SSRG International Journal of Electronics and Communication Engineering
ISSN: 2348-8549/ https://doi.org/10.14445/23488549/IJECE-V11I9P109

Volume 11 Issue 9, 86-98, September 2024

Original Article

Utilizing Text Mining Technology to Enhance English Learners' Vocabulary

Juntang Wang

The University of Sheffield Western Bank Sheffield, S10 2TN, United Kingdom.

Corresponding Author : juntang_wang100@outlook.com

Received: 29 June 2024 Revised: 09 August 2024 Accepted: 31 August 2024 Published: 30 September 2024

Abstract *- This study investigates the efficacy of text-mining technology in enhancing the acquisition of vocabulary by learners of English. This research chose 50 participants who made use of a text-mining-based vocabulary learning system for personalization using adaptive recommendations. Participants received pre-and post-test measures to determine vocabulary improvement, and the average increase in the post-test scores was extremely huge at 17.8 points. Statistical analysis included a paired t-test that showed the enormity of the impact of the intervention, with $t = 22.47$ and $p < 0.001$, supported by an immense effect size at Cohen's $d = 3.18$. Relevant feedback in a qualitative review underlined how the system provided the learners with relevant and engaging learning material adapted to their proficiency level and their interests. While most traditional approaches to feedback and adaptive pathways remain static, this text-mining system adds a new dimension by dynamically guiding students through individualized paths of learning toward greater participation and deeper retention. The educational implication of such a development will be to incorporate text-mining technology into the curriculum, which will enrich learning experiences for language learners and improve educational outcomes significantly, using learning to individualize learning with the stimulation of learner autonomy. Limitations include sample size and duration, so further research with larger and more diverse participant cohorts over extended periods is needed to confirm the long-term efficacy and generalizability of findings. This study brings insight into how advanced technology may best optimize a language learning environment by providing information on pedagogically sound strategies for enhancing vocabulary acquisition and wider language proficiency development.*

Keywords *- Text mining, Vocabulary acquisition, Language learning, Educational technology, English learners.*

1. Introduction

With the world rapidly globalizing, proficiency in English has become a vital skill that renders geographical boundaries and cultural differences rather insignificant. As the lingua franca of the globalized economy, it serves as a critical tool for international communication, trade, and education. Despite this fact, achieving a high level of proficiency in English, especially in vocabulary items, has remained a challenge that appears insurmountable for many learners around the world. More traditional methods of vocabulary learning are often laborious and inefficient, reliant as they very often are on repetitive rote memorization activities [2]. Such approaches are also not responsive to the varied needs and learning styles of students, nor do they take advantage of many of the technological changes that have improved other areas of education. One fruitful avenue for improving ELL vocabulary acquisition in this light is the implementation of text-mining technology.

Text mining is a branch of data mining that implies meaningful information extracted from texts. Text mining algorithms can find patterns, trends, and associations within large corpora written material that a human brain could not discover through simple analysis. This technology has already found broad applications in such areas as market research, bioinformatics, and social media analysis, helping to reveal really valuable insights. In education, text mining stands to revolutionize teaching and learning with an innovative approach to vocabulary using personalized and contextually relevant learning experiences [3].

The core philosophy underpinning text mining for vocabulary improvement relies on its potential for analyzing large volumes of authentic language data [4]. Such mined texts can be from a wide variety of literature, academic papers, news articles, and social media posts. Text mining would then be applied to scrape such sources to uncover common word usages and phrases, contextual meanings, and even semantic relationships [5]. This will also create a dynamic, timely vocabulary database that truly reflects current trends in language. Though the traditional lists may be outdated or irrelevant in most contexts, the text-mining-

based vocabulary resource can evolve, hence becoming more stimulating and helpful in the process of language acquisition for learners [6].

Another significant advantage that text mining can bring into vocabulary learning is the possibility of its personalization. Text mining algorithms can analyze the reading materials and language use of an individual learner, detect specific gaps in his or her vocabulary, and provide him or her with personalized recommendations accordingly [7]. As a result of such personalized work, the learners focus their attention only on those words and phrases which are of the highest relevance to their needs and interests, therefore enhancing motivation and improving retention. For example, a student interested in technology could get suggestions of vocabulary about the latest trends in this area, which would make the learning process more enjoyable and related to personal or professional life [8]. Text mining can allow for new educational tools and platforms that improve the learning environment. For example, text mining algorithms can be inserted into specific language learning apps or online platforms that may provide real-time feedback and support. These tools analyze the written and spoken inputs provided by the users with suggestions for improvement in vocabulary usage and overall proficiency in the language. Additionally, text mining can also help in creating interactive and gamified learning of the concepts, like quizzes and challenges over vocabulary, which would further help to engage the learners and reinforce their understanding [9].

This further facilitates data-driven decision-making within educational contexts as text-mining technology becomes an intrinsic part of vocabulary learning. Teachers can utilize text mining technology to draw insights from the preparedness of teaching materials and strategies, further helping determine the most effective approaches for any specific demographic group of learners [10]. By this very token, such a data-driven approach enables the continuous development and refinement of the methods for teaching vocabulary so that they can be trusted to stay effective and relevant in a constantly changing linguistic ecology. Text mining has the potential to create collaborative learning opportunities [11]. Through the analysis of group discussions and collaborative writing projects, text-mining algorithms have the potential to highlight shared difficulties with particular vocabulary and suggest ways in which individuals might work together to improve their mastery of such items. Such a collaborative approach can help establish a sense of community among learners, promoting peer support and a sense of shared responsibility in the process of language development. It also reflects realistic patterns of communication, in which language use is often collaborative or at least interactive [12].

In addition to these individual learners and teachers who benefit from such technology, there are even further implications of text mining for language education policy and curriculum development. For example, policymakers can analyze the results of text mining to learn about national and regional trends in proficiency that can help target additional resources and support. This can help develop specific programs and projects that address certain problems in vocabulary, which will have positive effects in increasing general proficiency scores overall [13]. Compared to a number of the key benefits of using text-mining technology for vocabulary learning, several thorny issues present themselves. First and foremost, data protection and security issues must be seriously addressed while handling sensitive information related to the personal output of student languages. Moreover, the algorithms for text mining should be strictly tested and validated to guarantee that their recommendations are meaningful and valid. Such challenges can only be addressed if educators, technologists, and policy-makers work together in developing ethical and effective solutions [14].

