Impact Factor: 9.9 ISSN-L: 2544-980X

Coca and it is Abilities to Constract Linguistic Dictionaries

Ramazonova Shaxnoza Yo'ldosh qizi [1]

Abstract: English is a difficult language for learning. Many rules of English grammar break down on application, and the common uses of words are often learned through direct observation. Therefore many writers have trouble telling whether something sounds right to Americans. The Corpus of Contemporary English (COCA) collects 520 million English words as they have been used in speech, in writing, on TV, in academic writing, and other areas between 1990 and 2015. COCA has many potential uses, including: (1) researching common American idioms (unique expressions); (2) double-checking prepositions and verbs; (3) comparing styles between spoken and academic American English; (4) finding the right word for each medium.

Key words: American, right word, corpus, writer, trouble, compare, idiom, speech, contemporary, rule, grammar, application, common, preposition, double-check, expression, style, spoken.

INTRODUCTION

Researching common American idioms Common American idioms can be researched with the default search for COCA, called a "List" search. a. Type a word or phrase into the search bar, and press Enter. COCA will show a list of matches in blue all-caps hyperlinks. On the right, the results will show the frequency of that word or phrase in COCA. ☆ For example, "helpmeet" b. Select the word or phrase hyperlink that best fits your interests. The selection will show every usage of that word or phrase within COCA. By default, COCA will display (from left to right) the year, the medium, the publisher, and a selection from that source that shows the word or phrase. ☆ For example, "2013 MAG AmSpect" c. Click on the year, medium, or publisher to see more from that selection. COCA will display the publication information about that source, as well as an extended quote containing the original selection.

COCA may look up American phrases that are only vaguely understood. For instance, you may search COCA for "star * banner" if you're looking for an expression related to the American flag that starts with "star" and finishes with "banner." As a "wildcard" for searches, the asterisk [*] indicates that any text that falls between "star" and "banner" will show up in COCA's results. With 276 hits in COCA, "Star-Spangled Banner" is the most often occurring result in this example, whereas "Star-and-Cross-Spangled Banner" only receives one hit.

METHODOLOGY

The study is experimental in nature. This type of research helps determine whether the COCA intervention has a causal effect on the experimental group of the study (Kothari, 2004).

Corpus Selection and Participants

The data for analysis was collected from 60 ESL undergraduates during one semester of teaching in a writing skills class. Participants are students from the 101 writing skills class (Freshman English), who were divided equally into a control and experimental group. The former group resorted to online dictionaries and thesauri to improve their word choice, whereas the latter group accessed COCA to

[1]BuxDU, Xorijiy tillar fakulteti, Xorijiy til va adabiyoti (ingliz tili) yo'nalishi talabasi

refine their word choice. The corpus consists of 300 essays, which were collected at the beginning, middle, and end of the semester. The writings comprise the pre and post writing tests and essays of three types: cause and effect, compare and contrast, and persuasion.

RESULTS

2. Double-checking prepositions and verbs Common combinations of prepositions and verbs can be researched with one additional step on the default search. a. Type a word into the search bar, SPACE, and select [POS] next to the search bar. COCA will replace [POS] with a drop-down menu listing parts of speech. ☆ For example, "learn " b. Select the part of speech that would fit your search term. If you typed in a verb and need to know a preposition, then select "prep.ALL" ☆ For example, "learn [i*]" c. Press Enter. COCA will show a list of matches in blue all-caps hyperlinks. On the right, the results will show the frequency of that word or phrase in COCA. ☆ For example, "learn about", 5436; "learn from", 5204; "learn as", 60 The most frequent results are most common among English-speaking Americans.

3. Comparing styles between spoken and academic American English COCA's frequencies can also show the relative appropriateness of a word for each section, such as the unpopularity of "stuff" in academic English. COCA's frequencies can also show changes in popularity, such as the rising popularity of "stuff" between the earliest section and the latest. a. Type a word in the search bar, and then select "Sections" and check the box next to "Sections." Press Enter. ☆ For example, "stuff" COCA will show the frequencies of the word according to each section ☆ For example, Spoken, 22584; Academic, 9098; 1990-94, 9230; 2010-2015, 15122 b. Click the number of any section. COCA will show the "List View" of that word within that section. These results can't show what is the right appropriate without some context, but COCA can show which terms appear more frequently in spoken or academic American English.

