (84.6 \times ASW)$, где ASL – средняя длина предложения в словах, ASW – средняя длина слова в слогах.

$$TTR \text{ (type-token ratio)} = \frac{t}{n}$$

, где t – количество тайпов (уникальных лексических единиц); n – общее количество токенов (словоформ) [16].

Например, если в тексте из 100 слов имеется 50 уникальных лексических единиц, то TTR будет равно 0,5 (50/100).

ЛР текста используется в лингвистике для анализа и оценки разнообразия лексики в тексте. Этот показатель может быть полезен при изучении языковых особенностей, культурных различий и стилей письма. Кроме того, ЛР можно использовать в качестве инструмента для оценки уровня грамотности и образованности автора. Коэффициент ЛР «служит мерой успешности речевого акта, например, при патологии речи или в ситуации говорения на иностранном языке» [3, с. 21]. Он также применим как мера прогресса овладения иностранным языком у обучающихся [17], то есть когда ЛР измеряется несколько раз на протяжении определенного срока.

Хотя ЛР может быть полезным инструментом для анализа текста, он имеет свои ограничения. Во-первых, он не учитывает морфологическую разновидность слов, что означает, что слова разных частей речи могут быть учтены как разные типы, даже если они имеют одинаковую основу и, соответственно, являются однокоренными. Во-вторых, результат расчета индекса ЛР зависит от длины измеряемого текста, у длинных текстов ЛР всегда ниже, чем у коротких. Соответственно при расчете данного параметра необходимо сравнивать тексты только одинаковой длины [3]. При очень коротких текстах индекс будет всегда высоким, поскольку в начале все токены уникальны, при большом объеме текста количество уникальных слов является почти постоянной, поскольку становится больше повторений [16]. Поэтому вместо синтагматической оси, мы получаем что-то похожее на гиперболу.

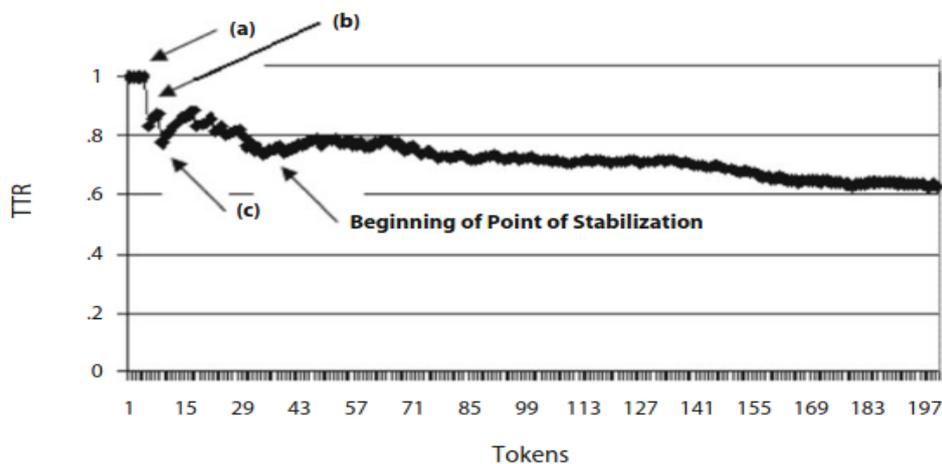


Рис. 1. Зависимость индекса TTR (ЛР) от длины текста

Данный график можно поделить на три части. В части, *a* ЛР равняется 1, поскольку в начале текста почти все лексические единицы уникальны. Как только начинают появляться повторения, в части *b* значение ЛР резко уменьшается. После этого снова наблюдается рост ЛР, что говорит о появлении новой порции уникальных единиц. Как мы видим, подобные колебания повторяются несколько раз до точки стабилизации (*point of stabilization*), после которой нет значительных изменений в количестве уникальных единиц. Мы видим, что пик разнообразия лексики содержится как раз до этой точки «насыщения».

Соответственно, для определения индекса ЛР нет смысла брать длинные тексты. При выборке текстов для анализа достаточно будет взять тексты объемом не более 300 слов [4].

Исследование

Материалом исследования послужил текстовый корпус учебников для 8–9 классов «Spotlight» (Английский в фокусе) под редакцией Ваулиной Ю. Е., Дули Дж. и др. и «Starlight» (Звездный английский) под редакцией Барановой К. М., Дули Д., Копыловой В. В. и др. объем которого составил 78 текстов 27215 словоупотреблений (см. таб. 1).

