Cultural and Religious Studies, October 2026, Vol. 14, No. 10, 637-651
doi: 10.17265/2328-2177/2026.10.002

A Corpus-Driven Synthesis of Lexical Co-selection Patterns: The Extended Unit of Meaning Framework for EFL Vocabulary Instruction*

YANG Huihui
Taizhou University, Taizhou, China

Traditional L2 vocabulary instruction has centred on memorising individual lexical items, leaving learners with little awareness of the co-selectional behaviour of words in use. Drawing on co-selection theory and the Extended Unit of Meaning model, this paper synthesises empirical corpus evidence on lexical co-selection into a four-dimensional analytical framework. It then turns to the Four Usages of corpus-driven instruction to design a four-stage teaching sequence: perceiving, analyzing, practicing, and transferring. Each stage is realized through activities such as concordance-line induction, pattern examination, and human-AI (artificial intelligence) collaborative writing. Operating as a closed loop, the teaching sequence moves learners from noticing co-selection patterns to deploying them in their own language use. The model thus carries co-selection theory into classroom practice, giving EFL (English as a Foreign Language) teachers a workable scheme for developing depth of L2 vocabulary knowledge.

Keywords: co-selection, extended unit of meaning, vocabulary teaching, depth of vocabulary knowledge

Introduction

L2 vocabulary competence constitutes the core of L2 ability, and its level of development directly constrains overall language proficiency (Nation, 2001); only when learners reach a certain vocabulary size threshold can effective reading be realised (Laufer, 1992). Vocabulary knowledge is generally divided into two dimensions, breadth and depth (Nation, 2001; Read, 2000). Vocabulary depth concerns the contextual use of words, which is well represented by the four dimensions of the Extended Unit of Meaning: co-selection patterns of lexis such as collocation and colligation (Sinclair, 1996; 2004), extending further to semantic preference and semantic prosody (Bednarek, 2008); sensitivity to semantic prosody has thus become a commonly adopted indicator for measuring learners' depth of vocabulary knowledge (Dushku & Paek, 2021). By contrast, traditional EFL (English as a Foreign Language) vocabulary teaching has long adhered to rote memorisation of meaning and form, giving priority to vocabulary size while leaving depth-oriented training relatively underdeveloped, which causes an imbalance between vocabulary breadth and depth in instructed L2 learning (Nation, 2013). Corpus linguistics supplies an alternative path. Research advocates drawing on authentic concordance lines to guide learners to induce collocational and colligational patterns (O'Keeffe, McCarthy, & Carter, 2007; Meunier & Reppen, 2015). Through such practice, learners can further uncover the semantic preference and semantic prosody manifested

* **Acknowledgements:** This study was supported by the Philosophy and Social Sciences Research Program for Universities in Jiangsu Province (No. 2021SJA2155).
YANG Huihui, M.A., Lecturer, School of Foreign Languages, Taizhou University, Taizhou, China.

within the Extended Units of Meaning built around node words (Sinclair, 2004; Bednarek, 2008). Vocabulary teaching that remains confined to word-level practice, pairing forms with meanings and growing vocabulary size, contributes little to develop depth of vocabulary knowledge, which partly accounts for the persistent lexical difficulties encountered by many EFL learners (Nation, 2001; 2013; Schmitt, 2008). From a pedagogical perspective, the root of this problem lies in the fact that traditional teaching treats words as isolated carriers of meaning and overlooks the usage-based nature of language: Meaning is invariably realised through the co-selection relations among words, grammatical structures, and context.

The model proposed here grows out of earlier empirical work in co-selection theory. The model converts theory into practice at three levels-a teaching philosophy drawn from co-selection theory's view of language, teaching procedures drawn from the analytic framework of the Extended Unit of Meaning, and classroom tasks drawn from the concordance-line analysis that supports them. Its purpose is to make learners see that words acquire their meaning through the contexts they occur in, and that context is where communication actually happens. The model accordingly aims to move learners from memorising single words to using them.

Prior studies within co-selection theory and the Extended Unit of Meaning framework have uncovered numerous observable lexical behaviours. Yet most of them examine individual node words, and the results do not build a systematic framework that teachers can apply directly in class. This study therefore brings these scattered strands of evidence together through corpus-driven integration, building a unified framework for analysis and instruction and narrowing the gap between co-selection theory and pedagogical practice.

Theoretical Foundations

Co-selection Theory

Co-selection theory is a central theoretical claim of Neo-Firthian corpus linguistics, originally advanced by Sinclair (1991) and fully systematized in 1996. The theory holds that meaning is not carried by isolated words; rather, it is jointly constructed and realised in real communication through the co-selection of lexis, grammar, semantics, and pragmatics across multiple levels (Sinclair, 1991; 1996; 2004).

Co-selection theory describes four levels of co-selection relations. (1) Lexis-lexis co-selection: The choice of a specific lexical item constrains the selection of other lexical items causing several words to be chosen together and thus forming collocations. For instance, the high-frequency collocates of *solve* include *crisis*, *conflict*, *difficulty*, *dilemma*, and *problem*, whereas *question* and *issue*, though belonging to the same semantic field, do not frequently co-occur with *solve* (Wei, 2012). (2) Lexis-grammar co-selection: Lexical behaviour is constrained by the syntactic structures in which a word occurs, and those structures are in turn constrained by the words that frequently occur within them; the node word cannot be analysed independently of the syntactic structure within which it occurs. For example, the verb *brook* is never found in affirmative structures but is consistently confined to negative ones (*will not brook*, *don't brook*) (Sinclair, 1996). (3) Lexis-semantics co-selection: The restriction does not stop at individual words, since a node word often attracts a whole semantic field. In the pattern *find it + adjective + to*, the open slot accepts a limited set of adjectives expressing *difficulty* or *evaluation*, including *difficult*, *easy*, *clear*, *impossible*, and *necessary* (Francis, 1993). (4) Lexis-pragmatics co-selection: Once these selections are made, the unit as a whole tends to take on a stable evaluative or attitudinal colouring. *Set in* is such a case, its subjects being typically unwelcome states of affairs—*rot*, *decay*, *despair*, and *malaise* (Sinclair, 1991).

