

А. И. Габитов, Л. Х. Габитова, Л. Х. Шаяхметова

ОЦЕНКА СЛОЖНОСТИ АНГЛОЯЗЫЧНОГО ТЕКСТА НА ОСНОВЕ ОТДЕЛЬНЫХ ГРАММАТИЧЕСКИХ ЭЛЕМЕНТОВ С ИСПОЛЬЗОВАНИЕМ ФОРМУЛ УДОБОЧИТАЕМОСТИ

Казанский (Приволжский) федеральный университет, г. Казань, Россия

Аннотация. Современные средства и формулы оценки сложности и удобочитаемости англоязычных текстов используют схожие методы оценивания, основанные на анализе ряда лексических элементов. Выбор характеристик и параметров текстов, оценка их влияния на сложность текстов не имеют общепринятой нормы, которая могла бы использоваться повсеместно. Многие факторы, влияющие на сложность текстов, не подвергаются анализу, в числе таких – сложность отдельных грамматических категорий. Сложность лексических единиц оценивается в отрыве от некоторых грамматических характеристик. Данная проблема требует отличного от привычной нормы подхода, учитывающего нагрузку грамматических элементов сложности на когнитивные способности человека, такие как память и внимание. Существуют направления лингвистики, изучающие обработку входных данных, которые подтверждают влияние грамматики текстовых материалов на сложность. Определенные грамматические категории способны вызвать большую трудность у обучающихся английскому языку, нежели другие, вызывая перегрузку ограниченных когнитивных ресурсов человека. В данном исследовании мы предприняли попытку создать англоязычные текстовые материалы определенного характера, содержащие грамматические элементы сложности, и проанализировать их при помощи формул оценки сложности и удобочитаемости. Результаты анализа были сопоставлены с результатами модифицированных вариантов с изменением сложности за счет грамматических элементов. Контрастивный анализ текстов проводился с применением 9 формул оценки сложности текста и удобочитаемости, в числе которых формула Флеш-Киндейда, Индекс Колман-Лиану и индекс удобочитаемости Фрая.

Ключевые слова: *грамматическая сложность, контрастивный анализ, обработка входных данных, параметры текста, сложность текста*

A. I. Gabitov, L. Kh. Gabitova, L. Kh. Shayakhmetova

ASSESSING COMPLEXITY OF ENGLISH TEXTS BASED ON INDIVIDUAL GRAMMATICAL ELEMENTS WITH THE USE OF READABILITY FORMULAE

Kazan (Volga Region) Federal University, Kazan, Russia

Abstract. Modern tools and formulas for assessing the complexity and readability of texts use similar assessment methods based on the analysis of a number of lexical elements. The selection of characteristics and parameters of texts in English language, assessment of their influence on the complexity of texts do not have a generally accepted norm that could be used everywhere. Many factors influencing the complexity of texts are not analyzed, including the complexity of individual grammatical categories. The complexity of lexical units is assessed in isolation from some grammatical characteristics. This problem requires an approach that is different from the usual norm, taking into account the load of grammatical complexity elements on human cognitive abilities, such as memory and attention. There are branches of linguistics that study input processing that support the influence of the grammar of textual materials on complexity. Certain grammatical categories can cause greater difficulty for language learners than others, causing an overload of a person's limited cognitive resources. In this study, we attempted to create text materials of a certain nature containing grammatical elements of complexity and analyze them using a formula for assessing complexity and readability. The results of the analysis were compared with the results of modified versions with changes in complexity due to grammatical elements. Contrastive text analysis was carried out using a number of formulas for assessing text complexity and readability.

