

University Students' Comprehension of Multiple Meanings of English Words Across Contexts: An Investigation of Literal, Figurative, Idiomatic, and Grammatical Meanings

Ismail M. Mohammed 

English Department, Zawra Faculty of Education, University of Zawia, Zawia, Libya.

*Corresponding author's email: Ismail Mohammed | ismailzeino72@gmail.com

Received 11-09-2026 | Accepted: 01-10-2026 | Available online: 09-10-2026 | DOI:10.26629/uzjeps.2026.21

ABSTRACT

This study investigates university students' comprehension of multiple meanings of English words across different contextual uses, with particular attention to literal, figurative, idiomatic, and grammatical meanings. The study addresses the problem that knowledge of an English word may extend beyond its familiar or dictionary meaning, while learners may experience difficulty interpreting the same word when it occurs in different linguistic contexts. The study aimed to determine the level of students' contextual comprehension, examine whether significant differences existed among the four semantic categories, and compare the performance of third- and fourth-year university students. A quantitative descriptive-analytical research design was employed with a purposive sample of 45 students from the English Department, including 25 third-year and 20 fourth-year students. Data were collected through a contextual vocabulary test that initially contained 60 items; following statistical item analysis, 47 valid items were retained for the final analysis. Descriptive statistics, Shapiro-Wilk normality testing, Mann-Whitney U, Friedman, Spearman's rho, and Cronbach's alpha were employed using SPSS. The results showed a relatively good overall level of contextual lexical comprehension, with a mean score of 34.07 out of 47. Performance across the four semantic categories was relatively similar: literal meaning 72.44%, figurative meaning 70.62%, grammatical-context meaning 75.28%, and idiomatic meaning 72.00%. The Friedman test showed no statistically significant differences among the four categories, $\chi^2 = 2.242$, $p = .524$. Likewise, no statistically significant difference was found between third- and fourth-year students in the total test score ($U = 212.50$, $Z = -0.858$, $p = .391$), although fourth-year students obtained a higher mean rank. A significant difference was found only for Word 3 ($p = .039$). The test demonstrated high internal consistency (Cronbach's $\alpha = .939$), and strong positive correlations were found between the four semantic categories and the total test score ($r_s = .874-.952$, $p < .001$). The findings suggest that students possess relatively good but heterogeneous contextual lexical knowledge and that academic progression alone may not be sufficient to produce significant growth in vocabulary depth. The study therefore recommends

explicit and systematic instruction that exposes learners to multiple contextual uses of the same lexical items in order to strengthen contextual comprehension and lexical depth.

Keywords: Multiple Word Meanings, Contextual Meaning, Vocabulary Depth, Polysemy, Figurative Meaning, Idiomatic Meaning, Grammatical Meaning, University Students.

فهم طلبة الجامعة لتعدد معاني الكلمات الإنجليزية باختلاف السياقات: دراسة للمعاني الحرفية والمجازية والاصطلاحية والنحوية

إسماعيل محمد زينو محمد

قسم اللغة الإنجليزية، كلية التربية زوارة، جامعة الزاوية، ليبيا

*المؤلف المراسل: إسماعيل محمد ismailzeino72@gmail.com

استقبلت: 2026-09-11 | قبلت: 2026-10-01 | متوفرة على الانترنت | 2026-10-09

استقبلت: 2026-09-11 | قبلت: 2026-10-01 | متوفرة على الانترنت | 2026-10-09 | DOI:10.26629/uzjeps.2026.21

ملخص البحث

تبحث هذه الدراسة في مدى فهم طلبة الجامعة لتعدد معاني الكلمات الإنجليزية باختلاف استخداماتها السياقية، مع التركيز بصفة خاصة على المعاني الحرفية والمجازية والاصطلاحية والنحوية. وتتناول الدراسة مشكلة تتمثل في أن معرفة الكلمة الإنجليزية قد تتجاوز معناها المألوف أو المعجمي، في حين قد يواجه المتعلمون صعوبة في تفسير الكلمة نفسها عندما ترد في سياقات لغوية مختلفة. وهدفت الدراسة إلى تحديد مستوى الفهم السياقي للمفردات لدى الطلبة، والتحقق مما إذا كانت هناك فروق ذات دلالة إحصائية بين الفئات الدلالية الأربع، ومقارنة أداء طلبة السنة الثالثة والسنة الرابعة في الجامعة. وقد اتبعت الدراسة تصميمًا بحثيًا وصفيًا تحليليًا كميًا، باستخدام عينة قصدية مكونة من 45 طالبًا وطالبة من قسم اللغة الإنجليزية، بواقع 25 من طلبة السنة الثالثة و20 من طلبة السنة الرابعة. وجمعت البيانات من خلال اختبار للمفردات في سياقات مختلفة، تكون في صورته الأولى من 60 فقرة؛ وبعد إجراء تحليل إحصائي لفقرات الاختبار، تم الاحتفاظ بـ 47 فقرة صالحة للتحليل النهائي. واستخدمت الدراسة الإحصاءات الوصفية، واختبار شابيرو-ويلك للتحقق من اعتدالية التوزيع، واختبار مان-ويتني U، واختبار فريدمان، ومعامل ارتباط سبيرمان، ومعامل ألفا كرونباخ، وذلك باستخدام برنامج SPSS. وأظهرت النتائج مستوى جيدًا نسبيًا من الفهم المعجمي السياقي بصورة عامة، حيث بلغ متوسط الدرجات 34.07 من أصل 47. وكان أداء الطلبة في الفئات الدلالية الأربع متقاربًا نسبيًا؛ إذ بلغت نسبة الأداء في المعنى الحرفي 72.44%، وفي المعنى المجازي 70.62%، وفي المعنى النحوي السياقي 75.28%، وفي المعنى الاصطلاحي 72.00%. وأظهر اختبار فريدمان عدم وجود فروق ذات دلالة إحصائية بين الفئات الأربع، حيث بلغت قيمة $\chi^2 = 2.242$ ، وقيمة $p = 0.524$. وبالمثل، لم يُعثر على فرق ذي دلالة إحصائية بين طلبة السنة الثالثة وطلبة السنة الرابعة في الدرجة الكلية للاختبار ($U = 212.50$ ، $Z = -0.858$ ، $p = 0.391$)، على الرغم من حصول طلبة السنة الرابعة على متوسط رتبي أعلى. وقد ظهر فرق ذو دلالة إحصائية في الكلمة الثالثة فقط ($p = 0.039$). كما أظهر الاختبار اتساقًا داخليًا مرتفعًا ($\text{Cronbach's } \alpha = 0.939$)، وظهرت ارتباطات موجبة قوية بين الفئات الدلالية الأربع والدرجة الكلية للاختبار ($r_s = 0.874-0.952$ ، $p < 0.001$). وتشير النتائج إلى أن الطلبة يمتلكون معرفة معجمية سياقية جيدة نسبيًا، لكنها متباينة، وأن التقدم الأكاديمي وحده قد لا يكون كافيًا