These four co-selection relations are not mutually exclusive; each maps onto one analytical element of the Extended Unit of Meaning: lexis-lexis co-selection onto collocation, lexis-grammar co-selection onto colligation,

lexis-semantics co-selection onto semantic preference, and lexis-pragmatics co-selection onto semantic prosody. Taken together, they trace a cline of increasing abstraction, from surface lexical and grammatical patterning (collocation, colligation), through shared semantic features of the collocates (semantic preference), to the evaluative stance carried by the unit as a whole (semantic prosody). Semantic prosody occupies the outermost position on this cline, and its role in the framework is an organising one: It gives the four analytical elements of the extended unit of meaning their coherence, turning them from a list of descriptive categories into one integrated unit. The following sections explain this role in detail.

Extended Units of Meaning (EUM)

Co-selection theory explains how linguistic items co-select one another, yet an abstract theory can only be applied through concrete, operationalised means of analysis. Sinclair (1996; 2004) turned it into a workable model, the Extended Unit of Meaning (EUM), which challenges the traditional view that the individual word is the basic unit of meaning. As the Extended Unit of Meaning claims, meaning arises not from the isolated word but from an extended unit composed of a node and four co-selection relations: collocation, colligation, semantic preference, and semantic prosody (Sinclair, 1996). The node word is not counted among these relations, because co-selection is a relation between a node word and the items that co-occur with it. The node word is therefore the core word around which the four relations obtain, not one of the relations themselves. The node word is the invariable element of the Extended Unit of Meaning and its form is fixed, while collocation, colligation, and semantic preference vary within the limits that the core word and the semantic prosody jointly impose. Semantic prosody is the attitudinal or evaluative function carried by the whole unit. The five elements thus form a continuum from concrete lexical choices to abstract evaluative function, and jointly they constitute a complete unit of meaning. Xu (2020; 2026) discusses this issue against the broader background of developments within the field. He argues that linguistic forms respond to multiple contextual features simultaneously. These contextual features differ in their strength and tend to operate in combination. Two insights can be drawn from the corpus-based studies reviewed above. First, to produce lexical output comparable to that of native speakers, learners need sensitivity to co-selection constraints operating across multiple dimensions at the same time. Second, while individual case studies may offer rich, detailed observations, they do not by themselves tell teachers how to address specific vocabulary items in classroom settings. Such evidence therefore needs to be reorganised into usable pedagogical materials. Sections 3 and 4 illustrate this process of integration.

The lexical co-selection cases reviewed above furnish abundant material and inspiration for instructional design. The “Four Usages” principles advanced by Xu (2022)—authentic materials for genuine use, priority to high-frequency common words, words taught in combination, and flexible use in context—supply direct guidance for turning teaching philosophy into practice. In particular, teaching materials should draw on authentic corpora (authentic materials for genuine use); teaching content should draw on typical, frequent words by frequency criteria (priority to common words); the key to teaching is to put single words in phrasal settings so as to complete context and resolve ambiguity (words in combination); and teaching activities should be set in authentic vocational scenarios (flexible use in context).

A Corpus-Driven Synthesis of Lexical Co-selection Patterns

The co-selection theory, Extended Unit of Meaning, and the Four-Usage Principles discussed above mainly address two core questions for vocabulary instruction: what should be taught and which criteria should guide

teaching material selection. Nevertheless, clarifying appropriate research methodologies is still necessary if scattered corpus studies of lexical co-selection are to be consolidated into a unified instructional framework. This section therefore elaborates the rationale and operational procedures of instruction design behind corpus-driven synthesis.

This study seeks to systematically integrate some representative corpus-driven empirical research on lexical co-selection to develop a unified instructional framework for EFL vocabulary. The subsequent section first elaborates the rationale for adopting the corpus-driven synthesis approach and then describes in detail the four-dimensional coding criteria applied to standardise the analysis of all documented cases presented in Table 1.

Rationale for Synthesising Existing Corpus Case Studies

Existing vocabulary studies drawing on co-selection theory and the Extended Unit of Meaning (EUM) have achieved notable findings on co-selection patterns of specific words, yet much of this work displays fragmented analytical perspectives. Most prior studies target individual node words (e.g., *meet*, *deny*, or *brook*) and draw on diverse corpora and analytical methods. Therefore, the findings are hard to cross-validate or generalise and they do not yet provide the systematic, comprehensive knowledge base that EFL vocabulary instruction needs.

Existing studies diverge in corpus source, target words, learners'L1 backgrounds, and analytical criteria, so their statistics cannot be merged in a standard quantitative meta-analysis. Rather than attempt such a synthesis, this study re-analyses corpus cases through a unified Extended Unit of Meaning framework to identify co-selection patterns for the purpose of vocabulary teaching.

Four-Dimensional Coding Scheme

For comparable and replicable synthesis of diverse corpus-driven research cases, this study devises unified four-dimensional coding rules drawing upon the Extended Unit of Meaning (EUM). Applied to all cases summarised in Table 1, the coding rules avoid the interference of divergent analytical procedures and methods from original studies. In terms of collocation, high-frequency lexical items co-occurring with each node word are sorted and grouped according to their semantic relations. For colligation, fixed syntactic patterns, grammatical functions within syntactic constructions, and co-occurring functional grammatical markers associated with node words are systematically documented. In the dimension of semantic preference, the collocates sharing the same semantic field are identified and generalised. As for semantic prosody, holistic attitudinal meanings conveyed by phrase-level units (rather than isolated lexical items) are coded into positive, negative, neutral, or complex polarities.

This encoding system based on Extended Unit of Meaning can incorporate various empirical cases, which have diverse sources and inconsistent analytical standards, into a unified analytical framework. It further lays solid methodological groundwork for the present study to generalise lexical co-selection patterns and develop a practical EFL vocabulary-teaching model.

Co-selection Patterns Revealed by Corpus Evidence

The cases selected from different corpora assembled in Table 1 are not arbitrarily compiled. They are selected according to clear comparability criteria to enable valid implementation of subsequent four-dimensional coding and cross-case synthesis. Specifically, the selected cases need to satisfy all three conditions below.