Keywords: *grammatical complexity, contrastive analysis, input processing, text parameters, text complexity*

Введение. Специалисты в области английского языка сталкиваются с проблемой выбора и адаптации текста на английском языке для определенных задач. Суть проблемы заключается в корректной оценке текста на предмет его сложности. Имеется множество теоретических подходов к оценке сложности англоязычного текста, предложенных такими учеными, как Р. Флеш, Дж. Кинкейд, М. Коулман, Т. Л. Лиану, Э. Дейл, Ж. Чалл, которые предоставляют возможность измерения общей удобочитаемости текста на основе различных параметров текста и их взаимосвязи [1, с. 15]. Однако выбор данных средств ставит специалистов в условия, где им приходится проводить самостоятельные исследования сложности англоязычного текста или довериться готовым инструментам оценки сложности текста, которые зачастую предлагают ранжировать тексты в соответствии с возрастной группой или уровнем языковой компетенции.

Актуальность исследуемой проблемы. Существует множество формул удобочитаемости, однако в них оценка сложности текста основывается на двух лингвистических характеристиках, коими являются длина предложения и сложность слов. Многие факторы, влияющие на сложность англоязычных текстов, не подвергаются анализу, в числе таких – сложность отдельных грамматических категорий. Есть направления лингвистики, изучающие обработку входных данных, которые подтверждают влияние грамматики текстовых материалов на сложность [5]. Данная проблема требует отличного от привычной нормы подхода к оценке сложности текстов, учитывающего нагрузку грамматических элементов сложности на когнитивные процессы человека, такие как память и внимание. Иными словами, выявляется необходимость в более глубоком исследовании выделенной проблемной области. Цель данного исследования заключается в определении критериев и особенностей оценки сложности англоязычных текстов с учетом определенных грамматических элементов и сопоставлении возможностей формул удобочитаемости, придерживаясь трактовки сложности текста как фактического или прогнозируемого результата конкретного читателя, выполняющего работу с текстом. В данном ключе для оценки сложности текста должны учитываться параметры самого текста и знания конкретного человека, работающего с ним, а именно слова, синтаксис и структура дискурса, длина слова и длина предложения, частота использования слов, языковая структура, повторение слов, связность текста, смысловая нагрузка текста, языковые условности и ясность, требования к иным знаниям, предъявляемые к читателю.

Материал и методы исследования. В данной работе использованы описательный метод, включающий наблюдение, сопоставление и обобщение, сбор фактического материала и его классификацию по формулам оценки сложности текстов (формула Флэш Кинкейда, формула Р. Ганнинга, индекс Колмана-Лиану, автоматизированный индекс удобочитаемости, формула FORCAST Grade Level, индекс удобочитаемости А. Рейгора, индекс удобочитаемости Фрая, новая формула Дейла-Чалл, формула Lensear Write); метод измерения; эксперимент, проведенный на основе двух вариантов нескольких текстов до и после модификации. Отрывки текстов были отобраны из аутентичных материалов художественных текстов повествовательной направленности.

Результаты исследования и их обсуждение. В качестве примера для выявления потенциальных проблем, возникающих с применением формул оценки сложности англоязычных текстов, можно привести формулу Дейла-Чалл. Неточность в оценке при использовании данной формулы проявляется в случае, когда значение будет иметь не только класс образования читателя, но и разница в гуманитарной или технической направленности. Таким образом, текст средней сложности, который большинство учеников

технической направленности посчитали бы скорее простым, оценивался бы формулой как неудобочитаемый [10, с. 21].

Оценка сложности грамматических элементов текста прорабатывалась учеными-лингвистами Р. Митчеллом, Р. Майлсом, М. Савиль-Тройке в области изучения второго языка, процесса развития любой языковой системы, отличной от родного языка человека. Данная междисциплинарная область исследований пришла к различным теориям, моделям и гипотезам, которые проливают свет на возможные объяснения явлений, наблюдаемых в процессе попыток получить более четкое представление о том, как происходит усвоение языка [9]. Одна из современных теорий усвоения второго языка берет свое начало в психологии и пытается объяснить это с когнитивной точки зрения. Согласно данной теории, овладение языком аналогично обучению любому сложному навыку. Для достижения результата задействуются психические процессы, такие как восприятие, поиск, организация и реорганизация ментальных представлений [10]. Рассматривая проблему с этой точки зрения, мы видим, что такие процессы, как внимание и память, представлены как ресурсы ограниченного характера. Данный принцип демонстрируется в теории обработки входных данных Ван Паттена [6, с. 23]. Согласно данной теории, добавление новых элементов в речь или текст напрямую влияет на повышение его общей сложности посредством увеличения нагрузки на когнитивные способности человека, задействованные в обработке.

