

Corpus-Based English Vocabulary Acquisition Research: A Case Study of University Students

Xuan Li, Luyun Liu

Officers College of PAP, Chengdu 610000, China

Abstract. This study investigates the impact of corpus-based methods on English vocabulary acquisition among university students. Utilizing a specialized corpus reflective of the university lexicon, the research employs pre- and post-test evaluations alongside qualitative feedback to assess the effectiveness of corpus engagement in vocabulary learning. The study's findings indicate that students exposed to corpus-based activities exhibit significant improvements in their command of university terminology. The enhanced performance is evident in their ability to recognize and use relevant vocabulary in context. Additionally, student feedback suggests a preference for corpus-based learning over traditional methods, highlighting its potential for more engaging and effective language education. This research underscores the value of integrating corpus analysis into teaching practices to meet the linguistic needs of university learners and enhance their professional communication skills.

Keywords: Corpus-Based, English Vocabulary, A Case Study of University Students.

1. Introduction

In the evolving landscape of English language education, the acquisition of vocabulary stands as a cornerstone of linguistic proficiency, especially for university students aspiring to thrive in an increasingly globalized world. The introduction of corpus linguistics, a method of studying language as expressed in corpora (or large, structured sets of texts), has revolutionized traditional approaches to language learning. This modern pedagogy leverages real-world language usage, extracted from vast databases of authentic text, to teach vocabulary in context. Such an approach is particularly pertinent for university students, who often require a practical and applicable command of English to excel in their specific fields. The corpus-based method not only enhances the depth of their vocabulary but also enriches their understanding of nuanced language usage, critical for professional and academic success. This article delves into the realm of corpus-based English vocabulary acquisition, with a focus on university students, examining how this innovative approach aligns with their unique learning needs. By exploring various research contexts, corpus types, and instructional methodologies, we aim to shed light on the effectiveness of corpus-based training in fostering a more robust and practical vocabulary skillset among university learners[1-2]. This intersection of linguistic theory and educational practice offers a promising pathway for university students to navigate the complexities of English vocabulary, equipping them with the tools to communicate effectively and confidently in their respective professional arenas.

2. Corpus Linguistics and English Vocabulary Learning

Corpus linguistics, an empirical approach to language study, utilizes real-world textual data to illuminate how language functions in various contexts. This branch of linguistics relies on corpora - extensive, structured sets of texts - which serve as reservoirs of authentic language usage. In the context of English vocabulary learning, especially for university students, corpus linguistics offers a pragmatic and relevant approach, aligning the process of learning with actual language use.

The fundamental premise of corpus linguistics is its emphasis on frequency and patterns of words in natural contexts. Unlike traditional vocabulary teaching, which often relies on decontextualized word lists and rote memorization, corpus linguistics encourages learners to engage with words as they are used in real-life situations. This exposure to language in context aids students in developing not

just a theoretical understanding of vocabulary but also practical skills in using these words appropriately and effectively [3].

One of the key contributions of corpus linguistics to vocabulary learning is its ability to highlight collocations - words that frequently appear together. For university students, understanding these collocations is crucial, as it enables them to grasp the nuanced meanings and uses of industry-specific terminology. For instance, a corpus of technical manuals or business correspondence can reveal common collocational patterns that are essential for students in those fields. This knowledge not only enriches their vocabulary but also enhances their comprehension and production of more natural, idiomatic English.

Moreover, corpus linguistics provides insights into the register-specific usage of words. Registers, or language varieties used in different social and professional contexts, often have distinct lexical choices. By analyzing corpora from various registers, students can learn the vocabulary that is most relevant to their university field. For example, the language used in academic journals differs significantly from that in casual conversation. University students, particularly those in technical or specialized fields, benefit from understanding these differences to communicate more effectively in their future workplaces.

Another advantage of corpus-based vocabulary learning is its adaptability to learner needs and interests. With the advent of digital corpora and user-friendly corpus analysis tools, students can explore language patterns related to their specific university areas. This personalized approach not only makes learning more engaging but also more relevant, as students can directly see the application of their learning in their field of study[4].

The corpus approach also aids in developing critical thinking and analytical skills. As students engage with corpus data, they learn to identify patterns, make inferences about language use, and critically assess the appropriateness of different lexical choices in various contexts. These skills are invaluable in a university setting, where precision and clarity of language are often paramount.