(1) The study draws upon co-selection theory and adopts the Extended Unit of Meaning as its analytical framework, focusing on lexical co-selection phenomena rather than broad discourse-level features.

(2) Grounded in authentic corpus evidence (native-speaker or learner data), it reports at least one target node, a single lexical item or a fixed multi-word unit with sufficient empirical detail across collocation, colligation, semantic preference, and semantic prosody that all four dimensions can be coded from the source text alone.

(3) Foundational studies that established the analytical framework are retained alongside the studies involving Chinese EFL learners or node words that permit L1-L2 comparisons with Chinese, so that the synthesis covers both the theoretical origins of co-selection research and its applicability to local teaching contexts.

Two foundational studies deserve a word of explanation. Sinclair (1996) and Hunston and Francis (2000) were not designed for Chinese learners, yet they are retained in Table 1 because they established the EUM as the analytic paradigm this study follows. The other native-speaker cases in the table supply the empirical exemplars against which the four dimensions can be coded.

After coding each case presented in Table 1 individually using the above-mentioned four-dimensional scheme, the co-selection features of node words have emerged for generalisation. This section first specifies the analytical elements of Extended Unit of Meaning, before sorting out concrete co-selection patterns revealed by corpus-derived evidence.

For teaching purposes, the four co-selection relations of the Extended Unit of Meaning can be treated as four aspects of word knowledge a learner must integrate. Collocation knowledge is knowing which words a *node* word habitually co-occurs with; colligation knowledge, the grammatical patterns it selects; semantic preference knowledge, the semantic field its typical collocates belong to; semantic prosody knowledge, the attitude or evaluation the whole unit conveys. Each kind can be learned from concordance lines, and together they define what it means to know a word deeply.

Table 1 brings together cases drawn from different corpora and languages, and it is obvious to find that no *node* word carries complete meaning by itself. Meaning emerges only when collocation, colligation, semantic preference, and semantic prosody work together as a whole one.

These findings clarify the co-selection constraints that shape lexical acquisition, but they do not yet say how a teacher should act on them. The next section converts the same dimensions into an instructional sequence, the "Four-Stage" Integrated model.

The evidence from corpora reaches the classroom as well. If lexical instruction holds meanings decontextualised and never shows how form and meaning co-select, it equips learners with lexical knowledge they cannot use in real communication. As the evaluative dimension within a lexical co-selection profile, semantic prosody constitutes a frequent source of interlanguage errors for EFL learners. Table 1 presents *DENY* as a representative example: This *node* word carries a prevailing negative semantic prosody among native English speakers. By contrast, its Chinese counterpart "否认" tends to present positive prosody. Chinese EFL learners tend to transfer such L1-driven prosodic properties into their English production (Lu, 2012), which highlights the value of cross-linguistic comparison for addressing deviant co-selection patterns.

Distinct senses of a word also display notable divergence in co-selection features, including changes in its prosodic polarity. This point finds support in Sun and He's (2026) analysis of three different senses of *meet*: the sense of *meet* denoting interpersonal interactions or physical contact conveys, in the coding adopted here, a neutral prosody; the compliance sense carries a positive prosody that marks the realisation of goals; and the sense describing confrontation with difficulties or deadlines shows a consistently negative prosody, regularly accompanied by negative markers such as fail to and not. Semantic prosody decides where one meaning-shift unit ends and the next begins. Meaning changes when the co-selection pattern adjusts as a whole, not when a

single word is replaced. Collocation and colligation may vary without making a new meaning-shift unit, so long as the prosody holds. The evaluative or attitudinal meaning of the lexical unit is carried by the prosody. Only when adjustments to co-selection patterns bring about shifts in core semantic prosody will a new meaning-shift unit emerge (Sun & He, 2026). Accordingly, vocabulary instruction ought to go beyond superficial instruction that merely introduces individual lexical items and their isolated collocations. Instead, instructors need to elaborate the full set of co-selection features for each individual lexical item. Colligation further shapes a word's evaluative or attitudinal meaning at the syntactic level, as illustrated by Boholm's (2012) investigation of *risk* and *danger*. Both words denote potential harm, yet their syntax differs. *Risk* typically keeps the affected person as the agent and is related to deliberate, goal-directed decisions, whereas *danger* (or *endanger*) separates the affected party from the agent. Learners who master such colligational contrasts can use near-synonyms with the right attitude and the right syntax at the same time.

The cases in Table 1 point to one consistent finding across the *node* words: Co-selection is highly language-specific, and semantic prosody dominates the co-selection profiles and is most easily distorted by L1 transfer. These observations suggest that EFL lexical instruction cannot be limited to rote memorisation of word senses and collocations; semantic prosody and its cross-linguistic differences ought to occupy a central position in teaching practice.

Table 1

Overview of Representative Corpus-Based Studies on Collocation, Colligation, Semantic Preference, and Semantic Prosody

Researcher(s) (Year)	Node word	Collocation	Colligation	Semantic preference	Semantic prosody	Pedagogical implications
Sinclair (1991, 1996)	brook	interference, disagreement	Negative structures: will not brook; don't brook	Nouns of intervention, objection	Negative prosody (refusal, uncompromising stance)	Targeted phrase-pattern recognition practice
Sinclair (1996, 2010)	sever	relations, connections, ties, links	sever + [all / possessive adj.] + N + with/between/from	Nouns of interpersonal and other connections	Negative prosody (decisive termination of ties)	Classroom tasks contrasting semantic prosody
Sinclair (2004)	naked eye	visible, invisible	Prep. + (the / possessive) + naked eye; N + BE + Adv. + Adj. + to/with the naked eye	Concepts of visibility	Negative prosody (difficulty of observation)	Concordance-based inductive perception training
Louw (1993)	utterly	failed, unattractive, absent, afraid, and other negative adjectives/verbs	Modifies adjectives and notional verbs	Negative states, undesirable traits	Negative prosody (wholly negative evaluation)	Semantic-prosody interpretation activities with literary texts
Francis (1993)	find / make	difficult, hard, easy (with find); clear (with make)	Verb + it + Adj. + to/that-clause	Concepts of difficulty, clarity, evaluation	Neutral prosody (objective, evaluative assessment)	Pattern-grammar sentence classification drills
Hunston & Francis (2000)	enlist, resign, volunteer, stand down, rank	soldier, chairman, director, candidate, member (occupation/role nouns)	V + as + n (noun slot = occupation/role)	Occupation, role, identity, function	Neutral prosody (role/identity definition)	Grouping tasks for same-pattern verbs
Boholm (2012)	risk / danger	decide, decision, choose, choice (stronger with risk); endanger, endangering	RISK as N/V; DANGER (N); ENDANGER (V); to risk NP/-ing; run/take the risk of -ing; agent–victim structures	risk: decision, agency, goal-directed action; danger: external Source, potential loss from outside the affected	Negative prosody (potential adverse outcome — shared by risk and danger); distinction lies in agency: risk = internal decision, danger = external cause	Awareness-raising on the agency-based risk/danger distinction for accurate risk communication

