IRA-International Journal of Education & Multidisciplinary Studies
ISSN 2455–2526; Vol.21, Issue 04 (Q4, 2025)
Pg. no. 168-187.
IRA Academico Research

Non-Linear Relationship Between Word Frequency and Second Language Vocabulary Acquisition

Lingyu Yi [#1]
[1]School of Foreign Studies, University of Science and Technology Beijing

#corresponding author
Type of Work: Peer Reviewed.
DOI: https://dx.doi.org/10.21013/jmss.v21.n4.p1

Review history: Submitted: August 23, 2025; Revised: Oct 18, 2025; Accepted: Oct 30, 2025

How to cite this paper:
Yi, L. (2025). Non-Linear Relationship Between Word Frequency and Second Language Vocabulary Acquisition. IRA-*International Journal of Education & Multidisciplinary Studies* (ISSN 2455-2526), 21(4), 168-188. https://dx.doi.org/10.21013/jmss.v21.n4.p1

© IRA Academico Research.
This work is licensed under a Creative Commons Attribution-NonCommercial 4.0 International License, subject to a proper citation to the publication source of the work.

Disclaimer: The scholarly papers as reviewed and published by IRA Academico Research are the views and opinions of their respective authors and are not the views or opinions of IRA Academico Research. IRA Academico Research disclaims any harm or loss caused due to the published content to any party.

IRA Academico Research is an institutional publisher member of *Publishers International Linking Association Inc. (PILA-CrossRef), USA*. Being an institutional signatory to the *Budapest Open Access Initiative, Hungary*, the content published by IRA Academico Research is available under Open Access. IRA Academico Research is also a registered content provider under *Open Access Initiative Protocol for Metadata Harvesting (OAI-PMH)*.

This paper is peer-reviewed following IRA Academico Research's Peer Review Program.

Lingyu Yi /0009-0002-5015-7906

ABSTRACT

Word frequency has long been recognized as a key factor in second language vocabulary acquisition (SLVA), yet its role across different lexical categories warrants further investigation. This study aims to examine the mechanisms by which word frequency shapes SLVA. A total of 100 third-year female undergraduates from a foreign language university, equally divided between English majors and non-English majors, participated in the study. Focusing on prepositions, polysemous words, and homophones, the research combined questionnaire data with quantitative analyses to assess the frequency–acquisition relationship. The results show that the frequencies of polysemous words and homophones are positively associated with acquisition outcomes, whereas the frequency of prepositions displays a negative effect. Cross-lexical analyses and computer-generated curves further reveal a significant nonlinear relationship between word frequency and SLVA. Chi-square tests confirm that the frequency effect is consistent across both English majors and non-English majors. These findings contribute to a more nuanced understanding of frequency effects in SLVA and provide pedagogical implications for vocabulary instruction.

Keywords: Word frequency, Nonlinear relationship, Second language vocabulary acquisition

1. Introduction

Vocabulary acquisition has long been recognized as a central concern within the field of second language acquisition (SLA). While substantial progress has been made in both theoretical modeling and empirical investigation, scholars consistently emphasize that vocabulary constitutes the foundation of language competence and communicative ability (Gass et al., 2008). Indeed, some have argued that SLA, to a large extent, is synonymous with vocabulary learning (Gass et al., 2008). This underscores the necessity of a sustained scholarly focus on second language vocabulary acquisition (SLVA).

Within SLVA research, word frequency has emerged as one of the most influential variables. Frequency effects are generally understood as the tendency for high-frequency words to be processed and acquired more efficiently than low-frequency words (Gollan et al., 2011). A robust body of evidence supports the facilitative role of frequency in vocabulary acquisition, establishing it as a key predictor of lexical development in a second language. However, despite this consensus, recent studies suggest that the frequency effect may not operate uniformly across all lexical items and contexts, raising questions about its stability and generalizability.

One important limitation in the existing literature concerns the extent to which frequency effects vary across lexical categories. While frequency has been examined in relation to general vocabulary learning, comparatively fewer studies have systematically compared its impact on specific types of words, such as prepositions, polysemous words, and homophones. Given that these lexical categories differ substantially in semantic complexity, functional load, and acquisition difficulty, it is reasonable to hypothesize that the role of frequency may be conditioned by lexical type rather than being universally positive or linear. Addressing this gap is essential for advancing a more nuanced understanding of frequency effects in SLVA.

The present study seeks to investigate the relationship between word frequency and SLVA with a particular focus on prepositions, polysemous words, and homophones. Drawing on frequency data from the Corpus of Contemporary American English (COCA) and combining questionnaire surveys

with quantitative analyses, the study examines whether word frequency exerts positive or negative effects across these categories and whether the relationship is linear or nonlinear. By systematically exploring these issues, the research aims to refine theoretical accounts of frequency effects in vocabulary acquisition and to generate pedagogical implications for vocabulary instruction in second language contexts.

2. Literature Review

2.1 The vocabulary acquisition in second language

Research on SLVA has become increasingly interdisciplinary and multidimensional, reflecting both the theoretical expansion and methodological diversification of the field. Building upon early descriptive studies, current research commonly employs questionnaires, experimental designs, and corpus-based approaches to investigate how vocabulary is learned in a second language. Broadly, the existing scholarship can be grouped into four major strands.

The first strand centers on incidental vocabulary acquisition in L2 learning. Studies in this area investigate the processes and outcomes of vocabulary acquired without direct instructional focus, often through extensive reading or exposure to input. Analyses of recent international trends (Yang & Luo, 2022) reveal that incidental acquisition has emerged as a prominent research hotspot, with increasing scholarly attention directed toward its effectiveness and pedagogical implications. This suggests that vocabulary acquisition is no longer seen as solely intentional, but also as deeply embedded in naturalistic and task-based learning environments.

The second strand emphasizes the relationship between SLVA and instructional strategies. Drawing on frameworks from cognitive psychology and sociolinguistics, these studies highlight the role of teaching methods in shaping vocabulary outcomes. Research demonstrates how innovations in classroom practices (Murphy et al., 2017; Ma, 2023), the use of data-driven learning in academic English contexts (Zare & Delavar, 2023), conditions of incidental versus intentional learning (Sok & Han, 2020), and cross-modal approaches to lexical complexity (Yoo & Kim, 2023) collectively illustrate the interplay between pedagogy and vocabulary development. These findings underscore that teaching strategies not only facilitate lexical growth but also mediate how learners engage with vocabulary across modalities and learning contexts.