لإحداث نمو ذي دلالة في عمق المفردات. وعليه، توصي الدراسة بتقديم تعليم صريح ومنهجي يعرض المتعلمين للاستخدامات السياقية المتعددة للمفردات نفسها، بهدف تعزيز الفهم السياقي للمفردات وعمق المعرفة المعجمية.

الكلمات المفتاحية: تعدد معاني الكلمات، المعنى السياقي، عمق المفردات، تعدد المعاني، المعنى المجازي، المعنى الاصطلاحي، المعنى النحوي، طلبة الجامعة.

Introduction

The phenomenon of polysemy in the English language is one of the most prominent challenges facing second language (L2) learners, as lexical knowledge of a word is not limited only to recalling its direct lexical meaning, but also extends to understanding its multiple contextual functions. Understanding vocabulary is the basis of linguistic competence and production, but meaning changes and is shaped based on the context in which the word appears. Whether it is a literal context, a metaphorical context, an idiomatic context, or a grammatical function.

Recent linguistic literature (e.g. Nation, 2013; Schmitt, 2010) confirms that possession of verbal wealth goes beyond mere “vocabulary size” to “vocabulary depth.” In the academic context, university students are expected to deal with complex texts that require unpacking rhetorical metaphors, idiomatic constructions, and functional uses of words. While the literal meaning of a word is easy to recognize, non-literal meanings pose challenges that requires higher-level linguistic processing (Littlemore, 2009).

Although English vocabulary is taught extensively at undergraduate levels, a recurring observation is that students' ability to comprehend multiple meanings decreases significantly when a word moves from its familiar literal context to figurative, idiomatic, or grammatical contexts. Therefore, many students tend to apply the initial literal meaning of the word in all contexts and ignore the rest of the other connotations of the word. This is a major flaw in the general understanding of academic texts and the reason that weakens their comprehension and analytical efficiency.

Many previous studies have addressed students' competence in distinguishing between literal and figurative meanings in general (e.g., Boers, 2004), or focused on idiomatic expressions separately. However, there is a clear lack of comparative studies that comprehensively and simultaneously examine learners' ability to process the four meanings together (literal, figurative, idiomatic, and grammatical) of the same words across different contexts, and determine the nature of statistical differences in reception and comprehension between these contexts among undergraduate students.

Significance of the Study: This study derives its importance from two aspects:

- **Theoretical importance:** Enriching the linguistic literature related to the depth of lexical knowledge and theories of contextual processing of multiple meanings among English learners as second / foreign language approaches.
- **Practical significance:** Providing field insights for university curriculum designers and English language teachers to develop teaching strategies that focus on “interconnected

contextual meaning” rather than memorizing vocabulary in isolation, which enhances students’ analytical skills.

Research objectives: This study mainly seeks to achieve the following objectives:

1. Measuring the level of university students’ understanding of the multiple meanings of English words across four different contexts (literal, figurative, idiomatic, and grammatical).
2. Detect whether there are statistically significant differences in the average students’ comprehension scores due to a change in context (type of meaning).
3. Identify the types of meanings that are most difficult and most easily understood by students.

Research Questions & Hypotheses:

1. What is the level of university students’ understanding of the multiple meanings of English words in the contexts: literal, figurative, idiomatic, and grammatical accuracy?
2. Are there statistically significant differences at the significance level ($\alpha = 0.05$) in students’ understanding of meanings depending on the type of context?

The statistical hypotheses of the study showed that the null hypothesis states there are no statistically significant differences at the 0.05 level in students’ comprehension of English word meanings across the four contextual categories: literal, figurative, idiomatic, and grammatical.

Based on the above, this study examines in an evaluative and analytical manner the extent to which university students are able to comprehend and deconstruct the multiple meanings of English words across literal, figurative, idiomatic, and grammatical contexts, with the aim of providing an accurate scientific vision that contributes to diagnosing the strengths and weaknesses of the learners’ contextual lexical understanding.

1. Literature review & Theoretical framework

1.1 Theories and Models

1. Nation’s Depth of Vocabulary Knowledge Framework, 2013.

Explanation and application: Nishi (Nation, 2013) divides vocabulary knowledge into three main dimensions: form, meaning, and use. The dimension of meaning branches out into understanding lexical and contextual relationships, with models holding that the learner may know the basic literal form and meaning, but fail to recognize the contextual and figurative constraints or grammatical functions of the word. (Nation, 2013; Schmitt, 2014).

2. Conceptual Metaphor Theory (CMT)

Explanation and application: Developed by Lakoff and Johnson (1980), it assumes that metaphor is not just a rhetorical tool but rather a cognitive system that links the source domain (Source Domain) and the target domain (Target Domain). This theory helps explain how students transfer literal meaning to figurative or idiomatic meaning across different contexts. (Lakoff & Johnson, 1980; Littlemore, 2009).