Cunningham & Fillmore (1995)	use a firearm	shoot, injure, kill	use + (a) + firearm/gun	Action nouns of bodily harm	Negative prosody (violence, bodily harm)	Real-scenario communicative transfer practice
Lu (2012)	DENY (L1 English)	allegation, right(s), access, charge	DENY + NP/S; DENY + Sb. + Sth.	Negative nouns of accusation, rights	Negative prosody (opposition, refutation)	Distinguish L1-standard collocations; caution against literal-translation errors
Lu (2012)	deny (否认; L1 Chinese)	facts, laws, role, existence	Negative marker + deny + NP/S (double negation frequent)	Objective facts or laws (neutral/positive)	Positive prosody (euphemistic agreement, approval)	Chinese–English pattern contrastive materials
Lu (2012)	DENY (Chinese learners)	fact, importance, advantage(s), convenience	NEG + DENY + NP/S (frequent double negation)	Neutral, positive nouns	Positive prosody (approval, transferred from L1 Chinese)	Targeted collocation "error-spotting" correction exercises
Wei (2012)	manifest/display (展现)	scenes, prospects, talent, images	Spatiotemporal adverbial + manifest + NP	Positive nouns of scenery, traits	Positive prosody (appreciation, approval)	Chinese–English phrase contrastive writing practice
Louw (1993)	idea	faintest, slightest (degree adjectives)	(do) not have + the + degree adjective + idea	Concepts of cognitive blankness	Negative prosody (complete lack of knowledge)	Concordance-based semantic-prosody interpretation of fixed phrases
Sun & He (2026)	meet (Sense 1)	people, eyes, stones	N (person) + meet + (with) + N; N (thing) + meet	Interpersonal objects, physical entities	Neutral prosody (equal interaction)	Multi-dimensional induction of polysemous phrases
Sun & He (2026)	meet (Sense 2)	needs, standards, demands	N + meet + (adj) + need/standard	Target, criterion nouns	Positive prosody (goal attainment)	Distinguish co-selection differences across senses
Sun & He (2026)	meet (Sense 3)	challenges, deadlines	(Neg.) + meet + challenge	Pressure, problem nouns	Negative prosody (difficulty, non-	Negative-pattern semantic-prosody

Notes. L1 = first language; NP = noun phrase; S = sentence/clause; N = noun; V = verb; Prep. = preposition; Adv. = adverb; NEG = negative marker. Chinese node words are given in English translation with the original Chinese in parentheses. Columns follow the extended unit of meaning model (Sinclair, 1996; 2004; Hunston & Francis, 2000).

An Integrated "Four-Stage" Corpus-Driven Model for EFL Lexical Instruction

The cases of lexical co-selection compiled in this study furnish instructional design with both substantial materials and pedagogical insight (see Table 1). Drawing on co-selection theory, the present study proposes an integrated "Four-Stage" vocabulary teaching model within the Extended Unit of Meaning framework. The four stages are progressive phases of instruction: perception, analysis, practice, and transfer. The "integrated" character of the model lies elsewhere. With the node word as its analytical item, the four dimensions of co-selection knowledge are presented as a unified whole in every stage rather than taught separately. Pedagogically, the unit of meaning defines the goal of instruction. Corpus data supply the raw material, concordance analysis supplies the method, and context-grounded tasks carry the instruction into practice.

Hierarchical Learning Objectives of the Four-Stage Model

The depth of lexical knowledge finds its concrete realisation within Extended Units of Meaning, which integrates collocation, colligation, semantic preference, and semantic prosody into a unified whole. Moving from collocation through colligation and semantic preference to semantic prosody makes the analysis progressively more abstract. The learner shifts from visible lexical behaviour to hidden evaluative meaning. L2 lexical development follows the same order: Learners first notice overt recurring patterns, then unpack covert attitudinal meaning, and finally learn to examine corpus evidence on their own and produce appropriate output. Drawing on these developmental patterns as well as central arguments from co-selection theory, the present instructional

framework sets three hierarchical levels of learning outcomes that correspond to the natural progression of EFL learners' lexical growth: (1) Identification: Basic-level learning enables learners to spot node words' high-frequency collocations and typical colligational patterns, and acquire one or two core extended-unit structures. (2) Interpretation-application: At the developmental level, learners should generalise semantic preference, recognise inherent semantic prosody, and apply two or three meaning units accurately in context. (3) Autonomous exploration: At the extension level, learners are encouraged to retrieve concordance lines from corpora independently, develop full co-selection profiles of *node* words, and flexibly draw on different units of meaning for communicative output.

The three levels move from receptive identification to autonomous inquiry, in line with the growth of L2 lexical knowledge (Nation, 2001; 2013). Semantic prosody deserves special attention here because it develops slowly and does not always keep pace with general proficiency (Dushku & Paek, 2021).

