

Методология

АНАЛИЗ КАЧЕСТВА КВАНТОВАНИЯ УЧЕБНОГО ТЕКСТА

Александр Рыбанов,

Волжский политехнический институт (филиал)
Волгоградского государственного технического университета
rybanoff@yandex.ru

В статье рассматриваются вопросы анализа качества квантованного представления учебной информации на основе квантитативных параметров текста: средних частотах частей речи, используемых в тексте; индексов формальной удобочитаемости текста; коэффициентов лексического и синтаксического разнообразия текста. Процесс получения значений квантитативных параметров ориентирован на использовании библиотеки морфологического анализа phpMorphy.

Ключевые слова: квантование учебных текстов, учебный контент, квантитативные характеристики, индекс формальной удобочитаемости

Введение

Инструментальные средства разработки учебного контента отстают в развитии от систем дистанционного обучения¹ (СДО). В свою очередь, успешность применения СДО зависит от качества и эффективной организации учебного контента.

1

Рыбанов А.А.
Оценка качества текстов
электронных средств
обучения // Школьные
технологии. 2011. № 6.
С. 172–174.

2

Кащева А.В.
Квантитативные и каче-
ственные методы иссле-
дования в прикладной
лингвистике // Соци-
ально-экономические
явления и процессы.
2013. № 3 (049).
С. 155–162.

3

Журавлёв А.Ф.
Опыт квантитативно-ти-
пологического исследо-
вания разновидностей
устной речи // Разно-
видности городской уст-
ной речи. Сборник науч-
ных трудов. М.: Наука,
1988. С. 84–150.

4

Верхолин С.С.
К вопросу о лингвотео-
ретических основах ме-
тодик авторизации тек-
ста // Ученые записки
Забайкальского государ-
ственного университета.
Серия: Филология, ис-
тория, востоковедение.
2013. № 2 (49). С. 22–27.

5

Дудиков М.Ю.
Квантитативные характе-
ристики профессиональ-
ной коммуникации //

Существующие в настоя-
щее время СДО Moodle, Ilias,
Claroline, Atutor и др. не предо-
ставляют разработчикам дис-
танционных учебных курсов
возможность оценки качества
учебного контента. Между тем,
оценка учебного контента на-
правлена на выявление досто-
инств и недостатков учебной
информации и на принятие ре-
шения о необходимости, а так-
же оптимальных условиях его
использования в процессе дис-
танционного обучения. Одним
из направлений решения зада-
чи оценки качества учебного
контента систем дистанцио-
нного обучения является ис-
пользование методов квантита-
тивной лингвистики.

Постановка задачи

Квантитативная лингвистика
(quantitative linguistics) — одно
из направлений прикладной
лингвистики, которое занима-
ется изучением языка с помо-
щью статистических методов².
Преимуществом квантитатив-
ных методов изучения текстов
является их точность и одно-
значность результатов. Расчёт
квантитативных характерис-
тик текста необходим для ре-
шения следующих проблем:

- определение стилевых и жан-
ровых характеристик текстов,
с целью последующей их клас-
сификации³;

- изучение образцов текстов, с
целью установление авторства⁴;
- обучение языку специально-
сти^{5,6}.

Разработка учебного кон-
тента СДО включает в себя
развитие технологий проекти-
рования контента, таких как
квантование⁷ учебной инфор-
мации. Квантование — это раз-
деление учебной информации
на элементарные фрагменты
(учебные единицы, шаги, кад-
ры) различного назначения:
информационные, тренирую-
щие, контролируемые, управ-
ляющие. Одной из проблем
технологии проектирования
контента является формирова-
ние системы количественных
критериев для оценки качества
квантования учебной инфор-
мации⁸. Квантитативные ха-
рактеристики текста могут со-
ставлять основу данной систе-
мы критериев. Wiio O.A.⁹ пред-
лагает использовать квантита-
тивные характеристики для
оценки показателя сложности
текста: чем больше прилага-
тельных и наречий в тексте,
тем сложнее текст. Глагол —
самая живая часть речи. Частое
применение глаголов в спряга-
емых формах приводит к тому,
что предложения легко запо-
минаются и понимаются. В та-
ких предложениях связанные
слова находятся близко друг к
другу и их связи легко осозна-
ются. Глаголы способствуют
пониманию текста¹⁰.