Одни виды предложений труднее обрабатывать, чем другие, а некоторые предложения настолько сложны, что превышают возможности человеческого синтаксического анализатора. Яркой иллюстрацией ограниченных возможностей синтаксических анализаторов человека являются предложения с множественным встраиванием в центр. Согласно Н. Хомскому и Дж. Кимбаллу, предложения с множественными центральными вложениями трудно обрабатывать [2], [8]. В редких случаях в письменной речи могут встречаться предложения с двумя степенями встраивания в центр [7, с. 366]. Предложения с тремя или более степенями встраивания в центр становятся практически невозможными для понимания человеком. Помимо прочего существует эффект отсутствия глагольной группы (*missing VP effect*). Данный термин был введен Э. Гибсоном и Дж. Томасом [5] и обозначает небольшой класс грамматических иллюзий в предложениях, которые воспринимаются как грамматически правильные, несмотря на то, что содержат нарушения правил грамматики. Можно сделать вывод, что грамматические отклонения также вызывают увеличение сложности текста.

Возможности человеческого мозга для обработки синтаксиса ограничены. Способность к вычислениям в уме ограничена небольшим подмножеством чисел, способность вспоминать списки не связанных между собой элементов ограничена списками, состоящими не более чем из семи или восьми элементов, и т. д. Ограничения такого рода часто связывают с системой рабочей памяти [4]. Проясним задачи, которые необходимо выполнить, чтобы успешно анализировать предложения. По определению синтаксический анализатор берет слова с начала строки и создает для них синтаксическую структуру. Имея веские доказательства того, что синтаксический анализ выполняется человеком поэтапно, задачу парсеров можно разделить на две основные подзадачи [5, с. 8]. Во-первых, синтаксический анализатор должен сохранить синтаксическую структуру в памяти. Во-вторых, синтаксический анализатор должен интегрировать каждое слово входной строки в развернутую синтаксическую структуру, когда это слово встречается.

Относительно простым для отслеживания в тексте и определения влияния на сложность англоязычного текста грамматическим элементом можно назвать условные предложения. Основными сложностями в условных предложениях являются их форма, использование времен и связанное значение двух предложений [3, с. 27]. Основываясь на вышеупомянутых эмпирических исследованиях, можно предположить, что трудности, с которыми

сталкиваются обучающиеся при изучении английских условных предложений, связаны с синтаксической сложностью этих структур, а не с неспособностью учащихся делать условные рассуждения на английском языке [3, с. 29]. Трудности с освоением английских условных выражений изучающими английский язык как второй иностранный вызваны синтаксической сложностью целевых структур. Включение данного грамматического элемента в текст должно влиять на его сложность в значительной мере.

В качестве эксперимента мы предлагаем провести с использованием формул сложности контрастивный анализ оценки англоязычных текстов, содержащих условные предложения и не содержащих их. Нами были отобраны несколько отрывков из аутентичных текстов, содержащих в себе условные предложения. Для обеспечения валидности полученных данных параметры текстов, используемых в формулах сложности, должны были оставаться в равных прогнозируемых диапазонах. Все изначальные тексты имели приблизительно равный объем слов в диапазоне от 120 до 130. Каждый текст имел набор из 7 предложений, каждое из которых включало в себя от 12 до 26 слов.