Teaching Content Grounded in the Four Co-selection Dimensions of EUM

With the objectives fixed, the model then offers instructional input that matches them. Conventional vocabulary instruction often treats lexical items in decontextualised isolation, separating collocational, grammatical, and evaluative semantic relations from one another, which prevents learners from developing coherent awareness of co-selection. Accordingly, instead of arranging learning materials around discrete word entries, the present framework organises teaching content centred on Extended Units of Meaning, integrating four co-selection dimensions: collocation, colligation, semantic preference, and semantic prosody. Detailed knowledge components are laid out below:

(1) Collocational knowledge concerns the lexical partners that frequently co-occur with the *node* word, covering their co-occurrence frequency and association strength. For instance, Sinclair's *naked eye* case in Table 1 demonstrates the prepositions to and with compete for the slot immediately before the *naked eye*, each anchoring a strong collocational pattern. When teaching the verb *meet*, instructors should highlight collocational differences across its distinct Extended Units of Meaning: *meet* people for interpersonal encounters, *meet* needs/requirements for satisfying criteria, and *meet* challenges for tackling difficulties.

(2) Colligational knowledge covers the typical syntactic environments in which the *node* word occurs. As shown in Table 1, *brook* is largely restricted to negative-oriented patterns such as will not *brook* and do not *brook*; such syntactic constraints constitute its defining colligational property within the Extended Unit of Meaning. When teaching *deny*, it is necessary to distinguish the preferred colligational frames (*DENY*+ NP/S, *DENY* + Sb. + Sth.) by native speakers from deviant patterns produced by L1 transfer, namely NEG + *DENY* + NP/S among EFL learners.

(3) Semantic preference knowledge concerns the shared semantic features of the *node* word's collocates. For example, collocates of *sever*, such as relations, connections, ties and links, exhibit a tendency toward the semantic domain of interpersonal relations. Similarly, *cause* tends to attract collocates associated with negative outcomes (e.g. problems, damage, trouble), which serve as representative material for teaching semantic preference.

(4) Knowledge of semantic prosody concerns the evaluative or attitudinal meaning conveyed by a complete Extended Unit of Meaning. Nearly all cases summarised in Table 1 contain descriptions of semantic prosody. In particular, Lu (2012) conducted English-Chinese contrastive analysis to illustrate how mismatched positive-negative prosody gives rise to learner errors brought about by L1 transfer. In classroom practice, prosody-

matching tasks can be devised to allow learners to assign concordance lines containing the *node* word to their corresponding semantic prosody categories.

Implementing the "Four-Stage" Integrated Model

The preceding section elaborates key components of co-selection related teaching contents including collocation, colligation, semantic preference, and semantic prosody, and identifies multi-level contextual constraints for L2 lexical learning. Turning abstract co-selection principles into workable classroom routines, this section constructs a full instructional framework consistent with the hierarchical learning objectives established previously.

Before elaborating on each stage, this section identifies the intended users of the model and the prior knowledge learners are expected to possess. It targets upper-intermediate and advanced EFL learners. Dushku and Paek (2021) reported that semantic-prosody awareness is slow to develop and does not always rise in step with learners' general proficiency; because such awareness is central to the present model, the model is best suited to upper-intermediate and advanced learners. Less advanced learners can start with Stage 1 concordance observation to build awareness of high-frequency collocations. They may then move on to subsequent stages as their co-selection awareness develops, following a progressive scaffolding sequence from simple to complex.

The four stages of the EUM-based instructional model. One point needs to be made clear at the outset. The three-level objectives (identification, interpretive-application, and autonomous-exploration) do not map one-to-one onto the four instructional stages. Instead, each instructional stage integrates multiple objective dimensions, and learners progress across these hierarchical levels gradually as they move through the model. The two orders describe different things: The objectives are levels of learner competence, while the stages are a sequence of teaching procedures. Specifically, Stage 1 (perception) and Stage 2 (analysis) mainly lead to the "identification" and "interpretive-application" objective levels, helping learners build co-selection awareness and grasp the deeper evaluative meaning of units of meaning through observation and articulation. Stage 3 (practice) consolidates the cognitive gains of the first two stages into a retrievable productive ability. Stage 4 (transfer) returns the initiative to the learner, who independently draws on units of meaning in real communication and finally reaches the "autonomous-exploration" level. Objectives and stages thus form an outcome-procedure loop rather than a rigid sequence.

Stage one: Perception-observing concordance lines to build awareness of co-selection:

Learning objective: To expose learners to authentic corpus evidence and enable them to gain direct intuitions about co-selection behaviours of target lexical items.

Core activities:

Concordance line observation: The instructor extracts 20-30 concordance lines for *node* words from corpus resources such as COCA and BNC and distributes these excerpts to learners. Students are guided to examine each line sequentially, marking recurring lexical items appearing on the left-hand and right-hand positions of the *node* words. For example, Sinclair's *naked eye* from Table 1 serves as an effective illustration: Competing prepositions to and with in the immediate left slot of the pattern the *naked eye* present perceptible features of co-selection and are well-suited for learners.

"Word-partner spotting" activity: Learners circle words of high co-occurrence with the *node* words within concordance lines to build an initial awareness of which lexical items frequently co-occur with a given *node* word. This task helps them form preliminary impressions regarding which lexical items tend to appear alongside the *node* words.

Rationale for design: Examining concordance lines constitutes a central practice within Data-Driven Learning (DDL) (Johns, 1991; Hunston, 2002). When learners *meet* enough authentic instances, they can detect regular lexical behaviour on their own and begin to perceive co-selection. Prior studies highlight the merits of corpus-informed instructional materials for teaching collocations and colligations. Such materials offer examples with full contextual information and clear register features drawn from real-language data, in contrast to conventional teaching resources (Meunier & Reppen, 2015; O'Keeffe et al., 2007).

Stage two: Analysis—Generalising co-selection patterns into EUM knowledge:

Once learners have developed co-selection awareness, intuitive observations need to be elevated into an articulable knowledge framework, which marks the transition to the second instructional stage.

Learning objective: Under the instructor's guidance, learners systematise observational evidence into explicit statements of co-selection knowledge.

Core activities:

Teachers and learners pick out high-frequency collocates of *node* words extracted from concordance lines. They then use frequency ranking or collocational strength metrics such as MI scores to identify core collocations.

