VIETNAMESE STUDENTS' PERSPECTIVES ON THE COCA CORPUS

Nguyen Thi Dan Tam

Ho Chi Minh City University of Economics and Finance (UEF)
Vietnam

ABSTRACT

This study investigates how Vietnamese students in an English course connected to a British University in Ho Chi Minh City view the COCA (Corpus of Contemporary American English). It examines survey data from 15 students and interviews with 5 participants using a mixed-methods methodology. The results show that although students appreciate COCA for learning grammar and vocabulary, issues include cultural misalignment, information overload, and complicated interfaces. It may be surprising to learn that some students use COCA in unique ways outside of the classroom. The study provides useful insights for incorporating corpus technologies into EFL training while highlighting the necessity of localised adjustments and instructional support.

KEYWORDS: COCA corpus, corpus-based learning, Vietnamese EFL students, vocabulary acquisition, EFL teaching

1. INTRODUCTION

The study explores the opinions of Vietnamese students about the COCA (Corpus of Contemporary American English) corpus as an English language learning resource. Knowing how cutting-edge resources like COCA affect language learning is essential as the need for advanced English competence grows in both academic and professional settings. Although a lot of study has been done on corpus-based learning in Western educational settings, little focus has been placed on how Vietnamese students view and use these resources. Students enrolling in a course offered in Ho Chi Minh City, Vietnam, that is associated with a UK university are the subject of this study. Through examining the advantages and difficulties these students perceive; the study seeks to offer insightful information to teachers incorporating corpus-based resources into their courses.

2. LITERATURE REVIEW

2.1 Corpus-based language learning

It is becoming more widely acknowledged that corpus-based language learning is a cutting-edge method of teaching and learning languages. Learners can examine linguistic patterns, vocabulary, and grammar in real-world scenarios by using tools like COCA, which offer actual, contextualised instances of English usage. The advantages of corpus-based techniques in encouraging learner

autonomy and data-driven learning are highlighted by Boulton (2012) and Römer (2021). Flowerdew (2015) emphasises how useful corpus-based approaches are in specialised domains like professional and academic English. In linguistic analysis, Leech (2011) emphasises their importance in connecting theory and practice, but McCarthy and O'Keeffe (2010) explore how corpus-based methods help address real-world communicative needs.

2.2 Benefits of COCA for non-native learners

The COCA corpus is well known for covering a wide range of registers, including academic texts, spoken language, and fiction. As such, it is considered an invaluable resource for language learners. Davies (2008) shows how COCA facilitates the learning of phraseological patterns and high-frequency collocations. According to Breyer (2009), COCA gives students real-world language input so they can experiment with usage in different settings. Lin and Lee (2015) talk about how corpus tools like COCA help non-native speakers better understand idiomatic idioms and collocations, which can be difficult. Furthermore, Nation (2013) highlights the importance of corpora in vocabulary acquisition, specifically in locating high-frequency terms that are essential for language fluency.

2.3 Challenges in using COCA

Navigating COCA might be difficult for non-native learners, despite its advantages. According to Anthony (2013), two major obstacles are the complexity of the interface and the sophisticated level of examples. Mukherjee (2004) points out that learners' ability to understand certain COCA scenarios may be hampered by cultural unfamiliarity. In a similar vein, O'Keeffe, McCarthy, and Carter (2007) highlight the possible misalignment between corpus data and students' current learning requirements and propose guided instruction as a solution. Additionally, McEnery and Hardie (2012) contend that in order for students to effectively use the features of corpus tools, they need to be trained. According to Nguyen and Dang (2021), learners may find it more difficult to connect corpus data to their regular language use if there are no localised corpora available.

2.4 Corpus-Based Learning in Vietnam

Although still very small, corpus tool utilisation is steadily increasing in Vietnam. Vietnamese students are open to corpus-based learning, but they need a lot of help using resources like COCA, according to Nguyen and Nguyen (2019). While Tran (2020) highlights the necessity of curriculum integration to optimise the benefits of corpora, Pham and Hoang (2021) explore how teacher mediation improves the efficacy of corpus-based activities in Vietnam. Nguyen (2022) emphasises how COCA might improve vocabulary acquisition, but she also points out that many students find the technical elements too much to handle. According to this research, with appropriate guidance, COCA can play a significant role in advancing language learning in Vietnam.

