LEARN Journal: Language Education and Acquisition Research Network
ISSN: 2630-0672 (Print) | ISSN: 2672-9431 (Online)
Volume: 18, No: 1, January – June 2025

Language Institute, Thammasat University
https://so04.tci-thaijo.org/index.php/LEARN/index

Exploring the Lexical Approach for Vocabulary Learning Through AI-Driven Feedback

Montri Tangpijaikul

fhummtt@ku.ac.th, Department of Foreign Languages, Faculty of Humanities, Kasetsart University, Thailand

APA Citation:
Tangpijaikul, M. (2025). Exploring the lexical approach for vocabulary learning through AI-driven feedback. *LEARN Journal: Language Education and Acquisition Research Network, 18*(1), 1015-1038. https://doi.org/10.70730/SFNP1171

Received
01/08/2024

Received in revised form
27/12/2024

Accepted
06/01/2025

ABSTRACT

Despite the significant impact of the lexical approach for vocabulary learning, its classroom implementation has not been uniform. While related activities share the common Observe-Hypothesize-Experiment (OHE) elements, practitioners and researchers do not highlight how language input from the observing stage is turned into output and at what stage of learning it is likely to be most challenging for learners. The article reports on one classroom action research conducted on two groups of non-English-speaking university students. This present study investigates whether language proficiency plays a role and which stage of the OHE learning process deserves special attention. Different types of learning behaviors were found, and pedagogical implications that contribute to implementation of the approach are discussed. The paper argues that the OHE paradigm is not complete without immediate feedback, a component that has historically been impractical in formal EFL contexts and that AI-driven feedback should fill in the gap by enhancing the effectiveness of the lexical approach.

Keywords: Lexical approach, vocabulary, AI feedback, EFL, ELT

Introduction

Although the lexical approach is not widely accepted as a language learning syllabus due to its reliance on exposure to unlimited target language from authentic data (Racine, 2018; Thornbury, 1998), this paper argues that the Observe-Hypothesize-Experiment (OHE) paradigm, rooted in this approach (Lewis, 2002), can be effectively implemented in inductive vocabulary learning activities because it aligns with the natural sequence of language acquisition.

According to influential language learning theories, comprehensible input is important for language acquisition (Krashen, 1981), with comprehensible output also essential for the development of communicative competence (Swain, 1985). The OHE paradigm incorporates 'observation' of input from real language data, while comprehensible output takes the form of 'experimentation.' To indicate learning, learners should be able to construct language that is comprehensible to others. The cognitive processing involved in 'hypothesizing' bridges the divide.

While both receptive and productive knowledge are essential for vocabulary learning (Nation, 2022, p. 52), transitioning from the former to the latter can be challenging, particularly in the absence of interaction, as highlighted by the interaction hypothesis (Long, 1983). Studies conducted prior to the advent of AI technologies largely demonstrated the positive effects of corrective feedback on language acquisition (Carroll et al., 1992; Lightbown & Spada, 1990; Tomasello & Herron, 1989). However, these findings relied heavily on teacher intervention, which is often constrained by limitations in time, sufficiency, and sustainability.

Without immediate corrective feedback, learners have no clear indication of whether their language output is correct or appropriate. With the assistance of AI-generated immediate feedback, the issue of limited interaction can now be mitigated. Another question remains: Does the ability to observe, hypothesize, and experiment correlate with language proficiency

The objective of this study is to determine which stage in the learning process of the OHE paradigm poses the greatest challenges for different groups of learners and to explore how AI-generated feedback can enhance the lexical approach as an inductive vocabulary learning activity. These questions can be addressed only through research on learners' performance in observation-based activities.

Literature Review

The Lexical Approach

Grammatical units, such as word order, verb tense and sentence structure provide the framework for organizing lexical units into meaningful phrases and sentences. The two elements are complementary with language viewed as grammaticalized lexis (lexical units arranged in a grammatical order), a principle underlying the lexical approach (Lewis, 2002). Lexical knowledge contributes significantly to L2 language acquisition (Nation, 2005; Uchihara & Saito, 2019) and correlates with language proficiency (Gao, 2017; Masrai, 2022; Rafique et al., 2023). Approaches directly related to lexical knowledge have positive effects, especially on writing and speaking. This is because, instead of processing words separately, lexical chunking allows the brain to process larger amounts of language by bundling smaller units together and dealing with them as larger ones, thus serving as a 'shortcut for language processing,' resulting in fluency in output (Wang, 2021, p. 227).

The lexical approach's relevant cognitive processes start from observation of language input, followed by forming hypotheses based on that input and finally constructing novel output. This is referred to as the Observe-Hypothesize-Experiment (OHE) teaching paradigm, originally introduced by Michael Lewis (2002), as an alternative to a more traditional Present-Practice-Produce (PPP) model. It emphasizes learning lexical units in chunks, through exposure to authentic texts from language corpora, rather than memorizing isolated words or grammatical units. While grammarians may view *'If I were you, I'd …'* as two clauses, it can also be observed as a single lexical chunk based on its frequent co-occurrences (Racine, 2018, p. 3). The assumption that learners can make a critical examination of how language is formed is problematic, however. Without focused guidance at the phrasal and clausal level, mere language observation is analogous to throwing learners into a swamp of infinite number of words assembled in a text and expecting them to swim effectively.

While most previous studies on lexical chunks yields positive result (Du, 2016; Gao, 2017; Yang, 2015), classroom implementation of the lexical approach has not been uniform (Racine, 2018), and teaching practitioners have not shown how input from the observing stage is turned into output. For instance, Wang's (2021) experiment found that Chinese students used more lexical chunks across all linguistic categories (polywords, institutionalized expressions, phrasal constraints, sentence builders) after one semester of instruction using the lexical approach. While lexical knowledge increased as expected, it is unclear whether the students produced new lexical chunks appropriately, as grammatical errors were present.

