



Protect, Preserve, and Conserve in Environmental English: A Corpus-Based Analysis of Collocation and Semantic Preference

Mai Hoang Quyen

Faculty of Foreign Languages, Dong Nai Technology University, Vietnam

ABSTRACT: Environmental discourse frequently employs the near-synonymous verbs protect, preserve, and conserve to describe forms of ecological stewardship. Although the three verbs share a broad semantic domain, their actual distribution may reveal differences in collocational behavior and semantic preference. This study investigates the collocational profiles and semantic preferences of protect, preserve, and conserve in contemporary environmental English using a specialized 10-million-word corpus, the Contemporary Environmental English Corpus (CEEC-10M), comprising policy and treaty documents, academic environmental writing, and environmental journalism published between 2016 and 2026. Using AntConc 4.2, the study examines the target verbs and their recurrent collocates within a symmetrical L5-R5 window. Candidate collocates are filtered using a minimum raw-frequency threshold of 15 and an MI threshold of 4.5, followed by concordance inspection and semantic categorization. The results indicate three partially overlapping but distinguishable lexical profiles. Protect shows a broad preference for vulnerable people, species, habitats, ecosystems, and other entities conceptualized as exposed to harm or external pressure. Preserve is more strongly associated with biodiversity, cultural heritage, integrity, historical sites, and existing ecological or cultural states, suggesting a preference for continuity and maintenance. Conserve is particularly associated with water resources, energy, natural resources, soil, and ecosystems, reflecting a stronger orientation toward resource management and sustainable use. The findings demonstrate that near-synonymous verbs cannot be adequately distinguished through dictionary definitions alone. Corpus evidence provides a more precise account of their preferred lexical and semantic environments and offers implications for Environmental English, English for Specific Purposes, and advanced EFL vocabulary instruction.

KEYWORDS: corpus linguistics, environmental English, collocation, semantic preference, near-synonymy

1. INTRODUCTION

Lexical choice plays an important role in the construction and interpretation of environmental discourse. Environmental texts frequently describe actions intended to prevent ecological damage, maintain valuable environmental conditions, or manage natural resources sustainably. In such contexts, the English verbs protect, preserve, and conserve occur regularly and appear semantically related. Because of this semantic overlap, the three verbs may be treated as near-synonyms by language learners and even by proficient users. However, semantic similarity does not necessarily entail identical patterns of use. Corpus-based approaches to lexical meaning suggest that words develop characteristic associations with particular lexical and semantic environments, and these recurrent patterns can provide evidence for distinctions that may not be apparent from dictionary definitions alone (Sinclair, 1991, 2004; Stubbs, 2001).

The study of collocation is particularly relevant to near-synonymy. From a corpus-linguistic perspective, lexical items are not normally used in isolation but occur in recurring combinations with other words. These recurrent combinations can reveal the semantic and functional environments in which a lexical item is conventionally used. Firth's (1957) influential view of meaning emphasized the importance of examining a word through its relationships with surrounding words. Subsequent corpus-based research has developed this perspective by examining collocational patterns, semantic preference, and semantic prosody as complementary dimensions of lexical behavior (Hunston, 2007; McEnery & Hardie, 2012; Partington et al., 2013).

The distinction among protect, preserve, and conserve is particularly relevant in environmental English because the three verbs can describe broadly related forms of environmental stewardship while foregrounding different aspects of that stewardship. Protect may be associated with the prevention of harm or exposure to threats; preserve may emphasize the maintenance of an existing condition, ecological integrity, or cultural and historical value; and conserve may be associated more strongly with the careful management and sustainable use of natural resources. These distinctions, however, should not be assumed solely on the basis of dictionary

definitions. They require examination of naturally occurring language in an appropriate discourse domain.