3. Direct Access & Graded Salience Hypothesis

Explanation and application: The hypothesis (Giora, 2003) indicates that the most prominent and recurring meaning (often the literal meaning or the common meaning) is recalled first in the mind of the learner, regardless of the context, which explains the difficulty of learners moving to complex idiomatic and grammatical meanings with the same efficiency. (Giora, 2003; Verspoor & Lowie, 2003).

1.2 Literature Review

What is the Contextual Meaning, Contextual Comprehension, and the Meanings of Words Across Context?

Contextual meaning refers to the interpretation of a word, phrase, or sentence based on the situation, language, and surrounding text where it appears. It emphasizes the importance of context in understanding language, as the same words can have different meanings depending on their usage and surrounding elements. Grasping contextual meaning enhances comprehension and communication by allowing us to accurately infer intended meanings and nuances. It is the part of meaning that comes from the situation, not just the dictionary entry or sentence structure (Ayad et al., 2026).

A simple example is "It's cold in here." Semantically, that sentence describes temperature. But in a real conversation, you might hear it as a request to close a window, turn off a fan, or end a discussion. The literal content stays the same, but the contextual meaning changes because listeners draw on speaker intention, shared assumptions, and the immediate setting (Karrar et al., 2026).

Contextual comprehension refers to the ability to understand and interpret information based on the surrounding context, which significantly influences meaning and understanding in both reading and communication (Hmouma, 2025). This skill is crucial for effective reading and communication, as it helps individuals derive meaning from texts and conversations that may not be explicitly stated.

The meanings of words can vary significantly depending on the context in which they are used. This phenomenon is known as polysemy, where a single word can have multiple meanings that are connected through shared concepts or history (Abouzied, 2026b). For example, the word "bank" can refer to a financial institution or the flat surface of a riverbank. Similarly, the word "head" can describe the top part of the body, the person in charge of a company, or the foam on a beer. Understanding these nuances is crucial for effective communication and comprehension. (*Words in English with Different Contextual Meanings | Lingoda*)

Previous documented studies are divided into two axes:

A: learners' performance in multiple meanings and different contexts

1. Littlemore, J., Krennmayr, T., Turner, J., & Paterson, R. (2014): The study examined the extent to which university students comprehend figurative and idiomatic meanings in academic lectures. The results showed that students, even at advanced levels, face great difficulties in deconstructing figurative and grammatical meanings compared to direct literal meanings.

2. Azuma and Sharifi's study (2018): The study examined students' understanding of polysemous words across four different contexts. I concluded that the grammatical and idiomatic context requires greater cognitive effort compared to the literal context, and that distinguishing between meanings depends on the extent of repetition and linguistic exposure.

B: Differences between academic levels (third year vs fourth year)

1. Al-Zahrani's study (Al-Zahrani, 2019): The study examined the growth and development in vocabulary depth (Vocabulary Depth) among third-year and fourth-year students in the English language departments. The results indicated that there were statistically significant differences in favor of fourth-year students in literal and figurative meanings due to increased exposure to academic texts.

2. Ahmed and Hassan's study (Ahmed & Hassan, 2021): It dealt with learners' performance in idiomatic expressions and grammatical functional meanings. Its results showed that linguistic development gains stability (Plateau effect) in the late university stages, as no fundamental differences appeared between the very advanced levels in terminological and grammatical comprehension.

1.3 Critical Analysis

Context of criticism: The results of the current research showed that there are no statistically significant differences between third-year and fourth-year students in overall performance and in the contexts of the four meanings (literal, figurative, idiomatic, and grammatical).

1.4 Deconstruct and critique previous studies based on this finding

1. **The Intermediate/Advanced Plateau Phenomenon:** The result of this study contradicts the results of the study (Al-Zahrani, 2019), which assumed a continuous linear development between the third and fourth years. This inconsistency gives the study monetary value; The results of this study support what is known as "flatness of linguistic development" (L2 Plateau), whereby learners at the advanced university level reach a certain level of receptive proficiency at which gradual differences cease unless explicit teaching strategies for vocabulary depth intervene (Ahmed & Hassan, 2021).
2. **Lack of methodological focus and university curricula:** Analytical criticism of previous studies shows that university curricula in their third and fourth years often focus on "vocabulary size" and "literary/critical content" without providing explicit and systematic training in contextual lexical processing. Therefore, the lack of differences between the two years is due to the similarity in the nature of linguistic input (Input) and traditional assessment methods that treat vocabulary as univocal or literal meanings (Abouzied, 2026a).
3. **The dominance of literal meaning (Salience Effect):** The similar performance between the two years in idiomatic, grammatical, and figurative meanings is due to the fact that both third- and fourth-year students still rely on "literal recall" as a primary strategy, which led to equal levels of error and comprehension across the two levels.

1.5 Research Gap

The direct research gap is identified in the following points:

1. Limitations of simultaneous four-way comparison: Most previous studies (e.g. Littlemore, J., Krennmayr, T., Turner, J., & Paterson, R. 2014; Boers, 2004) focused on pairwise comparison only (literal vs metaphor, or literal vs idiomatic). This study fills the gap by examining the four simultaneous dimensions (literal, figurative, idiomatic, grammatical) of the same words.

2. Developmental Plateau Gap: Previous literature has contradicted whether additional academic exposure (fourth year compared to third year) automatically leads to a deepening of understanding of multiple meanings. This study fills this field gap to prove or deny the longitudinal development of vocabulary depth among university students.