3. RESEARCH METHODS

To get a complete picture of students' opinions on COCA, this study uses a mixed-methods approach.

3.1 Participants

Fifteen students who are enrolled in a course delivered in Ho Chi Minh City and associated with a British institution are among the participants. These pupils' levels of English proficiency and their knowledge with corpus-based learning resources differ. Five students were also chosen to participate in in-depth interviews in order to offer qualitative observations.

3.2 Data Collection

All 15 participants were given a standardised questionnaire with 27 items on a 5-point Likert scale. Topics covered in the poll included familiarity with COCA, perceived advantages, difficulties, and desires for incorporating it into the curriculum. Five students participated in semi-structured interviews to learn more about their perspectives and experiences with COCA. Their usage patterns, particular difficulties they faced, and recommendations for bettering instructional strategies were the main topics of the questions.

3.3 Data Analysis

Descriptive statistics were used to examine the survey results in order to find trends and patterns. To gain a deeper knowledge of student viewpoints, interview transcripts were subjected to thematic analysis in order to identify recurrent themes.

3.4 Ethical Considerations

Participants received assurances that their answers would be anonymous and confidential. Prior to the survey and interviews, informed consent was acquired. This methodological approach guarantees that the study contributes to the larger conversation on corpus-based language acquisition in Vietnam by capturing the depth and breadth of student viewpoints on the COCA corpus.

4. FINDINGS

4.1 General Perceptions of COCA

According to the survey results, the majority of students had favourable opinions regarding the COCA corpus's ability to aid in their language acquisition. According to 73% of students, COCA is a helpful tool for comprehending language use in everyday situations. Eighty percent of respondents said that COCA enhanced their vocabulary, especially in recognising collocations and high-frequency words. 40% of students said that COCA lacked examples that were pertinent to Vietnamese cultural contexts, despite its apparent advantages. This suggests that there may be a localisation gap for EFL learners in Vietnam.
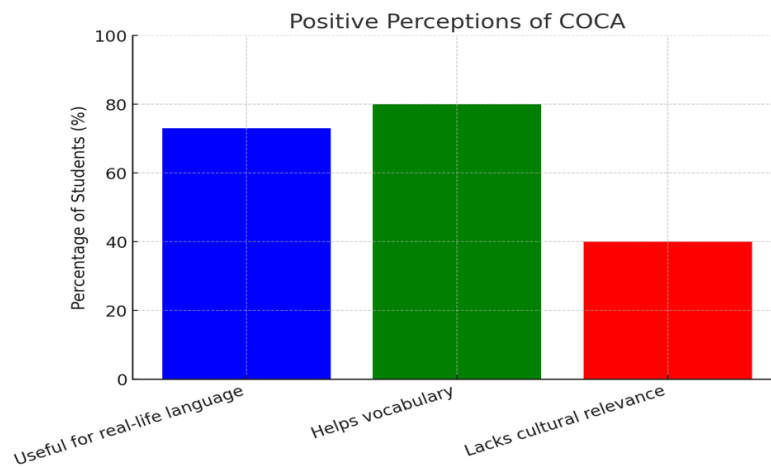
Figure 1. Positive perceptions of COCA

"COCA is excellent for learning American English, but it's difficult to find examples that compare to how we might use English in Vietnam, such as when creating emails or presentations," one student observed.

