Received: 09 Mar. 2026 | Revised: 10 Mar. 2026 | Accepted: 13 Mar. 2026 | Published online: 31 Mar. 2026

RESEARCH ARTICLE

Contemporary Education and Teaching Research
2026, Vol. 7 (3)68-76
DOI: 10.61360/BoniCETR262019790301

A Lexical Learning Case Study of Multimodality Discourse Analysis

Jun Lai[1,*] & Rui Xue[1]
[1]*School of International Studies, Sun Yat-sen University, P.R. China*

Abstract: The contemporary era of information technology has given rise to a novel approach to word retention: the utilisation of English vocabulary memory applications. With the support of mobile devices, English vocabulary memory applications have realised the convenience of word memorisation and attracted many users. The integration of multimodal theory into digital vocabulary applications has become an increasingly interesting topic in social science. However, limited research has explored systematic multimodal discourse analysis to examine the semantic construction through a complex interaction of different resources. The current study conducted a multimodal analysis of an English vocabulary memory application (No Need to Memorise Words) and analysed its functions to explore the use of multimodal means. The study employed the methodology of questionnaires to examine the overall perception of college users towards the APP and their evaluation of the multimodal means utilised in the application. The research and survey evaluated the effectiveness of multimodal means for word memorisation from both scientific and practical levels. Through the specific case study, the research investigated the APP's multimodal design across key learning activities. The analysis results showed that this APP mobilises users' auditory and visual senses, and facilitates word memorisation in both visual and auditory modalities through various means and resources, such as text, images, and sounds. The findings also show that this APP has a wide audience and its simple operation, as well as functions combined with multimodal means, effectively promoted word memory. The study contributes to the literature on multimodal learning strategies by providing a comprehensive analysis of how multimodal resources construct meaning in vocabulary learning. In addition, practical implications are also discussed for pedagogy applications.

Keywords: Multimodality, discourse analysis, vocabulary memory application, word memory

1. Introduction

Vocabulary is a central component of learners' language competence, since it provides lexical information to convey meaning (Wilkins, 1972). In the era of artificial intelligence (AI), advances in information and communication technologies have produced numerous digital tools, which are designed to facilitate vocabulary learning. The widespread availability of mobile vocabulary applications (APP) brings more convenience for learners by presenting words through combinations of text, audio, video, or animation. Various forms of multimodal presentations provide aids by enhancing motivation, understanding and memorisation to different extents. The powerful functions in these APPs make vocabulary acquisition more convenient and attract more learners to use them. These APPs highlight different components of learning, including word forms, pronunciations, and meanings of lexical items, etc. They also utilise various methods, such as spaced repetition, multimedia prompts, self-testing, etc., to facilitate independent study and self-regulated learning.

Despite the growing popularity of these Apps, empirical evidence within college contexts on the practical usage and contribution to vocabulary acquisition remains incomplete. For college and university students, learning autonomy is more obvious compared to earlier educational stages. Moreover, AI and digital tools have evidently become a routine part of their study practices. Therefore, it is crucial to evaluate which APP features are effective pedagogically and how learners can utilise them efficiently for the sustainment and development of vocabulary.

The current research investigates the efficacy of a specific vocabulary-learning APP with the name of

Corresponding Author: Jun Lai
School of International Studies, Sun Yat-sen University, P.R. China

"No Need to Memorise Words". This study focuses on this APP's multimodal functions and examines three specific questions: Which features employ multimodal devices? How do these multimodal elements facilitate memorising vocabulary? What are learners' perceptions of this APP's multimodal design? The goal of this research is to provide suggestions for pedagogical practices on selecting and using digital vocabulary tools. Furthermore, it aims to provide empirical evidence with college context about multimodal design principles supporting vocabulary retention.