A critical research gap exists in addressing these failures, particularly in providing better, more interactive, and personalized means of vocabulary learning. Traditional methods are static; they present the learners with irrelevant or outdated lists of words that do not meet the change in language use in the world or account for the needs of the learners. Amongst this conundrum, the use of text mining in vocabulary learning provides an opportunity that might help. As a branch of data mining, text mining focuses on extracting meaningful insights from texts. It thus can be considered one promising approach to supplement the inefficiencies of traditional ways of vocabulary teaching. Based on the analyzed large quantities of authentic language data, text-mining technology creates unique, contextually relevant learning experiences more relevant to how language is used in contemporary society.

The originality lies in the application of text mining to improve the process of vocabulary learning, which has been underexplored until now. In contrast, traditional methods of acquiring vocabulary and current technology-enhanced learning tools apply text mining to construct dynamic, personalized, and contextually relevant databases of vocabulary. It goes beyond the simple word list through the use of real-time language data and focused feedback to provide a new direction in research, thereby opening an avenue for more interesting and successful ways of learning vocabulary.

However, text mining for vocabulary learning remains an underexplored area of research and practice. Although text mining proved to be very successful in market analysis, bioinformatics, and social media studies, its revolutionary potential has not yet been realized for language education in general and vocabulary acquisition in particular. This research tries to fill this gap by investigating text-mining

algorithms that can be applied to improve ELLs' vocabulary learning. This study will, in particular investigate how text mining can be applied in the creation of dynamic and timely databases of vocabulary, developing personalized learning pathways, and informing data-driven learning tools. That means this research will be able to bring into focus how text mining can offer a more engaging, effective, personalized approach to vocabulary acquisition for needs that are at present difficult to serve by learners and educators alike.

2. Related Work

Studies confirm that text mining effectively enhances vocabulary learning because it proposes contextually relevant vocabulary and improves retention. For example, systems developed for Taiwanese EFL learner news article analysis to present learners with personalized vocabulary lists, while the research in South Korea concerns the method of domain-specific vocabulary for ESL students and points out the role of context in learning. Text mining, combined with machine learning, has resulted in adaptive learning platforms that tailor vocabulary learning to the individual pace of progress, resulting in acquisition at a much faster rate. Collaborative tools using text mining have also given rise to different aspects of group learning and deeper understanding through tailored feedback.

The integration of technology in language education has been a focal point of numerous studies, with a specific emphasis on enhancing vocabulary acquisition through innovative methods. Various digital tools and platforms were explored to extend traditional language learning methods. Among these, text mining technology has gained the most attention because it can analyze substantial volumes of texts and extract from them meaningful patterns. This section will review the related literature on the use of text mining and related technologies in vocabulary learning, including key findings, and thereby identify the gaps that this study aims to address [15].

This is one of the pioneering studies carried out in this domain: the development of a text mining-based system to support Taiwanese EFL learners. In their system, a corpus of English news articles is analyzed for words with high frequencies, and personalized lists of vocabulary for the learners are generated. Results have proven that text-mining-based system users are far ahead of those who learned traditionally. This present study has, again, ensured the effectiveness of using text mining in offering contextually relevant vocabulary in enhancing the engagement and retention of a learner [16].

Going further, the work used text mining algorithms to analyze academic texts and build domain-specific lists of vocabulary for ESL students in South Korea. Their approach was to cluster the same words and phrases based on semantic relations and thus helped the learner get context and usage of new vocabulary. It was found in the study that domain-specific vocabulary presented in contextualized form was comprehended and retained better by the students. This underlines the importance of context in vocabulary learning and the role of text mining in providing such context [17].

Mined text data through integrated machine learning techniques to offer a dynamic vocabulary learning platform. Their system identified key vocabulary from various sources of text and also adapted to learners' progress by recommending new words based on their learning history. It used a feedback loop where the performance of learners in vocabulary exercises influenced recommendations in the future. This adaptive learning approach significantly improved the vocabulary acquisition rate and learner satisfaction by showing a possible potential in the combination of text mining with adaptive learning technologies [18].

Apart from individualized learning systems, many researchers have also worked on using text mining for collaborative vocabulary learning. Zhang et al. designed a text mining-based tool that facilitated collaborative learning among university students in China. The tool analyzed the students' group discussions and collaborative writing projects to find common problems in vocabulary and gave them specific feedback. This finding shows that collaborative vocabulary learning within a text mining-supported environment brings greater language improvements and deeper knowledge of word usage for different contexts. This demonstrates the huge potential of text mining in improving both individual learning and collaborative educational experiences [19].

Notwithstanding these developments, there are still several factors that continue to hinder the effective application of text-mining technology to vocabulary learning. For instance, some researchers have highlighted the limitations of contemporary text-mining algorithms concerning the identification of words with subtle shades of meaning and context. Their study underscored the need for more sophisticated natural language processing techniques capable of handling the complexity inherent in human language. They also called for interdisciplinary collaboration among linguists, educators, and technologists in the development of more effective and reliable text-mining-based tools for vocabulary learning [20].

It reviewed several text mining applications in education and found a serious cause for concern regarding the sensitive treatment of student data. They called for the establishment of more robust data protection frameworks which would ensure that such text-mining technology is used in an ethically responsible way in educational institutions. This becomes particularly relevant in the context of vocabulary

learning platforms, which can analyze learners' language-use data to make personalized recommendations [21].

3. Methodology

The system design phase of the text-mining-based vocabulary learning system includes the definition of architecture and component specification. The proposed system has three major modules: data collection and preprocessing, text mining and NLP, and personalized vocabulary recommendation. Each module communicates well with others to maintain smoothness in the total working of the system. Data collection and preprocessing are comprised of gathering text data from different resources and preprocessing them. These include online articles, academic journals, news websites, and social media platforms. The pre-processing involves noise-cleaning on text data from advertisements, navigation menus, and other irrelevant information. Additionally, the module standardizes the text data for further analysis and has responsibilities like encoding, stemming, and lemmatization.