2. Methodology

a. Research Design: This study relied on the descriptive and analytical approach of a quantitative nature (Quantitative Descriptive Research Approach). Because it is appropriate for the nature of the study that aims to measure the level of understanding of students in multiple contexts, compare performance between academic levels, and reveal the nature of the correlational relationships between the different semantic categories of English vocabulary.

b. Study population and sample: The study population consisted of students in the English Department at the College of Arts (Al-Jameel District) for the academic year 2025/2026 AD. The study sample was selected using a purposive sample method, as it included all third-year students (25), and fourth-year students (20), so that the total size of the study sample was (45) participants.

c. Research Instrument and Development: To collect the necessary data, a controlled paper test was designed according to the following procedural steps:

1. Vocabulary selection: (25) English words commonly used in daily conversations were identified, and the approved Cambridge Dictionary was consulted to study their connotations and contextual meanings.

2. Vocabulary filtering: Words whose meanings ranged from three to five meanings were excluded, and the study was limited to words that the dictionary showed as having four different meanings and contexts. They are (15) main words, namely: (Trunk, Broad, Chair, Exam, Object, Enhance, Sentence, Subject, Reduce, Conduct, Mission, Operation, School, Board, Tree).

3. Preparing test items: The test was designed to include four different contexts for each word drawn from the same dictionary, to measure the student's ability to perceive meaning according to context.

4. Validity of the test: The measurement tool was presented in its initial form to an arbitrator specialized in linguistics in the English Department, who expressed his observations and approved the validity of the test and its ability to measure students' levels and abilities in understanding multiple meanings of vocabulary through the presented contexts.

5. Structure of test items and statistical refinement: The number of raw items in their initial form reached (60) items. After applying the test to the exploratory/main sample and examining the data to verify its integrity, (13) invalid items were excluded, bringing the final number of items subject to statistical analysis to (47) items, distributed as follows: (20) items for literal meanings, (10) items for idiomatic meanings, (9) Clauses for figurative meanings, and (8) items for grammatical meanings.

6. Practical application procedures: The testing field experiment was implemented after fully coordinating with the students and taking their informed consent for voluntary participation in the research. After collecting and correcting the test papers, the scores were recorded and entered into the Statistical Package for the Social Sciences (SPSS).

d. Statistical Treatment Analysis: To achieve the objectives of the study and answer its questions, the following statistical methods were used via the (SPSS) program:

1. Descriptive statistics: Means, standard deviations, frequencies, and percentages were calculated; To describe students' grades, their level of performance in the four semantic categories, and the overall grade.

2. Normality Test: The Shapiro-Wilk and the Kolmogorov-Smirnov tests were performed, and the results revealed that the data did not follow a normal distribution (the probability value was $p = 0.002$ for Shapiro-Wilk, and $p = 0.007$ for Kolmogorov).

3. Non-parametric Tests: Based on the result of the moderation test, non-parametric tests were relied upon in the analysis, and they included: Mann-Whitney U Test to compare between study groups, Spearman's Rank Correlation and Pearson Correlation test to measure correlations, Friedman Test to detect differences between semantic categories, and Cronbach's Alpha to measure the reliability and internal consistency of the test items.

The results of statistical analyzes (Pearson and Friedman) showed high internal consistency and consistency for all items; There were strong positive and statistically significant correlation coefficients between the four semantic categories and the total score, their values ranging at (0.952, 0.906, 0.874, 0.923), and all of which were significant at the high level of significance ($P = 0.001$).

3. Results

Table 1 — Test characteristics, reliability, and item analysis

N = 3 rd (20) + 4 th (25)	Items (Qs)	Mean total score	Cronbach's Alpha	SD	Minimu m	Mean	Maximu m	Corrected Item-Total Correlation	
								Low	High

45	47	34.07	.939	10.565	13	34.07	47	.305	.720
----	----	-------	------	--------	----	-------	----	------	------

The results of Table (1) indicate that N=45, Number of items = 47, Mean = 34.07, SD = 10.565, Minimum = 13, Maximum = 47, Cronbach's Alpha = .939, and Corrected Item-Total Correlation = .305 to .720. These are very strong results.

Table 2—Performance for words and the four semantic categories

Measure	Mean	Std. Deviation	Mean percentage
Word_1	3.38	1.134	84.44%
Word_2	2.18	1.173	72.59%
Word_3	3.07	1.405	76.67%
Word_4	1.64	.645	82.22%
Word_5	2.24	1.171	74.81%
Word_6	.60	.495	60.00%
Word_7	2.93	1.304	73.33%
Word_8	2.09	.996	69.63%
Word_9	2.93	1.232	73.33%
Word_10	2.71	1.576	67.78%
Word_11	.49	.506	48.89%
Word_12	2.67	1.243	66.67%
Word_13	2.84	1.147	71.11%
Word_14	1.07	.939	53.33%
Word_15	3.22	1.166	80.56%
Literal meaning	72.44	23.875	72.44%
Figurative meaning	70.62	24.750	70.62%

Grammatical context meaning	75.28	25.351	75.28%
Idiomatic meaning	72.00	22.422	72.00%

The table shows the performance of the 15 words, then the four semantic categories. Averages of words are calculated out of four, such as: Word 1 = $3.38 = 84.44\%$, Word 11 = $.49 = 48.89\%$, Word 14 = $1.07 = 53.33\%$. As for the rest of the words, they are counted from 1-3 because some words have one context, others have two contexts, and the last have three contexts. Therefore, the semantic categories are classified as follows: Literal = 72.44% , Figurative = 70.62% , Grammatical = 75.28% , and Idiomatic = 72.00% .