Предложения были модифицированы грамматически и лексически. Набор грамматических структур, содержащийся в текстах, был идентичен, тексты были написаны с использованием одинакового набора времен английского языка. Помимо условных предложений тексты включали в себя страдательный залог. Лексическая сложность представляла собой большую проблему ввиду различий конкретных формул в подходе к оценке сложности слов. Тексты содержали в себе набор из относительно простых и относительно сложных слов в соотношении 85 % к 15 % соответственно. В качестве оценочной базы был использован стандарт Common European Framework of Reference. Следующим этапом было составление вариантов текстов с изменением предложений, содержащих в себе условия. Была предпринята попытка сохранения семантического смысла предложений с одновременным удалением из них грамматического элемента условия. В результате одно сложное предложение было разделено на два простых с сохранением лексики. После изменения тексты стали короче на 6–7 слов, но получили на 3 предложения больше.

Нами был отобран ряд формул оценки сложности англоязычных текстов. Несколько формул можно объединить в одну категорию, где оцениваются числовые параметры текстов, отличия заключаются в конечных числовых константах. Тест Flesch-Kincaid Grade Level Test использует формулу, которая зависит от длины предложения и длины слова. А. Рейгор и Э. Фрай дополнительно к подсчету параметров текста используют графики для сопоставления результатов. Более детальный анализ предлагают формулы Lensear Write и Дейла-Чалл. В них предпринята попытка учета лексической сложности текста. Самой сложной для анализа системой оценки является уровень владения языком CEFR, она учитывает следующие переменные: среднее количество простых слов, среднее количество слов в предложении, среднее количество местоимений в предложении, среднее количество слогов в слове, среднее количество предлогов в предложении.

Данные, приведенные в таблице 1, показывают явное упрощение сложности англоязычных текстов по формулам сложности. Согласно формуле Флеш-Кинкейда, уровень сложности снизился с 1,4 до 2,6 пункта. Gunning Fog Index схожим образом показал упрощение сложности текстов от 2,1 до 2,8 пункта. Работающий схожим образом, но учитывающий также длину слов в буквах индекс Коулмана-Лиау показал упрощение, однако в меньшей степени – со значениями от 0,3 до 0,8 пункта. В то же время формула Automated Readability Index, которая также учитывает количество символов в словах, показала значительно большее значение упрощения сложности – от 2,0 до 3,3 пункта. В процентном соотношении данное значение приобретает 53–75 %. Формула FORECAST показывает обратный результат, согласно ей происходит усложнение текстов от 0,2 до 0,7 пункта.

Таблица 1 – Данные анализа сложности

	textA	textA_ changed	textA_ difference	textB	textB_ changed	textB_ difference	textC	textC_ changed	textC_ difference
Flesch-Kincaid Grade Level	8,40	5,80	2,60	10,10	7,90	2,20	9,30	7,90	1,40
Gunning Fog Index	10,80	8,00	2,80	12,60	10,10	2,50	11,90	9,80	2,10
Coleman-Liau Index	6,80	6,00	0,80	10,00	9,30	0,70	8,60	8,90	0,30
Automated Readability Index	7,10	3,80	3,30	9,40	6,20	3,20	8,10	6,10	2,00
FORCAST Grade Level	9,40	9,60	0,20	10,20	10,60	0,40	10,20	10,90	0,70
Raygor Readability	6,00	4,00	2,00	7,00	6,00	1,00	7,00	7,00	0,00
Fry Readability	7,00	6,00	1,00	10,00	11,00	1,00	9,00	9,00	0,00
New Dale-Chall Score	4,50	4,30	0,20	5,00	4,90	0,10	4,80	5,10	0,30
Lensear Write	75,20	82,10	6,90	68,30	75,00	6,70	68,60	72,20	3,60
CEFR Level	B2	B2		C1	C1		C1	C1	

Мы наблюдаем снижение сложности текстов. Формула Фрая, установившая значения для изначальных текстов выше, чем формула А. Рейгора, дает значение разницы от 0 до 1 пункта. Формула А. Рейгора показывает уменьшение от 0 до 2 пунктов. Формула Дейла-Чалл показывает уменьшение сложности от 0,1 до 0,3 пункта. Анализ по формуле Lensear Write показывает увеличение сложности от 3,6 до 6,9 пункта, что не является существенным, если взять его в процентном диапазоне, которое будет равно 5–9 %. Касательно уровня владения языком CEFR тексты были оценены на уровни B2 и C1, после модификации оценка не изменилась.