Актуальной является задача автоматизированного подсчёта значений количественных характеристик текста, с целью их последующего использования для оценки качества квантования учебного контента. Программная реализация автоматизированного определения ряда количественных характеристик текста возможна на основе библиотеки морфологического анализа *phpMorphy*, реализованной на платформе РНР. Библиотека *phpMorphy* направлена на решение следующих задач¹¹:

- лемматизация (получение нормальной формы слова);
- получение всех форм слова;
- получение полуграмматической информации для слова (часть речи, падеж, спряжение и т.д.);
- изменение формы слова в соответствии с заданными грамматическими характеристиками;
- изменение формы слова по заданному образцу.

Функции данной библиотеки могут быть эффективно использованы при автоматизации процесса анализа качества квантования учебного контента.

Количественные характеристики учебного контента

Среди существующих количественных характеристик текста рассмотрим следующие:

- количественные характеристики употребления частей речи;
- количественные характеристики удобочитаемости текста;
- количественные характеристики разнообразия текста.

Библиотека *phpMorphy* поддерживает обработку текста на следующих языках: русский, английский, немецкий. С помощью библиотеки *phpMorphy* могут быть определены следующие низкочастотные количественные характеристики текста, рассчитываемые на основе сведений о средних частотах частей речи, используемых в тексте:

1) *индекс аналитичности (analyticity index)* — отношение числа служебных слов к общему числу слов в тексте;

2) *индекс глагольности (verb index)* — отношение числа глаголов к числу слов в тексте;

3) *индекс субстантивности (substantive index)* — отношение числа существительных к числу слов в тексте;

4) *индекс адъективности (adjective index)* — отношение числа прилагательных к числу слов в тексте;

5) *индекс местоименности (pronoun index)* — отношение числа местоименных слов к числу слов в тексте;

6) *индекс автосемантической (autosemanticity index)* — отношение числа значащих слов;

Методология

Вестник Челябинского государственного университета. 2009. № 35. С. 63–67.

6

Агеев В.В.,
Сергеев В.М.,
Яковлева Е.И.

Средства оптимизации лингводидактики для сокращенных форм обучения // Вестник Нижегородского университета им. Н.И. Лобачевского. 2011. № 3–1. С. 37–43.

7

Аванесов В.С.

Применение заданий в тестовой форме и квантованных учебных текстов в новых образовательных технологиях // Педагогические измерения. 2012. № 2. С. 75–91.

8

Рыбанов А.А.

Количественные метрики для оценки качества квантования учебной информации // Педагогические измерения. 2013. № 4. С. 3–12.

9

Wii O.A.

Readability. Compression and Readership. Acta Universitatis Tamperensis, 1968, vol. 22 (A), p. 161.

10

Flesh R.

The Art of Plain Talk.- New York: Harper and

7) *индекс незнаменательности (unmomentous words index)* — отношение числа незнаменательных слов к числу слов в тексте;

8) *индекс именной лексики (nominal lexicon index)* — отношение суммы чисел существительных и прилагательных к числу слов в тексте.

Обозначения частей речи в библиотеке phpMorphy представлены в табл. 1.

Низкоуровневые количественные характеристики текста могут быть выражены через обозначения частей речи библиотеки phpMorphy следующим образом (COUNT_WORDS — количество слов в тексте):

1) *индекс аналитичности:*

$$\text{Analyticity_index} = (\text{PMY_RP_PREP} + \text{PMY_RP_CONJ} + \text{PMY_RP_PARTICLE}) / \text{COUNT_WORDS};$$

2) *индекс глагольности:*

$$\text{Verb_index} = (\text{PMY_RP_INFINITIVE} + \text{PMY_RP_VERB} + \text{PMY_RP_ADVERB_PARTICIPLE} + \text{PMY_RP_PARTICIPLE_SHORT}) / \text{COUNT_WORDS};$$