Table 3 - Statistical comparisons between third- and fourth-year students

Variable	Third Mean Rank	Fourth Mean Rank	Mann-Whitney U	Z	Asymp.Sig. (2-tailed)	Effect size	Classification
Word_1	23.96	21.80	226.000	-.686	.493	.102	
Word_2	23.64	22.20	234.000	-.415	.678	.062	
Word_3	19.92	26.85	173.000	-2.060	.039	.307	
Word_4	23.36	22.55	241.000	-.265	.791	.040	
Word_5	23.94	21.83	226.500	-.642	.521	.096	
Word_6	25.70	19.63	182.500	-1.817	.069	.271	
Word_7	21.88	24.40	222.000	-.685	.493	.102	
Word_8	21.36	25.05	209.000	-1.002	.316	.149	
Word_9	20.48	26.15	187.000	-1.531	.126	.228	
Word_10	22.00	24.25	225.000	-.623	.533	.093	
Word_11	21.90	24.38	222.500	-.725	.468	.108	
Word_12	20.56	26.05	189.000	-1.446	.148	.216	

Word_13	21.98	24.28	224.500	-.614	.539	.092	
Word_14	24.26	21.43	218.500	-.789	.430	.118	
Word_15	22.68	23.40	242.000	-.208	.835	.031	
Total literal meaning	21.56	24.80	214.000	-.828	.408	0.123	Small effect
Total figurative meaning	20.76	25.80	194.000	-1.298	.194	0.193	Small effect
Total grammatical context meaning	22.18	24.03	229.500	-.485	.628	0.072	Negligible effect
Total idiomatic meaning	22.58	23.53	239.500	-.243	.808	0.036	Negligible effect
Total Qs score	21.50	24.88	212.500	-.858	.391	.128	Small effect
Friedman test for (4) semantic categories	Chi-Square	P. Value	-	-	-	-	-
	2.242	.524	-	-	-	-	-

This result means that: Third-year Mean Rank = 21.50, Fourth-year Mean Rank = 24.88, $U = 212,500$, $Z = -.858$, $p = .391$, and the Effect size of Total Qs scores = .128. So, there are no statistically significant differences between third- and fourth-year students in the total test score. The difference in the direction of the average ranks favors the fourth year, but it is not a statistically significant difference. Friedman test shows $\chi^2 = 2.242$, and $p = .524$. That means, there are no statistically significant differences between students' performance in the four semantic categories.

Table 4 — Examination of normal distribution

Variable	Shapiro-Wilk Tests of Normality (A)			Mann-Whitney Test Statistics (B)		
	Shapiro-Wilk			Mann-Whitney U		
	Statistic	df	Sig.	U	Z	Asymp.Sig. (2-tailed)
Total literal meaning	.893	45	.001	214.000	-.828	.408

Total figurative meaning	.899	45	.001	194.000	-1.298	.194
Total grammatical context meaning	.827	45	.000	229.500	-.485	.628
Total idiomatic meaning	.925	45	.006	239.500	-.243	.808
Total Qs score	.910	45	.002	212.500	-.858	.391

The present table states that the Shapiro–Wilk test (.002) showed departure from normal distribution, so Mann–Whitney was used as a better test for up normality distribution.

Table 5 — Evidence of internal validity and correlations between semantic categories

Nonparametric Correlations

			Total literal meaning	Total figurative meaning	Total grammatical context meaning	Total idiomatic meaning	Total Qs scores
Spearman's rho	Total literal meaning	Correlation Coefficient	1.000	.803**	.792**	.788**	.952**
		Sig. (2-tailed)	.	.000	.000	.000	.000
	Total figurative meaning	Correlation Coefficient	.803**	1.000	.855**	.875**	.923**
		Sig. (2-tailed)	.000	.	.000	.000	.000
	Total grammatical context meaning	Correlation Coefficient	.792**	.855**	1.000	.816**	.874**
		Sig. (2-tailed)	.000	.000	.	.000	.000
	Total idiomatic meaning	Correlation Coefficient	.788**	.875**	.816**	1.000	.906**
		Sig. (2-tailed)	.000	.000	.000	.	.000
	Total Qs scores	Correlation Coefficient	.952**	.923**	.874**	.906**	1.000
		Sig. (2-tailed)	.000	.000	.000	.000	.

The table explained the Internal consistency of items. Spearman's rho and the results are very strong: Literal $\times$ Total = .952, Figurative $\times$ Total = .923, Grammatical $\times$ Total = .874, and Idiomatic $\times$ Total = .906. And all of them: $p < .001$. This is strong evidence that the four categories are related to the overall test score.

4. Discussion

Table 1 — Test characteristics, reliability, and item analysis

Interpretation and discussion:

The results of Table (1) indicate that the 47-item test was applied to a sample of 45 third- and fourth-year students. The arithmetic mean of the total score was 34.07 out of 47 degrees, with a standard deviation of 10.565, while the students' actual scores ranged between 13 and 47 degrees. These results indicate that there is a relatively noticeable variation in students' levels in the test of knowledge of multiple meanings of a single word, which allows distinguishing different levels of performance within the sample.

Cronbach's Alpha internal consistency coefficient also showed a high value of $\alpha = .939$, a value that indicates a very high level of internal consistency between the test items and provides strong evidence of the reliability of the instrument in measuring the target construct. The corrected correlation coefficients between each item and the total score also ranged between .305 and .720, which indicates that the items are appropriately related to the total test score, and that they contribute consistently to measuring the target ability. Overall, these results support the validity of the test in terms of reliability and internal consistency, and provide an appropriate basis on which to base its results in subsequent analyses.

Table 2—Performance for words and the four semantic categories

Interpretation and discussion:

Table (2) shows that there is variation in the level of students' performance on the fifteen words, which indicates that the students' ability to deduce the appropriate meaning of the word was not equal across different words and contexts. Word 1 achieved the highest performance rate of 84.44%, followed by Word 15 at 80.56%, and Word 4 at 82.22%, while the lowest performance levels appeared in Word 11 at 48.89%, Word 14 at 53.33%, and Word 6 at 60.00%. These differences reveal that the difficulty of identifying the appropriate meaning was not only related to the word itself, but may also be related to the context in which the word was mentioned and the multiple connotations it has.