The third strand addresses lexical-semantic relationships in L2 acquisition. Rather than treating vocabulary as isolated units, scholars have increasingly examined the semantic networks and lexical associations underlying L2 vocabulary growth. Research in this area argues for the inclusion of richer measurement indicators and interdisciplinary perspectives to capture the complexity of learners' lexical development (Lan, 2023). This line of inquiry highlights that vocabulary acquisition is not merely quantitative but also qualitatively dependent on the learner's ability to integrate words into broader semantic systems.

The fourth strand encompasses interdisciplinary and integrative approaches to SLVA. Drawing on biolinguistics, sociolinguistics, and psycholinguistics, these studies examine how individual and contextual factors condition vocabulary learning. Examples include investigations into the influence of gender (Ma & Li, 2024), age (Saito, 2024), and working memory (Jenkins & Anderson, 2021). Methodological innovations also feature prominently, with techniques such as annotated

360° image tasks (Papin & Kaplan, 2024) and eye-tracking (Nassif et al., 2022) offering new insights into the dynamics of lexical processing. These integrative approaches reflect the field's methodological sophistication and its commitment to capturing SLVA as a multifaceted phenomenon.

Collectively, these four strands demonstrate that research on SLVA has expanded significantly in both scope and depth. Scholars worldwide have advanced the study of incidental learning, teaching approaches, lexical semantics, and interdisciplinary methodologies, thereby enriching the theoretical landscape of vocabulary research. Nonetheless, gaps remain in terms of experimental design and systematic implementation, particularly in domestic scholarship. Future research should aim to strengthen the theoretical underpinnings of SLVA while simultaneously enhancing its practical applications, thereby advancing a more comprehensive understanding of vocabulary acquisition in a second language.

2.2 Frequency effects in second language vocabulary acquisition

In SLA, vocabulary learning is often understood as the result of repeated exposure to words in meaningful and comprehensible contexts (Krashen, 1988). Within this framework, word frequency has been identified as a key determinant of acquisition. The distribution of lexical frequency typically follows predictable statistical patterns, commonly referred to as word frequency distributions (Piantadosi, 2014). This suggests that frequency is not merely a descriptive measure of language use but also a central factor shaping how learners process and acquire vocabulary.

Research on word frequency in SLA has developed along several complementary dimensions, drawing insights from psycholinguistics, corpus linguistics, and applied linguistics. From a psycholinguistic perspective, studies have demonstrated that frequency supports vocabulary learning by influencing lexical storage, retrieval, and retention. For instance, investigations into learners' mental lexicons (Yang & Chen, 2022), processing mechanisms (Zheng & Chang, 2019), and forgetting rates (Pan et al., 2023) confirm the pervasive role of frequency in shaping vocabulary outcomes. From a corpus-based perspective, frequency analysis has been widely adopted as a methodological tool. Corpus studies, such as the Japanese Reading Materials Corpus (Li & Ding, 2023), have provided benchmarks for examining the distributional properties of lexical items and their pedagogical implications.

A third strand of research highlights the application of word frequency in professional and everyday contexts. In domain-specific areas such as Business English, frequency serves as an important indicator of communicative competence (Lin et al., 2018). Advances in technology have further expanded methodological possibilities, with techniques such as functional magnetic resonance imaging (fMRI) demonstrating correlations between word frequency and the processing of animacy information in nouns (Rundle et al., 2018). These studies illustrate that frequency effects are not only observable in classroom and corpus settings but also measurable in cognitive and neurological processes, underscoring the multidimensional significance of frequency in SLA research.

Thus, previous research consistently supports a positive association between word frequency and vocabulary acquisition. At the same time, emerging evidence suggests that this relationship may not be strictly linear across different lexical categories or learning contexts. This unresolved issue

highlights the need for further empirical work, thereby motivating the present study's focus on the potentially nonlinear and category-specific effects of word frequency in SLVA.

2.3 The present study

Although previous studies have consistently demonstrated a positive correlation between word frequency and language acquisition abilities, examining its effects on reading comprehension (Sui et al., 2024), alphabetic processing (Kuperman et al., 2024), and comprehension processes more broadly (Uchihara et al., 2023), this linear pattern does not appear to hold uniformly across all lexical categories. Such findings suggest that the relationship between frequency and acquisition may be more complex than traditionally assumed, particularly when different word types are taken into account.

Against this backdrop, the present study seeks to advance a more nuanced understanding of the mechanisms through which word frequency influences English as a second language (ESL) vocabulary acquisition. To achieve this aim, the research combines corpus-based frequency data with questionnaire surveys to investigate how frequency distributions shape the acquisition of different lexical categories. Special attention is given to polysemous words, which represent content vocabulary, and prepositions, which function as grammatical words, in order to assess whether frequency effects vary across these categories.

Based on the theoretical discussion and prior empirical findings, the study formulates the following expectations. First, a general consistency is anticipated between overall ESL vocabulary acquisition and word frequency, confirming the facilitative role of frequency. Second, acquisition patterns are expected to vary across lexical categories, with polysemous words and prepositions potentially exhibiting divergent frequency effects. Third, individual differences among participants are predicted to exert limited influence on the overall outcomes, thereby pointing to a degree of universality in the observed patterns. Through this design, the study aims not only to refine theoretical accounts of frequency effects but also to provide empirical evidence with pedagogical relevance for vocabulary instruction.

3. Method

3.1 Participants

The study recruited 100 third-year female university students from a foreign language university to reduce potential gender-related variation. The sample comprised 50 English majors and 50 non-English majors, thereby ensuring both diversity and representativeness. To guarantee the validity of the questionnaire data, participants' English proficiency was pre-assessed. All participants had passed the College English Test Band 4 (CET-4), which indicated sufficient competence for completing the questionnaire. In addition, all participants were native speakers of Chinese and learned English as their L2, providing a clear and consistent linguistic background for the study.

3.2 Corpus of Contemporary American English

Although the frequency distribution of words generally follows Zipf's law (Piantadosi, 2014), retrieving precise frequency information requires the use of corpus resources. To obtain more robust insights into ESL vocabulary acquisition, this study employed the COCA as the primary

source for word frequency analysis. After systematic data processing, the results presented in Table 1 were generated to meet the analytical needs of the study.