Таблица 1

Обозначения частей речи в библиотеке phpMorphy

Константа	Описание
PMY_RP_NOUN	существительное
PMY_RP_ADJ_FULL	прилагательное
PMY_RP_ADJ_SHORT	краткое прилагательное
PMY_RP_INFINITIVE	инфинитив
PMY_RP_VERB	глагол в личной форме
PMY_RP_ADVERB_PARTICIPLE	деепричастие
PMY_RP_PARTICIPLE	причастие
PMY_RP_PARTICIPLE_SHORT	краткое причастие
PMY_RP_NUMERAL	числительное (количественное)
PMY_RP_NUMERAL_P	порядковое числительное
PMY_RP_PRONOUN	местоимение-существительное
PMY_RP_PRONOUN_PREDK	местоимение-предикатив
PMY_RP_PRONOUN_P	местоименное прилагательное
PMY_RP_ADV	наречие
PMY_RP_PREDK	предикатив
PMY_RP_PREP	предлог
PMY_RP_CONJ	союз
PMY_RP_INTERJ	междометие
PMY_RP_PARTICLE	частица
PMY_RP_INP	вводное слово
PMY_RP_PHRASE	фразеологизм

3) индекс субстантивности:

$$\text{Substantive_index} = \text{PMY_RP_NOUN} / \text{COUNT_WORDS};$$

4) индекс адъективности:

$$\text{Adjective_index} = (\text{PMY_RP_ADJ_FULL} + \text{PMY_RP_ADJ_SHORT}) / \text{COUNT_WORDS};$$

5) индекс местоименности:

$$\text{Pronoun_index} = (\text{PMY_RP_PRONOUN} + \text{PMY_RP_PRONOUN_PREDK} + \text{PMY_RP_PRONOUN_P}) / \text{COUNT_WORDS};$$

6) индекс автосемантической:

$$\text{Autosemanticity_index} = 1 - \text{Unmomentous_words_index};$$

7) индекс незначительности:

$$\text{Unmomentous_words_index} = ((\text{PMY_RP_PREP} + \text{PMY_RP_CONJ} + \text{PMY_RP_PARTICLE}) + (\text{PMY_RP_PRONOUN} + \text{PMY_RP_PRONOUN_PREDK} + \text{PMY_RP_PRONOUN_P})) / \text{COUNT_WORDS};$$

8) индекс именной лексики:

$$\text{Nominal_lexicon_index} = (\text{PMY_RP_NOUN} + \text{PMY_RP_ADJ_FULL} + \text{PMY_RP_ADJ_SHORT}) / \text{COUNT_WORDS}.$$

Среди количественных характеристик удобочитаемости текста можно выделить *среднюю длину слова в слогах* и *среднюю длину предложения в словах*.

Количественные характеристики разнообразия текста описываются коэффициентами лексического и синтаксического разнообразия. Поскольку коэффициент — величина не абсолютная, а относительная (в определенном промежутке величин), можно в известных границах пренебречь длиной сопоставляемых текстов. Теоретический интерес представляет и исследование внутренней «динамики» текста с точки зрения сопоставления коэффициентов на разных его участках и в их отношении к общему для всего текста коэффициенту.

Коэффициент лексического разнообразия формируется из отношения числа лексем к общему числу слов текста:

$$K_{\text{лекс}} = L/C, \quad (1)$$

где $K_{\text{лекс}}$ — коэффициент лексического разнообразия; C — число слов в тексте (единицы между пробелами). Чем выше значение $K_{\text{лекс}}$, тем выше лексическое разнообразие текста.

Коэффициент синтаксического разнообразия вытекает из отношения числа предложений к числу слов данного текста:

$$K_{\text{синт}} = 1 - П/С, \quad (2)$$

где $K_{\text{синт}}$ — коэффициент синтаксического разнообразия; $П$ — число предложений; $С$ — число слов в тексте. Чем боль-

ше значение П/С, тем многочисленнее в целом предложения данного текста, а следовательно, выше возможность разнообразия синтаксических отношений между словами в составе отдельного предложения.