At the level of the four semantic categories, the average performance in the literal meaning reached 72.44%, in the figurative meaning 70.62%, in the grammatical context 75.28%, and in the idiomatic meaning 72.00%. These results show a relative convergence in performance levels between the four categories, with a slight descriptive advantage for grammatical context, while figurative meaning recorded the lowest performance percentage. This result initially indicates that students were able to deal with different types of meaning to similar degrees, although moving from literal meaning to figurative or idiomatic meaning may represent an additional challenge in some contexts. However, determining whether these descriptive differences are

statistically significant requires referring to the results of the Friedman test presented in Table (3).

Table 3 - Statistical comparisons between third- and fourth-year students

Interpretation and discussion:

Table (3) shows the results of the Mann–Whitney U test used to detect differences between third- and fourth-year students in the performance of the fifteen words and semantic categories and the total test score. The results indicate that there are no statistically significant differences between the two groups in most words, as the statistical significance values exceeded the .05 level. The only exception was Word 3, where the result showed a statistically significant difference between the two groups, $U = 173.00$, $Z = -2.060$, $p = .039$, with an effect size of .307, which indicates a relatively small to medium effect.

At the level of semantic categories, no statistically significant differences emerged between third- and fourth-year students in literal meanings ($p = .408$), figurative meanings ($p = .194$), grammatical context ($p = .628$), or idiomatic meanings ($p = .808$). There was also no statistically significant difference in the total test score ($U = 212.50$, $Z = -.858$, $p = .391$). Although the average rank for the total score was higher among fourth-year students (24.88) compared to third-year students (21.50), this difference did not reach the level of statistical significance.

This result indicates that the progression from the third to the fourth year did not result in a clear statistical superiority in the general ability to interpret multiple meanings of a word within the contexts used in the test. This may mean that knowledge of multiple meanings does not necessarily develop spontaneously with academic progress, but may require more focused exposure and education to multiple meanings, context, and the use of vocabulary in diverse linguistic situations.

Table 4 — Examination of normal distribution

Interpretation and discussion:

The results of examining the assumption of a normal distribution showed that some variables did not meet the assumption of a normal distribution, as some of the statistical significance values did not exceed the approved level of .05. There were also indications of a deviation from the normal distribution in some variables from the usual standard values for some words, such as the Word 3 in the fourth-year group, for kurtosis of 6.628 (Shapiro-Wilk test). Based on this result, nonparametric statistical tests were used when comparing the two study groups, especially the Mann-Whitney test, as it is appropriate for comparing two independent groups when the assumption of normal distribution is not met. The choice of nonparametric tests is consistent with the nature of the data and the sample size (45 students), and also enhances the soundness of the statistical conclusions shown in Table (3).

Table 5 — Evidence of validity and correlations between semantic categories

Interpretation and discussion:

The evidence presented in Table 5 indicates that there is an appropriate level of content validity for the test. The target vocabulary was chosen based on its common use, reference to its different meanings and connotations, and then building multiple contexts for it, before presenting the tool to a specialist in the field of linguistics to verify its suitability for measuring students' ability to recognize multiple meanings of a single word in different contexts. This procedure provides qualitative evidence that test content is related to the construct it aims to measure.

Spearman correlation coefficients also showed strong, statistically significant positive relationships between the four semantic categories and the total test score. The correlation between the literal meaning and the total score was $r_s = .952$, between the figurative meaning and the total score, $r_s = .923$, between the grammatical context and the total score, $r_s = .874$, and between the idiomatic meaning and the total score, $r_s = .906$, and all of these correlations are significant at the $p < .001$ level. These results indicate that students' performance in each of the semantic categories is strongly related to their overall performance on the test, which supports the internal consistency of the measured construct and indicates that the four categories represent interconnected components within the general ability targeted in the study.

Overall discussion of all results

This part is very important in research; Because we do not want to merely redescribe the tables, but rather we want to answer: What do the combined results mean?

Proposed general discussion:

The overall results of the study reveal a good level of ability of third- and fourth-year students to recognize multiple meanings of the English word in different contexts, with a clear variation in performance from one word to another. The average total score for the test was 34.07 out of 47, while the analysis of word performance showed variation in the percentages of correct answers, as the percentages ranged between 48.89% and 84.44%. This disparity indicates that knowledge of the English vocabulary does not represent fixed or one-dimensional knowledge, but rather is affected by the learner's ability to determine the meaning imposed by the contextual environment of the vocabulary.

This result gains special importance in light of the nature of the test, which does not merely measure recognition of the word or its common meaning, but rather tests the student's ability to move between different connotations of the same word depending on the context. This is clearly evident in the contrast between words, as some vocabulary words achieved high levels of performance, while other words presented greater difficulty for students. This suggests that having a basic meaning of a word does not necessarily guarantee the ability to interpret its meaning in all contexts.

Although there were descriptive differences between the four semantic categories, performance levels were relatively similar; The percentages reached 72.44% for literal meanings, 70.62% for figurative meanings, 75.28% for grammatical context, and 72.00% for idiomatic meanings. Importantly, the Friedman test did not reveal statistically significant differences between these categories, $\chi^2 = 2.242$, $p = .524$. Therefore, the apparent differences in percentages represent

descriptive differences and not statistical differences that can be generalized to the study population.

On the other hand, the comparison between third- and fourth-year students using the Mann–Whitney U test showed that there were no statistically significant differences in the total score of the test, $p = .391$, as well as the absence of significant differences in most words and semantic categories. Despite the higher average ranks in favor of fourth-year students, this superiority was not enough to prove the existence of a statistical difference. This can be explained by the fact that general academic exposure to the English language and progression from one school year to another does not necessarily lead to a proportionate growth in the ability to interpret multiple meanings of a word, especially if this linguistic aspect does not receive explicit and systematic training within educational practices.

In contrast, statistically significant differences emerged in Word 3 only, which indicates that some vocabulary or contexts may be more sensitive to individual or academic differences between learners than others. Since the rest of the words did not show similar differences, the result for Word 3 should not be interpreted as evidence of general superiority for fourth- or third-year students, but rather it should be treated as a result specific to this word and the contexts associated with it.