2. Literature Review

2.1 Previous studies of multimodality learning and memory

In the domain of behaviour psychology, the learning process is regarded as long-lasting changes in behaviour evoked by learners' experiences (Anderson, 2000; Chance, 2003). Learners explore the world and understand the meanings by constructing their own comprehension through various interactions. Different types of interactions introduce distinct categories of modes that are socially or culturally shaped (Kress & Van Leeuwen, 2002). Kolb and Whishaw (2003) further classified five modalities, or five groups of senses, including visual, auditory, tactile, olfactory and gustatory. Each modality comprises its own sub-modalities and is primarily managed by the corresponding cortical region in the brain. Proper transitions among various modalities are beneficial for learners to absorb the essence of knowledge and further reinforce memory (Gu, 2007). Moreover, various modalities can support and also complement each other. Thus, learning through multimodality becomes more efficient.

Multimodal learning is well-documented in cognitive psychology theories, such as Paivio's Dual Coding Theory (Clark & Paivio, 1991; Paivio, 1986). The framework posits that cognition is comprised of two distinct but interconnected subsystems, with one for processing verbal information, such as text and speech, and the other for non-verbal information, such as sounds and images. Both channels encode information in a complementary way, thus strengthening memory and boosting recall. Building upon the dual coding theory, other researchers further raised the cognitive theory of multimedia learning, which specifies that learners possess specific channels for visual and auditory processing. Due to the limited cognitive capacity, each processing channel requires more active resources to integrate multimodal inputs (Mayer, 2001; Weston, 1992).

However, previous literature remains inconclusive towards the inherent relationship between multimodal learning and memory. Some studies believed that multimodal methods could focus learners' attention on the essence of knowledge, further enriching memory, whereas others argued that it causes dispersion and interference of attention, instead (Anderson, 2000; Lefrancois, 2019; Mayer, 2001; Nelson & Fivush, 2000). For instance, recent studies revealed that although multimodal methods could produce similar retention as traditional interfaces do, the processing, in fact, introduced higher cognitive workloads and demands (Elgohary & Chen, 2025). Therefore, it is worth exploring further the inherent complexity between multimodal learning and memory, particularly regarding the question of whether multimodal presentation facilitates vocabulary acquisition and memory.

2.2 Previous studies of multimodality means used in vocabulary memory applications

Building upon the framework of multimodal learning and memory, a growing body of research has investigated how these theories are applied in practical applications. Many empirical studies have shown significant advantages yielded by multimodal learning strategies. For instance, a recent eye-tracking study investigated multimodal lexical encoding processes, with participants being equally distributed into four conditions, including Text-only Condition, Text+Image Condition, Text+Sound Condition, and Unimodal Condition (Tang & Chen, 2025). The study raised a new multimodal coupling strength index, which quantified cross-modal integration efficacy. Their results showed significantly stronger coupling in audiovisual conditions, presenting superior recall accuracy in the trimodal condition compared with the unimodal condition. It further supports the robust effect of audiovisual synergy in vocabulary retention. Moreover, another recent English-learning experiment examined the differences between unimodal and bimodal learning strategies by comparing the Reading-only Condition with the Reading+Listening Condition (Malone, 2025). The study found that the bimodal strategy could produce equal or even superior learning performances and influenced both the quality and quantity of lexical processing. Similar studies believed that the research based on the application of multimodal theory could optimise the traditional education pattern and uplift

the capacity of students in the aspect of English learning. Overall, these studies highlighted the potential benefits of applying multimodal theory to language learning and teaching. They suggested that exploring the combination of multiple modalities can improve vocabulary acquisition and retention, as well as human-computer interaction. Additionally, the application of multimodal theory could provide new insights for educators to develop effective teaching strategies for students of various ages and language proficiency levels.

Regarding the effects of the multimodality method in Vocabulary Memory Application, the previous studies revealed mainly in two aspects. From the perspective of pedagogy, most literature suggests that the usage of the multimodality method in vocabulary memory application could help teachers pay attention to the development needs of different students, teach students in accordance with their aptitude and play the role of guidance, promotion and supervision. The teaching strategy of making full use of multimodal means to assist students in regard to vocabulary memory can help teachers improve their ability to apply modern educational information technology and lead them to constantly update their subject literacy and teaching ability. With the usage of the multimodality method, the identity of the teacher has changed from a character of imparting to a participant, which has narrowed the distance between students and helped to improve the teacher-student relationship.