Function words, which serve grammatical purposes, include prepositions, conjunctions, and determiners, and represent a closed class that admits few new members during language development. By contrast, content words (i.e. nouns, verbs, adjectives, and adverbs) form an open class, capable of continual expansion (Fernández & Cairns, 2010). As shown in Table 1, among the 100 most frequently used words in English, 57 are function words and 43 are content words. Specifically, the function word group comprises 13 prepositions, 11 pronouns, 9 conjunctions, and 9 articles, while the content word group includes 22 verbs, 11 adverbs, 7 nouns, and 3 adjectives. This distribution offers a clear overview of the relative proportions of function and content words in high-frequency English vocabulary.

Function words, serving grammatical functions in language, encompass categories such as prepositions, conjunctions, and determiners, constituting a closed class of words. This type of vocabulary has a relatively fixed number and does not easily accommodate new members during language development. In contrast, content words, including nouns, verbs, adjectives, and adverbs, form an open class with the potential for unlimited expansion in number (Fernández & Cairns, 2010). Data from Table 1 reveal that, among the top 100 most frequently used words, function words account for 57, while content words make up 43. Further breakdown shows that these 57 function words include 13 prepositions, 11 pronouns, 9 conjunctions, and 9 articles, among others; the 43 content words consist of 22 verbs, 11 adverbs, 7 nouns, and 3 adjectives. These data provide an intuitive understanding of the distribution ratio between function words and content words within high-frequency English vocabulary.

3.3 Questionnaire

The primary data for this study were collected through a carefully designed questionnaire. Given that prepositions constitute the largest subset of function words and verbs dominate content words, the questionnaire specifically targeted these two categories. Additionally, for content words, the instrument emphasized polysemous words and homophones to examine the impact of word frequency on SLVA.

The questionnaire comprised 19 items. The first two items assessed participants' academic backgrounds, distinguishing English majors from non-English majors. Items 3 to 18 (hereinafter Q3–Q18) formed the core of the instrument and were organized into three thematic sections: Q3–Q4 focused on polysemous words, Q5–Q13 on homophones, and Q14–Q18 on prepositions. Of these, 16 items were directly designed to investigate the relationship between word frequency and SLVA, all adapted from example sentences in *Oxford Advanced Learner's English-Chinese Dictionary* (9th Edition). A distractor item was strategically placed between Q10 and Q11 to monitor participants' attention, though it was excluded from the official scoring.

To facilitate comprehension, the questionnaire instructions were presented in participants' native language (Chinese), while all answer options were provided in English (L2). For unfamiliar vocabulary items, brief Chinese prompts were included to ensure accurate understanding. The questionnaire was administered online via a secure link, providing participants with a relaxed

environment. Response times were monitored to enhance data validity, and all participants provided informed consent for the use of their responses in the study.

3.4 Analysis of data

To prepare the questionnaire responses for computational analysis, all 100 participants' textual data were converted into a digital format. Binary encoding was applied for demographic items (Q1–Q2), with "1" indicating English majors and "0" indicating non-English majors. Similarly, Q3–Q18 were encoded such that "1" represented a selected option and "0" an unselected option. It is important to note that this process preserved participants' original responses without modification, ensuring data integrity.

The core analysis focused on three lexical categories: polysemous verbs, homophones, and high-frequency prepositions. For polysemy detection, two verbs were selected: *tell* (ranked 92nd in frequency) and *want* (68th). Q3 assessed *tell*, with one participant providing an exact match to the standard answer, while 67 participants' responses included the correct option. Q4 assessed *want*, with seven exact matches and 82 responses encompassing the correct answer.

For homophones, three words were examined: *fair*, *lead*, and *bank*, each analyzed across relevant parts of speech. *Fair* as an adjective (Q6) was ranked 1242nd; 82 participants correctly identified its meaning, a decrease of 13 from the initial comprehension in Q5. As a noun (Q7, 7791st), 80 participants acquired the intended meaning, an increase of 16 from Q5. For *lead*, verb meaning (Q9, rank 349) was correctly selected by 57 participants, while its noun meaning (Q10, rank 1575) was mastered by 73 participants. *Bank* as a noun (Q12, rank 695) was correctly understood by 74 participants, and as a verb (Q13, rank 9627) by 68 participants. These results indicate variable acquisition patterns aligned with word frequency and part-of-speech distinctions.

High-frequency prepositions were examined through phrasal verb selection tasks (Q14–Q18). The 13 prepositions included *of, in, to, for, with, on, at, out, down, off, away, up,* and *over*, all ranking within the top 142. Participants' responses showed variability in mastery: three participants answered all five questions correctly, 12 answered four correctly, 33 answered three correctly, 23 answered two correctly, 22 answered one correctly, and seven did not answer any correctly. These findings highlight that even among high-frequency function words, acquisition patterns differ considerably.

In summary, the data processing and analysis procedures provided a reliable foundation for examining the relationship between word frequency and ESL vocabulary acquisition. The results reveal both consistency and variability in participants' performance across lexical categories, supporting further investigation into potential nonlinear effects of word frequency.

3.5 Data from responses of 100 participants

Table 2 provides a summary of the questionnaire responses from all 100 participants, disaggregated by English majors and non-English majors. This presentation allows for a comparative overview of participants' performance across the two groups, facilitating subsequent analysis of potential differences in vocabulary acquisition patterns.

4. Results

The word frequency effect refers to the phenomenon whereby high-frequency words are processed more efficiently than low-frequency words (Gollan et al., 2011), suggesting a generally positive correlation between word frequency and L2 vocabulary acquisition. Both external and internal factors influence this process. External factors primarily include language input and instructional strategies, whereas internal factors encompass learners' native language, age, cognitive abilities, motivation, and learning strategies.

From an internal perspective, high-frequency words tend to be processed more rapidly than low-frequency words. Repeated exposure strengthens learners' memory traces, facilitating faster and more accurate vocabulary acquisition (Sui et al., 2024). Consequently, in questionnaire responses, participants are more likely to correctly select the meanings of high-frequency words. However, the data also indicate that word frequency does not always correspond directly to acquisition outcomes; in some cases, high-frequency exposure does not guarantee accuracy, revealing instances of negative correlation. This suggests that while frequency can accelerate acquisition, it does not uniformly predict precision in L2 vocabulary learning.

4.1 Inter-word analysis

Inter-word analysis examines linguistic units that are not restricted to a single word class, such as homographs, which share the same form but possess multiple parts of speech and lexical meanings. Analysis of the questionnaire data revealed that the relationship between word frequency and L2 vocabulary acquisition is not fixed across word classes, suggesting an unstable linear relationship. Participants displayed varied acquisition patterns for different word types, although overall, L2 vocabulary acquisition tended to correlate with word frequency rankings.