Observing Sentences from Online Dictionaries

Among the sources available, many language teaching practitioners use dictionaries. Incorporating dictionaries in EFL classrooms leads to substantial improvement in use of collocations (Chen, 2016) and target vocabulary (Alsadoon, 2021). Online dictionaries are preferred because word definitions can be retrieved quickly with authentic sample sentences (Fauzanz et al., 2022). Most of the participants in a study conducted by Chiu & Liu (2013) acknowledged that online dictionaries assisted them to access the definitions of unfamiliar words quickly and remember them effectively. Dwaik (2015) found that language learners using online dictionaries appeared to have a higher average reading proficiency than those using printed dictionaries. Similarly, Trinh et al. (2021) found that learners have a strong preference for online dictionaries over printed ones because online dictionaries offer more sample sentences, collocations and grammatical patterns. Sample sentences in online dictionaries are taken from language corpora, and some are simplified to allow deeper understanding of the ways in which a word can be used. By providing sample sentences that demonstrate usage in various situations, learners can infer meaning through context clues provided by the surrounding text.

When learners have doubts about the meaning of L2 words and how to use them, sample sentences can help by serving as writing models (Tarp et al., 2017). The use of sample sentences in online dictionaries allows for meaningful input and output enabling them to write better in English, as learners can actively engage with the language through reading activities (Boonmoh, 2021). By reading sample sentences, learners could internalize the target vocabulary and adapt it in writing new sentences.

Most importantly, the accessibility and convenience of online dictionaries make them an ideal language input source both in and out of the classroom. Learners can access a wealth of comprehensible sample sentences from any device connected to the internet. Learner autonomy can thus be promoted in activities where online dictionaries are used (Chiu & Liu, 2013).

AI Feedback: the Last Piece of the Puzzle in the OHE Paradigm

Although the assumption that reading can help learners incorporate new vocabulary into their writing seems logical (Kim & Kim, 2022), there is no guarantee that target words will be used accurately. Recognizing a word in a sentence does not necessarily lead to the ability to appropriate use (Hsu, 2013). To effectively transfer new vocabulary from input to correct output, it is important for learners to receive feedback. The 'Interaction Hypothesis' (Long, 1983) holds the idea that language proficiency is fostered through the

negotiation of comprehensible input. Long's interaction hypothesis complements Krashen's input hypothesis by positing that while input is crucial, interactions and negotiations between communicators play a vital role in facilitating language acquisition (Long, 1994). Feedback, being fundamental to interaction, is crucial for learners to discern the correctness, appropriateness, and comprehensibility of their output.

Several studies have shown that written corrective feedback (WCF) plays a facilitative role in language learning (Zabihi & Erfanitabar, 2021), but identification of comprehensive errors is both discouraging and less effective than focused or selective WCF, where the area of attention is manageable in size (Lee et al., 2023). Another problem is the timing of feedback. Imagine a teacher returning assignments days later when students have moved on to new tasks. Memory limitations hinders recall of past goals and the delay in feedback discourages effective amendments (Galbraith & Vedder, 2019; Ranalli & Yamashita, 2022). Late feedback may be as unhelpful as no feedback at all. It is essential, therefore, to design writing tasks that provide learners with timely feedback (Webb, 2009), and focus more on selective errors in manageable portions at the phrasal or clausal level.

Automated feedback from artificial intelligence (AI) can supplement teacher feedback (Ranalli, 2018), especially for checking accuracy (Zhang, 2021), in writing (Huang & Renandya, 2020). Today automated written corrective feedback (AWCF) tools such as Grammarly and Quillbot can provide instant feedback to writers. With natural language processing (NLP) based on machine learning algorithms to simulate human-like conversation, chatbots such as ChatGPT can provide feedback, not only corrective grammatical notes but also commentaries on rhetorical concerns (Kohnke et al., 2023). This capacity of AI is particularly advantageous for language learning, as it facilitates prompt error identification and rectification, accelerating the acquisition process.

A recent review by Zhai & Ma (2023) analyzed 26 studies and identified positive effects of AI-driven feedback on writing quality. Similarly, Fu et al. (2022), in their review of 48 studies on AI, reported overall positive outcomes for student writing. Notably, their review highlighted a greater number of studies examining the impact of AI feedback on second language writing compared to first language writing. In an EFL context, AI tools have been found to effectively enhance vocabulary learning efficiency (Wang et al., 2024) while providing learners with interactive and comprehensible input (Alsadoon, 2021). Research is needed on how AI feedback can assist effective vocabulary tasks, building on earlier studies on explicit vocabulary instruction (Lee, 2003), the effectiveness of reading-integrated writing tasks, and their impact on vocabulary transfer from input to output (Lee & Muncie, 2006; Kim & Kim, 2022).

This study engages two groups of university students learning English as a foreign language in observation-based reading-integrated writing tasks and poses two questions based on the implementation of the OHE paradigm as follows:

1. Do students of higher proficiency use the observe-hypothesize-experiment (OHE) paradigm more effectively in their vocabulary learning than those of lower proficiency?

The first research question generates the following null hypotheses:

1.1 There is no significant difference between the students of higher proficiency and those of lower proficiency in their ability to carefully observe (O) sample sentences.
1.2 There is no significant difference between the students of higher proficiency and those of lower proficiency in their ability to extract hypotheses (H) based on the observed sample sentences.
1.3 There is no significant difference between the students of higher proficiency and those of lower proficiency in their ability to experiment (E) with new sentences from the observed linguistic patterns.

2. What different learning paths are observed in the way students use the observe-hypothesize-experiment (OHE) paradigm in their vocabulary learning?

Attention is given particularly to the middle of the OHE learning process, i.e., how learners extract their hypotheses from the sentence samples. This is important, as a mistaken hypothesis can lead to incorrect output in the experimental stage.

Research Methodology

Participants

A group of 60 Thai university students was randomly selected from two classes: one from a remedial English class and the other from an English major class. The 30 students in the first group, who come from different faculties and major subjects, are considered to have lower English proficiency, as they did not meet the university's English admission requirement. Their TOEFL scores were less than 450 (paper-based), 133 (computer-based), or 45 (internet-based), and they were therefore enrolled in

the remedial English course. The second group is homogeneous and clearly demonstrates higher language proficiency, as all its members are in their third year of studying in the English major program. The two groups were given the same observation-based sentence creation task to complete within 3 classes of 1.5 hours each. Details of the task are shown in the data collection and analysis. To ensure that the students understood how to perform the task, two 1.5 hour pre-sessional training classes were provided.