From the perspective of the students' capacity for memorising vocabulary, mobile applications utilise a combination of audio, visual, and interactive features, as well as spaced repetition and gamification techniques, to help users memorise English words efficiently. Compared to students with good academic performance, students with poor academic performance would get more significant improvement with the help of multimodal means. The coordination of different modalities could uplift the students' enthusiasm for word memory, lead them to improve their ability to understand and practice English vocabulary and stimulate their imagination to find their own ways to remember words.

Nevertheless, several gaps remain in the previous literature. For instance, most existing studies focus on testing a specific modality rather than conducting a comprehensive analysis of the interaction among multiple modalities. Furthermore, little research has applied discourse analysis to examine the construction of meaning through the interplay of visual, verbal and other elements. In addition, most of the participants of previous studies were limited to the range of junior high school and senior high school, so the influence of applying multimodal means in vocabulary memory applications on college students deserves to be explored. Consequently, the present study addresses these gaps by conducting a comprehensive multimodal discourse analysis of a specific mobile application, No Need to Memorise Words, to examine how the design features contribute to the vocabulary learning of college students.

2.3 The present study: a multimodality discourse analysis of a specific APP

Although substantial research has investigated multimodal learning strategies, limited attention has been given to the systematic analysis of the integration between multiple modalities and a real mobile APP. The current study presents a case study of the APP No Need to Memorise Words, exploring an alternative analysis method different from the traditional approaches.

Our study utilises multimodal discourse analysis, based on the theoretical framework developed by Kress & Van Leeuwen (2001). In the late 1970s, foreign scholars constructed a theory of Critical Discourse Analysis (CDA) based on discourse analysis and text linguistics. The theory of CDA is based on Halliday's Systemic Functional Linguistics and studies the relationship between the surface form of discourse and the deeper social structure and power. Fowler (1979) phrased that CDA aims to reveal various assumptions and beliefs that are not easily perceived by the reader. Kress & Van Leeuwen (2001) started studying tone, music, visual images, sculpture, painting, and architecture in the late 1980s. They developed Social Semiotics, which is a separate field of study that investigates how multimodal signs, including language, shape communication and meaning in social contexts. Kress and Leeuwen (1996) studied the relationship between modality and media and specifically explored the phenomenon of multimodal means to express meaning regularly. The grammatical framework they constructed for analysing visual images provides a theoretical framework and an analytical approach for multimodal discourse analysis. According to Kress and Leeuwen (2002), "Multimodal discourse is a discourse that incorporates multiple communication modalities (e.g., sound, text, image, etc.) to convey information." Royce (2002) explored the relationship between images and words in advertising, arguing

that different symbols are complementary in multimodal discourse and the multimodal symbols are synergistic in second language teaching. Norris (2004) developed her own framework for multimodal analysis and used this theoretical framework to analyse the identity construction process of two women and the production of associations in multi-person communication. Baldry and Thibault (2006) provided an exhaustive description of multimodal discourse transcription and analysis methods, including linguistic description, mapping, graphic combination, picture presentation, and how multimodal discourse analysis methods can be used to assist in teaching and learning in an e-learning background.

Li (2003) conducted a study on Kress and Van Leeuwen's Social Semiotic analysis approach for multimodal discourse, emphasising its significance in enhancing comprehension of the socio-semiotic features of language and its relevance for the creation and comprehension of multimodal discourse, as well as English language pedagogy. Furthermore, Hu and Dong (2006) studied multimodalization and the multimodal construction of meaning from theoretical and practical perspectives. Gu (2007) proposed to analyse multimodal discourse from both content and media levels, and used his own multimodal analysis framework to analyse a university celebration ceremony. Zhang (2009) proposed a comprehensive theoretical framework for multimodal discourse analysis, including the principles of modality selection, and explored the application of multimodal discourse theory to the practice of foreign language teaching. This approach studied multimodal discourse in terms of the main aspects of visual communication, such as image, mood, composition, colour and print layout, based on Kress and Van Leeuwen's theory of visual communication in the framework of Social Semiotics.