For the word *fair* (Q5–Q7), both its adjective and noun meanings were assessed. In Q5, 95 participants acquired the adjective meaning, whereas 64 acquired the noun meaning, indicating greater familiarity with the adjective. Initially, 36 participants did not recognize the noun sense referring to an amusement park. By Q7, 80 participants correctly identified this noun meaning, suggesting successful acquisition over the course of the questions. According to COCA, the adjective *fair* (ranked 1242nd) is more frequent than the noun *fair* (ranked 7791st), and acquisition patterns corresponded positively with frequency.

For *lead* (Q8–Q10), both verb and noun meanings were examined. In Q8, *lead* as a verb (ranked 349th) was fully acquired by 14 participants, and as a noun (ranked 1575th) by 10 participants, indicating a positive correlation with frequency. However, in Q9 and Q10, which tested verb (*to connect or guide*) and noun (*a clue or guide*) meanings, correct responses increased to 57 and 73 participants, respectively. This increase, despite relatively lower frequency for the noun, suggests instances of a negative correlation. Overall, no strictly linear relationship between word frequency and acquisition was observed for *lead*.

For *bank* (Q11–Q13), noun and verb meanings were also analyzed. In Q11, only 13 participants fully acquired each meaning, indicating limited impact of part of speech on initial acquisition. In subsequent questions, 74 participants correctly identified the noun meaning (*a type of cloud*), and

68 the verb meaning (*to accumulate or heap up*). COCA frequencies rank the noun at 695th and the verb at 9627th, confirming that higher frequency was associated with better acquisition.

In summary, the acquisition of homographs generally demonstrates that higher word frequency is linked to improved acquisition, though variability exists across specific lexical items. These findings indicate that while a positive correlation often emerges between word frequency and L2 vocabulary acquisition, the relationship is not uniformly linear.

4.2 Intra-word analysis

Intra-word analysis focuses on specific word categories, including verbs, nouns, and prepositions. Analysis of the survey data indicates that while correlations between word frequency and SLVA vary across word classes, consistent patterns emerge within individual categories.

Verbs. The verb items included *tell* (ranked 92nd), *want* (68th), *lead* (349th), and *bank* (9627th) across Q3, Q4, Q9, and Q13. Among 100 participants, acquisition counts were 67, 82, 43, and 17, respectively, demonstrating that higher-frequency verbs tend to be acquired more successfully. For polysemous verbs, *tell* and *want* were examined in Q3 and Q4, with one and seven participants fully acquiring the meanings, respectively. These results confirm that at the polysemy level, higher frequency correlates positively with acquisition outcomes, supporting a generally positive relationship between word frequency and L2 verb learning.

Nouns. The noun items, *fair* (7791st), *lead* (1575th), and *bank* (659th), corresponding to Q7, Q10, and Q12, exhibited less consistent patterns. Acquisition counts among participants were 82, 57, and 73, respectively. Further examination using Q5, Q8, and Q11 revealed initial acquisition counts of 57 (*fair*), 21 (*lead*), and 28 (*bank*), with subsequent increases of 23, 52, and 46 participants, respectively. These data indicate that acquisition outcomes for nouns do not consistently align with word frequency, suggesting a nonlinear relationship between frequency and L2 noun learning.

Prepositions. Thirteen high-frequency prepositions were tested across Q14–Q18. Although individual prepositions sometimes showed positive correlations (e.g., *off*, *in*, and *to*), the overall pattern did not hold. For instance, *to* (ranked 12th) was correctly acquired by 13 participants, whereas *down* (ranked 119th) was acquired by 77 participants, indicating a negative correlation. Overall, the data reveal that absolute linearity between word frequency and preposition acquisition is absent.

In summary, verb acquisition generally exhibits a positive correlation with word frequency, whereas nouns and prepositions display nonlinear or inconsistent patterns. These findings suggest that across word classes, the relationship between word frequency and SLVA is complex and not strictly linear.

4.3 Nonlinear relationship figures

Although the nonlinear relationship between word frequency and second language vocabulary acquisition has been extensively discussed in previous sections, this study further visualizes the association by analyzing word frequency data for thirteen questionnaire items alongside corresponding participant acquisition patterns. Using Excel, Figure 1 was generated to illustrate these dynamics. The thirteen words, ordered from lowest to highest frequency, are as follows: *bank*

(v., 3963), *fair* (n., 5272), *lead* (n., 59773), *fair* (a., 77775), *bank* (n., 138673), *lead* (v., 285740), *off* (prep., 479459), *down* (prep., 820294), *tell* (v., 1119692), *want* (v., 1671524), *to* (prep., 9232572), *in* (prep., 15670692), and *of* (prep., 23159162). Corresponding acquisition counts for participants are: 17, 82, 57, 27, 73, 43, 62, 77, 67, 82, 13, 27, and 51, respectively.

In Figure 1, the solid line represents the empirically observed relationship between word frequency and acquisition, while the dashed line depicts a linear prediction automatically generated by a computational algorithm based on the data. Notable deviations between the two lines provide visual evidence against a simple linear relationship. The linear prediction is entirely data-driven, emphasizing that discrepancies are not artifacts of manual plotting.

Within the word frequency range of 1 to 5,000,000 occurrences, the empirical relationship exhibits pronounced fluctuation peaks. To facilitate detailed examination, this segment was extracted and enlarged in Figure 2, without altering any data points. Figure 2 offers a clearer perspective on how low-frequency words influence vocabulary acquisition, highlighting the complexity of acquisition patterns in this range.

Furthermore, for the high-frequency prepositions *in*, *to*, and *of*, the empirical acquisition data diverge substantially from the predicted linear trends (Figure 1). Specifically, participant acquisition counts for *in* and *to* (27 and 13, respectively) fall below predicted values, whereas *of* (51) exceeds its linear prediction. These deviations provide additional robust evidence for the nonlinear and category-specific effects of word frequency on second language vocabulary acquisition.

5. Discussion

Although prior studies suggest a positive correlation between word frequency and second language vocabulary acquisition (Sui et al., 2024), empirical evidence indicates that this relationship is not universally linear. Specifically, the association between word frequency and SLVA demonstrates nonlinear characteristics, influenced by multiple interacting factors. The following sections examine the primary contributors to this nonlinearity.

5.1 The first cause of nonlinear relationships: context

One of the main factors contributing to the nonlinear relationship is context. Research has shown that contextual cues can enhance the processing efficiency and acquisition likelihood of low-frequency words (Desai et al., 2020). This effect is particularly evident in the acquisition of homographs.