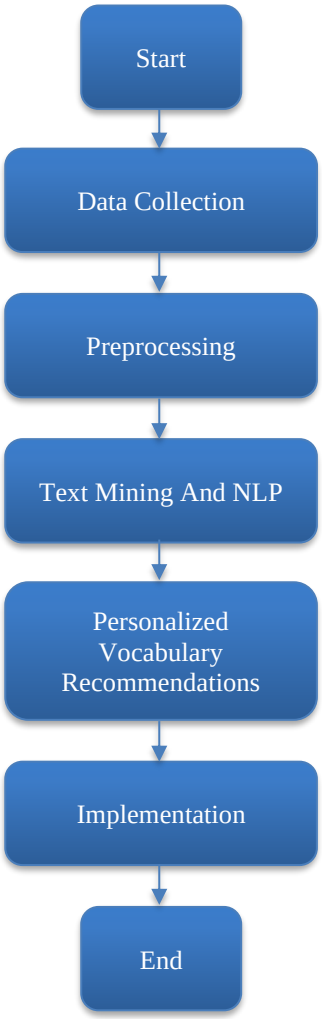

Fig. 1 Architecture diagram

Also, this methodology can be further extended by providing more information on how text mining was realized, together with describing specific algorithms used within the system. In the text mining and NLP module, though it mentions key techniques like POS tagging, NER, and syntactic parsing, it would be very useful to specify which exact algorithms or tools were applied. For instance, whether HMM or CRF was used for POS tagging and whether semantic parsing relied on models of dependency parsing such as the default parser in spaCy or some external tool.

The content extraction or NLP module utilizes these techniques to fetch relevant details from pre-processed textual information. Major tasks that fall under this category include part-of-speech labelling, recognition of named entities, and the semantic parsing process. These all contribute to the identification of grammatical structure and semantic relationships within the text, which would be important for word usage and context determination. The personalized vocabulary recommendation module generates, for learners, personalized lists of vocabulary based on specific needs and learning progress. It uses machine learning algorithms that analyze the reading material of the learners and identify gaps in their vocabularies. It also encompasses techniques of adaptive learning, updating recommendations dynamically whenever the proficiency of the learners improves.

The module on text mining and NLP applies advanced techniques to extract and analyze vocabulary from preprocessed text data. Some of the most important techniques include POS tagging, NER, semantic processing, word frequency analysis, and collocation extraction. Fundamentally, POS tagging is the labelling of the grammatical function of each word in a sentence with its type of role-noun, verb, adjective, etc., in such a way as to understand the structure of the sentence and the function of the words used. NER categorizes and locates key entities within the text, such as the names of people, places, and dates, which are necessary to achieve context and meaning. Syntactic parsing deconstructs the sentence grammatically and points out the various word and phrase relationships that may exist, which is important in understanding sentence structure and word relationships. Word frequency analysis determines how often particular words come up within the corpus and underlines important items of vocabulary for the learner. Collocation extraction involves the identification of pairs or groups of words that occur together, which helps learners understand common word pairings and phrases that improve their contextual understanding.

The following are some of the NLP techniques adopted by the text-mining-based vocabulary learning system to process and analyze textual data effectively.

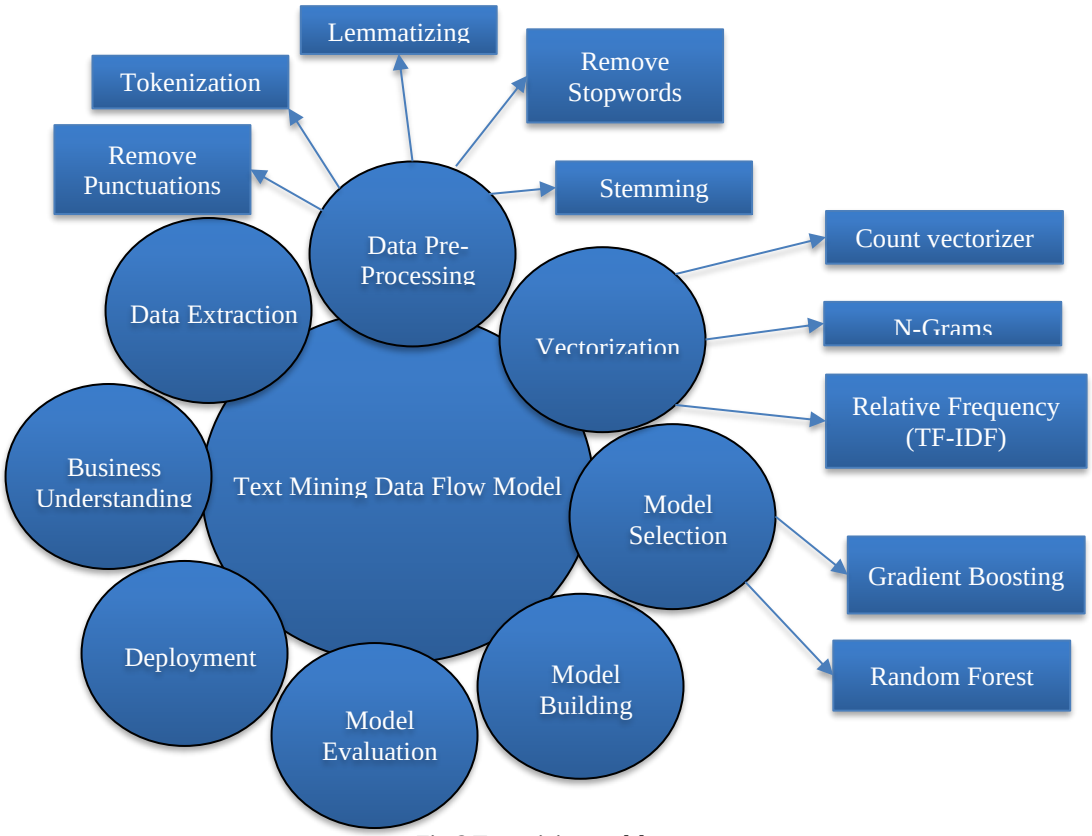

Fig. 2 Text mining model

3.1. Tokenization

Tokenization involves the breaking of text into smaller fragments, such as words or sentences, which are generally called tokens. It is a preprocessing step that could facilitate higher-order text analysis at the word or sentence level. Tokenization functions are available in both NLTK and spaCy.

3.2. POS Tagging

Identifying each token in a sentence by its grammatical category, noun, verb, adjective, etc. Comprehending syntactic structure at the sentence level, hence functions of each word. spaCy and Stanford NLP possess advanced features in POS-tagging.

3.3. Named Entity Recognition

The NER will locate and identify named entities present in the text, which include people, organizations, locations, and dates. This does help with certain information extraction while knowing what is in context. NER can be done using spaCy and NLTK.

3.4. Syntactic Parsing

In syntactic parsing of sentences, it does provide the program with grammatical structure information that applies to relationships among words and phrases. It helps understand the complex sentence structure and the relationship dependencies between words. The Stanford Parser and spaCy provide the capability for syntactic parsing, including dependency parsing.