Word Selection

As the groups were from classes using different coursebooks, it was necessary that the selected words were part of the lessons of both. Focusing on the vocabulary exercises consisting of 240 target words from the remedial class and 460 target words from the English major class, it was found that 95 of these words are shared by both coursebooks. For academic purposes, words not found on Coxhead's (2000) Academic Word List (AWL) were excluded. The remaining words were then purposely shortlisted to 2 adjectives, 2 nouns, and 2 verbs, ranked by frequency and shown in uppercase in Table 1 below, as content words that function as adjectives, nouns, and verbs are the most looked up in online dictionaries by learners (Boonmoh, 2021).

Table 1

Top Frequency Ranks of the 95 Target Words Shared by the Two Coursebooks

Frequency ranks	ADJ	N	V
1	DIVERSE	dependence	appeal
2	explicit	extent	CONDUCT
3	independent	FEATURE	DISPOSE
4	INTENSE	motive	intend
5	remote	perspective	require
6	spectacular	PURSUIT	specialize
7	subject	resident	suppose

Table 1 lists content words that appear frequently in the two coursebooks, arranged in grammatical categories. In the adjectival category, 'diverse' and 'intense' were selected as they can modify either mental or material processes, e.g. diverse identity vs. diverse group of workers, and intense fear vs. intense color. In the nominal group, 'feature' was selected because it can also function as a verb, e.g. special feature vs. feature a report while 'pursuit' was chosen for its use in collocations, e.g. in pursuit of. It is of interest to find how they are used and adapted in sentence creation. In the

verbal class, 'conduct' was selected because it is a transitive verb that is typically followed by specific nouns, e.g. experiment, study, research while 'dispose' was chosen for its use with the preposition 'of.' It is of interest to find out how they will be adapted to new contexts. These 6 selected words were put in a task worksheet along with 9 other words. Students were not told which words would be the focus of the study.

Data Collection and Analysis

Students were instructed to complete an individual observation-based sentence creation task by filling out four columns for each keyword in a worksheet. In the first column, with the assistance of an online dictionary, the meaning of the given word is written as they understand it, either in English or Thai. In the second column (observation), they choose a sample sentence from the dictionary that relates to the meaning provided in the first column. In the third column (hypothesis), they extract the lexical and grammatical patterns of the word observed in the chosen sentence. In the fourth column (experiment), they create new sentences, not exceeding 15 words, and with no more than five consecutive words taken from the sample sentence, based on their hypotheses. In the final step, they are asked to input their sentences into ChatGPT (version 3.5) using prompts such as 'Check:' or 'Check grammar:' to receive feedback on their sentences. An example from the vocabulary worksheet is shown in Table 2.

Table 2

Example of the Vocabulary Worksheet

Word	1. Observation (O)	2. Hypothesis (H)	3. Experiment (E)	Remarks (AI feedback)
UTILIZE (v.) Use, apply	The vitamins come in a form that is easily utilized by the body.	Be + utilized (V3) + by	This substance is utilized by doctor to treat certain diseases.	The word 'doctors' should be plural. (ChatGPT)
CONDUCT				
DISPOSE				
DIVERSE				

The three elements are sequential, E is based on H; H on O (O → H → E). To review the student's written performance step-by-step, each element was considered separately to identify where mistakes or misunderstandings occur. The following codes were used.

O = Observation
H = Hypothesis
E = Experiment
C = Correction
✓ = Correct
× = Incorrect
* = Ungrammatical
? = Unclear meaning

In the observation column, students rewrote one sentence chosen from the dictionary. Sentences were marked as correct unless it was replicated incorrectly, (e.g. with misspellings), in which case it was marked as incorrect.

In the hypothesis column, students extracted hypotheses from the sentence sample in the form of a lexical collocation or a grammatical pattern, with or without frame slots. Hypotheses were marked as correct or incorrect.

In the experiment column, students created new sentences based on their hypotheses and have them checked by ChatGPT. A sentence was marked as correct if there were no errors related to the hypothesis. Should an error related to the hypothesis be found, the sentence is marked as incorrect. If errors were found in parts of the sentence not directly related to the hypothesis, the sentence was considered ungrammatical and marked with an asterisk. If the sentence was semantically unclear, ambiguous, or nonsensical, it was marked with a question mark. If the sentence contained both grammatical and semantic errors, both an asterisk and a question mark were used.

To answer the first research question, a two-sample t-test is used to determine whether the frequencies of correct responses from the higher proficiency group significantly differ from those of the lower proficiency group. To answer the second research question, different types of written performance are observed and categorized, with their salient patterns exemplified and discussed.

To maintain anonymity, codes were employed in the analysis. For instance, student number 8 from the lower (L) proficiency group is coded as L08, while student number 20 from the higher (H) proficiency group is coded as H20.

Results and Discussion

In Table 3, the performances of the higher and lower proficiency groups are compared. Scores are derived from the output of 6 vocabulary items multiplied by the students from each group (6 * 30 = 180).

Table 3

Comparison of Correct Responses from Higher and Lower Proficiency Groups for Each Stage

	High proficiency	Low proficiency	Statistical significance between the two groups (*p*-value)
Correct observation (Mean)	178/180 (5.93)	165/180 (5.50)	0.0105
Correct hypothesis (Mean)	149/180 (4.96)	107/180 (3.53)	0.0002
Correct experiment (Mean)	155/180 (5.16)	130/180 (4.33)	0.0061

As can be seen, for the observation stage, the higher proficiency group outperforms the lower proficiency group, with the difference noticeable in nearly half of the 30 individuals in the population, and statistically significant with the *p*-value less than 0.05. Null hypothesis 1.1 (no significant difference in performance of the observation task between the students of higher and lower proficiencies) can thus be rejected.

Consistent with this, the ability to extract linguistic patterns from sample sentences in the hypothesizing stage is greater in the higher proficiency group. This difference is greatest in significance compared to the other parts of the OHE paradigm, as can be seen by the raw frequency count, the mean score and the p-value. The null hypothesis is therefore rejected. In other words, the students of higher proficiency can extract lexical and grammatical patterns and come up with their language hypotheses (H) significantly more effectively than those of lower proficiency. This is probably because the skill in hypothesizing a linguistic pattern requires knowledge of how lexical and grammatical items are related, and this is more commonly taught explicitly in a higher-level grammar class (Andrews, 2007).