For example, the word *lead* exhibits distinct frequency patterns depending on its part of speech: as a noun, it ranks 1,575th, whereas as a verb, it ranks 349th. In Q8, participants were asked to identify the part of speech without sentence context, resulting in 39 participants correctly acquiring the verb meaning and 35 acquiring the noun meaning, reflecting a preliminary positive correlation between frequency and acquisition. However, when participants were provided with full sentences containing *lead* in Q9 and Q10, acquisition increased markedly to 57 (verb) and 73 (noun), demonstrating the pivotal role of contextual information in facilitating correct vocabulary comprehension. For instance, in Q9, "The wire led to a speaker," participants applied grammatical rules and semantic cues to accurately select the verb meaning "to connect," as defined in the *Oxford Advanced Learner's English-Chinese Dictionary* (9th Edition).

Similarly, the homograph *bank* shows a striking context effect. As a noun, it ranks 659th; as a verb, it drops to 9,627th. In Q11, without contextual sentences, 33 participants acquired the verb meaning and 22 the noun meaning, indicating inconsistencies with word frequency. When contextual sentences were provided in Q12 and Q13, acquisition increased to 74 (noun) and 68 (verb). Notably, the noun meaning exhibited the largest gain, likely due to the sentence cue "a mass of clouds," which enabled participants to disambiguate meaning based on context.

These examples indicate that, for homographs, contextual cues reduce the cognitive processing cost of low-frequency words and enhance acquisition rates. Consequently, context represents a critical factor contributing to the nonlinear relationship between word frequency and SLVA. Further research is needed to determine whether similar context-driven effects are observed across other word classes.

5.2 The second cause of nonlinear relationships: frequent exposure

Prepositions represent a critical component in this study, and their frequency distribution and acquisition outcomes warrant careful consideration. Among the five prepositions examined—*of*, *to*, *in*, *down*, and *off*—*down* is particularly noteworthy. Despite ranking fourth in frequency among the five and 119th overall, it was acquired by the largest number of participants. This pattern significantly affects the observed correlation between preposition acquisition and word frequency. For instance, in the sentence "The rain came _____ in torrents," participants were asked to select one of *out*, *at*, *down*, or *up*. Given that rain moves from top to bottom, *down* emerges as the semantically optimal choice, highlighting the influence of context on preposition selection. Such context-specific cues complicate the straightforward interpretation of frequency effects.

Additionally, prior research has established that frequent exposure to a word can enhance memory and facilitate faster, more accurate vocabulary acquisition (Sui et al., 2024). However, findings from Q16 challenge this expectation. In this question, participants encountered prepositions such as *on*, *with*, *out*, and *in*, all of which hold high frequency rankings. The context provided "The company deals _____ computer software" elicited 53 responses for *with* and only 27 for *in*. Despite *in* having a higher overall frequency (ranked 7th) than *with* (ranked 17th), participants favored the phrasal verb *deal with*, likely due to its greater exposure in actual usage. According to COCA data, *deal with* occurs 72,092 times, whereas *deal in* appears only 3,316 times, creating a substantial exposure advantage for *with*. This discrepancy indicates a negative correlation between frequency and acquisition in this context, particularly at the level of phrasal verbs.

When examining individual prepositions, the number of participants selecting *in* remains lower than those choosing *with*, despite *in*'s higher frequency. Nevertheless, this negative correlation is specific to certain comparisons and does not generalize across all prepositions in the study. Consequently, frequent exposure can interact with semantic and contextual factors to produce nonlinear effects, underscoring the complexity of the relationship between word frequency and vocabulary acquisition.

5.3 Universal applicability of research findings

To minimize potential confounding variables, the present study selected a balanced sample of 50 English majors and 50 non-English majors, all matched in grade and gender to ensure homogeneity.

Nevertheless, participants' professional background remains a potential factor influencing vocabulary acquisition outcomes. To examine this effect, the study employed chi-square tests to quantitatively assess the impact of professional background on questionnaire responses.

The 16 questionnaire items were categorized into three groups: Q5–Q7, Q8–Q10, and Q11–Q13. This grouping was based on content refinement: Q6 and Q7 refined Q5 (*fair*), Q9 and Q10 refined Q8 (*lead*), and Q11 and Q12 refined Q13 (*bank*). For the chi-square analysis, six representative questions were selected: Q3, Q4, Q5, Q8, Q11, and Qn (Q14–Q18). All responses had been preprocessed using binary encoding for computational analysis.

Using SPSSAU software, data from these six questions were input and organized into Table 3. The research hypothesis posited that professional background would not significantly affect survey outcomes. Table 3 shows that, except for Qn ($p < 0.05$), all other questions yielded p-values greater than 0.05 (SPSSAU Project, 2024), indicating no significant effect of professional background.

To investigate the anomaly in Qn, an in-depth analysis was conducted. It was observed that responses to Q16, involving the preposition *in*, might disproportionately influence Qn results. After excluding Q16 data and re-running the chi-square test (yielding Qn–16), Table 4 showed p-values greater than 0.05, confirming that professional background no longer significantly affected the results. Specifically, in Q16, 27 participants correctly acquired *in*, with 23 English majors and 4 non-English majors, which accounted for the initially observed discrepancy. This outcome highlights the sensitivity of statistical tests to individual items while still supporting the overall reliability of the data.

In conclusion, the chi-square analysis confirms that participants' professional backgrounds did not significantly influence the majority of the survey outcomes. This finding reinforces the validity and generalizability of the study's results, supporting the robustness of the observed patterns in second language vocabulary acquisition across different learner profiles.

6. Conclusion

Given the central role of vocabulary acquisition in second language learning, this study investigated the intricate relationship between word frequency and SLVA across three specific categories: prepositions, polysemous words, and homophones. By focusing on these categories, the research addresses both open-class and closed-class lexical items, thereby capturing a more nuanced perspective on vocabulary learning mechanisms. The use of a balanced sample of 100 third-year female undergraduates, equally divided between English majors and non-English majors, ensures the representativeness and reliability of the findings, while controlling for potential confounding variables such as gender and educational level.

The findings demonstrate a general trend of consistency between overall SLVA performance and word frequency, confirming the relevance of frequency effects in language learning. However, a more detailed intra- and inter-word analysis reveals that this relationship is not strictly linear. In particular, variations emerge across word classes: verbs generally show a positive correlation between frequency and acquisition, while nouns and prepositions exhibit nonlinear or even negative patterns. These results challenge the traditional assumption of a universal linear

relationship, suggesting that word frequency interacts with other linguistic and cognitive factors in complex ways.