Pattern identification: Learners generalise the typical colligational patterns of *node* words and distinguish native-speaker frequent syntactic patterns from syntactic structures commonly misused by L2 learners. As shown in Table 1, the word *brook* has distinct colligational characteristics. It appears almost exclusively in negative syntactic environments and helps illustrate the co-selection principle that lexical items cannot be separated from their syntactic contexts.

Extended-unit construction: Working with concrete concordance lines, the instructor guides learners to generalise a node word's semantic preference and identify its semantic prosody. For example, learners compare different uses of *meet* drawn from concordance lines: when a teacher meets parents…, products *meet* standards, and …fail to *meet* challenges. This comparison helps learners understand how the same word diverges into distinct Extended Units of Meaning across different co-selection patterns. This aligns with the core findings reported by Sun and He (2026), who identify three meaning-shift units for *meet* in Table 1.

Contrastive awareness raising: When clear divergences appear between English and Chinese co-selection patterns, as illustrated by the conflicting semantic prosody features of *deny* and its Chinese equivalent "否认" (Lu, 2012, see Table 1), specific contrastive activities need to be included. These activities help learners develop sensitivity to negative L1 transfer.

Rationale for design: The Extended Unit of Meaning offers a procedure for discovering and generalising co-selection patterns systematically (Sinclair, 1996). With teacher support, learners start from concrete concordance lines and draw up abstract characterisations of collocations, colligations, semantic preference, and semantic prosody. The design is deliberately inductive: Learners move from observed lines to general statements about a word's co-selection behaviour.

Stage three: Practice-consolidating co-selection knowledge through task-based activities:

Learning objective: Controlled and semi-controlled tasks, which respectively emphasise accuracy and limited learner output, are adopted to facilitate learners' internalisation of previously acquired Extended Units of Meaning.

Core activities:

Modified gap-fill exercises: Instead of deleting isolated single words, these tasks remove key components within whole co-selection patterns built around Extended Units of Meaning. For example:

The company _____ _____ the new safety standards set by the government. (expected output: met + requirements)

Such items oblige learners to take both *node*-word choices and collocational appropriateness into simultaneous consideration.

Error-detection tasks: Learners are presented with sentences where prosodic clashes or collocational errors occur and asked to diagnose and revise these sentences. Consider the sentence "The research caused great benefits to the community". Learners are guided to spot a semantic prosody clash in the given instance: The negative semantic prosody from uses of *cause* conflicts with the positive attitudinal meaning carried by benefits, and propose alternatives such as brought about or led to. Drawing on L1 transfer misuse of *deny* summarised in Lu (2012) from Table 1, comparable error-identification drills can also be constructed.

Matching-reorganisation tasks: Components of Extended Units of Meaning are separated and rearranged. Learners reassemble these fragmented lexical items and justify their decisions. For example, learners match three meaning units of *meet* (*meet* friends, *meet* standards, *meet* challenges) to suitable contexts, while identifying their respective semantic prosody: interpersonal encounter, goal achievement, and challenging handling.

Rationale for design: Error-detection activities draw insights from Xu's (2026) notion of AI (artificial intelligence)-generated content identification tasks. Learners can identify non-native usages only after they have built knowledge of co-selection. Matching-reorganisation exercises reinforce the cognitive principle that patterns are inextricably linked to meaning (Hunston & Francis, 2000).

Stage four: Transfer-creative application within authentic contexts:

Learning objective: This stage shifts learners from acquiring knowledge about co-selection to deploying such knowledge. Learners are expected to flexibly apply Extended Units of Meaning in authentic communicative tasks.

Core activities:

Corpus-exploration project: Each learner selects a self-chosen target word of personal interest, for instance, key words from academic studies or their speciality. Using online corpora such as COCA and BNC, learners retrieve relevant concordance lines and independently produce a short analytical report on the Extended Unit of Meaning of the chosen *node* word. The output should summarise collocation, colligation, semantic preference, and semantic prosody, following analytical conventions illustrated by representative cases in Table 1.

Discourse-level gap-filling: Learners are given a short passage (e.g., an abstract or a review) containing multiple slots reserved for complete Extended Units of Meaning. Drawing on contextual information, they select the most suitable extended unit of meaning from provided options to fill each slot.

Human-AI collaborative writing: Learners produce a draft with AI tools. Rather than passively accepting AI-generated output, they act as final-decision makers (Xu, 2026), checking the draft against corpus evidence to see whether its collocational patterns and semantic-prosody effects match those of the target *node* words, and revising any non-idiomatic expressions. For example, as a prosody exercise, learners may ask two things: Does *cause* here presents a positive evaluation? Does *meet*, if placed in a negative context, get taken "asachieving a goal" when it should be taken as "handling a challenge"?

Law or literature-themed situational tasks (extension activity): For advanced-level learners, corpus cases from Table 1, namely the legal semantic analysis of using a firearm and literary interpretation of utterly, can be introduced. Learners simulate drafting legal documents or literary critiques to experience practical applications of co-selection knowledge within specialised domains.

Rationale for design: The design of the transfer phase embodies the principle of context-embedded practical application (Xu, 2022). Corpus-exploration projects build learners' capacity for independent inquiry. Human-AI collaborative writing, for its part, puts the learner rather than the tool in charge of judging and revising output (Xu, 2026), a role that matters in the large-language-model era.

To visualise the full implementation sequence of the proposed instructional framework, the complete teaching workflow is presented in Figure 1 below.