In line with the above findings, the ability to create new sentences based on observation of sentence samples is also greater in the higher-level group as shown by the higher scores with the p-value less than 0.05. Null hypothesis 1.3 is therefore also rejected, meaning that the higher proficiency

students significantly outperform lower proficiency peers in sentence creation based on the observation of given sentence samples. This is not surprising as the grammatical knowledge of a higher-level group is expected to be greater.

What is more interesting is a comparative view of the learners' performance at different stages across the two proficiency groups, enabling us to see when mistakes occur during the learning process.

Figure 1

Correct Output Between the 2 Groups at Different Stages

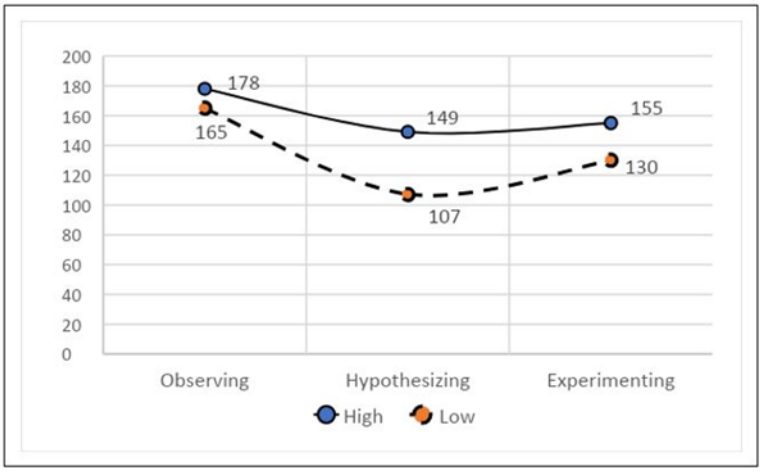


Figure 1 suggests that the initial stage of observing sample sentences is less problematic for learners compared to the other two stages. In this stage, learners simply replicate the chosen sentence, although errors may still occur due to carelessness or inattention, leading to incomplete sentences or misspellings.

The most significant difference between the two groups lies in their ability to deduce linguistic rules from the observations. This is crucial for learning, as incorrect hypotheses can lead to errors in the experimentation. However, the data also reveal that sentences produced in the experimenting stage are not always directly linked to incorrect hypotheses. In other words, some students manage to create new sentences correctly, even when they have not accurately extracted linguistic patterns from the sample sentence.

In dealing with the second research question to better understand the OHE learning process, this study identifies eight possible learning paths based on learners' favorable and unfavorable behaviors. The following

section presents patterns observed, and examples of correct and incorrect newly constructed sentences produced by students from both groups.

Successful Cases

Correct Hypothesis and Correct Output (✓✓✓)

[L09: ✓✓✓]
O: I enjoy *outdoor pursuits*, like hiking and riding. (✓)
H: [*adj.* + *pursuit(s)*] (✓)
E: My mom loves *indoor pursuits*, like cooking and baking. (✓)

The above example illustrates a situation in which the learner successfully derives the correct hypothesis from the sample sentence and creates a new sentence based on this. By replacing lexical items with new ones, the learner generates a different meaning while maintaining the original sentence structure. This illustrates the process in which the learner comprehends and applies the grammatical pattern, incorporating new vocabulary to create a sentence with a distinct meaning.

Correct Hypothesis and Correct Output with Grammatical Problems

[L02: ✓✓✓*]
O: Her eyes are her *best feature*. (✓)
H: [*attributive adj.* + *feature*] (✓)
E: Her hair color is *interesting feature*. (✓*)
C: Her hair color is an *interesting feature*.

In the student's newly created sentence, the indefinite article "an" is needed to clarify that hair color is one of the person's many features. The noun "feature" is a singular countable noun, which requires an article before it. This grammatical error is not directly related to the formulaic structure of the hypothesis, therefore the hypothesis is assumed correct. This ungrammaticality during the experimental stage is perhaps a sign of the student's developing interlanguage (Harmer, 2001, pp. 99-100).

[L02: ✓✓✓*]
O: Students from countries *as diverse as* Columbia and Lithuania use … (✓)
H: [*as diverse as*] (✓)
E: Plants from SE Asia *as diverse as* Thailand and Indonesia is abandant. (✓*)
C: Plants from SE Asia *as diverse as* Thailand and Indonesia are abundant.

Errors were found in subject-verb agreement, where "is" was used instead of "are," and the word "abundant" was misspelled in the example. However, the usage of "diverse" did not pose any issues. Regarding the grammatical point in question, the student successfully relied on the original context and crafted a new sentence similar to the original sample.

Correct Hypothesis and Correct Output with Some Semantic Problems

[H07: ✓✓✓?]
O: The experiments *were conducted by* scientists. (✓)
H: [*BE* + *conducted* + *by*] (✓)
E: Our study plans *were conducted by* him. (✓?)

[H29: ✓✓✓*?]
O: The battalion had been preparing to *conduct operations* in the same area. (✓)
H: [*conduct* + *something*] (✓)
E: Our factory always ready to *conduct our machines* for good products. (✓*?)

Although students in the higher group create new grammatical sentences during the experiment stage, they face challenges with word choice. In H07, passive construction is commonly used with the word 'conduct,' and it is conventional to say conduct operations as the observed example of H29 shows. However, the verb 'conduct' typically refers to carrying out an activity, such as research or an experiment. It is not commonly used in the context of study plans or how machines are operated. In H07, the word 'prepared' is preferred, while in H29, the word 'operate' is favored. The issue with H29 is the word 'something' in her hypothesis, which is too broad. Here, 'something' specifically refers to an activity to be carried out. This suggests that hypothesizing desirable formulaic patterns requires collocational knowledge, which is a challenge shared by EFL learners of all proficiency levels (Namvar, 2012; Lateh et al., 2021).