3. Methodology

3.1 Research target

In the present research, we have targeted a specific APP, which is called "No Need to Memorise Words", a language-learning application launched in 2014. It was specifically designed to improve learners' vocabulary acquisition without the need for rote memorisation. Unlike traditional vocabulary-building techniques that require learners to memorise lengthy lists of words, this APP employs advanced algorithms to create a more interactive and engaging learning experience. It focuses on contextual learning, which helps learners connect new vocabulary with real-life situations and experiences. This approach promotes a deeper understanding of the language and encourages learners to use new vocabulary in practical ways. This APP provides a range of interactive exercises and games designed to make the learning experience more fun and enjoyable. In 2023, Apple's iOS App Store in China ranked it 19th in the Education category and 4th among all vocabulary memory applications.

This APP comprises four main components: the Homepage, Dashboard, School and My Collection, all of which are designed to be simple and user-friendly without being overly complex. The Homepage is the central hub and is divided into four sections: Check-in, Learning, Reviewing, and Personal Home. The Check-in button allows users to sign in daily, while Learning and Reviewing are the most frequently used functions. The part of the school consists of High-quality Courses and Learning Strategies, which offer both free and paid courses, such as video tutorials on vocabulary, specialised oral training courses, sentence writing for postgraduate entrance exams, and sharing of learning experiences and knowledge points. The part of My Collection records learning progress.

3.2 Method

The current study designed a questionnaire to evaluate learners' attitudes towards this APP. The questionnaire used a Likert scale, with five response levels: "Full agreement, "Agreement", "Uncertainty", "Disagreement", "Full disagreement". Our participants were 230 students from 22 different majors at colleges in mainland China. Most students (73) were in their freshman year, accounting for 31.7% of the total sample. There were also 49 Sophomores (21.3%), as well as 55 juniors (23.9%) and 53 seniors (23%), respectively.

The questionnaire consists of three parts. The first part is the opinion of the helpfulness of multimodal means of the application. The second part is the overall view and the habit of using the vocabulary memory APPs. The third part is their perspective on the effect of multimodal means applied in the application.

4. Analysis and Discussion

4.1 Multimodality discourse analysis

According to Zhang (2009), systemic functional linguistics offers a theoretical framework for the analysis and study of multimodal discourse.

Multimodal discourse can be analysed at several levels, including cultural, contextual, content, and expressive levels. For instance, the cultural level is the foundation of multimodal communication, as communication traditions, forms, and techniques are shaped by culture. Without the cultural dimension, situational contexts lack explanatory power. The cultural level includes two aspects: ideology, which refers to human thinking patterns, philosophies, life habits, and social rules, and genre, which refers to communicative procedures or structural potentialities that concretely realise this ideology. Moreover, the contextual level governs the communication process, including the scope, tone, and manner of discourse, which must also realise the chosen genre in a specific communicative mode. Furthermore, the content level comprises discourse meaning and form. Discourse meaning includes ideational, interpersonal, and textual meanings that are influenced by the scope, tone, and mode of discourse. At the formal level, the formal features of different modalities are interrelated and embody discourse meaning. Each modality has its own formal system, such as visual grammar, auditory grammar, and tactile grammar. In the end, the media level includes verbal and non-verbal media. The media forms that accompany language in linguistic media play a complementary and reinforcing role in the transmission of linguistic meaning, while non-linguistic media include both physical and non-physical media. Non-physical media mainly consists of instrumental and environmental components.

Based on Zhang's (2009) theoretical framework of multimodal discourse, we analysed this APP at four dimensions, including cultural, contextual, content, and media dimensions, respectively.

4.1.1 Cultural level analysis

Designers and college users share the same cultural context. At the mindset level, the designer's objective is to create an application that helps college users learn and memorise words. College students are the target users and beneficiaries of the application. Therefore, both parties share a common mindset of downloading and using this APP for vocabulary memorisation. In terms of living habits, college users aim to use the vocabulary memorisation application to form a routine of word memorisation, while the designer's goal is to encourage college users to download and use the application's functions. Hence, both parties can establish communication based on shared living habits. Furthermore, within the cultural context, college students need to learn and memorise words in order to pass exams, and the application provides a means to fulfil this need. Thus, college users and application designers follow the same set of social rules.