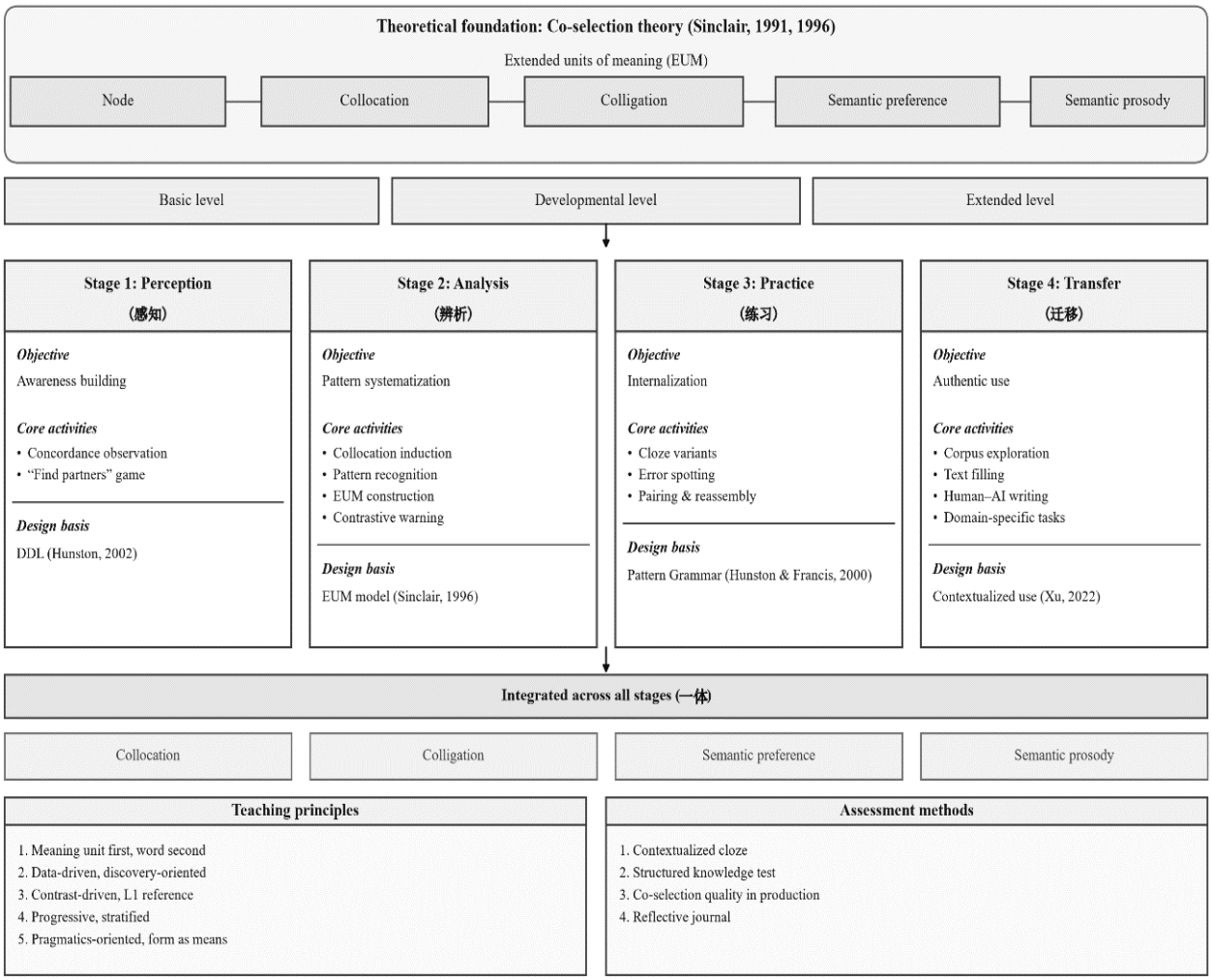


Figure 1. The "Four-Stage Integrated" vocabulary teaching model based on co-selection theory.

Drawing on Sinclair's co-selection theory and the Extended Units of Meaning model (Sinclair, 1991; 1996), the present study constructs an integrated four-stage vocabulary-teaching model informed by four co-selection dimensions: collocation, colligation, semantic preference, and semantic prosody. Encompassing the full instructional sequence of perception, analysis, practice, and transfer, this model constitutes a closed-loop instructional design system for EFL lexical development.

Figure 1 illustrates both the static structure and procedural flow of the proposed model. Nevertheless, its practical feasibility and pedagogical effectiveness remain to be tested within classroom settings. The next section therefore considers the conditions under which the model can be implemented and the ways it can be tested in class.

Application and validation of the instructional model. This section turns to classroom application. It specifies where and under what conditions the model can be used, and how its teaching effect might be verified in follow-up studies.

Validation of the framework can be carried out along three distinct dimensions:

First, pedagogical experimental validation. Two parallel classes are selected for comparison. The experimental class adopts the "Four-Stage" integrated teaching model, while the control class receives conventional vocabulary instruction dominated by definition explanation and example illustration. Through pre-test and post-test comparisons, this study examines the effectiveness of the proposed model in improving learners' productive vocabulary performance in terms of accuracy, collocational naturalness, and contextual appropriateness.

Second, longitudinal tracking based on learner corpus data. Students' writing samples collected before and after the intervention are used to construct a longitudinal learner corpus. This corpus is analysed to trace the developmental changes of learners' co-selection knowledge, with a particular focus on changes in semantic prosody accuracy; the coding of co-selection patterns follows the English-Chinese contrastive procedure used by Lu (2012).

Third, learner reflection and self-assessment. Interviews and questionnaires are adopted to investigate learners' acceptance of the proposed model and their cognitive shifts in vocabulary learning strategies, namely, whether they have transformed from rigid word memorisation to usage-based lexical learning.

Operationalising these three approaches calls for suitable measures. In the pedagogical experiment, this study adopts Read's (2000) framework for evaluating vocabulary depth. It employs a set of tasks including collocation recognition, extended-unit-of-meaning gap-filling, and semantic prosody judgement to quantify improvements in learners' productive accuracy and contextual appropriateness in the experimental group. For the longitudinal corpus analysis, learner texts are annotated for co-selection patterns following Lu's (2012) English-Chinese contrastive coding, and the rate of semantic prosody errors serves as the core indicator of development. For reflective assessment, a vocabulary learning strategy questionnaire is administered to examine whether learners achieve fundamental cognitive transformation from rigid word memorisation to usage-based lexical learning.

Conclusion

Co-selection theory, this study has argued, offers a usable account of what to learn, how to learn, and how to use lexical items. Its translation into the classroom is the "Four-Stage" integrated model proposed here. The aim of vocabulary teaching is not to store word forms and meanings in memory but to let learners control Extended Units of Meaning and the contexts they fit into. Firth's (1957) remark that "You shall know a word by the company it keeps" captures the point exactly: Knowing a word means knowing its collocational world. The task of the model is to build that kind of knowledge systematically.