Further analysis indicates that contextual cues and exposure frequency play pivotal roles in shaping SLVA outcomes. For instance, the acquisition of homographs such as *lead* and *bank* is significantly facilitated by contextualized sentence exposure, highlighting the importance of semantic and syntactic environments in mediating frequency effects. Similarly, the frequency of occurrence in authentic language use influences preposition acquisition, as seen in phrasal verbs where lower-frequency items can sometimes be acquired more effectively due to salient contextual or collocational cues. These findings collectively underscore the multifaceted nature of vocabulary learning, demonstrating that frequency alone cannot fully account for SLVA patterns without considering context and exposure dynamics.

Finally, the universality of these findings is supported by the chi-square analysis, which shows that participants' academic majors do not significantly influence the observed trends. This suggests that the nonlinear effects of word frequency on SLVA are not confined to learners of a particular academic background, enhancing the generalizability of the study's conclusions.

In sum, this research contributes both theoretically and practically to the field of second language acquisition. Theoretically, it advances our understanding of the complex, nonlinear interplay between word frequency, context, and exposure in vocabulary learning. Practically, it offers evidence-based insights for language instruction, highlighting the need to integrate contextualized exposure and targeted vocabulary strategies to optimize SLVA outcomes. Overall, the study emphasizes that effective vocabulary acquisition in second language learning is shaped by a dynamic interaction of frequency, context, and learner engagement, rather than by word frequency in isolation.

Acknowledgment

The author is indebted to the editor and anonymous reviewers for many valuable comments.

References

Desai, R. H., Choi, W., & Henderson, J. M. (2020). Word frequency effects in naturalistic reading. *Language, Cognition & Neuroscience, 35*(5), 583–594. New York, NY: Routledge.

Fernández, E. M., & Cairns, H. S. (2010). *Fundamentals of psycholinguistics*. New York, NY: John Wiley & Sons, Ltd.

Gass, S. M., Behney, J., Plonsky, L., & Selinker, L. (2008). *Second language acquisition: An introductory course* (3rd ed.) [M]. New York, NY: Routledge.

Gollan, T. H., Slattery, T. J., Goldenberg, D., Van Assche, E., Duyck, W., & Rayner, K. (2011). The frequency drives lexical access in reading but not in speaking: The frequency-lag hypothesis. *Journal of Experimental Psychology: General, 140*(2), 186–209.

Jenkins, K. C., & Anderson, R. T. (2021). A preliminary study of the relationship between working memory, language experience, and grammatical skill in dual language learners. *International Journal of Bilingualism, 25*(6), 1718–1734.

Krashen, S. D. (1988). *Second language acquisition and second language learning*. Englewood Cliffs, NJ: Prentice Hall.

Kuperman, V., Schroeder, S., & Gnetov, D. (2024). Word length and frequency effects on text reading are highly similar in 12 alphabetic languages. *Journal of Memory and Language, 135,* 104497.

Murphy, A., Franklin, S., Breen, A., Hanlon, M., McNamara, A., Bogue, A., & James, E. (2017). A whole class teaching approach to improve the vocabulary skills of adolescents attending mainstream secondary school, in areas of socioeconomic disadvantage. *Child Language Teaching and Therapy, 33*(2), 129–144.

Nassif, L., Huntley, E., & Mohamed, A. (2022). Attention to verbal morphology in L2 Arabic reading: An eye-movement study. *Foreign Language Annals, 55*(3), 769–792.

Papin, K., & Kaplan, R. (2024). A study of vocabulary learning using annotated 360° pictures. *Computer Assisted Language Learning, 37*(5–6), 1108–1135.

Piantadosi, S. T. (2014). Zipf's word frequency law in natural language: A critical review and future directions. *Psychonomic Bulletin & Review, 21*(5), 1112–1130.

Rundle, M., Coch, D., Connolly, A., & Granger, R. (2018). Dissociating frequency and animacy effects in visual word processing: An fMRI study. *Brain and Language, 183*, 54–63.

Saito, K. (2024). Age effects in spoken second language vocabulary attainment beyond the critical period. *Studies in Second Language Acquisition, 46*(1), 1–25.

Sok, S., & Han, Z. (2020). A study of L2 vocabulary acquisition under incidental and intentional conditions. *VIAL-Vigo International Journal of Applied Linguistics, 17*(17), 113–140.

Sui, L., Woumans, E., Duyck, W., & Dirix, N. (2024). The word frequency effect in first- and second-language reading by Chinese and Dutch bilinguals. *Bilingualism: Language and Cognition, 28*(1), 1–16.

The SPSSAU Project. (2024). *SPSSAU* (Version 24.0) [Computer software]. https://www.spssau.com

Uchihara, T., Webb, S., Saito, K., & Trofimovich, P. (2023). Frequency of exposure influences accentedness and comprehensibility in learners' pronunciation of second language words. *Language Learning, 73*(1), 84–125.

Yoo, H., & Kim, H. (2023). A cross-modal analysis of lexical sophistication: EFL and ESL learners in written and spoken production. *IRAL – International Review of Applied Linguistics in Language Teaching*.

Zare, J., & Delavar, K. A. (2023). A data-driven learning focus on form approach to academic English lecture comprehension. *Applied Linguistics, 44*(3), 485–504.

兰茜. (2023). 二语三大词汇语义关系发展研究综述. *外语教育*, 23(00), 101–112. [Translation]: Lan, X. (2023). A review of the research on the development of three major L2 lexical-semantic relations. *Foreign Language Education, 23*(00), 101–112.

李秀红, 丁鸽. (2023). 系统科学理论下阅读语篇实意词词频分析与应用. *系统科学学报*, 31(01), 130–136. [Translation]: Li, X., & Ding, G. (2023). Analysis and application of word frequency of notional words in reading text under the theory of system science. *Chinese Journal of Systems Science, 31*(01), 130–136.

林竞, 李传芳, 余军. (2018). 商务英语专业词汇门槛的设定——基于Range对2006—2015年《金融时报》和《财富》商务英语新闻的分析. *厦门大学学报(哲学社会科学版)*, (2), 165–172. [Translation]: Lin, J., Li, C., & Yu, J. (2018). Lexical thresholds for the business English major: Range-aided analysis on FT & *Fortune* business articles from 2006–2015. *Journal of Xiamen University (Arts & Social Sciences), 2*, 165–172.