4.2 Challenges in Using COCA

Although most people valued the resource, linguistic and technological obstacles stood out as key obstacles. According to 67% of pupils, navigating COCA's interface was challenging. Because of the advanced level of the examples, 53% of respondents found it difficult to understand the grammatical patterns given. Many students (33%) said they felt "intimidated" by the abundance of material that COCA provides, calling it "information overload." As a result, some people stopped using COCA entirely unless their instructor specifically instructed them to.
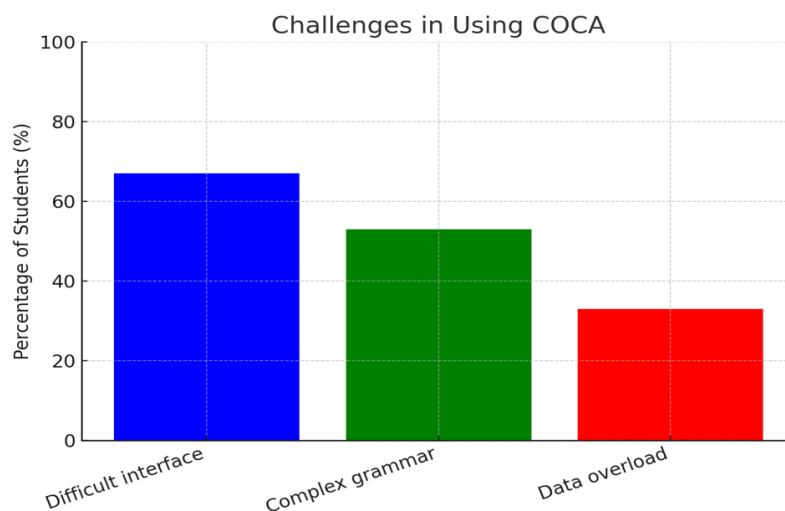

Figure 2.Challenges in using COCA

"I see too many results when I search for something, and I don't know how to pick the most important ones," one student clarified. It feels too much.

3. Frequency and nature of usage

The results revealed a great deal of diversity in the frequency and purposes of students' use of COCA. The main purposes of COCA for 47% of students were vocabulary development and word collocation comprehension. Just 13% used COCA to enhance their writing abilities, compared to 27% who investigated its grammatical patterns. An intriguing pattern surfaced in which students used COCA for non-academic purposes, such deciphering song lyrics or investigating colloquial terms to improve their conversational abilities.
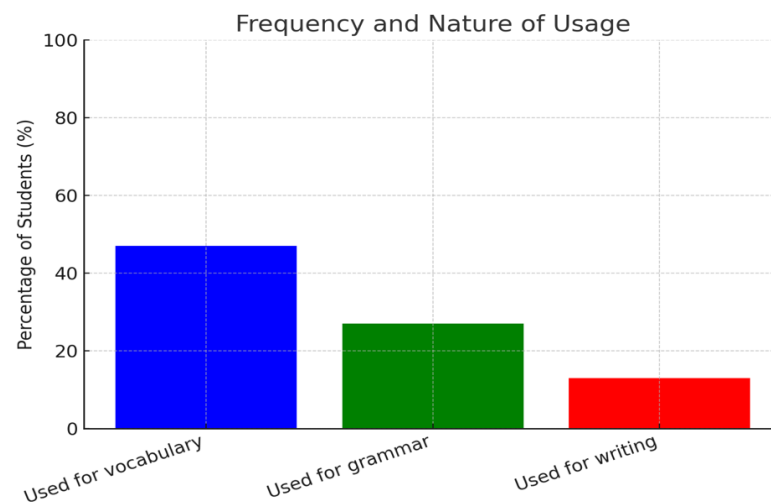

Figure 3. Frequency and nature of usage

"I use COCA to check idioms in songs or movies," one student shared. It's entertaining and improves my comprehension of lingo.

4. Impact on Learning Outcomes

Although they also pointed to limits, students reported observable progress in certain language areas.According to 67% of students, COCA helped them better comprehend register and collocations.
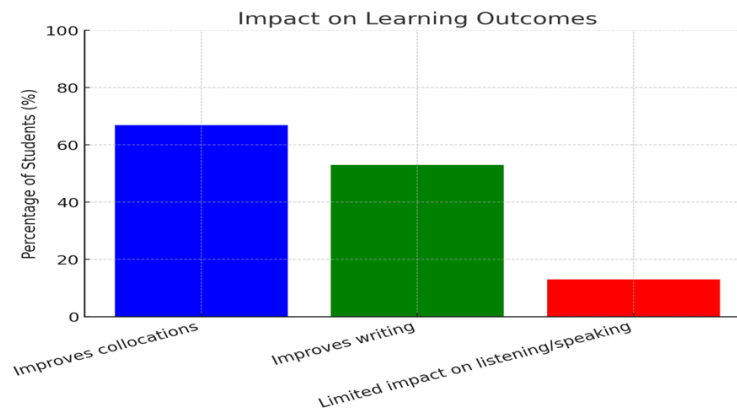
Figure 4. Impact on learning outcome

After using COCA, 53% of respondents said they were able to write in English more naturally. Thirteen percent of students said that COCA had little effect on their speaking and listening abilities, claiming that these abilities needed more interaction-based learning than corpus-based resources.