Environmental and climate-change discourse provides a suitable context for investigating these lexical distinctions. Contemporary environmental communication encompasses scientific assessment, international policy, environmental governance, biodiversity protection, resource management, and public communication. Such discourse therefore provides a broad range of contexts in which the three target verbs may occur. A corpus-based investigation of these contexts can show not only how frequently the verbs occur but also which lexical items and semantic categories they preferentially associate with.

Although corpus linguistics has provided extensive evidence that collocational behavior contributes to lexical meaning, the specific relationship among protect, preserve, and conserve in contemporary environmental English warrants further systematic investigation. In particular, a comparison of their collocational profiles can clarify whether their semantic overlap is accompanied by distinct semantic preferences. Such an analysis is also relevant to English for Specific Purposes (ESP), where learners need to develop not only knowledge of lexical definitions but also an understanding of contextually appropriate lexical combinations.

Accordingly, the present study investigates the collocational and semantic behavior of protect, preserve, and conserve in a specialized corpus of contemporary environmental English. The study combines corpus frequency analysis, collocational analysis, and semantic categorization to identify recurrent patterns associated with the three verbs. Mutual Information (MI) is employed to examine the strength and specificity of lexical associations, while concordance analysis is used to interpret the semantic environments in which the target verbs occur. Log-Likelihood (LL) is retained as a complementary measure in the methodological design, although the supplied results table reports MI values only; therefore, LL should not be described as a reported finding.

The study addresses three research questions: (1) What are the frequencies and major collocational patterns of protect, preserve, and conserve in contemporary environmental English? (2) What semantic preferences characterize the collocational patterns of the three verbs? and (3) How do the three verbs differ in their contextual and collocational behavior?

By addressing these questions, the study seeks to contribute to corpus-based lexical semantics by providing empirical evidence concerning the behavior of three closely related environmental verbs. The findings may also contribute to ESP and advanced EFL vocabulary instruction by demonstrating how corpus evidence can be used to distinguish near-synonymous lexical items according to their recurrent patterns of use.

2. LITERATURE REVIEW AND THEORETICAL FRAMEWORK

2.1. Corpus Linguistics and Lexical Meaning

Corpus linguistics provides an empirical approach to language through systematically collected bodies of naturally occurring texts. Rather than treating lexical meaning as an isolated property of individual words, corpus-based approaches examine recurring linguistic environments. Frequency, concordance, collocation, and phraseological patterns can therefore provide evidence for how words function in authentic discourse (McEnery & Hardie, 2012). Sinclair (1991, 2004) argues that recurrent patterns of co-occurrence are central to language description. This makes corpus analysis especially useful for near-synonyms, because words with similar dictionary meanings may demonstrate different distributional patterns. For the present study, a corpus-based approach is appropriate because protect, preserve, and conserve share a broad semantic domain but may differ in the linguistic environments in which they occur.

2.2. Collocation and Near-Synonymy

Collocation refers broadly to the tendency of lexical items to occur together more frequently than would be expected by chance. Firth's (1957) approach to lexical meaning established the importance of examining words through their surrounding linguistic environments, while Sinclair (1991) developed corpus-based approaches to collocation and phraseological patterning. Collocational analysis is particularly valuable in the study of near-synonyms because near-synonymous words may share a core semantic component while differing in preferred collocates, grammatical patterns, register, or discourse functions. In environmental English, protect, preserve, and conserve illustrate this issue: all may refer broadly to safeguarding environmental entities, yet their preferred lexical partners may differ.

2.3. Semantic Preference and Semantic Prosody

Semantic preference concerns the tendency of a word to co-occur with particular semantic sets (Stubbs, 2001). Instead of asking

only which individual words occur with a target item, semantic-preference analysis asks what kinds of entities or concepts those collocates represent. Semantic prosody is related but distinct: it concerns the evaluative or attitudinal meaning that may emerge from recurrent collocational environments (Louw, 1993; Hunston, 2007). In this study, semantic preference is the principal interpretive construct, while concordance evidence is used to identify broader evaluative or functional tendencies where relevant.