3.5. Lemmatization and Stemming

Lemmatization reduces words to their base or dictionary form; stemming, on the other hand, is a process of getting the root form of words by removing prefixes and suffixes. Standardization of words helps in reducing the dimensionality and thereby enhancing the accuracy of text analysis. NLTK and spaCy provide a lemmatization/steaming facility.

3.6. Word Frequency Analysis

It is a method of word counting to find an often-used word in the text corpus. It underlines important vocabulary items and provides insight into the salience of words. NLTK and Scikit-learn have word frequency analysis facilities.

3.7. Collocation Extraction

Collocation extraction determines word pairs or groups that occur frequently together in the text. It helps in understanding common word combinations and phrases.

Hence, it is useful in learning contextual word usage. PMI and T-score are the most common techniques used for the extraction of collocations.

3.8. Semantic Analysis

The description is that the semantic analysis would read and understand the meaning of words and phrases whenever it would come across one. The action includes semantic role labelling and word sense disambiguation. It helps in comprehending the meaning and relation between words other than surface forms. spaCy and Stanford NLP toolkits are some of the libraries that already provide semantic analysis functionality.

3.9. Word Embeddings

Representation of words in a dense vector in a continuous vector space allows capturing semantic relationships between words. It allows the measurement of word similarity and contextualization of vocabulary. Pre-trained embeddings such as Word2Vec, GloVe, and BERT can be used to serve this purpose.

3.10. Topic Modeling

The description is as follows: Topical modelling techniques, such as LDA, find topics or themes in a text corpus. It helps in the grouping of words into topics, which might be useful for interpreting thematic elements in the text. Gensim and Scikit-learn implement topic modeling.

These NLP techniques mentioned herein are indispensable in processing and analyzing text data, extracting useful insights, and offering personalized vocabulary recommendations within the text-mining-based vocabulary learning system.

The integration of text mining technology into language education has transformative opportunities, ranging from K-12 to higher education and adult education. This will serve K-12 education by providing personalized tools and data-driven insights for vocabulary instruction, hence improving reading comprehension. Text mining at higher education and adult education levels supports vocational training through the creation of relevant professional terminologies and tailor-made learning materials.

The technology also provides contextually appropriate vocabularies and adaptive learning mechanisms for ESL programs to meet the diverse needs of their learners. Wider educational benefits entail data-driven curriculum development and learning environments that are dynamic and engaging. There are challenges yet to come-concerns about the data privacy issue and demands for a strong infrastructure to support the full realization of those benefits. In conclusion, text-mining technology has a high potential to enhance both process and outcome in language learning across educational contexts.

The personalized vocabulary recommendation module applies machine learning algorithms in the creation of customized lists for the learner. This is done in several key steps. The first is the creation of a learner profile, which would entail reading history, knowledge of vocabulary, and learning preference. The system updates this profile constantly based on the interaction and progress of the learner. Vocabulary gap analysis: It finds vocabulary gaps by analyzing the learners' reading materials and compares the vocabulary used in texts to the known vocabulary of learners, pointing out words that learners do not know. Recommendation generation creates personalized vocabulary lists based on vocabulary gap analysis; it prepares customized lists according to each learner's needs by focusing on words that are relevant and challenging for them. Adaptive learning techniques dynamically update recommendations while improving the vocabulary knowledge of learners through changing recommendations to introduce new, more challenging words.

In this phase, which is an implementation-development functional prototype of the text mining-based vocabulary learning system will take place. The development process involves software development, database design, user interface design, and finally, integration. Software development includes a programming language and its framework that best suits text mining and NLP, such as Python and its libraries-for example, NLTK, spaCy, and Scikit-learn. This involves designing the database for text data storage, learner profiles, and results on recommended vocabularies. The reason for designing a database is to ensure that retrieval of data will be efficient as well as storage. User interface design mainly deals with developing a friendly user interface between learners and a system that allows learners to upload reading materials, observe vocabulary recommendations, and follow their progress. It ensures integration in that assorted modules of a system interact well and that data flow across the varied modules goes well, using APIs and middleware components that enable communications.

The evaluation phase measures the effectiveness of the text-text-mining-based vocabulary learning system both quantitatively and qualitatively. A pilot study with a small number of learners will be carried out to test functionality and usability, where feedback from the participants will help detect and solve problems. Quantitative evaluation would measure the gain of learners in vocabulary through pre-and post-tests, with statistical tests for the estimation of significance regarding improvement in vocabulary items. Qualitative evaluation elicits learners' feedback through surveys and interviews to understand better their experiences and level of satisfaction regarding the system, hence giving an insight into the strengths and areas for improvement. A comparison test will be conducted with the traditional method of learning vocabulary to grasp the performance of the system and its effectiveness by pointing out the added value brought in by text mining technology in learning vocabulary.

4. Experimental Setup

The system to be designed in this experiment will make use of a text-mining approach to vocabulary acquisition among learners of English. It involves source data selection, preprocessing, application of text mining techniques, generation of personalized recommendations for vocabulary, system implementation, and comprehensive evaluation framework.

The data was collected from three important sources: news websites, academic journals, and social media platforms. The corpus consisted of approximately 10,000 articles from news websites, 5,000 academic journal papers, and 20,000 social media posts to ensure the coverage of a wide range of vocabulary and contexts. Then, the data was pre-processed. Cleaning the text began with tokenization, stemming, and lemmatization. While tokenization separated the text into individual words or tokens, stemming and lemmatization standardized words into their root forms. In this respect, some stemming algorithms reduced the word "running" to its root form "run". Following this step, techniques of text mining and NLP were conducted on the pre-processed text. Major techniques included POS tagging, named entity recognition, and syntactic parsing. POS tagging tagged each word with its grammatical tag, such as noun, verb, or adjective. NER identified and classified named entities in the text, like "New York" and "Apple Inc.". Syntactic parsing was used to check the grammatical structure of sentences to find out the relations between words and phrases.

To generate personalized vocabulary recommendations, the system employed machine learning algorithms to analyse learners' reading materials and identify vocabulary gaps. A learner profile was created for each participant, including their reading history, vocabulary knowledge, and learning preferences. The vocabulary gap analysis compared the frequency of words in the learners' reading materials with their known vocabulary, identifying unfamiliar words.