COCA is excellent for writing and grammar, but it doesn't truly help me comprehend accents or talk effectively, according to one student.

5. CULTURAL AND CONTEXTUAL RELEVANCE

According to the students' feedback, there is a gap in how COCA meets their learning requirements in the Vietnamese setting. Sixty percent of students concurred that learners exposed to British and Australian English as well as other English dialects found COCA's emphasis on American English to be somewhat restrictive. 47% of respondents said that the COCA examples did not adequately correspond with the real-world communication situations that Vietnamese learners would encounter, including academic presentations or job interviews. A localised corpus or further instructions on how to modify COCA's examples for Vietnamese situations were requested by a few pupils.
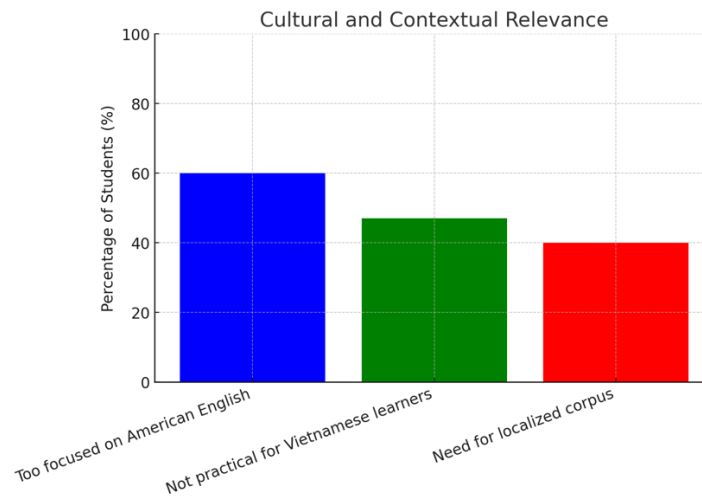
Figure 5. Cultural and context relevance

"I wish there was a Vietnamese-English corpus," said one pupil. The expressions in COCA can feel too American at times and aren't appropriate for our communication style.

In summary, most people agree that COCA is a helpful tool, especially for learning collocations and increasing one's vocabulary. Technical problems, information overload, and language barriers make it difficult for students to utilise the tool efficiently. By creatively utilising COCA to explore idiomatic phrases and song lyrics, some students show an innovative side to learning. COCA has minimal impact on speaking and listening, but it enhances writing and grammar. More locally relevant examples are required in order to meet the particular needs of Vietnamese learners.

5. DISCUSSIONS

The study's conclusions complement and add to the knowledge offered by the literature review. In line with earlier studies (e.g., Davies, 2008; Breyer, 2009; Lin & Lee, 2015; Nation, 2013), our findings show that Vietnamese students place a high value on the COCA corpus because it helps them learn new words and understand grammatical patterns and collocations. According to Boulton (2012) and Römer (2021), the majority of students agreed that COCA offered real-world language examples that enhanced their educational experience, reinforcing the pedagogical advantages of the corpus-based approach.

Simultaneously, our research revealed issues that reflect those expressed by Mukherjee (2004) and Anthony (2013). One detail that contributes to the current discussion on technical hurdles in corpus utilisation is the fact that students reported feeling overwhelmed by the amount of data and having trouble navigating the corpus interface. Our results further imply that the frightening character of large data sets may discourage self-directed investigation unless appropriately scaffolded, despite the literature's emphasis on the necessity of teacher mediation to overcome these obstacles (McEnery & Hardie, 2012; O'Keeffe et al., 2007).

The cultural and contextual relevance of the COCA examples is a significant difference from earlier research. Our study shows that 40% of students believed the instances did not fit Vietnamese cultural contexts, despite previous works highlighting the value of real material. This finding echoes requests in the literature (Nguyen & Nguyen, 2019; Pham & Hoang, 2021) for more localised corpus resources. Additionally, the creative application of COCA for non-academic contexts—like examining song lyrics and colloquial expressions—indicates a new trend in learner agency that hasn't gotten much attention in earlier studies.