The reliability results support these conclusions, as the Cronbach's Alpha coefficient reached .939, with the corrected correlation coefficients between the items and the total score ranging between .305 and .720. These values provide strong evidence of the internal consistency of the test and indicate that the items perform to an appropriate degree within the overall instrument. There were also high correlations between the semantic categories and the total score, which ranged between .874 and .952. They are all significant at $p < .001$, provides additional support that the four categories are interrelated components within the general ability that the test measures.

Based on the overall results, it can be said that the study provides evidence that the sample members' knowledge of multiple meanings of the English word was relatively good but heterogeneous across vocabulary and contexts. The results also indicate that the main difficulty does not necessarily consist of one type of meaning, as no statistical differences appeared between the literal, figurative, idiomatic and grammatical context categories, but rather it appears that the nature of the word and the context in which it is used play a clearer role in explaining the difference in performance.

The present findings support the distinction between vocabulary size and vocabulary depth proposed in previous vocabulary research. Although students demonstrated a generally good level of performance, their comprehension varied considerably across lexical items, suggesting that knowing a word does not necessarily imply knowing all of its contextual uses. This finding is consistent with Schmitt's (2014) view that vocabulary depth represents an important dimension of lexical knowledge that cannot be reduced to the number of words learners know.

5. Conclusion

This study investigated university students' ability to comprehend multiple meanings of English words across four contextual dimensions: literal, figurative, idiomatic, and grammatical meanings. The study was conducted with 45 undergraduate students, comprising 25 third-year and 20 fourth-year students in the English Department. Using a 47-item contextual vocabulary test, the study examined students' overall level of contextual lexical comprehension, differences between academic levels, and the relationships among the four semantic categories.

The findings indicate that the participants demonstrated a generally good level of ability to interpret multiple meanings of English words, although their performance varied considerably across individual words and contexts. The mean total score was 34.07 out of 47, with scores ranging from 13 to 47. At the category level, students obtained relatively similar performance rates in literal meaning (72.44%), figurative meaning (70.62%), grammatical-context meaning (75.28%), and idiomatic meaning (72.00%). Although grammatical-context meaning showed the highest descriptive percentage and figurative meaning the lowest, the Friedman test revealed no statistically significant differences among the four semantic categories, $\chi^2 = 2.242$, $p = .524$. Thus, the apparent differences in percentages should be interpreted as descriptive rather than statistically significant differences.

With respect to the first research question, the findings suggest that students possessed a relatively good but heterogeneous ability to comprehend multiple contextual meanings. Their performance was not uniform across the fifteen target words, with correct-response rates ranging from 48.89% to 84.44%. This variation suggests that contextual lexical comprehension cannot be explained solely by whether a meaning is literal, figurative, idiomatic, or grammatical. Rather, the particular lexical item and the contextual conditions in which it occurs appear to play an important role in determining comprehension.

Regarding the second research question, the findings provide no evidence of statistically significant differences in overall contextual lexical comprehension between third- and fourth-year students. The Mann–Whitney U test showed that the difference in total test performance was not statistically significant ($U = 212.50$, $Z = -0.858$, $p = .391$), despite the higher mean rank obtained by fourth-year students. Similarly, no significant differences were found between the two academic levels in literal, figurative, grammatical, or idiomatic meanings. The only significant difference occurred for Word 3 ($U = 173.00$, $Z = -2.060$, $p = .039$), which should be regarded as a word-specific finding rather than evidence of a general academic-level advantage.

The study therefore suggests that progression from the third to the fourth university year does not necessarily result in a statistically significant improvement in students' ability to interpret multiple meanings of words in context. This finding is consistent with the broader distinction between vocabulary size and vocabulary depth, whereby knowing more words does not necessarily mean knowing those words deeply enough to recognize their different contextual uses. Schmitt (2014), for example, emphasizes that vocabulary depth concerns how well words are known and that depth can provide information beyond vocabulary size.

The study also provides evidence supporting the reliability and internal consistency of the research instrument. Cronbach's alpha was .939, while corrected item–total correlations ranged

from .305 to .720. In addition, strong positive correlations were found between each semantic category and the total test score, ranging from $r_s = .874$ to $r_s = .952$, all statistically significant at $p < .001$. These findings indicate that the four semantic dimensions are strongly interconnected components of the broader construct of contextual lexical comprehension.

Theoretically, the findings contribute to research on vocabulary depth and contextual lexical knowledge by demonstrating that students' knowledge of a word should not be viewed as a single, fixed meaning. Instead, lexical knowledge involves the ability to recognize and interpret meanings according to contextual use. This interpretation is compatible with the distinction between vocabulary size and depth emphasized in vocabulary research (Nation, 2013; Schmitt, 2014), and with cognitive and contextual perspectives on figurative and polysemous meaning.

Pedagogically, the findings suggest that university vocabulary instruction should move beyond presenting isolated dictionary definitions and should provide systematic opportunities for students to encounter and analyze the same lexical items in different contexts. Classroom activities can incorporate contrasting literal and figurative uses, idiomatic expressions, grammatical functions, contextual inference tasks, corpus-based examples, and repeated exposure to polysemous words. Such practices may help learners develop greater depth of lexical knowledge and improve their ability to infer intended meanings in academic reading, writing, and translation.

Overall, the study concludes that the participants had a relatively good level of contextual understanding of multiple English word meanings, but this knowledge was uneven across individual words and contexts. No statistically significant differences were found among the four semantic categories or between third- and fourth-year students in overall performance. Consequently, the principal educational implication is not that one particular semantic category represents a universal source of difficulty, but that contextual lexical knowledge requires explicit, systematic, and integrated instruction. Further research with larger and more diverse samples, additional academic levels, and longitudinal designs is needed to determine how contextual vocabulary depth develops over time and how explicit instructional interventions can enhance learners' ability to interpret multiple meanings.

