Fluctuations of text complexity: the case of Basic State Examination in English

Diana Gizatulina[1*]*, Farida* Ismaeva[1]*, Marina* Solnyshkina[1]*, Ekaterina* Martynova[1], and *Iskander* Yarmakeev[1]

[1]Kazan Federal University 18 Kremlyovskaya str, Kazan, 420008 Russian Federation

Abstract. Text complexity as a research problem is equally relevant in linguistics and education since its solution provides an algorithm to match readers of certain categories to texts. Numerous studies have been conducted to identify quantitative and qualitative parameters that affect text complexity in ESOL. However, text complexity range within one proficiency level remains a research niche. The current study is aimed at identifying the range of text complexity fluctuations within one proficiency level and their appropriateness for readers. We conduct a multi-factor analysis and contrast 66 English texts for Basic State Examination (OGE) in readability, average sentence length, word length, number of verbs and nouns, cohesion and lexical diversity. The text features computed with the two online services, Text Inspector and Coh-Metrix, provide slightly different quantitative results though consistent in qualifying the range of text complexity fluctuations as high. The research findings refute the hypothesis of a linear nature of text complexity growth in the textbooks designed to increase students' proficiency and confirm the lack of correlation between the revealed and claimed complexity of texts. The algorithm suggested by the authors can be useful for textbook writers and test developers selecting reading material for any proficiency level.

1 Introduction

The problem of reading comprehension has been a focus of numerous studies in foreign language learning as it is viewed to be of high importance for teachers, textbook writers, exam materials developers and students [1]. It dates back to the middle of the 20th century when globalization of education followed by academic mobility triggered demand for objective language assessment of students and learning materials. Over years the attention directed to the problem reflected an ever-increasing interest of researchers in defining skills and abilities of students of every level. Yet, the texts used in testing received much less attention and were in the majority of cases provided with no more than the so-called lexical minimums [2]. Neither morphological, nor lexical features of texts employed to examine test takers of different proficiency levels have been validated and recommended to textbook writers. As it is

* Corresponding author: giz-diana@mail.ru

demonstrated in this paper, descriptive (readability), morphological and lexical text parameters constitute a type which can be assigned to texts of different proficiency levels.

The modern linguistic paradigm defines a text as a model for relevant use of language and requires its age, cognition and language appropriateness [3]. Hence, the task of selecting a classroom book implies availability of information on readability, language proficiency level and curriculum appropriateness of all the books offered. Nowadays, textbook writers and curriculum developers are also expected to sequence reading texts 'from simple to difficult', avoiding fluctuations and sharp leaps or drops in text readability. In general, text complexity assessment implies labeling a text with a certain educational level, such as a grade, readers' age or a CEFR proficiency level. Unfortunately, the research conducted in the area shows that in many cases textbooks fail to correspond students' cognitive and language abilities [4]. There is no recommended algorithm of text complexity assessment, including explicit guidelines on text features range and fluctuation within one proficiency level, either. Thus, the current study aimed at identifying the range of complexity fluctuations of texts selected to teach one proficiency level, BSE (OGE), and their appropriateness for readers of the claimed proficiency level, fills the research niche. We also offer an algorithm which could be of further use in similar educational contexts. Hence, our *central research question* relates to the nature of a text type used to assess students in BSE (OGE). The *second research question* refers to the character of readability progress in three sources used in the research. The hypothesis behind this part of the study is that texts are to be sequenced appropriately from less to more difficult (readable).

The study was designed and conducted in four stages: 1) Compiling BSE (OGE) Corpus; 2) classification of texts based on the types of reading in BSE (OGE); 3) computing text metrics: average sentence length, average word length, readability (Flash-Kincaid grade), verbs count, nouns count, referential cohesion, lexical diversity; 4) identifying the range of the above mentioned fluctuations influencing text complexity.