In conclusion, the integration of corpus linguistics into English vocabulary learning represents a significant shift from traditional, often more theoretical approaches. For university students, this shift is particularly beneficial. It bridges the gap between language learning and practical application, offering a more tailored, context-driven, and interactive learning experience. By harnessing the rich insights provided by real-world language use, corpus linguistics equips university students with a robust, functional vocabulary that resonates with their professional aspirations and needs.

3. Corpus-Based Language Pedagogy (CBLP)

The Corpus-Based Language Pedagogy (CBLP) represents a paradigm shift in language education, integrating the insights of corpus linguistics into the classroom. This innovative approach to teaching and learning is grounded in the use of authentic language data derived from corpora — extensive, structured collections of real-world texts. CBLP stands out for its emphasis on evidence-based instruction, where language teaching is informed by actual language usage, rather than prescriptive rules or isolated vocabulary lists.

The heart of CBLP lies in its four foundational design principles. The first principle, 'detecting lexical errors', involves using corpora to identify and correct common language mistakes. By analyzing patterns in a corpus, educators can pinpoint frequent errors in vocabulary usage, grammar, or syntax. This approach is particularly beneficial for university students, who often need to communicate with precision in their fields. Through corpus analysis, they can learn the correct usage of industry-specific terminology, reducing the likelihood of misunderstandings in professional contexts [5].

The second principle, 'observing and analyzing language patterns', encourages students to engage with language data directly. This involves students examining corpora to understand how certain words or phrases are used in different contexts. For instance, a student in a health-related university program might study medical journals or patient case reports to grasp the specific jargon and

colloquialisms used in healthcare settings. Such activities not only enhance vocabulary acquisition but also foster deeper comprehension of how language functions in their area of expertise.

The third principle, 'data-driven learning', shifts the role of learners from passive recipients to active investigators of language. In this approach, students are encouraged to explore corpora independently, formulating hypotheses about language use and then testing these against real-world examples. This method cultivates a more exploratory and autonomous learning attitude, empowering students to take charge of their language learning process. For university students, this translates into a more personalized and relevant learning experience, as they can focus on language relevant to their university subjects.

The fourth principle, 'using corpora for authentic material creation', involves educators leveraging corpus data to create learning materials that reflect real language use. Unlike traditional textbooks, which may present language in a contrived or outdated manner, corpus-based materials offer current, authentic examples of language. This is especially critical for university students, who need to be familiar with the language as it is currently used in their industry. By working with materials that mirror real-world communication, students gain practical exposure that directly translates to their future workplaces.

CBLP not only enhances vocabulary acquisition but also develops critical language skills. Through interaction with authentic language data, students improve their ability to analyze, infer, and apply language in various contexts. This aligns closely with the needs of university students, who require not just theoretical knowledge but practical language skills that can be immediately applied in their professional lives[6].

In essence, Corpus-Based Language Pedagogy transforms the way language is taught and learned. By grounding language education in the realities of language use, CBLP offers a more dynamic, relevant, and effective approach to vocabulary learning. For university students, this approach not only facilitates language acquisition but also equips them with the skills to navigate the linguistic demands of their chosen careers successfully.

4. Case Study: University Students and English Vocabulary Acquisition

The integration of corpus linguistics into language education, particularly for university students, has opened new pathways for effective English vocabulary acquisition. This case study delves into how corpus-based methods enhance the learning experience and outcomes for these students, offering insights into the practical application and benefits of this approach in university settings.

University students often face unique challenges in language learning. They need to acquire not just general English vocabulary but also specialized terminology relevant to their specific fields. Traditional methods of vocabulary teaching, which may focus on broad, general language skills, often fall short in meeting these specific needs. This is where corpus-based methods come into play, offering a tailored approach that aligns with the practical language requirements of university learners [7-8].

A key aspect of corpus-based learning for university students is its focus on real-world language use. By analyzing language as it appears in authentic texts - such as technical manuals, business reports, or industry-specific publications - students gain exposure to the vocabulary and language structures that are most relevant to their field. This relevance is crucial for maintaining student engagement and ensuring that the language skills they acquire are directly applicable to their university contexts.

For instance, consider university students in the healthcare sector. A corpus-based approach might involve studying medical journals, patient case reports, and healthcare guidelines. This immersion in industry-specific language helps students understand how medical terminology is used in context, aiding them in developing both a theoretical and practical understanding of these terms. Such exposure is invaluable for students who will need to communicate effectively in medical settings, where precision and clarity are paramount.