Количество словоупотреблений в подкорпусах

Таблица 1

Класс	Количество словоупотреблений
8 кл. «Spotlight» (Английский в фокусе) под редакцией Ваулиной Ю.Е., Дули Дж. и др.	5552
8 кл. «Starlight» (Звездный английский) под редакцией Барановой К.М., Дули Д., Копыловой В.В. и др.	6361
9 кл. «Spotlight» (Английский в фокусе) под редакцией Ваулиной Ю.Е., Дули Дж. и др.	6448
9 кл. «Starlight» (Звездный английский) под редакцией Барановой К.М., Дули Д., Копыловой В.В. и др.	8854
	27215

Выбор данных линеек учебников обусловлен тем, что они подготовлены в соответствии с требованиями ФГОС для основной школы, включены в Федеральный перечень учебников, рекомендованных к изучению в школе¹, и, соответственно, учитывают требования обязательной образовательной программы до 2022 г.

Для расчета ЛР был использован лингвистический анализатор сложности англоязычного текста TextInspector². Данный – анализатор используется для подготовки учебных материалов, тестов и оценки работ учащихся. Он также дает возможности рассчитать процент использования академической лексики и лексики среднего уровня владения английским языком B1 по шкале CEFR, напрямую коррелирующие с ЛР (см. таб.2).

¹ Федеральный перечень учебников на 2021–2022 учебный год. <https://4ege.ru/documents/56987-federalnyy-perechen-uchebnikov-na-2018-2019-uchebnyy-god.html>

² <https://textinspector.com/workflow/26B3C092-C663-11ED-A5DF-6F3D603959EA>

Параметры сложности текстов из учебников
«Spotlight» и «Starlight»

Таблица 2

	8 класса «Spotlight»	9 класса «Spotlight»	8 класса «Starlight»	9 класса «Starlight»
TTR (ЛР)	0,58	0,60	0,65	0,68
Количество академической лексики (%)	2,33%	2,60%	2,51%	2,86%
Количество слов уровня В1 по CEFR (%)	11,10%	11,87%	12,230%	15,11%

Согласно полученным данным, параметры сложности учебных текстов, анализируемых учебников, в значительной степени отличаются друг от друга. Наиболее высокую степень сложности имеют тексты из учебников «Starlight» под редакцией Барановой К. М., Дули Д., Копыловой В. В. и др. Соответственно, тексты из учебника «Spotlight» под редакцией Ваулиной Ю. Е., Дули Дж. демонстрируют наименьшую степень сложности.

ЛР текстов учебника для 8 класса «Starlight» выше на 10 %, чем в учебнике одноименного класса «Spotlight». Что касается учебников для 9 класса, то индекс ЛР в учебнике «Starlight» так же имеет тенденцию к повышению в среднем на 11 % сравнительно с учебником «Spotlight».

Индексы ЛР текстов незначительно увеличиваются от класса к классу, что свидетельствует об постепенном увеличении сложности и расширении диапазона словарного запаса. Прослеживаются дидактические принципы доступности и посильности, которые предполагают, что материал представляется в соответствии с возрастными и интеллектуальными возможностями учащихся, а его усвоение не вызывает у них непреодолимых трудностей [6].

При этом более высокий параметр сложности ЛР текстов учебников «Starlight» позволяет предположить, что тексты содержат высокий процент лексики уровня В1 и академической лексики.

ВЫВОДЫ

Исследование позволило выявить, что такой параметр сложности текста, как Лексическое разнообразие, учебника «Starlight» незначительно выше показателей учебника «Spotlight», что говорит о незначительном усложнении учебника. Тот факт, что учебник «Starlight» сложнее, небезоснователен, поскольку данная серия учебников рекомендована для обучения курса там, где делается упор на углубленное изучение иностранного языка. Учебник «Spotlight» является центральным элементом учебно-методического комплекта линейки «Английский в фокусе» и базовым учебником для общеобразовательных организаций.

Однако отметим, что значение данных параметров недостаточно для определения «идеального учебника», поскольку рассчитанные параметры сложности не отражают структуру, оформление и содержательные признаки учебных материалов.