Incorrect Hypothesis but Correct Output

[L02: ✓×✓*]
O: Every month he *must dispose of* the oil his restaurant uses to fry … (✓)
H: [*modal V.* + *despose of* + *N.*] (×)
E: I *must dispose of* the garbage my house everyday. (✓*)
C: I *must dispose of* the garbage from my house every day.

The misspelling in H is simply due to carelessness. The new sentence is ungrammatical in two ways: the missing preposition 'from' in front of 'my house' and the adverb 'every day' that should be written as separate words. Disregarding some minor grammatical mistakes not related to the lexical pattern of the word under study, this case is considered successful because the word 'dispose' is used in the new context appropriately and semantically consistent with that shown in the sample sentence. From the sample sentence, the student seems to understand that 'dispose' is used together with an object that needs to be eliminated.

[L03: ✓×✓]
O: How you choose to *conduct your private life* is your own business! (✓)
H: [*conduct* + *pronoun*] (×)
E: I have to *conduct my research*. (✓)

The observed sentence, taken from the Cambridge Online Dictionary, is rewritten correctly. The word 'conduct' is followed by a noun phrase, which consists of the possessive adjective 'your' and another noun phrase 'private life.' However, the student was not aware that possessive adjectives should be considered part of the noun phrase, resulting in the incomplete hypothesis of 'conduct + pronoun' instead of 'conduct + noun (phrase).' Nevertheless, the correct sentence in the output suggests that she actually understands the grammatical pattern but struggled to express her understanding clearly in the hypothesis. Perhaps the ability to extract correct hypotheses from observed sentences needs to be taught explicitly at the sentence level (Andrews, 2007; Hochman & Wexler, 2017).

Unsuccessful Cases

Incorrect Observation Leading to Incorrect Hypothesis and Output

[L21: ×××]
O: This car *has* some excellent design *feature*. (×)
H: [*HAVE* + … + *feature*] (×)
E: This bike *has* good performance *feature*. (×)
C: This bike *has* good performance *features*.

In the above example, a mistake is found at the observation stage where the student failed to notice the plurality of the word 'features' in the sample sentence. Since the study requires students to rewrite the observed sentence exactly as they see it, we might expect no mistakes to be found at this stage. However, though rare, this is still possible. When a mistake occurs

at an initial stage, as here where the student was not aware that 'feature' should be plural when referring to a vehicle that has more than one feature, this results in the word being extracted in its singular noun form instead of the plural which leads to incorrect written output.

Incorrect Hypothesis Leading to Incorrect Output

[L01: ✓××*?]
O: Lowell, Massachusetts, is noted for its *diverse* ethnic *communities*. (✓)
H: [*BE* + *diverse*] (×)
E: Thailand, in Bangkok, there is a huge *diverse of* tourists. (××*?)
C: In Bangkok, Thailand, there is a huge number of *diverse tourists*.

While the adjective 'diverse' is used in the sentence sample in its attributive function, modifying the following noun, the extracted lexical pattern in the hypothesis shows 'diverse' in its predicative function, preceded by the verb 'to be.' This student seems to be confused about the different adjectival functions of 'diverse' and has also misused the adjective as a noun. The meaning in the propositional statement is also ambiguous, as Bangkok is part of Thailand. The sentence created by the student is therefore marked as (?) suggesting that it is problematic in terms of content organization as well.

As demonstrated in this example, it is crucial to address the phenomenon of mistaken hypotheses leading to incorrect output. This issue merits attention because we cannot assume the student will produce correct output without the ability to extract a hypothesis from the observed.

[H15: ✓××*]
O: Space is not the place within which things *are disposed*, but … (✓)
H: [*BE* + *disposed (V3)*] (×)
E: Those toys and mess *were disposed* after … (×)
C: Those toys and the mess *were disposed of* after …

The student appears to be confused between 'be disposed,' which can be replaced by 'be arranged,' and 'be disposed of,' which means 'be eliminated.' This confusion resulted in the output being inconsistent with their hypothesis and rendered it ungrammatical, as the latter phrase requires the preposition 'of.' Through increased exposure to a diverse range of sample sentences where the keyword is used in a variety of constructions, students should become more aware of these nuances.

[H16: ✓××?]
O: Her sense of humor *disposed me* to like her. (✓)
H: [*someone* + *dispose* + *someone*] (×)
E: Her behavior *disposes her teacher* badly. (×?)
C: Her behavior *disposes her teacher to* view her badly.

The above student's newly invented sentence was rather awkward and might lead to confusion. This arose from the incomplete hypothesis which should have been 'dispose + someone + infinitive with to'. The missing infinitive resulted in ambiguity in the experiment stage.

Correct but Incomplete Hypothesis Leading to Incorrect Output

[L01: ✓✓×*?]
O: He *was* young and *intense*. (✓)
H: [*BE* + *intense*] (✓)
E: He *was intense* mind to read book for this exam. (×*?)
C: He has an *intense mindset* to read books for this exam.

The hypothesis extracted from the sentence sample is correct, that is intense as a predicative adjective. However, the new sentence created has intense in its attributive modification of the noun mind. The derived sentence is therefore not consistent with the extracted linguistic pattern in the hypothesis. This suggests that the ability to recognize patterns in a sentence does not guarantee a thorough understanding of the pattern being observed. Specifically, predicative and attributive adjectives have distinct syntactic positions. This shows that the OHE paradigm still needs corrective feedback for learning to take shape.

Another point to note is the uncommon phrase 'intense mind', where 'intense mindset' is preferred to make clear the idea of someone being highly committed to a task. Finally, the context implies that the person is reading multiple books to prepare for the exam, so the plural form books should be used.

Correct Hypothesis, but Incorrect Output

[L13: ✓✓×*]
O: He suddenly felt *an intense pain* in his back. (✓)
H: [*an intense* + *N.*] (✓)
E: He has *accident intense* at crossroads. (×*)
C: He had an *intense accident* at the crossroads.