Измерение квантитативных характеристик учебного контента

При автоматической обработке текста возможна ситуация, когда для словоформы функция определения части речи возвращает несколько значений. Например, в библиотеке `phpMorphy` функция `getPartOfSpeech` возвращает для слова 'PROGRAM' массив со значениями частей речи:

```
var_dump($morphy->getPartOfSpeech('PROGRAM')); // array('СУЩЕСТВИТЕЛЬНОЕ', 'ПРИЛАГАТЕЛЬНОЕ', 'ГЛАГОЛ')
```

Поэтому значение каждой квантитативной характеристики текста должно описываться величиной погрешности её расчёта.

Введем следующие обозначения для процесса автоматического определения количества слов в тексте T , относящихся к части речи k :

- 1) η_k – количество однозначных определений части речи k ;
- 2) μ_k – количество неоднозначных определений части речи k .

3) θ_k – количество слов части речи k в тексте T .

Распределение вероятностей частей речи в тексте T неизвестно, поэтому, согласно *принципу недостаточного основания Лапласа*, при автоматическом распознавании частей речи нет причин считать их различными.

Основываясь на принципе недостаточного основания, полагаем, что

$$\eta_k + \Delta_k \leq \theta_k \leq \eta_k + \mu_k - \Delta_k.$$

Исходя из этого примем, что

$$\theta_k = \eta_k + \frac{\mu_k}{2},$$

тогда абсолютная погрешность Δ_k при автоматическом определении части речи k :

$$\Delta_k = \frac{\mu_k}{2},$$

а относительная погрешность δ_k при автоматическом определении части речи k :

$$\delta_k = \frac{\Delta_k}{\theta_k} \cdot 100 \% = \frac{\mu_k}{2\eta_k + \mu_k} \cdot 100 \%.$$

Проведем анализ качества квантования учебной информации на примере произведения А.П. Чехова «Белолобый»¹². Введем следующие обозначения для представления текста произведения А.П. Чехова «Белолобый»: T – исходный текст, T' – квантованный текст.

12
Верещик И.
Квантование текста и
разработка заданий в
тестовой форме (на при-
мере произведения
А.П. Чехова) // Педаго-
гические измерения.
2012. № 1. С. 98–105.

Результаты определения частей речи в исходном и квантованном текстах представлены в табл. 2 и 3.

Таблица 2

Результаты определения частей речи слов в исходном тексте T

Часть речи	η_k	μ_k	θ_k	Δ_k	δ_k
PMY_RP_NOUN	429	138	498	69	13,855
PMY_RP_ADJ_FULL	102	40	122	20	16,393
PMY_RP_ADJ_SHORT	5	66	38	33	86,842
PMY_RP_INFINITIVE	36	5	38,5	2,5	6,494
PMY_RP_VERB	285	59	314,5	29,5	9,380
PMY_RP_ADVERB_PARTICIPLE	37	4	39	2	5,128
PMY_RP_PARTICIPLE	15	4	17	2	11,765
PMY_RP_PARTICIPLE_SHORT	3	2	4	1	25
PMY_RP_NUMERAL	9	9	13,5	4,5	33,333
PMY_RP_NUMERAL_P	0	3	1,5	1,5	100
PMY_RP_PRONOUN	97	98	146	49	33,562
PMY_RP_PRONOUN_PREDK	0	0	0	–	–
PMY_RP_PRONOUN_P	28	69	62,5	34,5	55,200
PMY_RP_ADV	44	219	153,5	109,5	71,336
PMY_RP_PREDK	0	32	16	16	100
PMY_RP_PREP	203	35	220,5	17,5	7,937
PMY_RP_CONJ	1	254	128	127	99,219
PMY_RP_INTERJ	0	170	85	85	100
PMY_RP_PARTICLE	28	120	88	60	68,182
PMY_RP_INP	0	4	2	2	100
PMY_RP_PHRASE	0	4	2	2	100

На основе значений Δ_k и δ_k определим погрешности автоматического получения значения квантитативной характеристики β для текста T:

- абсолютная погрешность $\Delta\beta$;
- относительная погрешность $\delta\beta$;

$$\Delta\beta = \frac{1}{2 \cdot W} \sum_{i \in P} \mu_i;$$

$$\delta\beta = \frac{\sum_{i \in P} \mu_i}{2 \cdot \sum_{i \in P} \eta_i + \sum_{i \in P} \mu_i},$$

где P – множество частей речи, использующихся при определении квантитативной характеристики β .