4.1.2 Contextual level analysis

The software provides a virtual contextual setting for the process of "word memorisation". Within this context, users can learn and review words, practice dictation, and even use virtual currency earned from memorising words to purchase value-added services. In this contextual setting, the process of vocabulary memorisation becomes a task-oriented process that defines the tone and mode of discourse in the user's interaction with this APP. This approach enhances the entertainment value and sense of achievement for users, thereby increasing the attractiveness of the software.

4.1.3 Content level analysis

The analysis of content levels is categorised into two principal parts: the meaning level and the form level.

At the level of discourse meaning, this APP is designed to present the conceptual and textual meaning of words through vocabulary albums, example sentences, and other means. Based on a unified standard, a fixed program is developed for the design of vocabulary-learning. For each word, Chinese explanations, pronunciations, collocations, example sentences, and root as well as affixes are provided. The conceptual meaning of words is conveyed through Chinese explanations, while the collocations and example sentences illustrate the textual meaning of words.

At the level of form, the formal features of different modalities in this APP are interrelated and jointly embody the meaning of words. Since media only serve as a means of meaning transmission, it can only directly express meaning through organisation and schematization at the formal level. Each modality has its own system at the formal level, such as visual grammar and auditory grammar. The visual grammar system includes vocabulary symbols that already have an assigned meaning and a set of rules, called grammar, that combine the symbols into more complex structures. In this APP, the visual grammar system organises the textual medium at the formal level, assigns each word a specific meaning, and organises the words into collocations and examples to express more complex meanings. Similarly, the auditory grammar system in this APP assigns each word a specific pronunciation and organises the words into example sentences that are expressed in

sound form. Visual grammar and auditory grammar can be unified to convey the complete meaning of words. For example, the presentation of the word "grapple" in this APP includes both visual and auditory forms. In the visual form, vocabulary and word meanings, phrases and phrase meanings, example sentences and sentence meanings are presented one-to-one. In the auditory form, the pronunciation of the words and the example sentences are also broadcast one-to-one. By combining the visual and auditory modalities, the learner can fully understand the discourse meaning of the word "grapple".

Various modalities of discourse are designed to reflect the overall meaning intended by the speaker. In the process of communication, one mode may not be sufficient to express the meaning of the communicator clearly. In such cases, other modes can play a reinforcing, complementary, moderating, or synergistic role to express the meaning more fully and help the listener understand the purpose of the discourse. According to Zhang (2009), a common multimodal discourse pattern involves a single mode of discourse that is insufficient for expressing its intended meaning and therefore needs to be complemented by another mode.

In this APP, the relationship between the visual and auditory modalities is a "complementary relationship", in which the visual presentation of text symbols alone is insufficient to convey the full meaning of the word, but with the help of sound symbols, the user can fully receive the meaning of the word and its pronunciation. Within the category of complementary relations, there are reinforcing and non-reinforcing relations. In reinforcing relations, one modality is the main communicative form, while the other or more forms reinforce it. In contrast, non-reinforcing relations indicate that the two communicative modalities are missing and complementary to each other, especially the combination of auditory and visual. The non-reinforcing relation involves several forms of coordination, association, and crossover. Zhang (2009) explained that in the coordination relation, multiple modalities collaborate to convey the complete meaning intended by the communicator, and the absence of any one of these modalities would result in incomplete communication. For example, a television program combines images with sound. Both modalities are necessary for the audience to fully understand the content of the program. In a joint relationship, different types of media in the same modality can form a joint relationship and come together to represent meaning. For instance, in a multimodal discourse format, such as a wildlife documentary, the animal's voice and the narrator's voice combine to embody the meaning realised by the sense of hearing. Moreover, multiple modalities can cross to embody meaning. For example, a communicator can explain the process, rationale, and reasons for their actions while working, thereby creating a phenomenon in which two modalities cross to embody the overall meaning.

In this APP, the coordination relationship within the non-reinforcement relation is the dominant multimodal relationship. During the word presentation process, textual information is received through vision, and sound information is received through hearing. The combination of the two modalities is necessary to fully convey the meaning of the word.