Even though the hypothesis is correct, with 'intense' functioning as an attributive adjective modifying the head noun that follows, the word 'intense' is placed after the noun in the experimental sentence. This is likely due to L1 interference since adjectives in the Thai language follow rather than precede the noun they modify. Another indication of this lower-level learner's interlanguage is the incorrect use of the present tense. This is probably because the notion of tenses in the Thai language is realized through lexical markers suggesting time, rather than integrated into sentences as a grammatical system as in English (Na Phuket, 2016). While generating the correct hypothesis from the input is important, this example demonstrates that there is no guarantee that a correct hypothesis derived from language input will lead to correct output.

In light of the findings, learning vocabulary through sentence observation within the framework of the lexical approach appears to enable learners to understand the meanings of new words and the basic lexical rules governing their use in sentences. The underlying assumption is that once learners correctly form hypotheses based on their observations, they can generate new and creative sentences. Unfortunately, not all learners can accurately extract linguistic patterns, leading to incorrect use of the target language, as can be seen in their learning paths (learner's footprints), e.g. ✓✕✕* or ✓✓✕*? shown in the findings. Another challenge remains at the experimental stage concerning semantic appropriateness or naturalness. Both grammatical and semantic anomalies can be attributed to the fact that errors are an inherent part of learners' interlanguage (Harmer, 2001, pp. 99-100).

While Wang (2021) identified positive correlations between the application of the lexical approach and learners' proficiency levels, this study adds that the most significant differentiating factor between higher and lower proficiency learners is their ability to extract linguistic patterns from their observations. Among the six selected words, 'dispose' proved to be the most problematic for learners, even within the higher proficiency group. Eleven out of thirty sentences created by the higher-level students contained errors in the use of this word, primarily due to incorrect rule formation. This difficulty is understandable, as the word can be used to describe either mental or material processes, supports both active and passive constructions, and requires collocational knowledge, such as the phrase 'be disposed of.' Beyond form and meaning, grammatical and collocational knowledge are essential components of knowing a word (Nation, 2022, p. 65; Schmitt, 2020, p. 33), with collocational knowledge found to be the most challenging aspect for EFL learners (Lateh et al., 2021; Namvar, 2012). The issue lies in how these abilities can be enhanced without explicitly teaching lexical and grammatical patterns and immediate feedback received from learners' output.

Since sentences and vocabulary form the building blocks of all writing, it has been suggested that learners focus on sentence grammar when working with words to derive new meanings (Hochman & Wexler, 2017). In a study comparing explicit and implicit grammar instruction, Andrews (2007) found that explicit instruction is significantly more effective than implicit instruction for mastering complex sentence structures. This idea is strongly suggested at the stage of forming hypothesis. Though the OHE teaching paradigm can be effective for vocabulary learning, incorrect or incomplete hypotheses are possible, which may lead to incorrect output.

Additionally, AI-generated immediate feedback can help correct these mistakes (Fu et al., 2022; Huang & Renandya, 2020; Kohnke et al., 2023; Zhai & Ma, 2023; Zhang, 2021), enabling vocabulary learning to take shape especially in activities where the OHE paradigm is applied.

Conclusion and Pedagogical Implications

To understand a word fully, it is crucial to grasp its parts of speech and various grammatical functions. Misusing these functions can lead to grammatical errors and ambiguity. To promote comprehensible output in vocabulary learning, students should be provided with the opportunity to observe how the word they are studying is used in grammatical sentences. This is where teaching basic phrase structure rules may be necessary (Andrews, 2007; Hochman & Wexler, 2017). For example, it is important to know that some verbs commonly appear in their passive construction, e.g., 'be conducted by,' [BE + V3 + by] or that an adjective can be part of a noun phrase [Det + Adj + N], or a verb phrase [BE + Adj].

To create new sentences, learners should not rely totally on imitation, as it can lead to plagiarism. That is why students in this study were instructed not to reproduce phrases containing more than five consecutive words from the sentence samples. There is a fine line between plagiarism and syntactic creativity, which arises from reconstructing the recursive linguistic patterns that enable the production of infinite sets of words (Chomsky, 1965). Distinguishing between them is crucial, and it is the responsibility of the teacher to raise awareness of this distinction. Therefore, while it is essential for learners to recognize linguistic patterns in the language input, in the final stage of vocabulary learning, they should have the opportunity to creatively use language in their own contexts, drawing from the comprehensible input they have observed.

This study has highlighted various types of incorrect output made by learners. The presence of mistakes in the OHE exercise is not inherently problematic however, as they are expected as part of learning (Harmer, 2001, pp. 99-100). The most significant issue lies in whether learners recognize and

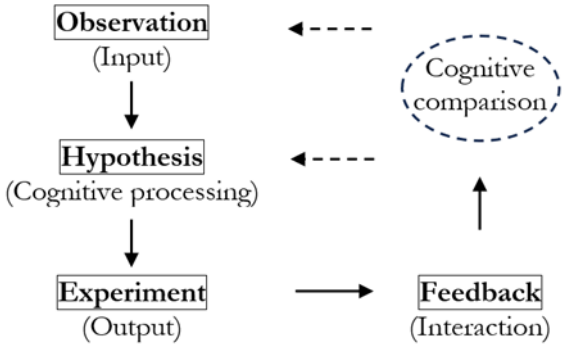
address these mistakes. This study therefore supports focused corrective feedback as part of learner development, an idea consistent with the findings of Zabihi & Erfanitabar (2021) and Lee et al. (2023).

At the initial stage of language learning, receiving feedback is crucial as it aids in refining learners' understanding of lexical items (Long, 1983), akin to how children acquire their first language through the positive and negative feedback they receive from caregivers (Krashen, 1981). Without feedback, students may not know their output is incorrect or how it can be enhanced. Without feedback, the OHE paradigm can descend into habit formation, as learners may lack awareness of the appropriateness of their output, thereby impeding their progress. Receiving timely feedback has historically been a challenge for learners of English, especially where English is used primarily within the confines of a classroom (Zabihi & Erfanitabar, 2021). This challenge is exacerbated in large language classes, where immediate feedback from the teacher is often impractical.