Furthermore, corpus-based learning facilitates a deeper understanding of collocations - words that frequently appear together - which is vital in mastering the nuances of professional language. For example, in the field of information technology, understanding the difference between 'software development' and 'software design' is crucial. Through corpus analysis, students can grasp these subtleties, enhancing their ability to communicate accurately and professionally.

The adaptability of corpus-based methods to different university areas is another significant advantage. Students can explore corpora related to their specific fields, making their learning experience more personalized and relevant. This targeted approach not only aids in vocabulary acquisition but also fosters greater interest and motivation among students, as they can see the direct application of their learning efforts.

An added benefit of this method is the development of autonomous learning skills. As students engage with corpus data, they learn to identify patterns, make inferences, and critically evaluate language use. These skills are crucial not only for language learning but also for their professional development, encouraging a more analytical and independent approach to problem-solving and communication[9].

In conclusion, the case of university students and English vocabulary acquisition through corpus-based methods illustrates the effectiveness of this approach in catering to the specific needs of these learners. By providing exposure to authentic, industry-specific language, enhancing understanding of professional terminology and collocations, and fostering autonomous learning, corpus-based methods equip university students with the practical language skills they need to succeed in their respective fields. This case study underscores the value of adapting language education to the practical realities and requirements of different university areas, ensuring that students are well-prepared to meet the linguistic demands of their future careers.

5. Learners' Perceptions and Experiences

The implementation of corpus-based learning methods in English language education has significantly impacted learners, particularly in how they perceive and experience the process of vocabulary acquisition. This section delves into the perceptions and experiences of learners, especially university students, who have been exposed to corpus-assisted vocabulary learning activities. Understanding these perceptions is crucial in evaluating the effectiveness of corpus-based methods and in tailoring them to better suit learner needs.

Learners' engagement with corpus-based learning activities typically reveals a range of positive perceptions. Many students describe these activities as 'interesting', 'fun', 'innovative', and 'autonomous'. This positive feedback is a stark contrast to traditional rote learning methods, which are often perceived as tedious and unengaging. The interactive nature of corpus-based learning, which allows students to explore and analyze real language data, contributes to this heightened interest and enjoyment. This is particularly significant for university students, who often require motivation to engage with English vocabulary that is directly relevant to their professional fields.

One of the key aspects appreciated by learners is the sense of autonomy and empowerment that corpus-based activities offer. Unlike traditional methods where learning is teacher-centered and prescriptive, corpus-based learning places students in the driver's seat. They become investigators of language, exploring corpora to discover patterns, collocations, and usages that are relevant to their needs and interests. This learner-centered approach not only fosters a deeper understanding of the language but also cultivates critical thinking and independent learning skills[10].

Additionally, students often report that corpus-based activities make learning easier and more effective. The exposure to authentic language usage helps in contextualizing vocabulary, making it easier for learners to remember and apply new words and phrases. For university students, this contextuality is invaluable. By seeing how specific terminology is used in real-world scenarios relevant to their field, students can better grasp the practical application of the language they are learning.

However, it's not without its challenges. Some students may initially find the approach daunting, particularly those who are accustomed to more traditional learning methods. The shift from a structured, teacher-led classroom to a more exploratory, learner-driven environment can be challenging. Overcoming this barrier often involves guided exploration, where educators initially lead students through corpus analysis before gradually handing over more autonomy[11].

The integration of technology in corpus-based learning also presents both opportunities and challenges. While digital corpora and analysis tools have made language data more accessible, some learners may struggle with the technical aspects of these tools. Effective incorporation of corpus-based activities, therefore, necessitates not only linguistic guidance but also support in using technological resources[12].

In conclusion, learners' perceptions and experiences of corpus-based English vocabulary learning are predominantly positive, marked by a sense of engagement, autonomy, and effectiveness. These experiences are particularly beneficial for university students, who find the context-driven and industry-specific focus of corpus-based learning highly relevant and useful. The challenges noted, primarily around the initial adjustment to a new learning style and the use of technology, underscore the need for supportive and adaptable teaching practices. Overall, the learner-centered, interactive, and authentic nature of corpus-based learning represents a significant advancement in language education, aligning closely with the needs and preferences of modern learners.

6. Conclusion

In conclusion, the exploration of corpus-based methods in English vocabulary learning, especially for university students, reveals a transformative approach in language education. This method, grounded in real-world language usage, offers numerous advantages over traditional rote learning techniques. It aligns vocabulary acquisition with practical applications, making learning more relevant and effective for students in university settings.