Two limitations of this study need to be acknowledged. First, the framework is built on co-selection findings already reported in the literature; no learner-corpus evidence was collected for the present intervention, so the model's generalisability to local EFL classrooms has yet to be demonstrated. Second, its effectiveness is at present argued from design rather than from data. The next steps are to test the "Four-Stage" model in quasi-experimental teaching designs, and to examine whether it transfers to English for Specific Purposes contexts such as academic and business English.

References

Bednarek, M. (2008). Semantic preference and semantic prosody re-examined. *Corpus Linguistics and Linguistic Theory, 4*(2), 119-139. Retrieved from https://doi.org/10.1515/cllt.2008.006

Boholm, M. (2012). The semantic distinction between "risk" and "danger": A linguistic analysis. *Risk Analysis, 32*(2), 281-293. Retrieved from https://doi.org/10.1111/j.1539-6924.2011.01668.x

Cunningham, C. D., & Fillmore, C. J. (1995). Using common sense: A linguistic perspective on judicial interpretations of "use a firearm". *Washington University Law Quarterly, 73*(3), 1159-1214.

Dushku, S., & Paek, Y. (2021). Investigating ESL learners' awareness of semantic prosody across proficiency levels. *Language Awareness, 30*(3), 234-256. Retrieved from https://doi.org/10.1080/09658416.2020.1871360

Firth, J. R. (1957). *Papers in linguistics 1934-1951*. Oxford: Oxford University Press.

Francis, G. (1993). A corpus-driven approach to grammar: Principles, methods and examples. In M. Baker, G. Francis, and E. Tognini-Bonelli (Eds.), *Text and technology: In honour of John Sinclair* (pp. 137-156). Amsterdam and Philadelphia: John Benjamins. Retrieved from https://doi.org/10.1075/z.64

Hunston, S. (2002). *Corpora in applied linguistics* (2012 Cambridge paperback ed.). Cambridge: Cambridge University Press. Retrieved from https://doi.org/10.1017/CBO9781139524773

Hunston, S., & Francis, G. (2000). *Pattern grammar: A corpus-driven approach to the lexical grammar of English*. Amsterdam and Philadelphia: John Benjamins. Retrieved from https://doi.org/10.1075/scl.4

Johns, T. F. (1991). Should you be persuaded: Two samples of data-driven learning materials. In T. Johns and P. King (Eds.), *Classroom concordancing* (ELR Journal 4, pp. 1-16). Birmingham: University of Birmingham.

Laufer, B. (1992). How much lexis is necessary for reading comprehension? In P. J. L. Arnaud and H. Béjoint (Eds.), *Vocabulary and applied linguistics* (pp. 126-132). New York: Macmillan. Retrieved from https://doi.org/10.1007/978-1-349-12396-4 12

Louw, B. (1993). Irony in the text or insincerity in the writer? The diagnostic potential of semantic prosodies. In M. Baker, G. Francis, and E. Tognini-Bonelli (Eds.), *Text and technology: In honour of John Sinclair* (pp. 157-176). Amsterdam and Philadelphia: John Benjamins. Retrieved from https://doi.org/10.1075/z.64.11lou

Lu, J. (2012). A corpus-based contrastive study of the patterning features of Chinese EFL learner English. *Modern Foreign Languages, 35*(1), 70-78.

Meunier, F., & Reppen, R. (2015). Corpus versus non-corpus-informed pedagogical materials: Grammar as the focus. In D. Biber (Ed.), *The Cambridge handbook of English corpus linguistics* (pp. 498-514). Cambridge: Cambridge University Press. Retrieved from https://doi.org/10.1007/9781139764377.028

Nation, I. S. P. (2001). *Learning vocabulary in another language*. Cambridge: Cambridge University Press. Retrieved from https://doi.org/10.1017/CBO9781139524759

Nation, I. S. P. (2013). *Learning vocabulary in another language* (2nd ed.). Cambridge: Cambridge University Press. Retrieved from https://doi.org/10.1017/CBO9781139858656

O'Keeffe, A., McCarthy, M., & Carter, R. (2007). *From corpus to classroom: Language use and language teaching*. Cambridge: Cambridge University Press. Retrieved from https://doi.org/10.1017/CBO9780511497650

Read, J. (2000). *Assessing vocabulary*. Cambridge: Cambridge University Press. Retrieved from https://doi.org/10.1017/CBO9780511732942

Schmitt, N. (2008). Instructed second language vocabulary learning. *Language Teaching Research, 12*(3), 329-363. Retrieved from https://doi.org/10.1177/1362168808089921

Sinclair, J. (1991). *Corpus, concordance, collocation*. Oxford: Oxford University Press.

Sinclair, J. (1996). The search for units of meaning. *Textus, 9*(1), 75-106.

Sinclair, J. (2004). *Trust the text: Language, corpus and discourse*. London: Routledge. Retrieved from https://doi.org/10.4324/9780203594070

Sinclair, J. (2010). Defining the definiendum. In G. Schryver (Ed.), *A way with words: Recent advances in lexical theory and analysis* (pp. 37-48). Uganda: Menha Publishers.

Sun, H., & He, X. H. (2026). A corpus-based analysis of the meaning-shift unit of reciprocal verbs. *Foreign Languages and Translation, 35*(2), 51-59. doi:10.19502/j.cnki.2095-9648.2026.02.007

Wei, N. (2012). Theory of co-selection and the corpus-driven phraseology. *Journal of PLA University of Foreign Languages, 35*(1), 1-6.

Xu, J. J. (2020). Multifactorial contextual co-selection: The recent development of corpus linguistics. *Foreign Languages and Their Teaching, 42*(3), 1-10. doi:10.13458/j.cnki.flatt.004677

Xu, J. J. (2022). Lexical syllabus design of higher vocational English and its instructional implications informed by the Principles of Four Usages. *Foreign Language Education in China, 5*(1), 43-49.

Xu, J. J. (2026). A human-AI collaborative foreign language teaching method driven by humanistic and AI competency synergy. *Foreign Language World, 47*(2), 2-8. doi:10.26971/j.cnki.flw.1004-5112.2026.02.013