6. Limitations of the Study

Several limitations should be considered when interpreting the findings of this study. First, the sample was relatively small and consisted of 45 students from one English Department, which limits the generalizability of the findings to students in other institutions or educational contexts. Second, the study employed purposive sampling and included only third- and fourth-year students; therefore, the findings cannot be generalized to other academic levels without further investigation. Third, the study used a researcher-developed contextual vocabulary test consisting of a selected set of English words and contexts. Although the instrument demonstrated high internal consistency, the findings may partly reflect the characteristics of the selected lexical items and contexts. Finally, the study employed a cross-sectional rather than longitudinal design. Consequently, the absence of a statistically significant difference between third- and fourth-year students should not be interpreted as definitive evidence that vocabulary depth does not develop over time. Future longitudinal research with larger and more diverse

samples would provide stronger evidence regarding the development of contextual lexical knowledge.

7. Pedagogical Implications and Recommendations

Based on the findings of the study, several pedagogical implications can be proposed. First, English vocabulary instruction at the university level should place greater emphasis on vocabulary depth rather than vocabulary size alone. Teachers should expose learners to the same lexical items across a range of literal, figurative, idiomatic, and grammatical contexts.

Second, vocabulary activities should encourage students to infer meaning from surrounding linguistic and situational information rather than relying exclusively on isolated dictionary definitions. Contrastive activities that present the same word in different contexts may be particularly useful in developing contextual awareness.

Third, curriculum designers should consider incorporating systematic practice with polysemous words and contextual meaning into university English courses. Such practice can include authentic reading passages, corpus-based examples, contextual inference exercises, sentence comparison, and translation or writing tasks that require students to select the appropriate meaning according to context.

Fourth, the absence of statistically significant differences between third- and fourth-year students suggests that additional academic exposure alone may not guarantee substantial development in contextual vocabulary depth. Therefore, explicit instructional attention to multiple word meanings should be integrated across university levels rather than being assumed to develop automatically through general language exposure.

Finally, future research should examine whether explicit instructional interventions designed around contextualized vocabulary learning can produce measurable improvements in learners' ability to recognize and use multiple meanings of English words.

References

- Abouzied, A. S. (2026a). Alternative assessment in the educational process. *International Journal of Peer-Reviewed Multidisciplinary Research*, 2(1), 90–111.
<https://ijprmr.com/index.php/ijprmr/article/view/17/27>
- Abouzied, A. S. (2026b). Test-Wiseness Skills among Female Students at the Abu-Isa Faculty of Education. *International Journal of Peer-Reviewed Multidisciplinary Research*, 3(1), 01-23
- Ahmed, M., & Hassan, R. (2021). The plateau effect in L2 vocabulary depth acquisition among senior university students. *Journal of Pragmatics and Language Teaching*, 14(2), 112–129.
- Al-Zahrani, S. (2019). Assessing vocabulary depth development in EFL undergraduate students: A comparative cross-sectional study. *International Journal of Applied Linguistics*, 29(3), 405–420.
- Ayad, N., Alrumayh, S., Hasan, L., Omar, Z., Hmuma, M., Alouzi, K., Kasheem, M., & Masuwd, M. (2026). The integration of Islamic values in teaching and learning

- practices at the University of Zawia. *Cigarskruie: Journal of Educational and Islamic Research*, 3(2), 192–202. <https://doi.org/10.65190/cigarskruie.v3i2.472>
- Cambridge University Press & Assessment. (n.d.). *Vocabulary*. In *Cambridge Dictionary*. <https://dictionary.cambridge.org/dictionary/english/vocabulary>
- Fiveable. (2026, July). *Contextual meaning | Intro to semantics and pragmatics*. <https://fiveable.me/introduction-semantics-pragmatics/key-terms/contextual-meaning>
- Giora, R. (2003). *On our mind: Salience, context, and figurative language*. Oxford University Press.
- HarperCollins Publishers. (n.d.). *Vocabulary*. In *Collins English Dictionary*. <https://www.collinsdictionary.com/dictionary/english/vocabulary>
- Hmouma, M. (2025). Exploring Libyan EFL Undergraduates' Attitudes Towards AI-Driven English Learning Applications. *University of Zawia Journal of Educational and Psychological Sciences*, 13(2). Retrieved from <https://journals.zu.edu.ly/index.php/UZJEPS/article/view/1064>
- Karrar RKF, Abdelrady AH, Ibrahim DOO, Masuwd M, Nsaif RA, Alouzi K, Hmouma M, Zeat T, Aladi SHR and Alfallah B (2026). Automating language input: investigating the effect of AI-generated contextual flashcards delivered via Anki on EFL learners' sentence-level fluency. *Frontiers in Education*. 11:1862047. doi: 10.3389/feduc.2026.1862047
- Azuma, T., & Sharifi, A. (2018). Contextual effect on polysemous word recognition in L2 learners. *System*, 74, 88–98.
- Lakoff, G., & Johnson, M. (1980). *Metaphors we live by*. University of Chicago Press.
- Littlemore, J. (2009). *Applying cognitive linguistics to second language learning and teaching*. Palgrave Macmillan.
- Littlemore, J., Krennmayr, T., Turner, J., & Paterson, R. (2014). An investigation into metaphor uses at different levels of second language writing. *Applied Linguistics*, 35(2), 117–144. <https://doi.org/10.1093/applin/amt004>
- Lingoda. (n.d.). *Words in English with different contextual meanings*. <https://www.lingoda.com/blog/en/english-words-that-mean-something-different/>
- Nation, I. S. P. (2013). *Learning vocabulary in another language* (2nd ed.). Cambridge University Press.
- Schmitt, N. (2014). Size and depth of vocabulary knowledge: What the research shows. *Language Teaching*, 47(4), 481–502.
- Verspoor, M., & Lowie, W. (2003). Making sense of polysemous words. *Language Learning*, 53(3), 547–586.