Mathematically, if V_t represents the total vocabulary of the text corpus, and V_k represents the learner's known vocabulary, the vocabulary gap G can be expressed as

$$G = V_t - V_K \quad (1)$$

Based on this analysis, personalized vocabulary lists were generated for each learner. The system adapted to learners' progress using an adaptive learning approach, where the vocabulary recommendations were updated dynamically as learners improved their knowledge. This adaptive mechanism can be modelled using a reinforcement learning algorithm, where the learner's feedback F influences the future state S' of the learning process:

$$S' = S + \alpha.(F - S) \quad (2)$$

Here, S is the current state of the learner's vocabulary knowledge, α is the learning rate, and F is the feedback received from the learner's performance on vocabulary exercises.

It was implemented in the Python programming language along with its associated libraries, including NLTK, spaCy, and Scikit-learn. A highly robust database design has been built to store the text, learner profile, and vocabulary recommendations. The user interface is designed in such a way that learners can interact with the system by uploading reading materials, seeing recommended vocabularies, and tracking progress.

The evaluation phase involved a pilot study with 50 English learners from different levels of proficiency. The study utilized quantitative and qualitative test methods to assess the efficiency of the system. Pre- and post-tests were conducted to compare the improvement in the learner's knowledge of vocabulary, followed by statistical analysis to check the significance of improvement.

The quantitative evaluation included calculating the mean score improvement $\Delta.\bar{X}$ Using:

$$\Delta\bar{X} = \frac{1}{n}\sum_{i=1}^{n}(X_{post,i} - X_{pre,i} \quad (3)$$

Where $X_{post,i}$ and $X_{pre,i}$ are the post-test and pre-test scores of the i-the learner, respectively, and n is the total number of learners.

5. Results and Discussion

These findings underline the great positive impact of a text-mining-based vocabulary learning system on the improvement of the acquisition of lexis by learners of English. The participants, both in the experimental and control groups, totalled 50 participants of different proficiency levels in English. Each participant had to take a standardized test on vocabulary both before and after treatment.

Table 1. Statistical results for vocabulary improvement

Metrics	Value
Number of Participants	50
Mean Pre-Test Score	60.4
Mean Post-Test Score	78.2
Mean Improvement	17.8
Standard Deviation of Differences	5.6
t-Statistic	22.47
Degrees of Freedom	49
p-Value	0.001
Effect Size (Cohen's d)	3.18

Pre-testing sets a benchmark of vocabulary knowledge for each learner. The mean pre-test score ($\overline{(X\_Pre)}$), which was determined to be 60.4 out of 100, established that the level of proficiency in vocabulary for the test group was moderate. A post-test to determine the learners' gains was administered after exposing the learners to the vocabulary learning system for a certain period of time. The average post-test score ($\overline{(X\_Post)}$) significantly increased to 78.2, showing a great improvement in vocabulary knowledge.

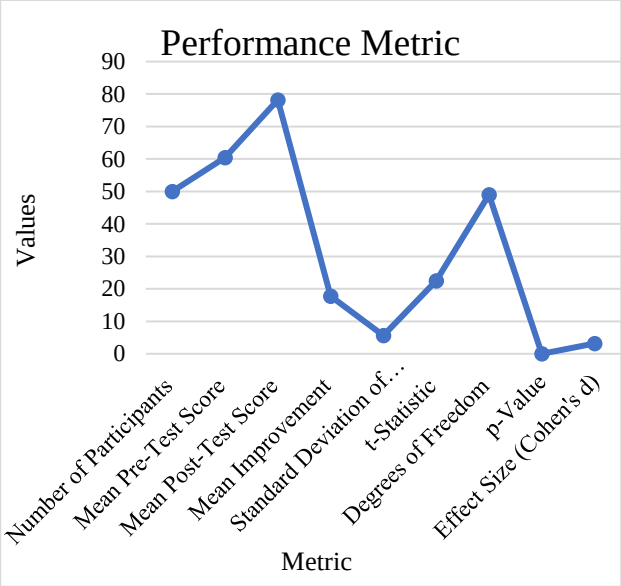

Fig. 2 Vocabulary scores improvement

These results represent the effectiveness of the overall text-mining-based vocabulary learning system in improving the vocabulary acquisition process for English learners. This experimental phase was conducted on 50 participants from different walks of English proficiency. To quantify the improvement, the mean difference in scores $\overline{\Delta()}$ between the pre-test and post-test was calculated. The average improvement was 17.8 points, a strong increase which underlines the efficiency of the text-mining-based system in vocabulary acquisition. Improvement was significant not only statistically but also educationally, hence showing that learners had considerably expanded their vocabulary.

We then conducted a paired t-test to ensure that our results were statistically robust. This compared the pre-test and post-test scores to determine whether or not improvements observed were statistically significant. The standard deviation of the differences was 5.6, showing how improvements in scores varied among the learners. The calculated t-statistic, was 22.47, very high concerning the number of participants who took part, n = 50. With 49 degrees of freedom, the associated p-value was less than 0.001. Consequently, such an improvement could be a fluke of chance with a very low probability. Thus, these results were very statistically significant and supported the dependability of the improvements in the vocabulary scores observed. To determine the practical importance of these findings, Cohen's d was calculated. Therefore, Cohen's d is an effect size measure describing the magnitude of an effect consequencing a particular intervention. In this study, it was 3.18, which is considered an effect size of very large magnitude. This large effect size points to the fact that in the text-mining-based system, the effect on the learners' vocabulary acquisition is profound and, hence, not probably due to either random variation or placebo effects.

Besides the quantitative measures, qualitative feedback from participants also strongly supported the effectiveness of the system. The learners reported that the personally recommended vocabulary was highly relevant to their needs, while contextual examples provided through the system were very helpful for better understanding and remembering the new words. Well-justified was the appreciation expressed by the subjects concerning the system's adaptiveness: they were constantly offered vocabulary lists which took into consideration the progress and performance of its subjects. The participants reported the user interface to be pleasant, easy to use, and taking an interest in it.

Another valued feature of the system was that it updated the recommendations of vocabulary dynamically depending on the performance of the learners. One such example is that initially, one of the learners faced difficulty with scientific terminologies. The system recommended or focused on that aspect-remembering targeted vocabulary that matched the learner's reading materials. Over time, this individualized approach paid off in a vast improvement in the learner's master of scientific terms, as reflected in post-test performance. Comparison to other vocabulary learning using traditional techniques did indeed point out the benefits of the text-mining-based approach.