Например, погрешности для индекса адъективности определяются как :

$$\begin{aligned} \delta_{\text{Adjective_index}} &= \\ &= \frac{\mu_{\text{PMY_RP_ADJ_FULL}} + \mu_{\text{PMY_RP_ADJ_SHORT}}}{2 \cdot W} \end{aligned}$$

$$\delta_{\text{Adjective_index}} = \frac{\mu_{\text{PMY_RP_ADJ_FULL}} + 2 \cdot (\eta_{\text{PMY_RP_ADJ_FULL}} + \eta_{\text{PMY_RP_ADJ_SHORT}}) + \mu_{\text{PMY_RP_ADJ_SHORT}}}{\mu_{\text{PMY_RP_ADJ_FULL}} + \mu_{\text{PMY_RP_ADJ_SHORT}}}$$

Таблица 3

**Результаты определения частей речи слов
в квантованном тексте T'**

Часть речи	η_k	μ_k	θ_k	Δ_k	δ_k
PMY_RP_NOUN	216	61	246,5	30,5	12,373
PMY_RP_ADJ_FULL	46	16	54	8	14,815
PMY_RP_ADJ_SHORT	3	33	19,5	16,5	84,615
PMY_RP_INFINITIVE	20	2	21	1	4,762
PMY_RP_VERB	140	32	156	16	10,256
PMY_RP_ADVERB_PARTICIPLE	14	1	14,5	0,5	3,448
PMY_RP_PARTICIPLE	8	2	9	1	11,111
PMY_RP_PARTICIPLE_SHORT	0	0	0	–	–
PMY_RP_NUMERAL	5	4	7	2	28,571
PMY_RP_NUMERAL_P	0	1	0,5	0,5	100
PMY_RP_PRONOUN	43	54	70	27	38,571
PMY_RP_PRONOUN_PREDK	0	0	0	–	–
PMY_RP_PRONOUN_P	16	37	34,5	18,5	53,623
PMY_RP_ADV	29	102	80	51	63,750
PMY_RP_PREDK	0	14	7	7	100
PMY_RP_PREP	103	22	114	11	9,649
PMY_RP_CONJ	0	126	63	63	100
PMY_RP_INTERJ	0	83	41,5	41,5	100
PMY_RP_PARTICLE	15	64	47	32	68,085
PMY_RP_INP	0	2	1	1	100
PMY_RP_PHRASE	0	1	0,5	0,5	100

Погрешности вычисления квантитативных характеристик исходного и квантованного текстов представлены в табл. 4.

Погрешности Δ_β , Δ_k , могут быть использованы для сравнительного анализа программ автоматической обработки текста по точности определения

частей речи слов и квантитативных характеристик.

Квантитативные характеристики удобочитаемости исходного и квантованного текстов приведены в табл. 5. Данные характеристики необходимы для определения индекса формальной удобочитаемости текста.

Таблица 4

Погрешности вычисления количественных характеристик употребления частей речи

Количественная характеристика	Исходный текст T			Квантованный текст T'		
	Значение	Δ_{β}	δ_{β}	Значение	Δ_{β}	δ_{β}
Индекс аналитичности	0,229	0,107	46,849	0,236	0,112	47,321
Индекс глагольности	0,216	0,019	8,959	0,211	0,019	9,227
Индекс субстантивности	0,261	0,036	13,855	0,259	0,032	12,373
Индекс адъективности	0,084	0,028	33,125	0,077	0,026	33,333
Индекс местоименности	0,109	0,044	40,048	0,11	0,048	43,541
Индекс автосемантической	0,662	0,151	22,809	0,654	0,16	24,312
Индекс незначительности	0,338	0,151	44,651	0,346	0,159	46,119
Индекс именной лексики	0,345	0,064	18,541	0,377	0,058	17,188

Таблица 5

Количественные характеристики удобочитаемости исходного и квантованного текстов

Количественная характеристика	Исходный текст T	Квантованный текст T'
Средняя длина слова в слогах	2,052	2,023
Средняя длина предложения в словах	14,264	12,614

Количественные характеристики исходного и квантованного текстов при разнообразии исходного текста приведены в табл. 6.