2 Material and Method

The research corpus comprises 66 texts used for training secondary school children for Basic State Examination in English (hereafter abbreviated as BSE (OGE)). We retrieved 36 texts from the Open Bank of assignments at FIPI website (Federal Institute for Pedagogical Measurements) [5], 24 texts from the textbook "BSE (OGE) - 2020. Angliiskii yazyk. 30 trenirovochnykh variantov ekzamenatsionnykh rabot dlya podgotovki k OGE" [6] and 6 texts - from "BSE (OGE) 2020. Angliiskii yazyk. Gotovimsya k itogovoi attestatsii" [7]. We mark the texts used as follows: F 1- 34 [5], G 1 – 24 [6], V1 –6 [7]. We use only the materials of Reading 2 in BSE (OGE) which FIPI claims to correspond to A2-B1 CEFR level [5]. While tested in Reading 2, test takers are expected to demonstrate their ability in 'interactive reading' [8], i.e. comprehension of "stretches of language of several paragraphs to one page" followed by a test (open or cloze test, answering questions, recalls etc.). Reliability and fairness of this part of BSE (OGE) is supposedly provided by selecting texts of only one type, i.e. narrative science-fiction or publicist.

The total size of the Corpus collected for the study is over 22000 tokens in 66 texts with the average length of the text being 333 tokens. A corpus of this size is viewed as representative enough to estimate lexical, morphological and descriptive features of one type of texts [9].

Text analysis was conducted with the help of two online tools, i.e. TextInspector [10] and Coh-Metrix (CM) [11]. The Text Inspector (TI) is an online server based on English Vocabulary Profile which enables to process a text of up to 410 words at a time [12]. TI tags each word with its CEFR level and provides statistics on the sentence count, an average sentence length, an average word length, and Type Token Ratio. Coh-Metrix analyzes texts on a wide range of dimensions and employs lexicon profilers, part-of-speech taggers, syntactic

parsers, corpora, latent semantic analysis, etc. Coh-Metrix evaluated the Coh-Metrix L2 Readability index, referential cohesion, narrativity etc [13].

While conducting TextInspector analysis we had to reduce texts to 410 words due to the limitations of the service. The clipped part did not exceed 15% of the original text, hence we maintained the absolute essential vocabulary of 50% required by D. Biber to provide representativeness of the text sample [9].

3 Analysis

Descriptive features, an average sentence length and average word length, were measured with both servers, TI and CM. The results are presented in Fig. 1.
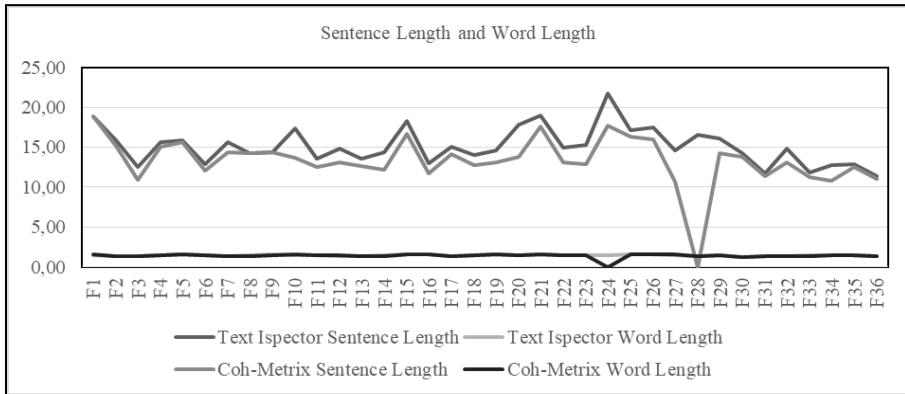

Fig.1. Fluctuations of sentence length and word length parameters.