Finally, our results show a more moderate effect on speaking and listening skills, even if the literature largely supports the effectiveness of COCA in enhancing writing and grammar. This disparity points to a possible weakness in corpus-based learning, indicating that although COCA is a useful tool for developing specific language skills, other interactive techniques could be needed to completely enable comprehensive language development.

All things considered, the study's conclusions support a large portion of the body of research on corpus-based language learning while also highlighting particular difficulties and creative usage patterns among Vietnamese EFL students. These observations highlight the necessity of localised adjustments and improved support when incorporating corpus tools such as COCA into a variety of educational settings.

6. LIMITATIONS AND CONTRIBUTIONS

The results may not be transferable to other contexts because there were only 15 survey respondents and 5 interviewees from a single course in Ho Chi Minh City. Using self-reported data only could result in biases that compromise the veracity of the insights. It's possible that participants' experiences with COCA were impacted by differences in their level of technological proficiency.

However, the study advances our knowledge of corpus-based learning in a non-Western setting by offering insightful information on how Vietnamese students view the advantages and difficulties of utilising COCA.Since many students thought that the American-centric content did not adequately address their practical communication needs, it highlights the necessity for localised corpus examples. The results show innovative, unconventional applications of COCA, like examining song lyrics and idioms, which open up new possibilities for student involvement.

The findings highlight the value of individualised help and guided instruction while providing useful recommendations for using corpus technologies into EFL programs.

REFERENCES

Anthony, L. (2013). *A critical look at software tools in corpus linguistics*. Linguistic Research, 30(SI), 141–161.

Boulton, A. (2012). Corpus consultation for ESP: A review of empirical research. In A. Boulton, S. Carter-Thomas, & E. Rowley-Jolivet (Eds.), *Corpus-Informed Research and Learning in ESP* (pp. 261–291). John Benjamins.

Breyer, Y. (2009). Learning and teaching with corpora: Reflections by student teachers. *Computer Assisted Language Learning, 22*(2), 153–172.

Davies, M. (2008). *The Corpus of Contemporary American English (COCA): One billion words, 1990–present*. Brigham Young University.

Flowerdew, L. (2015). *Corpus-based analysis in language for specific purposes (LSP) research*. Routledge.

Leech, G. (2011). Theoretical and empirical issues in corpus linguistics. *Linguistic Issues in Language Technology, 5*(3), 1–23.

Lin, A. M. Y., & Lee, P. (2015). Making visible the invisible: Native speaker norms in L2 education through corpus-based learning. *TESOL Quarterly, 49*(1), 25–48.

McCarthy, M., & O'Keeffe, A. (2010). *The Routledge handbook of corpus linguistics*. Routledge.

McEnery, T., & Hardie, A. (2012). *Corpus linguistics: Method, theory, and practice*. Cambridge University Press.

Mukherjee, J. (2004). English data-driven learning (DDL): Pedagogical opportunities and challenges. *International Journal of Applied Linguistics, 14*(2), 177–189.

Nation, I. S. P. (2013). *Learning vocabulary in another language*. Cambridge University Press.

Nguyen, T. M. T., & Nguyen, V. H. (2019). Exploring EFL learners' attitudes towards corpus-based activities in Vietnam. *Vietnam Journal of Language Studies, 35*(1), 56–74.

Nguyen, D. T. (2022). Exploring the impact of corpus tools on EFL learners' vocabulary acquisition in Vietnam. *Asian EFL Journal, 24*(3), 89–110.

Nguyen, M., & Dang, H. (2021). Challenges in integrating corpus tools in Vietnamese EFL classrooms. *ELT Journal, 75*(2), 235–246.

Pham, M. T., & Hoang, L. N. (2021). Teacher mediation in corpus-based English learning: Insights from Vietnamese contexts. *Asian Journal of English Language Teaching, 31*(2), 112–129.

Tran, N. H. (2020). The adoption of corpus-based learning in EFL classrooms in Vietnam: Challenges and potentials. *English Language Teaching, 13*(7), 98–106.