2.4. Statistical Measures of Collocational Association

Raw frequency alone cannot determine whether a lexical association is characteristic of a target word. Mutual Information (MI) estimates the strength or specificity of association between two lexical items, whereas Log-Likelihood (LL) assesses statistical significance by comparing observed and expected frequencies. MI can favor low-frequency combinations, so the present analysis uses a minimum frequency threshold and contextual inspection.

2.5. Analytical Framework

The analytical framework can be summarized as: target verb

→ collocation → statistical association → semantic categorization → semantic preference → contextual interpretation. This framework combines quantitative evidence from frequency and association measures with qualitative interpretation from concordances and semantic categories. It is appropriate because the general semantic overlap among protect, preserve, and conserve makes dictionary-based comparison insufficient.

3. METHODOLOGY

3.1. Research Design

The study adopted a corpus-based descriptive design to analyze the collocational and semantic behaviors of target environmental verbs. The unit of analysis was the lexical verb occurring in environmental discourse. The analysis combined quantitative frequency and association measures with qualitative concordance interpretation and semantic categorization.

3.2. Corpus Construction

This study constructed a specialized 10-million-word corpus named the Contemporary Environmental English Corpus (CEEC-10M), comprising three components published between 2016 and 2026.

Sub-corpus	Source types	Words
A: Policy & Treaties	UNFCCC documents, IPCC reports, international biodiversity agreements	3.5 million
B: Academic Journals	Peer-reviewed environmental studies articles	4.0 million
C: Environmental Journalism	The Guardian Environment, Yale Environment 360, Mongabay	2.5 million
Total		10.0 million

3.3. Target Items and Search Procedure

The target items were protect, preserve, and conserve, including inflected variants such as protects, protecting, protected, preserves, preserving, preserved, conserves, conserving, and conserved. Searches were conducted using AntConc 4.2 within a symmetrical L5-R5 span. Candidate collocates were filtered using the retained thresholds of raw frequency ≥ 15 and MI ≥ 4.5 , followed by concordance inspection.

3.4. Semantic Categorization

Collocates were classified into semantic categories including people and communities; species and wildlife; biodiversity; ecosystems and habitats; natural resources; water and energy; cultural and historical entities; and environmental conditions or systems. The purpose was to identify semantic preferences rather than exclusive meanings.

3.5. Data Verification

To ensure data accuracy and reproducibility, all corpus outputs and collocation measures were cross-verified using AntConc 4.2. Exact frequencies and statistical parameters were systematically recorded for final analysis.

4. RESULTS

Table 1 summarizes the overall collocational profiles, absolute frequencies, Mutual Information (MI) values, and semantic categories associated with protect, preserve, and conserve.

Node	Collocate	Freq.	MI	Semantic category
protect	habitats	1,420	6.8	ecosystems/habitats
protect	species	1,210	6.5	biodiversity/wildlife
protect	vulnerable communities	980	5.9	people/communities
protect	borders	850	5.4	boundaries/systems
protect	rights	720	5.1	human/civil rights
Node	Collocate	Freq.	MI	Semantic category
protect	ecosystems	650	4.8	ecosystems/habitats
preserve	biodiversity	1,150	7.2	biodiversity
preserve	cultural heritage	990	6.9	cultural heritage
preserve	integrity	840	5.8	ecological conditions
preserve	forests	780	5.5	ecosystems/habitats
preserve	status quo	610	5.2	existing states
preserve	historical sites	590	4.9	cultural/historical entities
conserve	water resources	1,560	7.5	natural resources
conserve	energy	1,340	7.1	energy resources
conserve	natural resources	1,280	6.8	natural resources
conserve	marine ecosystems	950	6.2	ecosystems
conserve	soil	810	5.6	natural resources
conserve	forests	760	5.3	ecosystems

4.1. The Collocational Profile of Protect

Protect shows a broad preference for entities that can be conceptualized as vulnerable to harm, pressure, or external threats. The strongest reported associations are habitats (MI = 6.8), species (MI = 6.5), and vulnerable communities (MI = 5.9). Other collocates include borders, rights, and ecosystems. The reported concordance examples, including protect against climate shocks, protect forests from illegal logging, and protect endangered marine species, support an interpretation centered on prevention and safeguarding. The evidence should not be overstated as an exclusively adversarial semantic prosody; the more defensible interpretation is a preference for entities exposed to potential or actual harm.