As we can see, with few exceptions, the tools provide fairly similar metrics of sentence and word length. The average sentence length is 16.1 (TI) or 14.3 (CM). Both tools computed the same word length - 1.44 (TI, CM). We also conclude that the descriptive metrics range is quite wide for the sentence length, i.e. 7.3 – 22.9 with the mean to be 14.4-17.0 words per sentence (See Fig.2). The word length varies from 1.24 to 1.64 syllables per word.
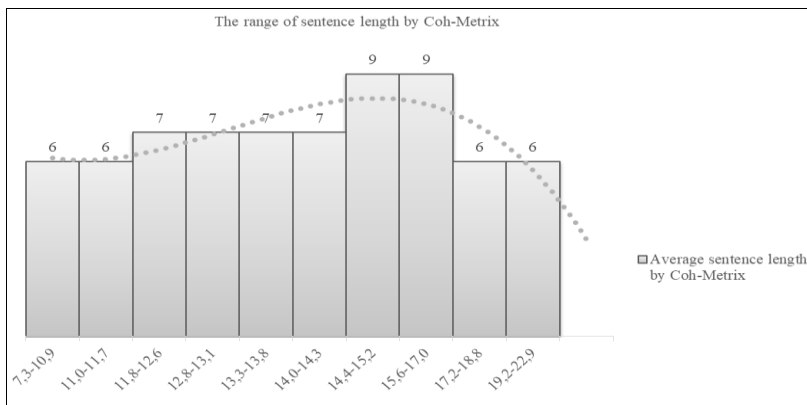

Fig.2. The range of sentence length (CM).

4 Readability

Based on the metrics calculated, i.e. sentence length and word length, the tools assess texts readability. TI computes Flash-Kincaid grade level (FKGL) with the help of the following formula, developed by Flash-Kincaid [14].

$$FKGL = (0,39 * ASL) + (11,8 * ASW) - 15,59, \tag{1}$$

where ASL is an average sentence length; ASW is an average number of syllables per word [14]. The index in this formula corresponds to the USA educational index: the values from 1 to 10 are considered to be appropriate for secondary school students, from 11 to 15 - for higher education; and the values from 16 to 20 correspond to complex scientific texts.

Coh-Metrix L2 Reading Index (CML2 x) is an index defining readability levels for non-native speakers of English [14]. It is measured with the following formula:

$$CML2 = -45.032 + (52.230 \ x \ CWO) + (61.306 \ x \ SSS) + (22.205 \ x \ CELEX), \tag{2}$$

where CWO, i.e. Content Word Overlap, denotes referential cohesion, SSS, i.e. Sentence Syntax Similarity, is "the portion of intersection tree nodes across all adjacent sentences as well as between all combinations across paragraphs [15], CELEX is an average frequency of occurrence of all words.

Flash-Kincaid formula is an indicator of readability that correlates with a specific age of a reader, CML2, however, is viewed by the developers as a feature reflecting cognitive and psycholinguistic abilities required to comprehend a text.

The average readability level for the texts under study is assessed as 7.9 by TI and 7.0 by CM which corresponds to 8 or 7 years of formal schooling respectively. Unfortunately, its lowest and highest points range from 4.31 to 14.63 (TI) and from 3.74 to 10.56 (CM) identifying a much wider readability range that is recommended by researchers [17] (Fig.3). We have to emphasize here that students' reading levels are never limited to one level, and textbooks readability range is supposed to include a student's zone of proximal development where reading texts are neither too difficult nor too easy so that a student is challenged but not frustrated [16]. The zone of proximal development, or the 'actual developmental level as determined by independent problem solving' and the 'potential development as determined through problem solving under adult guidance or in collaboration with more capable peers' [17] is extremely important as it provides students with a motivational context to study and develop. In our case the readability range is too wide to characterize the books under study as matching to potential readers. Similarly, CML2 average is 17.2, with the range between the lowest (6.29) and highest (27.69) levels is over 20 (Fig.3).
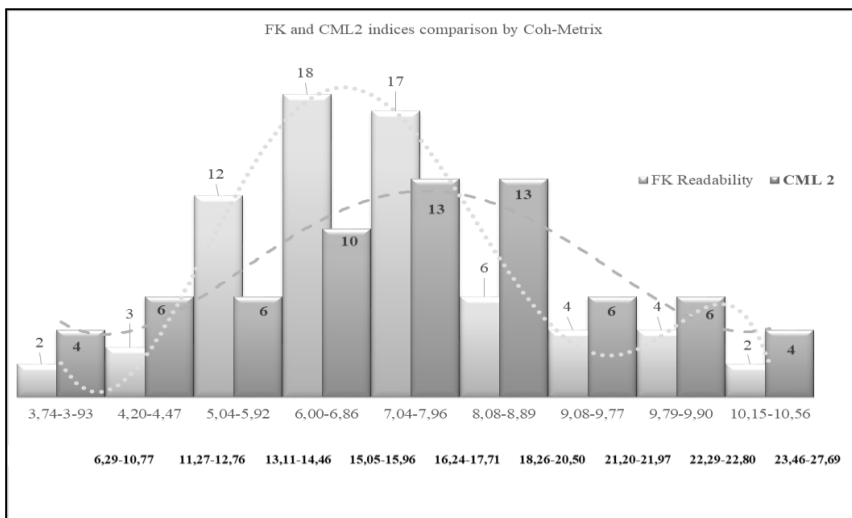
Fig. 3. FK and CML2 indices comparison by Coh-Metrix.