The integration of corpus linguistics into language education, particularly through Corpus-Based Language Pedagogy (CBLP), has shown to significantly enhance the depth and relevance of vocabulary learning. This approach not only aids in acquiring industry-specific terminology but also in understanding the nuanced usage of language in different contexts. The four foundational principles of CBLP – detecting lexical errors, observing and analyzing language patterns, data-driven learning, and using corpora for authentic material creation – provide a robust framework for effective language teaching and learning.

The positive impact of these methods is further evidenced in the perceptions and experiences of learners. Students have reported finding corpus-based learning activities to be engaging, fun, and more aligned with their learning needs. The autonomy and hands-on experience that these methods offer have been particularly beneficial, fostering a deeper understanding of language and developing critical thinking skills. However, the transition to this new mode of learning can pose challenges, necessitating guided support from educators and a gradual adjustment period for students.

For university students, the practical implications of corpus-based learning are profound. As these learners prepare for careers in various industries, having a strong command of relevant, industry-specific English vocabulary is invaluable. Corpus-based methods equip them with not just linguistic knowledge, but the skills to apply this knowledge effectively in their professional lives.

In summary, corpus-based English vocabulary learning represents a significant advancement in language education. It effectively bridges the gap between theoretical knowledge and practical application, particularly for university students. By harnessing the power of real-world language data, these methods offer a more dynamic, context-driven, and learner-centric approach to vocabulary acquisition. As the field of language education continues to evolve, the incorporation of corpus-based methods is likely to play an increasingly vital role in equipping learners with the language skills necessary for success in the modern world.

References

- [1] Zhang, H. & Shi, Y. Evolution of English language education policies in the Chinese mainland in the 21st century: A corpus-based analysis of official language policy documents[J]. *Linguistics and Education*, 2023, 76: 101190.
- [2] Crosthwaite, P., Luciana, & Wijaya, D. Exploring language teachers' lesson planning for corpus-based language teaching: A focus on developing TPACK for corpora and DDL[J]. *Computer Assisted Language Learning*, 2023, 36(7): 1392-1420.
- [3] Shen, C., Guo, J., Shi, P., Qu, S., & Tian, J. A corpus-based comparison of syntactic complexity in academic writing of L1 and L2 English students across years and disciplines[J]. *Plos one*, 2023, 18(10): e0292688.
- [4] Reynolds, B. L., Xie, X. S., Pham, Q. H. P. The potentials for incidental vocabulary acquisition from listening to computer science academic lectures: a higher education corpus-based case study from Macau[J]. *Frontiers in Psychology*, 2023, 14: 1-36.
- [5] Ma, Q., Yuan, R., Cheung, L. M. E., Yang, J. Teacher paths for developing corpus-based language pedagogy: a case study[J]. *Computer Assisted Language Learning*, 2022, 1(1): 1-32.
- [6] Tan, P. Construction of Multimodal University English Corpus Based on Functional Software[C]. *International Conference on Frontier Computing*, 2022, 1(1): 1335-1342.
- [7] Bacha, N. N. & Khachan, V. A Corpus-Based Lexical Evaluation of L1 Arabic Learners' English Literary Essays[J]. *International Journal of Arabic-English Studies*, 2023, 23(2): 415-442.
- [8] Kamrotov, M., Talalakina, E., & Stukal, D. Technical vocabulary in languages for special purposes: The corpus-based Russian economics word list[J]. *Lingua*, 2022, 273: 103326.
- [9] Lan, G., Zhang, Q., Lucas, K., Sun, Y., Gao, J. A corpus-based investigation on noun phrase complexity in L1 and L2 English writing[J]. *English for Specific Purposes*, 2022, 67: 114-126.
- [10] Zhang, R. Developing morphological knowledge with online corpora in an ESL vocabulary classroom[J]. *Frontiers in Psychology*, 2022, 13: 927636.
- [11] Ege, F., Yuksel, D., & Curle, S. A corpus-based analysis of discourse strategy use by English-Medium Instruction university lecturers in Turkey[J]. *Journal of English for Academic Purposes*, 2022, 58: 101125.
- [12] Boyko, S. A., Koltsova, E. A., & Spiridonova, V. A. Application of a corpus-based approach in teaching English for specific purposes to Master's degree students of engineering and technical majors[J]. *Global Journal of Engineering Education*, 2022, 24(1): 40-45.