Таблица 6

Количественные характеристики разнообразия исходного и квантованного текстов

Количественная характеристика	Исходный текст T	Квантованный текст T'
Коэффициент лексического разнообразия	0,306	0,355
Коэффициент синтаксического разнообразия	0,944	0,94

Лексическое разнообразие характеризует информационную насыщенность текста. Для квантованного текста, по сравнению с исходным, характерно уменьшение степени повторяемости словоформ. И, следовательно, коэффициент лексического разнообразия для квантованного текста несколько выше, чем для исходного (рис. 1).

Синтаксическое разнообразие проявляется в использовании разнообразных синтаксических средств — при квантовании коэффициент синтаксического разнообразия уменьшается. На рис. 2 для фрагментов квантованного текста №2 и №3 коэффициент синтаксического разнообразия квантованного текста больше, чем для исходного, что свидетельствует о необходимости повторной процедуры квантования для данных фрагментов.

Индекс формальной удобочитаемости учебного контента

Следует различать формальную удобочитаемость текста $R_{form}(I)$, являющейся функцией только от параметров самого учебного контента I , и его индивидуальную удобочитаемость $R_{ind}(I,u)$, которая зависит как от характеристик учебного контента I , так и от свойств читателя u . Для количественной оценки формальной удобочитаемости можно использовать индекс Ю. Тулдава¹³ или Флеша¹⁴.

Индекс Ю. Тулдава рассчитывается по формуле:

$$R(\bar{i}, \bar{j}) = \bar{i} \cdot \lg \bar{j}, \quad (3)$$

где $R(\bar{i}, \bar{j})$ — индекс формальной удобочитаемости (рис. 3), \bar{i} — средняя длина слова в словах, \bar{j} — средняя длина предло-

¹³ Тулдава Ю.А. Об измерении трудности текста. // Ученые записки Тартуского государственного университета, 1975, вып. 345. С. 102–119.

¹⁴ Rudolf Flesch. A new readability yardstick, Journal of Applied Psychology, 32, 1948. С. 221–233.

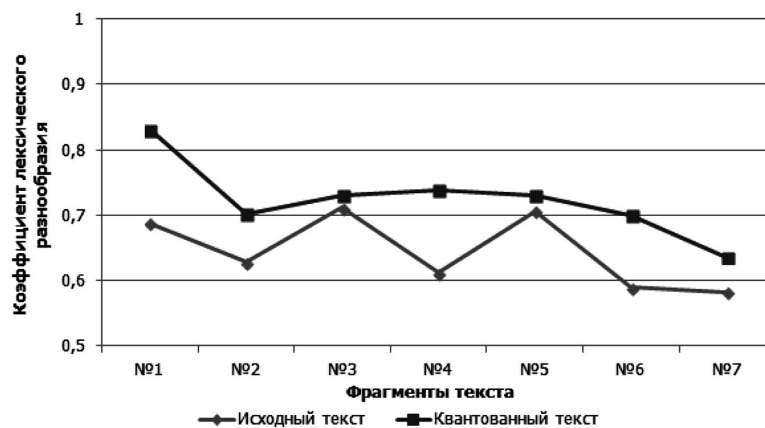


Рис. 1. Сравнительный анализ коэффициента лексического разнообразия по фрагментам текстов