We also tested texts readability sequencing measured with FKGL and CML2 in texts G1 – G24 [6] processed with CM (see Fig.4). The range of FKGL of the texts under study fluctuates from 3.74 in text G6 to 10.15 in G17 which makes the range wider than any natural zone of proximal development (Fig.4) [16].
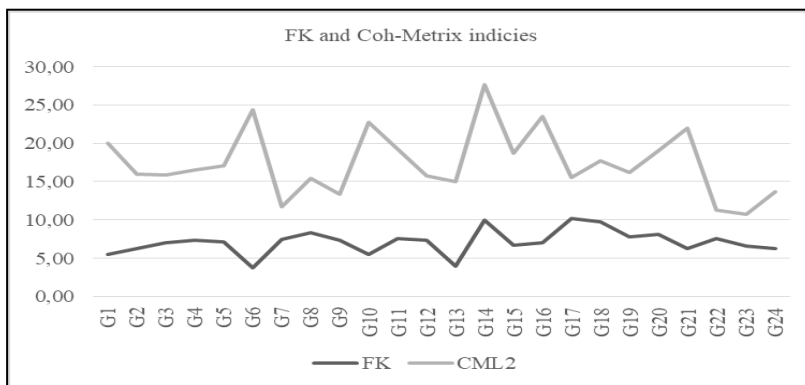

Fig.4. Text readability sequence in "BSE (OGE) – 2020".

The figure testifies to the texts being potentially demotivating for students as readability does not increase gradually but fluctuates randomly.

5 Morphological and lexical features

Verbs count and nouns count were estimated with TextInspector. The ratio of verbs to nouns (VNR) was calculated manually with the formula:

$$VNR = V{:}N * 100, \qquad (3)$$

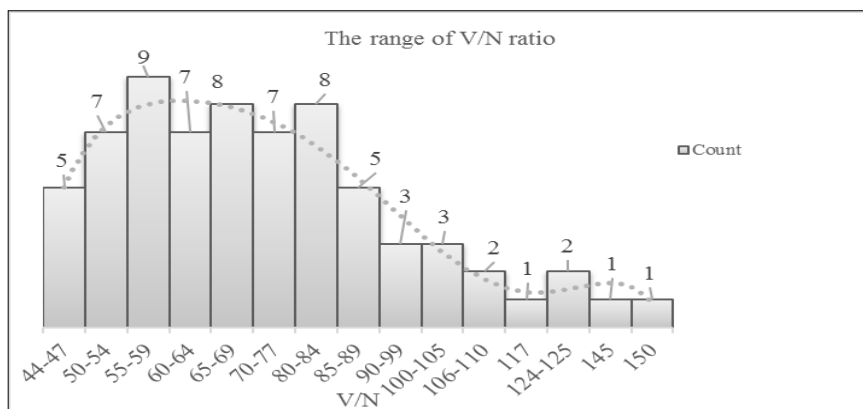where V stands for Verbs, N for Nouns, (see Fig. 5).

Fig.5. The range of V/N ratio.