The emerging AI-assisted technologies, such as ChatGPT, Grammarly, and QuillBot, boost autonomous language learning, as these tools can serve as learning companions or co-pilots (Kohnke et al, 2023; Wang et al., 2024; Zhai et al., 2023). In formal EFL contexts, the teacher's role should shift toward creating exercises that enable students to practice language skills and receive instantaneous feedback, rather than relying on delayed teacher responses. The integration of AI presents an opportunity to modify OHE, which typically concludes at the experimental stage. Combining OHE and AI-assisted tools leads to the sequence of activities shown in Figure 2.

Figure 2

OHE Paradigm Enhanced by AI Feedback



In this, language learning commences with comprehensible language input, from which linguistic features can be extracted. These extracted features serve as formulaic patterns that direct the construction of language output. AI-assisted tools aid in reshaping this output to enhance comprehensibility and accuracy, facilitating what Tomasello and Herron (1989) refer to as a 'cognitive comparison' between incorrect or unclear utterances and their correct and clearer counterparts. Subsequently, a new cycle of observation begins, centered around the received feedback. Thus, language teachers can enhance their classroom teaching by incorporating AI-assisted tools into their pedagogy. With AI-generated immediate feedback, the goal of vocabulary learning should focus on a sustainable process rather than the final product. With AI feedback, as shown in Figure 2, the cycle of language learning is completed with support from the input hypothesis (Krashen, 1981), the output hypothesis (Swain, 1985), and the interaction hypothesis (Long, 1983). This study provides a simple example of how AI tools can supplement and enhance well-established teaching practice. Further research should consider how students use the feedback from the AI tools.

About the Author

Montri Tangpijaikul: An Associate Professor in the Department of Foreign Languages, Faculty of Humanities, Kasetsart University, Thailand. He earned a Doctorate in Linguistics from Macquarie University. His research interests include discourse analysis, business English, corpus-based and lexical studies.

References

Alsadoon, R. (2021). Chatting with AI bot: Vocabulary learning assistant for Saudi EFL learners. *English Language Teaching 14*(6), 135-157.

Andrews, K. L. Z. (2007). The Effects of implicit and explicit instruction on simple and complex grammatical structures for adult English language learners. *TESL-EJ, 11*(2), 1-15.

Boonmoh, A. (2021). Use of dictionaries and online tools for reading by Thai EFL learners in a naturalistic setting. *Lexikos, 31*, 239-258. https://doi.org/10.5788/31-1-1645

Carroll, S., Swain, M. & Boberge, Y. (1992). The role of feedback in adult second language acquisition: Error correction and morphological generalizations. *Applied Psycholinguistics, 13*, 173-198.

Chen, Y. (2016). Dictionary use for collocation production and retention: A call-based study. *International Journal of Lexicography, 30*(2), 225–251. https://doi.org/10.1093/ijl/ecw005

Chiu, L. L. & Lui, G. Z. (2013). Effects of printed, pocket electronic, and

online dictionaries on high school students' English vocabulary retention. *Asia-Pacific Education Researcher, 22*(4), 619–634. https://doi.org/10.1007/s40299-013-0065-1

Chomsky, N. (1965). *Aspects of the theory of syntax*. Cambridge, MA: MIT Press.

Coxhead, A. (2000). A new academic word list. *TESOL Quarterly, 34*(2), 213-238. https://doi.org/10.2307/3587951

Du, Z. (2016). A corpus-based study on the structural and functional features of lexical chunks in Chinese academic English. *Technology Enhanced Foreign Language Education, 10*, 9-13.

Dwaik, R. A. A. (2015). English digital dictionaries as valuable blended learning tools for Palestinian college students. *English Language Teaching, 8*(11), 1-10. https://doi.org/10.5539/elt.v8n11p1

Fauzanz, A., Basthomi, Y. & Ivone, F. A. (2022). Effects of using online corpus and online dictionary as data-driven learning on students' grammar mastery. *LEARN Journal, 15*(2): 679-704. https://so04.tci-thaijo.org/index.php/LEARN/article/view/259945/176140

Fu, Q.-K., Zou, D., Xie, H. & Cheng, G. (2022). A review of AWE feedback: Types, learning outcomes, and implications. *Computer Assisted Language Learning, 37*(1-2), 179-221. https://doi.org/10.1080/09588221.2022.2033787

Galbraith, D. & Vedder, I. (2019). Methodological advances in investigating L2 writing processes: Challenges and perspectives. *Studies in Second Language Acquisition, 41*(3), 633-645. https://doi.org/10.1017/S0272263119000366

Gao, X. (2017). A comparable-corpus-based study on native English and Chinese academic writers' use of English lexical bundles. *Foreign Languages and Their Teaching, 3*, 42-50. https://doi.org/10.1515/CJAL-2019-0029

Harmer, J. (2001). *The practice of English language teaching*. Essex: Pearson Education Limited.

Hochman, J. C. & Wexler, N. (2017). One sentence at a time: The need for explicit instruction in teaching students to write well. *American Educators, 41*(2), 30-43.

Hsu, W. (2014). The effects of audiovisual support on EFL learners' productive vocabulary. *ReCALL, 26*(1), 62-79. https://doi.org/10.1017/S0958344013000220

Huang, S. & Renanya, W. A. (2020). Exploring the integration of automated feedback among lower-proficiency EFL learners. *Innovation in Language Learning and Teaching, 14*(1), 15-26. https://doi.org/10.1080/17501229.2018.1471083

Kim, S. Y. & Kim, K. S. (2022). Vocabulary transfer from reading to

writing: A comparison of essay writing and synchronous CMC. *The Electronic Journal for English as a Second Language, 26*(1), 1-21. https://doi.org/10.55593/ej.26101a8

Kohnke, L., Moorhouse, B. L. & Zou, D. (2023). ChatGPT for language teaching and learning. *RELC Journal, 54*(2), 537-550. https://doi.org/10.1177/00336882231162868

Krashen, S. D. (1981). *Second Language Acquisition and Second Language Learning.* Oxford: Pergamon Press Inc.