4.1.4 Media level analysis

Modalities refer to abstract systems, while media provide the material foundation for these modalities. Multimodal discourse media systems contain two types, including language and non-language. The language type is further subdivided into two types of media, paralinguistic and pure language. Non-language media can be classified into two types, namely body and non-body. Non-body media are further subdivided into tool and environment. In the learning mechanism of this APP, language media play a primary role. The system's example sentence pronunciation function includes not only the sound in pure language but also paralinguistic elements such as intonation, tone, accent, and background music, which facilitate conveying the example sentence's overall meaning, the target word's meaning, and the word's context. In the example sentence text presentation, the bolded target word is effective in emphasising its meaning. Non-language media also significantly contribute to this APP's learning mechanism. Users' bodily movements are limited during word learning and memorisation, so body media involvement is minimal. However, non-body media, such as mobile devices like phones and computers, are the primary tools for learning and dominate the construction of multimodal discourse. Non-body media is also evident in the user's environment. This APP's scenarios are generally in schools, where users can link the learned words with their surroundings and reinforce the word memorisation process in conducive learning atmospheres, such as classrooms

and libraries.

4.2 Data analysis

The study ulitised SPSS 29.0 for analysis.

Cronbach's alpha measures internal consistency reliability.

Overall, the results suggest that this APP has high levels of internal consistency reliability. All options show corrected item-total correlations above 0.7, indicating that each item measures the same underlying construct. Cronbach's alpha measures internal consistency reliability. Cronbach's alpha values are also high, ranging from 0.953 to 0.961, suggesting high item reliability when used together. The results suggest that this APP software is a reliable tool for learning and memorising vocabulary. Its various functions, such as multi-modal learning modes, example sentences, and pronunciation features, are all effective at helping users remember words. The results also suggest that the software is internally consistent and reliable, with each item measuring the same underlying construct.

Moreover, a reliability analysis of the usage habits of this APP was also conducted. Overall, the survey instrument appears to be quite reliable, with Cronbach's alpha values ranging from 0.936 to 0.945, indicating good internal consistency. All items have a corrected item-total correlation above 0.8, indicating that they are measuring the same construct of using habits. The squared multiple correlations are also relatively high, indicating that the items as a group are good predictors of overall usage habits. It is worth noting that item 4 ("I will recommend to friends, classmates, and family members") has a lower correlation with the overall survey score and a lower Cronbach's alpha value, suggesting that it may be less important or relevant to the construct of using habits than the other items.

In summary, the results of this reliability analysis suggest that the survey instrument used to gather data about the usage habits of this APP is a reliable and valid measure of this construct. The results can be used to guide future research and improvements to the app.

Moreover, the value of Cronbach's Alpha is 0.968, indicating a high degree of internal consistency among the items. The Cronbach's Alpha based on standardised items is also 0.968, indicating that the items are equally reliable after standardisation. The number of items in the set is 14. Overall, these results suggest that the set of items is highly reliable and consistent, and can be used effectively for measuring the construct of interest.

Furthermore, the KMO (Kaiser-Meyer-Olkin) measure is a statistic used to assess the sampling adequacy for factor analysis. In this case, the KMO value is 0.946, which is considered excellent, indicating that the data are highly suitable for factor analysis.

The Bartlett's Test of Sphericity is used to test the null hypothesis that the variables in the data set are uncorrelated, meaning that a factor analysis is not appropriate. In this case, the Bartlett's Test of Sphericity has an approximate chi-square value of 3686.490 with 91 degrees of freedom, and a significance level less than 0.001, which indicates that the null hypothesis should be rejected, and that factor analysis is appropriate for this data set.

Table 1 displays the results of a one-way ANOVA, examining the effect of grade level on perceptions of this APP. Within the dimension of the APP Functionality, mean scores range from 31.67 to 33.53. The F-value of 0.43 suggests no significant difference in the mean scores across the four grade levels ($p > 0.05$). Within the dimension of the APP Usage Habits, the mean score for each option ranges from 16.51 to 16.96. The F-value of 0.18 suggests that there is no significant difference in the mean scores between the four options ($p > 0.05$). Overall, the ANOVA results suggest that there are no significant differences in the mean scores between the different options within each category, indicating that this APP is effective for all users, regardless of their grade level or vocabulary learning habits.