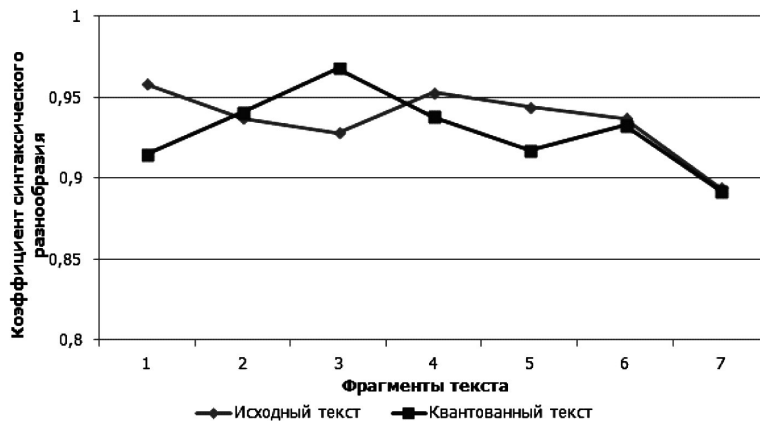


Рис. 2. Сравнительный анализ коэффициента синтаксического разнообразия по фрагментам текстов

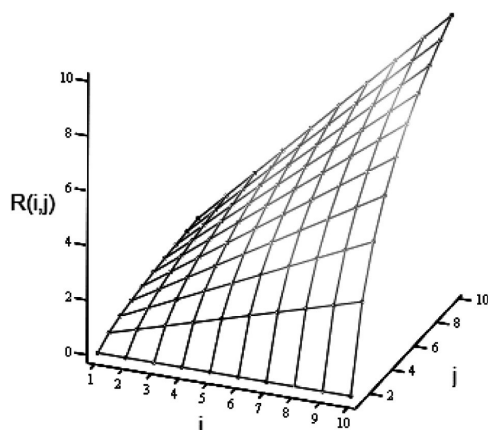


Рис. 3. Вид функции

жений в словах. Формула (3) разработана на основе закономерности, наблюдаемой в разных языках. Поэтому формула Ю. Тулдава предназначена для анализа текста на разных языках.

Индекс Флеша рассчитывается по формуле:

$$Fr(\bar{i}, \bar{j}) = 206.835 - \alpha_1 \bar{j} - \alpha_2 \bar{i}, (4)$$

где α_1, α_2 — коэффициенты, зависящие от языка (для английского языка — $\alpha_1 = 1,015, \alpha_2 = 84,6$ для русского языка — $\alpha_1 = 1,3, \alpha_2 = 60,1$). Соответствия между значениями индекса Флеша и лингвистическими переменными «уровень удобочитаемости» и «уровень образования» приведены в табл. 7.

Таблица 7

Лингвистические переменные «уровень удобочитаемости» и «уровень образования» для индекса Флеша $Fr(\bar{i}, \bar{j})$

Значение индекса Флеша $Fr(\bar{i}, \bar{j})$	Уровень удобочитаемости	Уровень образования
90–100	очень высокий	5 классов
80–90	высокий	6 классов
70–80	выше среднего	7 классов
60–70	средний	8–9 классов
50–60	ниже среднего	10–12 классов
30–50	низкий	вуз
0–30	очень низкий	выпускник вуза

Проанализируем, как изменилась формальная удобочитаемость квантованного текста по сравнению с исходным. В табл. 8 приведены значения индексов формальной удобочитаемости $R(\bar{i}, \bar{j})$ и $Fr(\bar{i}, \bar{j})$ для соответствующих фрагментов исходного и квантованного текста.

Индекс формальной удобочитаемости для квантованного текста равен 2,227, а для исходного текста равен 2,368, что свидетельствует о лучшем пред-

ставлении квантованного текста. В то же время сравнительный анализ индексов $R(\bar{i}, \bar{j})$ для исходного и квантованного текстов по фрагментам (рис. 4) позволил определить, что фрагменты квантованного текста №3, №7 требуют дальнейшего преобразования.