Список литературы

1. Бирюков Б. В., Тюхтин В. С. О понятии сложности // Логика и методология науки: Мат-лы IV Всесоюз. Симпоз. (Киев, Июнь, 1965 г.) – М., 1967. – С. 219–231.
2. Биличенко Н. Учебный текст и его разновидности. – Pandia, 2007. – Режим доступа: <https://pandia.ru/text/77/304/32260.php>. – (Дата обращения: 01.07.2023).
3. Захарова Е. Ю., Савина О. Ю. Лексическое разнообразие текста и способы его измерения // Вестник Тюменского государственного университета. Гуманитарные исследования. – 2020 – № 1 (21) – С. 20–34.
4. Казачкова М. Б., Галимова Х. Н. Лексическое разнообразие текста как параметр сложности текстов // Вестник марийского государственного университета. – 2021. – № 3 (43). – С. 384–390.
5. Казачкова М. Б., Галимова Х. Н. Сравнительный анализ сложности текстов учебников английского языка // Дискурс профессиональной коммуникации. – 2022 – № 4.– С.22–32.
6. Леонтьев А. А. Возрастные и индивидуальные особенности школьника–в в процессе обучения иностранному языку // Иностранные языки в школе.– 1976. – № 1. – С. 97–101.
7. Лернер И. Я. Критерии сложности некоторых элементов учебника: Проблемы школьного учебника. – М.: Просвещение, 1974. – Вып. 1. – С. 47–58.
8. Лексическое разнообразие – Lexical diversity. – Режим доступа: https://ru.wikibrief.org/wiki/Lexical_diversity. – (Дата обращения: 01.07.2023).
9. Пешкова Н. П., Яхиббаева Л. М. Первичность и вторичность как онтологические категории учебного текста // Вестник Южно-Уральского гос. университет. –2009. – №2 (135). – С. 56–60.
10. Сабина А. А. Учебный текст: структура и прагматика // Известия РГПУ им. А. И. Герцена– 2009. – № 97. – С. 222–225.
11. Солнышкина М. И., Замалетдинов Р. Р., Гафиятова Э. В., Гизатулина Д. Ю., Бегаева М. Н. Многофакторный анализ сложности текста // Иностранные языки в школе. – 2021. – № 2. – С. 28–34.
12. Солнышкина М. И., Казачкова М. Б., Харькова Е. В. Инструменты измерения сложности текстов на английском языке // Иностранные языки в школе – 2020. – № 3 – С. 15–21.
13. Тарева Е. Г. Учебник иностранного языка: сила воздействующей функции // Иностранные языки в школе.– 2021. – № 7. – С. 5–11.
14. Ушаков К. М. О критериях сложности учебного материала школьных предметов // Новые исследования в педагогических науках. – 1980. – № 2 (36). – С. 33–35.
15. Цетлин В. С. Дидактические требования к критериям сложности учебного материала // Новые исследования в педагогических науках. – 1980. – № 1 (35). – С. 30–33.
16. Fisher, D., Frey, N., & Lapp, D. Text complexity: Raising rigor in reading. – Newark, DE: International Reading Association, 2012. – 212 p.
17. Fergadiotis G., Heather H. W., Thomas M. W. Measuring lexical diversity in narrative discourse of people with aphasia // American Journal of Speech-Language Pathology. – 2013. – Vol. 22. №. 2 – P. 397–408.