马艳臣. (2023). 多视角下二语词汇习得与教学策略探究. *现代英语*, (05), 31–34. [Translation]: Ma, Y. (2023). Exploring second language vocabulary acquisition and teaching strategies from multiple perspectives. *Modern English*, (5), 31–34.

马拯, 李赞. (2024). 二语新颖词语义韵的习得：语境变异性及性别的影响. *现代外语*, 47(06), 1–12. [Translation]: Ma, Z., & Li, Z. (2024). Acquiring semantic prosody in L2 novel word learning: The effect of context variability and gender. *Modern Foreign Languages*, 47(6), 1–12.

潘克菊, 杨连瑞, 陈士法. (2023). 二语词汇保持与损耗影响因素探究. *西安外国语大学学报*, 31(4), 64–68. [Translation]: Pan, K., Yang, L., & Chen, S. (2023). Investigating factors affecting retention and attrition of second language vocabulary. Journal of Xi'an International Studies University, 31(4), 64–68.

杨启伟, 罗少茜. (2022). 国际二语词汇附带习得研究热点和趋势(2000—2021)——基于SSCI期刊的可视化分析. *外语界*, (6), 74–80. [Translation]: Yang, Q., & Luo, S. (2023). Research

trends and hotspots in incidental second language vocabulary acquisition (2000–2021): A visualization analysis based on SSCI journals. *Foreign Language World*, (6), 74–80.

杨雪, 陈建林. (2022). 英语作为二语和三语时学习者心理词库对比研究. *外语学刊*, (01), 92–98. [Translation]: Yang, X., & Chen, J. (2022). A contrastive study on the mental lexicon of learners when English is used as the second and third language. *Foreign Language Research*, (1), 92–98.

郑丽娜, 常辉. (2019). 中国学习者对英语规则动词和不规则动词的心理加工机制研究. *外语与外语教学*, (04), 78–89+148. [Translation]: Zheng, L., & Chang, H. (2019). Processing mechanism of English regular and irregular verbs by Chinese EFL learners. *Foreign Languages and Their Teaching*, (4), 78–89, 148.

Information Regarding Figures and Tables

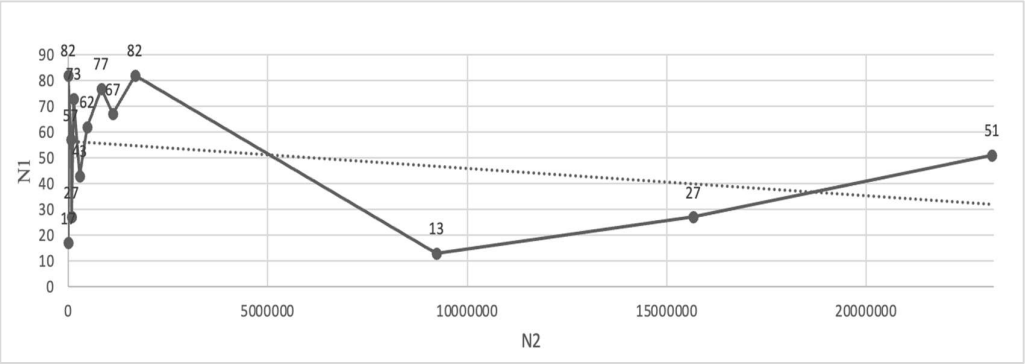

Figure 1. Nonlinear relationship presented by words in the questionnaire

Note: N1 = number of learners achieving acquisition; N2 = number of word frequency

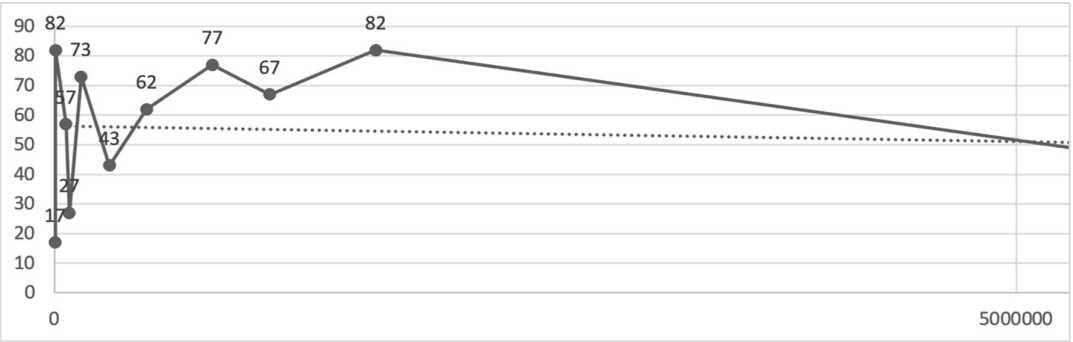

Figure 2. Nonlinear relationship (word frequency range: 1-5,000,000)

Note: N1 = number of learners achieving acquisition.

Table 1. The top 100 word frequency from COCA

Content words				Function words			
rank	term	pos	wf	rank	term	pos	wf
2	be	v.	32394756	1	the	art.	50033612
11	have	v.	10514314	3	and	conj.	24778098
15	do	v.	8186412	4	a	art.	24225478
26	say	v.	4096416	5	of	prep.	23159162
31	go	v.	3546732	6	to	t.	16770155
34	get	v.	3347615	7	in	prep.	15670692
37	can	v.	3091046	8	i	pron.	14217601
39	know	v.	2761628	9	you	pron.	12079413
47	will	v.	2372215	10	it	pron.	11042044
48	so	adv.	2369749	12	to	prep.	9232572
49	would	v.	2349400	13	that	conj.	8319512
50	make	v.	2290830	14	for	prep.	8194970
51	just	adv.	2270900	16	he	pron.	6467470
52	up	adv.	2108756	17	with	prep.	6442861
53	think	v.	2077762	18	on	prep.	6080156
54	time	n.	2018725	19	this	det.	5541440
56	see	v.	1958700	20	n't	neg.	5285354
59	out	adv.	1828593	21	we	pron.	5180711
61	come	v.	1802158	22	that	det.	5002963
62	people	n.	1800205	23	not	neg.	4655980
63	take	v.	1768822	24	but	conj.	4523086
64	year	n.	1729962	25	they	pron.	4503650
68	want	v.	1671524	27	at	prep.	4024079
69	how	adv.	1666469	28	what	det.	3807502
72	now	adv.	1601991	29	his	art.	3718978
74	other	adj.	1539952	30	from	prep.	3711425
75	could	v.	1529795	32	or	conj.	3420339