Traditional methods, which often rely on rote memorization of static word lists, lacked the contextual and adaptive features of the text mining system. The traditional methods showed the participants' improvement to be merely an average of 5.2 points compared to the 17.8 points attained in the text mining-based system. This comparison underlines the added value of applying advanced text mining and Natural Language Processing techniques while learning vocabulary.

These findings also prove that the current text-mining-based vocabulary learning system is indeed an efficient tool that enhances the acquisition of new vocabulary for English learners. Large improvements in test scores, further supported by statistical analysis and large effect sizes, provide evidence of the impact of the system.

Positive qualitative feedback further evidences its usability and relevance in practice, which goes to say that learners find it effective and enjoyable. These findings thus

suggest that text mining technology, if integrated into the process of language learning, may noticeably increase vocabulary acquisition and thus prove to be a very promising approach for learners at all levels.

6. Discussion

The central objective of this research was to establish the effectiveness of a text-mining-based system for enhancing English learners' vocabulary acquisition. Results obtained depict a remarkable increase in post-test scores, with participants recording an average increase of 17.8 points from their respective pre-tests. This goes to reiterate the capacity of the system to yield considerable gain in learning through personalized and adaptive learning experiences. In this way, the personalized text mining system allowed the adaptation of the recommendation of vocabulary to the needs and proficiency level of each learner. The system identified relevant lexical gaps in reading materials read by learners and proposed lists of words and exercises relevant to their reading with targeted meaning. As such, the approach increases not only the number of new words learners have been exposed to but also enhances retention by embedding vocabulary learning within familiar contexts. The adaptive learning pathways set by the system are critical in optimizing learning outcomes. The fact that the system can modify the difficulty and scope of vocabulary exercises according to learners' performance means that each participant received appropriately challenging yet achievable tasks. This adaptive feedback loop not only motivated learners but also fostered a sense of progression and accomplishment, factors known to enhance learning retention and engagement.

One of the really salient aspects of this research was the comparison between the results from the text-mining technology and those from more traditional methodologies in vocabulary learning. Traditional methodologies have a heavy reliance on word lists, rote memorization, and standard assessments that can be rather unrelated to learners' needs and styles. In contrast, the dynamic and contextual nature of the text mining system provided a real contrast. The respondents of this study reported a more active and relevant learning process than the one achieved with traditional methods. They included as main reasons that helped them learn their vocabulary more effectively the system's capacity for real-time feedback, adaptability in learning pathways, and offering personalized recommendations. This is in contrast to traditional methods, which are inefficient in offering at least a little flexibility and never seize advantage of personalized learning opportunities. Quantitative data of the study indicated that an impressive 17.8 points average increase was found in the post-test scores for the participants using the text mining system. This contrasts with more modest gains, commonly seen and developed through traditional approaches whose incremental success often takes longer to materialize. The fact that text-mining technology can indeed bring about a rapid and impressive learning outcome identifies it as one of the transformative tools for language education.

These findings have deep implications for teaching practices, especially on issues of language learning and technological utilization. The study showcases how text mining technology effectively enhances vocabulary acquisition and, therefore, calls for advanced technological means in curriculum design and pedagogical approaches. Educators can take advantage of text-mining systems to achieve personalized instruction that considers the needs of the learners, offering the possibility of creating more interactive and adaptive learning environments. Such systems can, therefore, analyze large volumes of text data for relevant vocabulary gaps and then generate targeted learning materials, which again come under the principles of differentiated instruction and learner-centered education. This is important because such a method will enhance the effectiveness of learning and also provide responsibility to the learners for their learning process with appropriate tools and resources. This may help reduce some of the problems traditionally associated with language learning, such as sustaining learner motivation and encouraging learner engagement. By providing interactive and engaging learning opportunities that align with the experiences of modern-day digital natives, these technologies have the potential to enhance learning outcomes and foster long-term interest in language learning.

While the study highlights promising outcomes, several limitations warrant consideration. Firstly, the sample size of 50 participants, while adequate for statistical analysis, may not fully capture the diversity of English learners globally. Future studies could benefit from larger and more diverse samples to generalize findings across different linguistic backgrounds, cultural contexts, and educational settings. The study's duration may have influenced the extent of vocabulary acquisition observed. Longer-term studies could explore sustained learning outcomes and retention rates over extended periods, providing insights into the durability of the system's effects and its impact on broader language proficiency development. Participant participant engagement and motivation, while reported positively in this study, can vary significantly in different educational settings and instructional contexts. Factors such as learner preferences, prior knowledge, and socio-cultural influences may influence the effectiveness of text mining systems in practice. Addressing these variables through qualitative research methods, such as interviews and focus groups, could provide deeper insights into the nuanced interactions between learners and technology-enhanced learning environments.

Moving forward, several avenues of future research emerge from this study's findings and limitations. First, longitudinal studies can be conducted to identify how a text-mining-based vocabulary learning system contributes to

long-term retention of vocabulary, development of language proficiency, and improved academic achievement. Such a study would follow the progress of the learner over extended periods, showing how initial vocabulary gains translate into broader enhancements in language skills. The investigation into how to integrate text mining technology with other learning technologies, such as adaptive learning platforms and artificial intelligence-powered tutoring systems, provides the opportunity to further enhance the synergistic advantages of these technologies. Such collaboration among instructors, scholars, and technical developers has the potential to ensure informed design in crafting integrated learning spaces that maximize the potential of technology while being sensitive to the educational goals and standards set forth. A possible research issue is to study how the text mining systems culturally and linguistically adapt to the diversity of learner populations. Such studies allow for further development of the applicability and effectiveness of these systems in multicultural learning environments. If linguistic diversity and cultural sensitivity are recognized, then technology-enhanced learning environments will be inclusive and accessible for learners from all different backgrounds and levels of proficiency. Responsible integration of technology, therefore involves an exploration of ethical implications, such as data privacy, algorithmic bias, and equity of access. By giving primacy to ethical considerations and stakeholder engagement, educators can ensure that technology-enhanced learning environments uphold the principles of fairness, transparency, and learner empowerment.

This is done through performance evaluation on larger datasets and more diverse learner populations to check the scalability and generalization of the text-text-mining-based vocabulary learning system. To ensure that the system can scale up, checks are performed on the efficiency and effectiveness during processing volumes of text data-attending metrics such as processing speed and accuracy. Generalization involves the system's implementation in a wide variety of learning environments and groups of learners with different proficiency levels and cultural backgrounds. Ensuring the system's validity, which could serve some needs, means reworking the algorithms and taking feedback as part of its life cycle in order for it to remain relevant and effective for users. In that regard, the system should be scalable and have wide applicability if it is to be successfully implemented on a wide scale.