Table 1 ANOVA Analysis of Each Category in Grade Dimension

Variables	Options	N	Mean	Std. Deviation	F	Sig
Functionality Survey	Freshman	73	32.51	8.54	0.43	0.73
	Sophomore	49	32.90	7.78		
	Junior	55	31.67	10.32		
	Senior	53	33.53	7.79		
Usage Habits Survey	Freshman	73	16.51	5.40	0.18	0.91
	Sophomore	49	17.18	4.55		
	Junior	55	16.67	6.48		
	Senior	53	16.96	5.20		

In addition, we perform a correlation analysis between two variables, the Functionality and the Usage Habits. The Pearson correlation coefficients indicate that there is a strong positive correlation between these two variables ($r=0.789$, $p<0.01$). This suggests that there is a relationship between the usage habits and the perceived usefulness of the APP. The high correlation coefficient indicates that individuals who use the app more frequently also tend to perceive it as being more useful. This finding is useful for the developers of the app, as it suggests that improving user habits could lead to increased perceived usefulness and satisfaction with the app.

5. Discussion and Conclusion

This study presented a multimodal discourse analysis of the mobile application No Need to Memorise Words, which is designed to assist college students in memorising English vocabulary. The analysis was based on Zhang's (2009) theoretical framework of systemic functional linguistics, which divided multimodal discourse into cultural, contextual, content, and media dimensions. At the cultural level, designers and college users shared the same cultural context and mindset, and followed the same set of social rules. At the contextual level, the software provided a virtual contextual setting for the process of vocabulary memorisation, which defined the tone and mode of discourse in the user's interaction with this APP. At the content level, the software presented the conceptual and textual meaning of words through vocabulary albums, example sentences, and other means. The formal features of different modalities in this APP were interrelated and jointly embodied the discourse meaning of words. The visual grammar system organised the textual medium at the formal level, while the auditory grammar system assigned each word a specific pronunciation and organised the words into example sentences that are expressed in sound form. The combination of visual and auditory modalities helped the learner fully understand the discourse meaning of words. The analysis demonstrated the effectiveness of using a multimodal approach to enhance the entertainment value and sense of achievement for users, thereby increasing the attractiveness of the software.

The credibility analysis evaluated the reliability of this APP's functions. The results suggested that the software has high levels of internal consistency reliability. All options had corrected item-total correlations above 0.7, indicating that each item is measuring the same underlying construct. The reliability analysis assessed the reliability of the survey instrument used to gather data about users' habits of the APP. The results suggested that the survey instrument is reliable and valid, with high Cronbach's alpha values and corrected item-total correlations above 0.8 for all items. It was worth mentioning that item 4 of "Using habits" has a lower correlation with the overall survey score and a lower Cronbach's alpha value, indicating that it might be less important or relevant to the construct of using habits than the other items.

In conclusion, the data analyses provided valuable insights into the distribution of students' grades, the reliability of the APP functions, and the reliability of the survey instrument used to gather data about user habits of the APP. The results could guide further research and improvements to the app and software.

The present study made significant contributions to the field of multimodality discourse analysis by investigating the effectiveness and reliability of the mobile application for assisting college students in English vocabulary memorisation. The findings of the study highlighted the importance of multimodal discourse in language pedagogy and the role of technology in enhancing the learning experience of students.

One of the major contributions of the study was the demonstration of the effectiveness of the multimodal approach in enhancing the entertainment value and sense of achievement for users, thereby increasing the attractiveness of the software. By combining visual and auditory modalities, the APP helped learners fully understand the discourse meaning of words, which was essential for improving their English proficiency. The study provided valuable insights into the design and implementation of multimodal applications for language learning and offered practical recommendations for improving the effectiveness and attractiveness of such applications.

Another significant contribution of the study was the evaluation of the reliability of the APP functions and the survey instrument used to gather data about user habits of the APP The results of the credibility analysis suggested that the software has high levels of internal consistency reliability and that the survey instrument was reliable and valid, which provided valuable information for future research and improvements to the app and software.