References

1. Biriukov B. V., Tiukhtin B. C. *O poniatii slozhnosti* [About the concept of complexity]. *Logika i metodologiya nauki: materialy IV- Vsesoiuz. simpoz.* Moscow, 1967, pp. 219–231.
2. Bilichenko N. *Uchebnyi tekst i ego raznovidnosti* [Educational text and its varieties]. Pandia, 2007. Available from: <https://pandia.ru/text/77/304/32260.php> (accessed 01 July 2023).
3. Zakharova E. Iu., Savina O. Iu. *Leksicheskoe raznoobrazie teksta i sposoby ego izmereniia* [Lexical diversity of the text and ways to measure it]. *Vestnik Tiimenskogo gosudarstvennogo universiteta. Gumanitarnye issledovaniia*, 2020, no.1 (21), pp. 20–34.
4. Kazachkova M. B., Galimova Kh. N. *Leksicheskoe raznoobrazie teksta kak parametr slozhnosti tekstov* [Lexical diversity of the text as a parameter of the complexity of texts]. *Vestnik mariiskogo gosudarstvennogo universiteta*, 2021, no. 3 (43), pp.384–390.
5. Kazachkova M. B., Galimova Kh. N. *Sravnitel'nyi analiz slozhnosti tekstov uchebnikov angliiskogo iazyka* [Comparative analysis of the complexity of texts of English textbooks]. *Diskurs professional'noi kommunikatsii*, 2022, no. 4, pp.22–32.
6. Leont'ev A. A. *Vozrastnye i individual'nye osobennosti shkol'nikov v protsesse obucheniia inostrannomu iazyku* [Age and individual characteristics of schoolchildren in the process of learning a foreign language]. *Inostrannye iazyki v shkole*, 1976, no. 1, pp. 97–101.
7. Lerner I. Ja. *Kriterii slozhnosti nekotorykh elementov uchebnika: Problemy shkol'nogo uchebnika* [Criteria for the complexity of some elements of the textbook: Problems of the school textbook]. Moscow, Prosveshchenie Publ., 1974, pp. 47–58.
8. *Leksicheskoe raznoobrazie* [Lexical diversity]. Available from: https://ru.wikibrief.org/wiki/Lexical_diversity (accessed 01 July 2023).
9. Peshkova N. P., Iakhibbaeva L. M. *Pervichnost' i vtorichnost' kak ontologicheskie kategorii uchebnogo teksta* [Primary and secondary as ontological categories of the educational text]. *Vestnik. IuzhnoUralskogo gosudarstvennogo universiteta*, 2009, no. 2, pp.56–60.
10. Sabinina A. A. *Uchebnyi tekst: struktura i pragmatika* [Educational text: structure and pragmatics]. Available from: https://lib.herzen.spb.ru/text/sabinina_97_222_225.pdf (accessed 04 September 2023).
11. Solnyshkina M. I., Zamaletdinov R. R., Gafiatova E. V., Gizatulina D. Ju., Begaeva M. N. *Mnogofaktornyi analiz slozhnosti teksta* [Multivariate analysis of text complexity]. *Inostrannye iazyki v shkole*, 2021, no. 2, pp. 28–34.
12. Solnyshkina M. I., Kazachkova M. B., Khar'kova E. V. *Instrumenty izmereniia slozhnosti tekstov na angliiskom iazyke* [Tools for measuring the complexity of texts in English]. *Inostrsnnie iazyki v shkole*, 2020, no. 3, pp.15–21.
13. Tareva E. G. *Uchebnik inostrannogo iazyka: sila vozdeistvuiushchei funktsii* [Foreign language textbook: the power of the influencing function]. *Inostrannye iazyki v shkole*, 2021, no 7, pp. 5–11.
14. Ushakov K. M. *O kriteriiakh slozhnosti uchebnogo materiala shkol'nykh predmetov* [About the criteria of the complexity of the educational material of school subjects]. *Novye issledovaniia v pedagogicheskikh naukakh*, 1980, no. 2 (36), pp. 33–35.
15. Tsetlin B. C. *Didakticheskie trebovaniia k kriteriiam slozhnosti uchebnogo materiala* [New research in pedagogical sciences] *Novye issledovaniia v pedagogicheskikh naukakh*, 1980, no. 1 (35), pp. 30–33.
16. Fisher D., Frey N., Lapp D. *Text complexity: Raising rigor in reading*. Newark, DE, International Reading Association, 2012. 212 p.
17. Fergadiotis G. Heather H. W., Thomas M. W. *Measuring lexical diversity in narrative discourse of people with aphasia*. *American Journal of Speech-Language Pathology*, 2013, vol. 22, no. 2, pp. 397–408.

**LEXICAL DIVERSITY AS A PARAMETER OF THE COMPLEXITY OF TEXTS
IN ENGLISH TEXTBOOKS
(ON THE EXAMPLE OF THE TEXTBOOKS "SPOTLIGHT" AND "STARLIGHT")**

Kazachkova M.B., Galimova K.N.

In the presented work, a comparative analysis of the complexity of text material was carried out through the Lexical Diversity Index of domestic textbooks "Spotlight" and "Starlight" (grades 8 - 9). These textbooks are approved and recommended by the Ministry of Education and Science of the Russian Federation and are used in Russian schools. To calculate lexical diversity, the linguistic complexity analyzer of English text TextInspector was used. Lexical diversity of a text means variability, richness, variety, and range of vocabulary implemented in the text. The research material was a text corpus of the textbooks, the volume of which amounted to 78 texts and 27,215 word usages. The study revealed: The lexical diversity of texts in the textbook for 8th grade "Starlight" is 10% higher than in the textbook "Spotlight". The Lexical Diversity Index of the text material of the textbook for grade 9 "Starlight" also tends to increase by an average of 11% compared to the textbook "Spotlight". Indicator of text complexity Lexical diversity of the "Starlight" textbook is slightly higher than that of the "Spotlight" textbook, which indicates a slight complication of the textbook.

Key words: educational text, text complexity, Lexical diversity, text corpus, token, type.