The results demonstrate that the mean of VNR, typical of this type of texts, varies from 50 to 85. Texts with the highest degree of narrativity, i.e. 99 – 150, are outliers and as such present less difficulty for test takers. Texts with VNR below 50 have the lowest narrativity. Cf. examples of texts with different VNR:

G17 (VNR=44): "*It was the first single by unknown group from Liverpool called the Beatles. It was the first of a number of big hits that would make John Lennon, Paul McCartney, George Harrison and Ringo Starr the most successful pop group the world has ever known. However, the road to success was not always easy*".

F30 (VNR=125): "*Once he had made his choice, he took out an axe to cut down the tree. I wanted to help my dad and brothers, but they always told me to stand back. When the cutting began, I tried to hold the tree up*".

Text F30 tells a story, with characters, events, places, and things that are familiar to most readers and consequently is less complex, while G17 contains many more nouns which make the text more informative and difficult to remember and reproduce.

Lexical features were assessed with both instruments, i.e. TI and CM. The measure of Textual Lexical Diversity (MTLD) demonstrates how 'lexically' rich is a text and is computed with the following formula:

$$MTLD = L/n, \qquad (4)$$

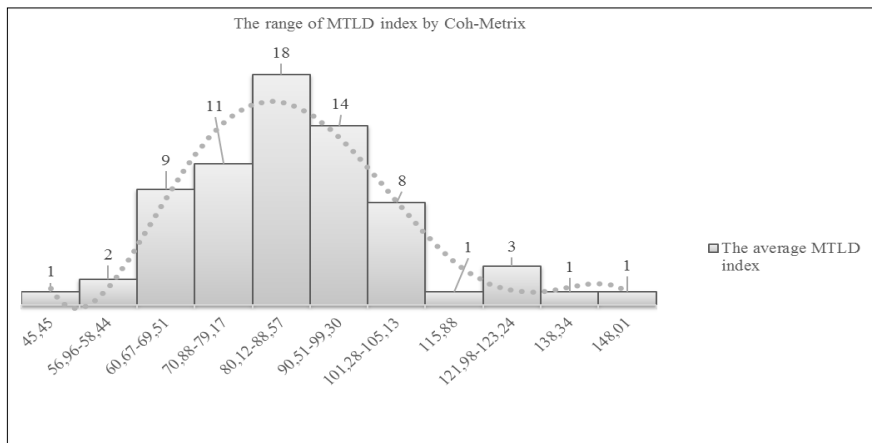where L is a number of word forms, n is a number of lines [14].

Fig.6. The range of MTLD index by Coh-Metrix.

MTLD index ranges from 45.45 to 148.01, however mean demonstrated in 41 texts tends to vary from 80 to 99, which is viewed as a type pattern for BSE (OGE) texts.

Referential cohesion, or Content Word Overlap, takes into account the proportion of words that intersect between pairs of sentences [14].
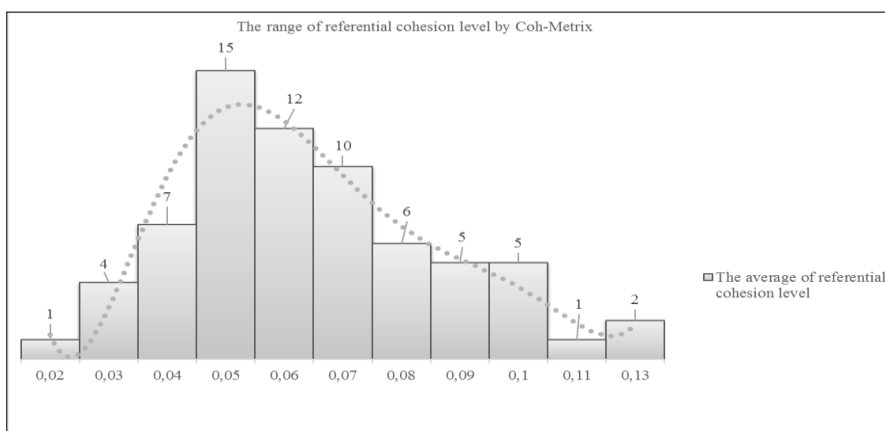

Fig. 7. The range of referential cohesion level by Coh-Metrix.