78	here	adv.	1413594	33	by	prep.	3372222
79	then	adv.	1344434	35	she	pron.	3188078
81	look	v.	1338475	36	my	art.	3106939
82	way	n.	1260011	38	as	conj.	2946119
83	more	adv.	1248955	40	if	conj.	2709809
86	thing	n.	1202004	41	me	pron.	2638743
87	well	adv.	1189096	42	your	art.	2577505
89	also	adv.	1142799	43	all	det.	2503556
91	use	v.	1126042	44	who	pron.	2493429
92	tell	v.	1119692	45	about	prep.	2427703
93	good	adj.	1111721	46	their	art.	2417058
95	man	n.	1091176	55	there	exclam.	1980173
96	day	n.	1068902	57	her	art.	1931189
97	find	v.	1051936	58	as	prep.	1880190
98	give	v.	1048189	60	one	num.	1816593

Table 1 (*continued*)

Content words				Function words			
rank	term	pos	wf	rank	term	pos	wf
100	new	adj.	1017175	65	him	pron.	1717209
				66	them	pron.	1701589
				67	some	det.	1684262
				70	when	conj.	1650353
				71	which	det.	1613281
				73	like	prep.	1583444
				76	our	art.	1467955
				77	into	prep.	1461573
				80	than	conj.	1342798
				84	these	det.	1223310
				85	no	art.	1206112
				88	because	conj.	1167024

90	two	num.	1139973

Note: pos = part of speech; wf = word frequency.

Table 2. Results of the Q3 to Q18

Question	Options	Total	N1	N2	Question	Options	Total	N1	N2
Q3	A	67	39	28	Q11	A	90	44	46
	B	76	41	35		B	76	42	34
	C	85	39	46		C	39	22	17
	D	27	11	16		D	61	32	29
Q4	A	82	42	40		E	33	17	16
	B	72	37	35	Q12	A	6	2	4
	C	73	39	34		B	10	3	7
	D	30	12	18		C	74	42	32
Q5	A	95	50	45		D	2	1	1
	B	41	27	14		E	8	2	6
	C	29	16	13	Q13	A	5	0	5
	D	64	40	24		B	6	1	5
Q6	A	82	47	35		C	11	4	7
	B	11	2	9		D	10	3	7
	C	4	1	3		E	68	42	26
	D	3	0	3	Q14	A	62	40	22
Q7	A	6	0	6		B	21	7	14
	B	8	3	5		C	9	1	8
	C	6	4	2		D	8	2	6
	D	80	43	37	Q15	A	9	2	7
Q8	A	88	45	43		B	10	1	9
	B	39	19	20		C	77	46	31
	C	85	46	39		D	4	1	3
	D	35	18	17	Q16	A	15	3	12
	E	46	29	17		B	53	23	30
Q9	A	24	9	15		C	5	1	4

	B	57	37	20		D	27	23	4
	C	11	4	7	Q17	A	76	40	36
	D	2	0	2		B	13	5	8
	E	6	0	6		C	11	5	6
Q10	A	8	3	5		D	0	0	0
	B	7	2	5	Q18	A	19	10	9
	C	3	1	2		B	18	7	11
	D	73	43	30		C	12	3	9
	E	9	1	8		D	51	30	21

Note: N1 = number of English major students; N2 = number of non-English major students.

Table 3. Analysis results of Chi-square test

Question	Q-options	PRO (%)		Total	P
		EMS	N-EMS		
Q3	Q3-A	39(30.00)	28(22.40)	67(26.27)	0.279
	Q3-B	41(31.54)	35(28.00)	76(29.80)	
	Q3-C	39(30.00)	46(36.80)	85(33.33)	
	Q3-D	11(8.46)	16(12.80)	27(10.59)	
Total		130	125	255	
Q4	Q4-A	42(32.31)	40(31.50)	82(31.91)	0.675
	Q4-B	37(28.46)	35(27.56)	72(28.02)	
	Q4-C	39(30.00)	34(26.77)	73(28.40)	
	Q4-D	12(9.23)	18(14.17)	30(11.67)	
Total		130	127	257	
Q5	Q5-A	50(37.59)	44(46.32)	94(41.23)	0.466
	Q5-B	27(20.30)	14(14.74)	41(17.98)	
	Q5-C	16(12.03)	13(13.68)	29(12.72)	
	Q5-D	40(30.08)	24(25.26)	64(28.07)	
Total		133	95	228	
Q8	Q8-A	45(28.66)	43(31.62)	88(30.03)	0.678
	Q8-B	19(12.10)	20(14.71)	39(13.31)	

	Q8-C	46(29.30)	39(28.68)	85(29.01)	
	Q8-D	18(11.46)	17(12.50)	35(11.95)	
	Q8-E	29(18.47)	17(12.50)	46(15.70)	
Total		157	136	293	
Q11	Q11-A	44(28.03)	46(32.39)	90(30.10)	0.916
	Q11-B	42(26.75)	34(23.94)	76(25.42)	
	Q11-C	22(14.01)	17(11.97)	39(13.04)	
	Q11-D	32(20.38)	29(20.42)	61(20.40)	
	Q11-E	17(10.83)	16(11.27)	33(11.04)	
Total		157	142	299	
Qn	Q14	40(27.78)	22(25.58)	62(26.96)	0.044*
	Q15	46(31.94)	31(36.05)	77(33.48)	
	Q16	23(15.97)	4(4.65)	27(11.74)	
	Q17	5(3.47)	8(9.30)	13(5.65)	
	Q18	30(20.83)	21(24.42)	51(22.17)	
Total		144	86	230	

Note: PRO = proportion of responses to options; EMS = English major students; N-EMS = non-English major students; P = *p*; * = $p < 0.05$.**Table 4.** Analysis results of Chi-square test for Q(n-16)

Question	Q-options	PRO (%)		Total	P
		EMS	N-EMS		
Q(n-16)	Q14	40(43.48)	22(29.73)	62(37.35)	0.088
	Q15	2(2.17)	7(9.46)	9(5.42)	
	Q17	40(43.48)	36(48.65)	76(45.78)	
	Q18	10(10.87)	9(12.16)	19(11.45)	
Total		92	74	166	

Note: PRO = proportion of responses to options; EMS = English major students; N-EMS = non-English major students; P = *p*.