Выводы. Как мы можем наблюдать, формулы оценки неоднозначно отреагировали на модификации текстов. Увеличение количества предложений с одновременным уменьшением слов и средней длины предложений в теории должно было дать уменьшение сложности по всем формулам, однако на практике мы видим иной результат. Некоторые формулы отреагировали на это увеличением сложности, в частности, формула FORECAST и Automated Readability Index, что мы объясняем уменьшением количества коротких слов после удаления условных предложений и замены времен, в них использовавшихся. Формулы, применяющие графики для определения уровня сложности, показали значительное отличие в оценке одних и тех же текстов. Формулы Флеша-Кинкейда, Р. Ганнинга, индекс Колман-Лиану и новая формула Дейла-Чалл были наиболее предсказуемы и показали ожидаемое снижение сложности. В то же время изначальная сложность текстов определяется ими с большими вариациями.

Формулы оценки текстов на сложность опираются главным образом на параметры, которые можно выразить числовыми значениями, реже – на сложность лексического наполнения. Подобные формулы не могут достоверно спрогнозировать изменение сложности текстов в зависимости от грамматических компонентов ни прямым, ни косвенным образом. Конкретный текст получает неидентичную оценку по разным формулам. Оценка сложности грамматической составляющей текстов требует иного подхода, учитывающего сложность обработки данного элемента человеком. Исследование данной проблемы должно продолжаться в направлении изучения когнитивных ограничений человеческого мозга при обработке грамматических элементов текста.

ЛИТЕРАТУРА

1. Солнышкина М. И., Казачкова М. Б., Харькова Е. В. Цифровые технологии измерения сложности текстов как инструмент управления качеством обучения чтению на английском языке // Иностранные языки в школе. – 2020. – № 3. – С. 15–21.
2. Chomsky N., Miller G. A. Introduction to the formal analysis of natural languages // Journal of Symbolic Logic. – 1963. – Vol. 33(2). – P. 299–300. – DOI 10.2307/2269904.
3. Chou Ch. L. Chinese speakers' acquisition of English conditionals: acquisition order and L1 transfer effects // Second Language Studies. – 2000. – Vol. 19(1). – P. 57–98.
4. Gibson E. Linguistic complexity: locality of syntactic dependencies // Cognition. – 1998. – Vol. 68. – Iss. 1. – P. 1–76. – DOI 10.1016/S0010-0277(98)00034-1.
5. Gibson E. and Thomas J. Memory limitations and structural forgetting: the perception of complex ungrammatical sentences as grammatical // Language and Cognitive Processes. – 1999. – Vol. 14(3). – P. 225–248.
6. Hashemnezhad H., Khalili Zangalani S. Input Processing and Processing Instruction: Definitions and Issues // International Journal of Applied Linguistics & English Literature. – 2013. – Vol. 2. – P. 23–27. – DOI 10.7575/ijalel.v.2n.1p.23.
7. Karlsson F. Constraints on multiple center-embedding of clauses // Journal of Linguistics. – 2007. – Vol. 43(2). – P. 365–392. – DOI 10.1017/S0022226707004616.
8. Kimball J. Seven principles of surface structure parsing in natural language // Cognition. – 1973. – Vol. 2(1). – P. 15–47. – DOI 10.1016/0010-0277(72)90028-5.
9. Korzen I. Are some languages more complex than others? On text complexity and how to measure it // Globe: A Journal of Language, Culture and Communication. – 2021. – Vol. 12. – P. 18–31.
10. Lines N. The Past, Problems, and Potential of Readability Analysis // CHANCE. – 2022. – Vol. 35(2). – P. 16–24. – DOI 10.1080/09332480.2022.2066411.