However, the current study still has limitations, since our participants are mainly college students,

and the research focus is on a single language learning APP. In the future, more research could explore the generalizability of the findings. To overcome the limitation of a limited sample size, future studies could include a larger and more diverse sample of English language learners. Additionally, comparative analyses between different language learning apps could provide a better understanding of the effectiveness of different approaches to language learning. Finally, investigating the long-term effects of using language learning apps on language acquisition and retention could add to the understanding of the effectiveness of these Apps.

Conflict of Interest

The authors declare that they have no conflicts of interest in this work.

Acknowledgement

This research was supported by the National Social Science Fund of China (25BYY029).

References

Anderson, J.R. (2000). *Learning and Memory: An integrated approach (2nd Ed.)*. New York: John Wiley & Sons.

Baldry, A. P., & Thibault, P. J. (2006). Multimodal Transcription and Text Analysis: A Multimedia Toolkit and Coursebook. London, Equinox.

Chance, P. (2003). *Learning and Behavior*. Thomson: Wadsworth.

Clark, J.M. & Paivio, A. (1991). Dual coding theory and education. *Educational Psychology Review, 3*(3), 149-210.

Elgohary, O., & Chen, Z. T. (2025). Attention, Action, and Memory: How Multi-modal Interfaces and Cognitive Load Alter Information Retention. *arXiv preprint arXiv:2509.05898*

Fowler, R., Hodge, B., Kress, G., & Trew, T. (1979). *Language and control*. London: Routledge.

Kolb, B. & Whishaw, I.Q. (2003). *Fundamentals of Human Neuropsychology*. New York: Worth Publishers.

Kress, G. & Van Leeuwen, T. (1996). *Reading Images: The Grammar of Visual Design.* London and New York: Routledge.

Kress, G. & Van Leeuwen, T. (2002). Colour as a semiotic mode: Notes for a grammar of colour. *Visual Communication, 1*(3), 348-368.

Lefrancois, G.R. (2019). *Theories of Human Learning: Mrs Gribbin's Cat* (7th ed.). Cambridge: Cambridge University Press.

Malone, J. (2025). Multimodal Effects of Processing and Learning Contextualized L2 Vocabulary. *Language Learning & Technology, 29*(3), 49-69.

Mayer, R.E. (2001). *Multimedia Learning*. Cambridge: Cambridge University Press.

Nelson, K. & Fivush, R. (2000). Socialization and Memory. In E. Tulving & Craik, F.I.M.(Eds.), *The Oxford Handbook of Memory* (pp.283-295). Oxford University Press.

Norris, S. (2004). Analyzing multimodal interaction：A methodological framework. London: Routledge.

Paivio, A. (1986). *Mental Representations: A Dual Coding Approach*. New York: Oxford University Press.

Royce, T. (2002). Multimodality in the TESOL classroom: Exploring visual-verbal synergy. *TESOL Quarterly, 36* (2), 191-205.

Gu, Y. G. (2007). On multimedia learning and multimodal learning. *China Academic Journal Electronic Publishing House, 114*, 3-12 (in Chinese).

Hu, Z.L. & Dong, J. (2006). How is meaning construed multimodally—A case study of a PowerPoint presentation contest, *Technology Enhanced Foreign Language Education, 109*, 3-12(in Chinese).

Li, Z.Z. (2003). Social semiotic approach to multimodal discourse. *Foreign Languages Research, 5*, 1-7 (in Chinese).

Tang, Q. & Chen, C. (2025). Vocabulary Retention Under Multimodal Coupling Strength Index (MCSI): Insights from Eye Tracking. *Applied Sciences, 15*(14).

Weston, T. (1992). *Atlas of Anatomy.* London: Marshall Cavendish Books Limited.

Zhang, D.L. (2009). On a synthetic theoretical framework for multimodal discourse analysis, *Foreign Languages in China, 1*, 24-30 (in Chinese).

How to Cite: Lai, J. & Xue, R. (2026). A Lexical Learning Case Study of Multimodality Discourse Analysis. *Contemporary Education and Teaching Research, 07*(03), 68-76.
https://doi.org/10.61360/BoniCETR262019790301