In the texts studied the metrics of this feature fluctuate from 0.02 to 0.13. The mean data is expected to rank from 0.05 to 0.07.

Thus, in this study we present BSE (OGE) text pattern defined by a combination of seven features which significantly differ from text features of other proficiency text patterns [3, 4]. The method of finding numerical attributes to text parameters improves the quality of text classification and may be applied to identify other proficiency levels text types.

6 Conclusion

Based on the multi-factor analysis conducted on 66 reading texts used in BSE (OGE) training we conclude the following: (1) text readability in the sources tested does not increase linearly but fluctuates within the rage of over 7 (FKGL) and 20 (CML2) education levels thus being in the majority of cases beyond the zone of students'/test takers' proximal development; (2) a

reading text pattern used in Reading 2. BSE (OGE) has the following linguistic metrics: readability (FKGL) – 7-7.9, average sentence length – 14.4-17.0 words per sentence, word length – 1.44 syllables per word, verbs to nouns ratio – 50-85, referential cohesion – 0.05-0.07, lexical diversity (MTLD) – 80-99. The text features computed with two online services, Text Inspector and Coh-Metrix, provide slightly different though consistent metrics in defining BSE (OGE) text pattern based on FKGL, average sentence length, word length, verbs to nouns ratio, referential cohesion and lexical diversity.

Acknowledgments

References

1. M. I. Solnyshkina, E. V. Harkova, M. B. Kazachkova, Journal of Language & Education, **6**, (2020)
2. M. Lukasik, Lingwistyka Stosowana, **23**, 3 (2017)
3. V. D. Solovyev, V. V. Ivanov, R. B. Akhtiamov, Journal of research in applied linguistics, **10** (2019)
4. A. I. Gabitov, M. I. Solnyshkina, L. Kh. Shayakhmetova, L. G. Ilyasova, S.A. Акbarova, Revista Publicando, **4**, 13 (2017)
5. FIPI: Federal Institute for Pedagogical Measurements. URL: http://os.fipi.ru/tasks/9/a (accessed: 25.08.2020)
6. L. M. Gudkova, O. V. Terent'eva, BSE-2020: Angliiskii yazyk: 30 trenirovochnykh variantov ekzamenatsionnykh rabot dlya podgotovki k obshchemu gosudarstvennomu ekzamenu: uchebnoe posobie (ACT, Moscow, 2019)
7. Yu. S. Veselova, BSE. Gotovimsya k itogovoi attestatsii: uchebnoe posobie (Intellekt-tsentr, Moscow, 2020)
8. H. D. Brown, Language Assessment: Principles and Classroom Practice (Pearson Education ESL, 2019)
9. D. Biber, Literary and Linguistic Computing, **8**, 4 (1993)
10. TextInspector: The professional web tool for analyzing texts. URL: https://textinspector.com/workflow (accessed: 20.04.2020)
11. Coh-Metrix: Analysis of text on cohesion and language. URL: https://link.springer.com/article/10.3758/BF03195564 (accessed: 20.04.2020)
12. The CEFR for English. URL: https://www.englishprofile.org/wordlists (accessed: 20.04.2020)
13. A. C. Graesser, D. S. McNamara, M. M. Louwerse, Behavior Research Methods, Instruments, & Computers, **36** (2004)
14. D. S. McNamara, A. C. Graesser, P. M. McCarthy, Z. Cai, Automated evaluation of text and discourse with Coh-Metrix (Cambridge University Press, Cambridge, 2014)
15. M. Azadnia, A. R. Lotfi, R. Biria, Research in English Language Pedagogy, **7**, 2 (2019)
16. Using Readability Levels to Guide Students to Books. Special report (Renaissance Learning, 2006)
17. L. S. Vygotsky, Mind in society: The development of higher psychological processes (Cambridge, MA: Harvard University Press, 1978)