Аналогичная ситуация наблюдается для индекса Флеша: для квантованного текста $Fr(\bar{i}, \bar{j})$ равен 68,855, а для исходного текста $Fr(\bar{i}, \bar{j})$ равен 64,966, что также свидетельствует о

Таблица 8

Индексы для фрагментов исходного и квантованного текстов

Фрагмент текста	$R(\bar{i}, \bar{j})$		$Fr(\bar{i}, \bar{j})$	
	Исходный текст	Квантованный текст	Исходный текст	Квантованный текст
Фрагмент №1	2,878	2,307	50,199	62,127
Фрагмент №2	2,392	2,360	66,516	69,305
Фрагмент №3	2,442	3,163	60,350	39,007
Фрагмент №4	2,717	2,318	56,182	70,342
Фрагмент №5	2,688	2,118	54,572	73,261
Фрагмент №6	2,462	2,176	62,777	76,150
Фрагмент №7	1,973	2,002	72,860	70,200

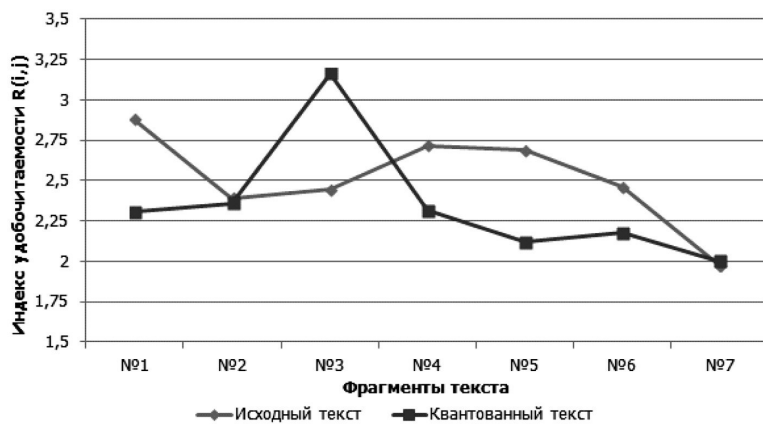


Рис. 4. Сравнительный анализ индекса удобочитаемости $R(i,j)$ по фрагментам текстов

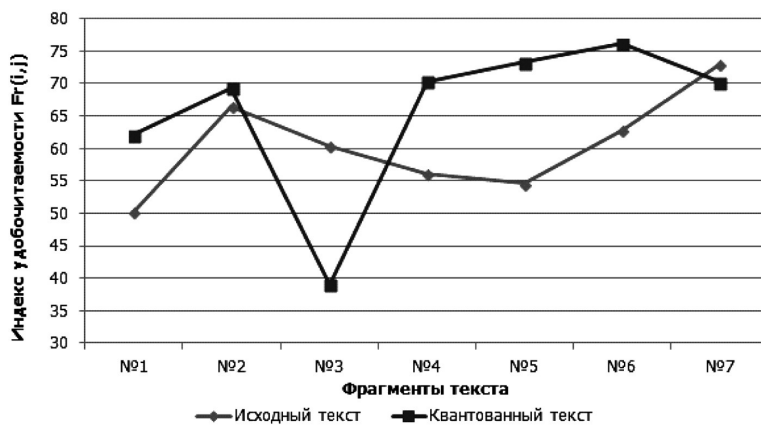


Рис. 5. Сравнительный анализ индекса удобочитаемости $Fr(i,j)$ по фрагментам текстов

лучшем представлении квантованного текста. В то же время, сравнительный анализ индексов сложности для исходного и квантованного текстов по фрагментам (рис. 5) позволил определить, что фрагменты квантованного текста №3, №7 тре-

буют дальнейшего преобразования.

Таким образом, коэффициент синтаксического разнообразия и индекс формальной удобочитаемости показывают, что фрагменты квантованного текста №3 и №7 требуют повторной процедуры квантования.

Заключение

Предлагаемая система количественных характеристик (1–4) учебного контента применима к слабоструктурированным текстам. Данная система критериев неприменима к формулам, таблицам, графическим и

мультимедийным объектам. Учитывая, что данные объекты, как правило, не подлежат количественному, система количественных характеристик (1–4) может успешно применяться в составе автоматизированных систем проектирования учебного контента.