Lateh, N. H. M., Shamsudin, S., Raof, A. H. A., Mohamed, A. F., Mahmud, N. Nasir, N. S. M., Hanapi, N. F. (2021). Learners' collocation use in writing: Do proficiency levels matter? *Indonesian Journal of Applied Linguistics, 11*(2), 418-426. https://doi.org/10.17509/ijal.v11i2.31632

Lee, I., Luo, N. & Mak. P. (2023). Issues of error selection for focused written corrective feedback in authentic classroom contexts. *RELC Journal, 54*(3), 616-629. https://doi.org/10.1177/00336882211028425

Lee, S. H. (2003). ESL learners' vocabulary use in writing and the effects of explicit vocabulary instruction. *System, 31*(4), 537-561. https://doi.org/10.1016/j.system.2003.02.004

Lee, S. H. & Muncie, J. (2006). From receptive to productive: Improving ESL learners' use of vocabulary in a post-reading composition task. *TESOL Quarterly, 40*(2), 295-320. https://doi.org/10.2307/40264524

Lewis, M. (2002). *The lexical approach: The state of ELT and a way forward.* Boston, MA: Thomson Heinle.

Lightbown, P. & Spada, N. (1990). Focus-on-form and corrective feedback in communicative language teaching: Effects on second language learning. *Studies in Second Language Acquisition, 12*, 429-448.

Long, M. H. (1983). Native speaker/non-native speaker conversation and the negotiation of comprehensible input. *Applied Linguistics, 4*, 126-141. https://doi.org/10.1093/applin/4.2.126

Long, M. H. (1994). On the advocacy of the task-based syllabus. *TESOL Quarterly, 28*(4), 782-790. https://doi.org/10.2307/3587562

Masrai, A. (2022). Lexical knowledge and L2 general language proficiency: collocational competence and vocabulary size as determinants of lexical knowledge. *Cognitive Processing, 24*(2), 289-300. https://doi.org/10.1007/s10339-022-01120-2

Na Phuket, P. R. & Bidin, S. J. (2016). Native language interference in writing: A case study of Thai EFL learners. *International Journal of Foreign Language Teaching & Research, 4*(16), 25-36.

Namvar, F. (2012). The relationship between language proficiency and use

of collocation by Iranian EFL students. *3L: The Southeast Asian Journal of English Language Studies, 18*(3), 41-52.

Nation, I. S. P. (2005). Teaching and learning vocabulary. In: E. Hinkel (Ed.), *Handbook of research in second language teaching and learning*. (pp. 581-595). Lawrence Erlbaum Associates.

Nation, I. S. P (2022). *Learning vocabulary in another language.* Cambridge: Cambridge University Press.

Racine, J. P. (2018). *Lexical approach*. In J. I. Liontas (Ed.), *The TESOL encyclopedia of English language teaching*. (pp. 1-7). Wiley Blackwell. https://doi.org/10.1002/9781118784235.eelt0169

Rafique, S., Waqas, A. & Shahid, C. (2023). The correlation between vocabulary knowledge and English language proficiency at undergraduate level. *Pakistan Journal of Humanities and Social Sciences, 11*(2), 1162-1171. https://doi.org/10.52131/pjhss.2023.1102.0422

Ranalli, J. (2018). Automated written corrective feedback: How well can students make use of it? *Computer Assisted Language Learning, 31*(7), 653-674. https://doi.org/10.1080/09588221.2018.1428994

Ranalli, J. & Yamashita, T. (2022). Automated written corrective feedback: Error-correction performance and timing of delivery. *Language Learning & Technology, 26*(1), 1-25. https://doi.org/10.1080/09588221.2018.1428994

Schmitt, N. & Schmitt, D. (2020). *Vocabulary in language teaching.* Cambridge University Press.

Swain, M. (1985). Communicative competence: Some roles of comprehensible input and comprehensible output in its development. In S. M. Gass and C. G. Madden (Eds), *Input in second language acquisition*. (pp.235-253). Newbury House.

Tarp, S., Fisker, K. & Sepstrup, P. (2017). L2 Writing assistants and context-aware dictionaries: New challenges to lexicography. *Lexikos, 27*, 494-521. https://doi.org/10.5788/27-1-1412

Thornbury, S. (1998). The lexical approach: A journey without maps? *Modern English Teacher, 7*(4), 7-13.

Tomasello, M. & Herron, C. (1989). Feedback for language transfer errors: The garden path technique. *Studies in Second Language Acquisition, 11*, 385-395.

Trinh, T. L. A., Tran, T. K. N., Vo, T. B. N. & Huynh, T. T. S. (2021). The difference effects of paper dictionaries vs. online dictionaries. *AsiaCALL Online Journal, 12*(3), 28-38. https://asiacall.info/acoj/index.php/journal/article/view/34

Uchihara, T. & Saito, K. (2019). Exploring the relationship between productive vocabulary knowledge and second language oral ability.

The Language Learning Journal, 47(1), 64-75. https://doi.org/10.1080/09571736.2016.1191527

Wang, J. (2021). The empirical study of lexical approach in college English classroom teaching and its effects on art major's writing. *English Language Teaching, 14*(12), 227-238. https://doi.org/10.5539/elt.v14n12p227

Wang, Y., Wu, J., Chen, F., Wang, Z., Li, J. & Wang, L. (2024). Empirical assessment of AI-powered tools for vocabulary acquisition in EFL instruction. *IEEE Access, 12*, 131892-131905. https://doi.org/10.1109/ACCESS.2024.3446657

Webb, S. A. (2009). The effects of pre-learning vocabulary on reading comprehension and writing. *Canadian Modern Language Review, 65*(3), 441-470. https://doi.org/10.3138/cmlr.65.3.441

Yang, Y. (2015). The use of lexical bundles in Chinese English majors' and American college students' writings. *Foreign Language World, 3*, 42-27.

Zabihi, R. & Erfanitabar, D. (2021). The revision effects of varying degrees of written corrective feedback explicitness on L2 learners' writings. *RELC Journal, 55*(1), 14-28. https://doi.org/10.1177/00336882211054649

Zhai, N. & Ma, X. (2023). The effectiveness of automated writing evaluation on writing quality: A meta-analysis. *Journal of Educational Computing Research, 61*(4), 875–900. https://doi.org/10.1177/07356331221127300

Zhang, T. (2021). The effect of highly focused versus mid-focused written corrective feedback on EFL learners' explicit and implicit knowledge development. *System, 99*, 1-16. https://doi.org/10.1016/j.system.2021.102493