4. Finding the right word for each medium COCA can help identify what patterns of use are appropriate in different media, including academic writing. For example, COCA can show that "stuff" is more commonly used in Academic Geology and Social Science publications than in Philosophy publications. a. Type a word in the search bar. For example, "stuff." b. Select a medium listed under "Sections." Press Enter. For example ACAD: Philosophy and ACAD:Geog/SocSci COCA will show the "tokens" or number of the word according to each medium ☆ For example, "Philosophy" has 128 "tokens"; "Geog/SocSci," 324 These results also can't show which words are always appropriate, but they can indicate what is more or less common in any of the designated fields.

METHODOLOGY

The study is experimental in nature. This type of research helps determine whether the COCA intervention has a causal effect on the experimental group of the study (Kothari, 2004).

Corpus Selection and Participants

The data for analysis was collected from 60 ESL undergraduates during one semester of teaching in a vocabulary skills class. Participants are students from the 101 writing skills class (Freshman English), who were divided equally into a control and experimental group. The former group resorted to online dictionaries and thesauri to improve their word choice, whereas the latter group accessed COCA to refine their word choice. The corpus consists of 300 essays, which were collected at the beginning, middle, and end of the semester. The writings comprise the pre and post writing tests and essays of three types: cause and effect, compare and contrast, and persuasion.

ANALYSIS

The study analyzes students' writings quantitatively and qualitatively to produce more valid results (Tangkiengsirisin, 2010).

Qualitative Analysis

Vocab Profile (VP) software was used for the descriptive analysis of the corpus. (VP) analysis is based on Laufer and Nation's lexical frequency list (1995) opposed to the essay word choice. It is helpful in displaying a detailed text breakdown with three different colors that categorize students' vocabulary into different word list levels: 1) blue (K1 level or words from the first 1,000 frequent word list), 2) green (K2 level or words from the second 1,000 frequent word list), and 3) yellow (academic word list AWL or words beyond the first 2,000 frequent word list). The software also calculates the proportion of low and high frequency words used by the writers (Luafer & Nation, 1995). The following tables clarifies how the software presented a descriptive lexical profile analysis for different rated samples.

The data is analyzed by Lu's linguistic computational tool (2010), which provides quantitative results for the lexical complexity components found in the writing samples. More particularly, the results are obtained from the Lexical Complexity Analyzer (LCA) with 18 of its indices: lexical sophistication (5 indices) and lexical variation (13 indices). The computational tool is helpful in making a reliable statistical analysis and directly providing the researcher with the discriminant function value of every component (Kim, 2014). Table 1 (as cited in Kim, 2014) gives a detailed picture of the subcomponents of lexical variation and lexical sophistication. It helps the researcher in identifying the prominent lexical features that underpin the writing proficiency of students.

RESULTS AND FINDINGS

The writing average of the experimental group in the three types of writing (persuasive, cause-effect, compare-contrast) was significantly higher than that of the control group. Table 2 shows the difference in mean between the two groups.

However, in order to study the difference in the use of vocabulary, the corpus of persuasive, cause-effect, and compare- contrast writings of both groups was analyzed separately by the Lexical

Complexity Analyzer. The computational analysis showed the ratio of every lexical subcomponent in every type of writing for both groups (See Appendix 1). Then, the total mean of every subcomponent ratio across the three types of writing was calculated and later compared with the one in the other group. Analytical comparison between the results of both groups showed the following:

CONCLUSION

The aim of the study was to make the academic writing process less challenging for the Lebanese undergraduates. The purpose was to help students overcome the demotivating challenges they face with the use of appropriate vocabulary. Hence, the choice was the implementation of COCA , an online corpus, which is used for the first time in Lebanon as a lexical reliable reference. The research investigated the effectiveness of COCA in helping students develop their word choice and improve their quality of writing. Findings suggest that experimental group students who used COCA in all types of writings had shown a better vocabulary competence in terms of using less frequently word choice and more variety of content words. The online corpus helped the students avoid repetition and use advanced word choice, especially vocabulary words beyond the first 2,000 basic words in the word list of Laufer and Nation (1995) and that of the British National Corpus (BNC). Consequently, it is suggested that lexical sophistication and lexical variety are predicative complex components of L2 writing proficiency. Thus, it is suggested that English teachers integrate the use of COCA corpus as a new vocabulary strategy in the writing process of the undergraduate academic writings. It will encourage writers to enhance their writing proficiency through the use of words beyond the list of the 2,000 frequent ones and through the rich variation of lexical choice. Finally, for future research, further investigation is recommended for exploring the effectiveness of COCA in developing the syntactic and morphological complexity of Lebanese undergraduate writing. The effectiveness of COCA in developing the syntactic and morphological complexity of Lebanese undergraduate writing.

USED LITERATURE:

1. Ai, Haiyang & Lu, Xiaofei (2013). A corpus-based comparison of syntactic complexity in NNS and NS university students' writing. In Ana Díaz-Negrillo, Nicolas Ballier, and Paul Thompson

(eds.), Automatic Treatment and Analysis of Learner Corpus Data, pp. 249-264. Amsterdam/Philadelphia: John Benjamins

2. Bulte', B., & Housen, A. (2015). Evaluating short-term changes in L2 complexity development. *Circulo de Linguistica Aplicada a la Comunicacion, 63,* 42-76. http://www.ucm.es/info/circulo/no63/bulte.pdf
3. Davies, M. (2009). The 385+million word corpus of contemporary American English (1990-2008+): Design, architecture, and linguistic insights. *International Journal of Corpus Linguistics, 14,* 159-190. http://dx.doi.org/10.1075/ijcl.14.2.02dav
4. Hinkel, E. (2003). Simplicity without elegance: Features of sentences in L1 and L2 academic texts. *TESOL Quarterly, 37,* 275-301. 10.2307/3588505
5. Housen, A., Kuiken, F., & Vedder, I. (2012). Complexity, accuracy, and fluency: Definitions, measurement and research. In A., Housen, F. Kuiken, & I. Vedder (Eds.). *Dimensions of L2 performance and proficiency: Complexity, accuracy, and fluency in SLA* (pp.1-17). Amsterdam: John Benjamins.
6. Jarvis, S. (2002). Short texts, best-fitting curves and new measures of lexical diversity. *Language Testing, 19,* 57-84. DOI: 10.1191/0265532202lt220o.
7. Kim, Y., J. (2014). Predicting L2 writing proficiency using linguistic complexity measures: A corpus-based study. *English Teaching, 69* (4), 27-51. DOI: 10.15858/engtea.69.4.201412.27.
8. Knoch, U. (2009). Diagnostic Writing Assessment: The Development and Validation of a Rating Scale. Frankfurt: Peter Lang GmbH.
9. Laufer, B., & Nation, P. (1995). Vocabulary size and use: Lexical richness in L2 written productions. *Applied Linguistics 16* (3), 307-322. Doi: 10.1093/applin/16.3.307
10. Lu, Xiaofei (2010). Automatic analysis of syntactic complexity in second language writing. *International Journal of Corpus Linguistics, 15*(4):474-496. DOI: 10.1075/ijcl.15.4.02lu
11. Lu, Xiaofei (2011). A corpus-based evaluation of syntactic complexity measures as indices of college-level ESL writers's language development. *TESOL Quarterly, 45*(1):36-62. DOI: 10.5054/tq.2011.240859.
12. Lu, Xiaofei & Ai, Haiyang. (2015). Syntactic complexity in college-level English writing: Differences among writers with diverse L1 backgrounds. *Journal of Second Language Writing, 29,* 16-27. DOI: 10.1016/j.jslw.2015.06.003
13. McNamara, D.S., Louwerse, M.M., McCarthy, P.M., & Graesser, A.C. (2010). Coh-Metrix: Capturing linguistic features of cohesion. *Discourse Processes, 47,* 292-330. DOI: 10.1080/01638530902959943.
14. Ortega, L. (2012). Interlanguage complexity: A construct in search of theoretical renewal. In B. Kortmann, & B. Szmrecsanyi (eds.), *Linguistic Complexity: Second Language Acquisition, Indigenization, Contact* (pp.127-55). Berlin: de Gruyter.