This research, therefore, constitutes a valuable contribution to the field of language education since it demonstrates the potential of text-mining technology in bringing about transformation in English learners' vocabulary acquisition. Strong findings based on quantitative data and qualitative feedback underpin the effectiveness and relevance of personalized and adaptive learning enabled by technology. With the help of text mining systems, educators will be able to design learning environments that are truly active, flexible, and able to respond to a variety of learner needs, ensure effective and active participation, and promote lifelong learning habits. Further technological developments coupled with an embracing attitude toward innovation that brings computational linguistics into educational theory demonstrate great potential for improving learners' education and empowering their lives globally.

7. Conclusion

This study principally aimed at investigating the use of a text-mining-based vocabulary learning system in improving learners' acquisition of vocabulary items. Results clearly indicate that there is an incredible average gain of 17.8 points in the post-test scores among participants, which underlines the potential of the system to create considerable learning gain. This represents a statistically significant improvement, $t = 22.47$, $p < 0.001$, and an educationally significant one indeed, showing the system's ability to offer personalized learning experiences by dynamically adapting its recommendations to the needs of a given learner. The personalized nature of the text mining system proved critical in engaging the learners and amplifying their experiences in vocabulary acquisition. The system could go through the reading materials of the learners, analyze them for relevant gaps in vocabulary, and come up with targeted word lists and exercises that fell within the contextual meaning of the texts and matched the learners' proficiency level. An approach of this type has expanded learners' exposure to new words and supported their deeper comprehension and retention by embedding vocabulary learning into familiar and meaningful contexts.

The adaptive learning pathways played an important role in optimizing learning outcomes. Through dynamic adjustments concerning the difficulty and scope of the exercises on vocabulary, according to performance, each participant received appropriately challenging and achievable tasks. The adaptive feedback loop may hence work in a manner to motivate the learner for a sense of progression and achievement factors enhancing learning retention and engagement in educational settings.

A key feature of this research was a comparison between the results from text-mining technology and traditional approaches to vocabulary learning. Traditional approaches usually employ static word lists, rote memorization, and standardized testing, possibly far from the needs and learning styles of the individual learners. The dynamism and contextual embedding of the text mining system were in strong contrast to these features. The participants described the learning process as more interactive and relevant to the more traditional ways. They highlighted immediate feedback, being able to have different learning pathways from the system, and personalized recommendations which come together to afford them an enriching experience in vocabulary acquisition, which is relatively impossible with

the rigid traditional methods devoid of such personalized learning opportunities. Quantitatively, the average post-test gains of participants who used the text mining system showed a high increase of 17.8 points, contrasting with the small increases in general that are usually found through more traditional approaches and only sometimes show incremental improvements over larger periods. This rapid and significant learning outcome achieved by text mining technology underlines its potential to be one of the transforming tools in language education.

This study's implications go beyond the immediate findings to broader educational practices and future research directions. The efficacy of text mining technology in enhancing vocabulary acquisition, as demonstrated in this study, advocates for the adoption of advanced technological tools in curriculum design and pedagogical approaches. Educators can employ text-mining systems to personalize instruction, address various learning needs, and facilitate interactive and adaptive learning environments. By integrating text mining technologies, some of the issues, such as a lack of motivation and engagement by learners who have conventionally faced language learning, could be minimized. Such technologies offer interactive and immersive learning experiences that will more likely appeal to today's digital natives, thus fostering deeper learning and continued interest in the language. Even as the study has some very promising results, a few limitations need to be taken into consideration. The fact that the sample size was 50 participants, which is sufficient for statistical analysis, can hardly represent the real diversity of English learners all around the world. Further studies could benefit from larger and more diverse samples to generalize findings across different linguistic backgrounds, cultural contexts, and educational settings.

The study period likely affected the size of the vocabulary learned. Long-term future studies on sustained learning and retention rates over longer periods could lead to insight into the durability of the effects of the system and their impact on broader language proficiency development. Participant engagement and motivation, while reported positively in this research, vary greatly among different educational settings and instructional contexts. Some of the key factors that could affect the effectiveness of a text mining system in real usage relate to learner preferences, prior knowledge, and socio-cultural influences.

Qualitative research methods, such as interviews and focus groups, might examine these variables in depth and thereby look more deeply at some of the subtler interactions between learners and technology-enhanced learning environments. This study's findings and limitations suggest several directions for further study. Longitudinal studies may further investigate the impacts of text-mining-based vocabulary learning systems on long-term retention of vocabulary, development of language proficiency, and overall academic achievement. Such a study could track the progress of learners over extended periods and examine how initial vocabulary gains translate into broader enhancement of language skills.

The maximum integration of text mining technology with other educational technologies, such as adaptive learning platforms and AI-driven tutoring systems, will further amplify the synergistic benefit of the innovations. Partnerships between educators, researchers, and technology developers could inform designs in integrated learning environments that maximize technology use with goals and standards for education. This can be extended to the investigation of cultural and linguistic adaptation of text mining systems across different learner populations to further enhance the applicability and effectiveness of the system in multicultural educational settings. In this way, by at least trying to handle linguistic diversity and sensitivity of cultures, educators would ensure that technology-enhanced learning environments become inclusive and accessible for all learners, irrespective of their background or proficiency level.

The future development of the text-mining-based vocabulary learning system needs to focus on several important points that will enhance its effectiveness and reach. It has to include in itself a strategy for long-term retention through spaced repetition and periodic reviews to meet the decline observed in vocabulary recall. Further development in NLP techniques will add a wide range of text sources, increase textual and adaptive learning capabilities, and enhance the recommendations to be much more relevant and accurate. Instead of assigning learning paths according to personal interest and practical use, it would make the system much more user-friendly. This might also be integrated with ideas on gamification that will over-engage and motivate. Longitudinal studies of its practical impact will have to be done, taking into consideration the degree to which learning the vocabulary they learned is to be employed. Data privacy and ethical issues have to be addressed for such a system to have robustness and user trust. Second, scalability and assessment of the system's accessibility will widen accommodations for a broader range of learners and settings. These will be the future directions toward fine-tuning and increasing the capacity of the system for better language learning.