Статья поступила в редакцию 22.11.2023

REFERENCES

1. Solnyshkina M. I., Kazachkova M. B., Har'kova E. V. Cifrovye tekhnologii izmereniya slozhnosti tekstov kak instrument upravleniya kachestvom obucheniya chteniyu na anglijskom yazyke // Inostrannye yazyki v shkole. – 2020. – № 3. – S. 15–21.
2. Chomsky N., Miller G. A. Introduction to the formal analysis of natural languages // Journal of Symbolic Logic. – 1963. – Vol. 33(2). – P. 299–300. – DOI 10.2307/2269904.
3. Chou Ch. L. Chinese speakers' acquisition of English conditionals: acquisition order and L1 transfer effects // Second Language Studies. – 2000. – Vol. 19(1). – P. 57–98.
4. Gibson E. Linguistic complexity: locality of syntactic dependencies // Cognition. – 1998. – Vol. 68. – Iss. 1. – P. 1–76. – DOI 10.1016/S0010-0277(98)00034-1.
5. Gibson E. and Thomas J. Memory limitations and structural forgetting: the perception of complex ungrammatical sentences as grammatical // Language and Cognitive Processes. – 1999. – Vol. 14(3). – P. 225–248.
6. Hashemnezhad H., Khalili Zangalani S. Input Processing and Processing Instruction: Definitions and Issues // International Journal of Applied Linguistics & English Literature. – 2013. – Vol. 2. – P. 23–27. – DOI 10.7575/ijalel.v.2n.1p.23.
7. Karlsson F. Constraints on multiple center-embedding of clauses // Journal of Linguistics. – 2007. – Vol. 43(2). – P. 365–392. – DOI 10.1017/S0022226707004616.
8. Kimball J. Seven principles of surface structure parsing in natural language // Cognition. – 1973. – Vol. 2(1). – P. 15–47. – DOI 10.1016/0010-0277(72)90028-5.
9. Korzen I. Are some languages more complex than others? On text complexity and how to measure it // Globe: A Journal of Language, Culture and Communication. – 2021. – Vol. 12. – P. 18–31.
10. Lines N. The Past, Problems, and Potential of Readability Analysis // CHANCE. – 2022. – Vol. 35(2). – P. 16–24. – DOI 10.1080/09332480.2022.2066411.

The article was contributed on November 22, 2023

Сведения об авторах

Габитов Азат Ирекович – старший преподаватель кафедры языковой и межкультурной коммуникации Казанского (Приволжского) федерального университета, г. Казань, Россия, <https://orcid.org/0000-0002-7995-5795>, gabit.azat@gmail.com

Габитова Лилия Хабировна – старший преподаватель кафедры языковой и межкультурной коммуникации Казанского (Приволжского) федерального университета, г. Казань, Россия, <https://orcid.org/0000-0002-7738-5901>, corbita@list.ru

Шаяхметова Лейсан Хабировна – кандидат филологических наук, доцент кафедры языковой и межкультурной коммуникации Казанского (Приволжского) федерального университета, г. Казань, Россия, <https://orcid.org/0000-0003-3166-5219>, habirshah@mail.ru

Author Information

Gabitov, Azat Irekovich – Senior Lecturer of the Department of Linguistic and Intercultural Communication, Kazan (Volga Region) Federal University, Kazan, Russia, <https://orcid.org/0000-0002-7995-5795>, gabit.azat@gmail.com

Gabitova, Liliia Khabirovna – Senior Lecturer of the Department of Linguistic and Intercultural Communication, Kazan (Volga Region) Federal university, Kazan, Russia, <https://orcid.org/0000-0002-7738-5901>, corbita@list.ru

Shayakhmetova, Leysan Khabirovna – Candidate of Philology, Associate Professor of the Department of Linguistic and Intercultural Communication, Kazan (Volga Region) Federal University, Kazan, Russia, <https://orcid.org/0000-0003-3166-5219>, habirshah@mail.ru