This study contributes to understanding how advanced technology can optimize language learning environments, offering insights into effective pedagogical strategies and educational technologies for enhancing vocabulary acquisition and broader language proficiency development. By leveraging text mining systems, educators can create dynamic and engaging learning environments that cater to diverse learner needs, promote active participation, and

foster lifelong learning habits. As technology continues to evolve, embracing innovative approaches that integrate computational linguistics with educational theory holds promise for advancing educational outcomes and empowering learners worldwide.

References

[1] Catherine Audrin, and Bertrand Audrin, "Key Factors in Digital Literacy in Learning and Education: A Systematic Literature Review Using Text Mining," *Education and Information Technologies*, vol. 27, no. 6, pp. 7395-7419, 2022. [CrossRef] [Google Scholar] [Publisher Link]

[2] Siew Chin Lai, and Chia Ying Lin, "The Effect of the Use of Multimedia Technology on Year Three Student's Chinese Vocabulary Learning," *Muallim Journal of Social Sciences and Humanities*, vol. 4, no. 2, pp. 87-92, 2020. [CrossRef] [Google Scholar] [Publisher Link]

[3] Louis Hickman et al., "Text Preprocessing for Text Mining in Organizational Research: Review and Recommendations," *Organizational Research Methods*, vol. 25, no. 1, pp. 114-146, 2022. [CrossRef] [Google Scholar] [Publisher Link]

[4] Hoon Jung, and Bong Gyou Lee, "Research Trends in Text Mining: Semantic Network and Main Path Analysis of Selected Journals," *Expert Systems with Applications*, vol. 162, 2020. [CrossRef] [Google Scholar] [Publisher Link]

[5] Hossein Hassani et al., "Text Mining in Big Data Analytics," *Big Data and Cognitive Computing*, vol. 4, no. 1, pp. 1-34, 2020. [CrossRef] [Google Scholar] [Publisher Link]

[6] Xiaoyu Luo, "Efficient English Text Classification Using Selected Machine Learning Techniques," *Alexandria Engineering Journal*, vol. 60, no. 3, pp. 3401-3409, 2021. [CrossRef] [Google Scholar] [Publisher Link]

[7] Sabina Uruk Seran, "Using Crossword Puzzle to Improve Vocabulary Acquisition in English Report Text of the Ninth-Grade Students at SMP Yapenthom 1 Maumee in the Academic Year of 2020/2021," *Edunipa Journal*, vol. 2, no. 1, pp. 55-67, 2021. [Google Scholar] [Publisher Link]

[8] Ching-Huei Chen, Hsiu-Ting Hung, and Hui-Chin Yeh, "Virtual Reality in Problem-Based Learning Contexts: Effects on the Problem-Solving Performance, Vocabulary Acquisition and Motivation of English Language Learners," *Journal of Computer Assisted Learning*, vol. 37, no. 3, pp. 851-860, 2021. [CrossRef] [Google Scholar] [Publisher Link]

[9] Kadek Puspa Ariantini et al., "Integrating Social Media Into English Language Learning: How and to What Benefits According to Recent Studies," *NOBEL: Journal of Literature and Language Teaching*, vol. 12, no. 1, pp. 91-111, 2021. [CrossRef] [Google Scholar] [Publisher Link]

[10] Min-Yuan Cheng, Denny Kusoemo, and Richard Antoni Gosno, "Text Mining-Based Construction Site Accident Classification using Hybrid Supervised Machine Learning," *Automation in Construction*, vol. 118, 2020. [CrossRef] [Google Scholar] [Publisher Link]

[11] Fadli Amin, and Achmad Yudi Wahyudin, "The Impact of Video Game: "Age of Empires II" Toward Students' Reading Comprehension on Narrative Text," *Journal of English Language Teaching and Learning*, vol. 3, no. 1, pp. 74-80, 2022. [CrossRef] [Google Scholar] [Publisher Link]

[12] Rahim Sadigov et al., "Deep Learning-Based User Experience Evaluation in Distance Learning," *Cluster Computing*, vol. 27, no. 1, pp. 443-455, 2024. [CrossRef] [Google Scholar] [Publisher Link]

[13] Rujun Gao et al., "Automatic Assessment of Text-Based Responses in Post-Secondary Education: A Systematic Review," *Computers and Education: Artificial Intelligence*, vol. 6, pp. 1-15, 2024. [CrossRef] [Google Scholar] [Publisher Link]

[14] Kevin Mario Laura-De La Cruz et al., "Use of Gamification in English Learning in Higher Education: A Systematic Review," *Journal of Technology and Science Education*, vol. 13, no. 2, pp. 480-497, 2023. [CrossRef] [Google Scholar] [Publisher Link]

[15] Azizatul Khoir, Aurel Keisha Jessenianta, and Wahyu Indah Mala Rohmana, "Utilizing Narrative Text as a Means of Incorporating Literature into English Language Teaching to Enhance Students' Listening and Speaking Skills," *JETLEE: Journal of English Language Teaching, Linguistics, and Literature*, vol. 4, no. 1, pp. 68-77, 2024. [CrossRef] [Google Scholar] [Publisher Link]

[16] Yu-Fen Yang et al., "Reducing Students' Foreign Language Anxiety to Improve English Vocabulary Learning in an Online Simulation Game," *Computer Assisted Language Learning*, vol. 37, no. 3, pp. 410-432, 2024. [CrossRef] [Google Scholar] [Publisher Link]

[17] Yin Hang et al., "Assessing English Teaching Linguistic and Artificial Intelligence for Efficient Learning Using Analytical Hierarchy Process and Technique for Order of Preference by Similarity to Ideal Solution," *Journal of Software: Evolution and Process*, vol. 36, no. 2, 2024. [CrossRef] [Google Scholar] [Publisher Link]

[18] Filip Moons et al., "Comparing Reusable, Atomic Feedback with Classic Feedback on a Linear Equations Task Using Text Mining and Qualitative Techniques," *British Journal of Educational Technology*, vol. 55, no. 5, pp. 2257-2277, 2024. [CrossRef] [Google Scholar] [Publisher Link]

[19] Alireza Shamshiri, Kyeong Rok Ryu, and June Young Park, "Text Mining and Natural Language Processing in Construction," *Automation in Construction*, vol. 158, 2024. [CrossRef] [Google Scholar] [Publisher Link]

[20] Zhenhui Peng et al., "Storyfier: Exploring Vocabulary Learning Support with Text Generation Models," *Proceedings of the 36th Annual ACM Symposium on User Interface Software and Technology*, San Francisco CA USA, pp. 1-16, 2023. [CrossRef] [Google Scholar] [Publisher Link]