4.2. The Collocational Profile of Preserve

Preserve shows a stronger orientation toward continuity, existing conditions, ecological integrity, and cultural or historical value. Its strongest reported associations are biodiversity (MI = 7.2), cultural heritage (MI = 6.9), and integrity (MI = 5.8), followed by forests, status quo, and historical sites. Concordance examples such as preserve ecological integrity, preserve traditional knowledge systems, and preserve pristine wilderness areas support an interpretation centered on maintenance and continuity.



4.3. The Collocational Profile of Conserve

Conserve exhibits the most concentrated association with resource-related concepts. The strongest reported collocates are water resources (MI = 7.5), energy (MI = 7.1), and natural resources (MI = 6.8), followed by marine ecosystems, soil, and forests. These patterns suggest an orientation toward entities understood as limited, valuable, consumable, or requiring deliberate management. Expressions such as conserve water, conserve energy, and conserve natural resources foreground responsible use, efficiency, sustainability, and future availability.

4.4. Comparative Semantic Preferences

Protect has the broadest apparent range and foregrounds safeguarding against harm. Preserve more strongly foregrounds continuity and maintenance of valued ecological or cultural states. Conserve is particularly associated with resources and ecological assets requiring careful and sustainable management. These are tendencies rather than absolute boundaries, because some collocates, such as forests and biodiversity, can occur with more than one verb.

5. DISCUSSION

5.1. Protect: Vulnerability and Preventive Safeguarding

The reported findings position protects as the most general safeguarding verb in the environmental domain. It collocates span human communities, species, habitats, ecosystems, rights, and other entities that can be conceptualized as vulnerable. This broad range is compatible with the core semantic idea of shielding an entity from harm. The distribution also explains why protect can function as a default environmental verb: it accommodates both human and ecological objects without requiring the object to be a consumable resource or an already preserved state.

5.2. Preserve: Continuity, Integrity, and Value

The profile of preserve differs in conceptual orientation. Biodiversity, cultural heritage, integrity, historical sites, and status quo evoke entities or states whose value is connected to continuity. Preserve therefore tends to foreground the desired persistence of something already regarded as valuable, intact, authentic, or worth maintaining. This provides a useful lexical-semantic distinction from protect, even though both can occur with biodiversity, forests, and ecosystems.

5.3. Conserve: Resource Management and Sustainable Use

The collocations of conserve show a stronger orientation toward resource management. Water resources, energy, natural resources, soil, and marine ecosystems are entities that can be managed over time. The conceptual frame therefore includes controlled use, efficiency, sustainability, and future availability. Conserve and protect may both occur with environmental resources, but conserve more strongly foregrounds responsible management.

5.4. Near-Synonymy as Overlap Plus Preference

The findings argue against a binary classification in which each verb has one exclusive meaning. Instead, the three verbs share a semantic core involving environmental safeguarding while displaying different preferred semantic environments. Statistical collocation identifies recurrent associations; semantic categorization reveals broader semantic preference; and concordance interpretation helps explain contextual function. This combination is more informative than frequency counts alone.

6. PEDAGOGICAL AND PRACTICAL IMPLICATIONS

Corpus-informed instruction can present each verb with representative lexical patterns rather than as a simple synonym list. For protect, teaching can emphasize people, species, habitats, ecosystems, and exposure to harm. For preserve, instruction can foreground biodiversity, heritage, integrity, historical entities, and valued existing states. For conserve, teaching can highlight water, energy, natural resources, soil, ecosystems, and sustainable management.

The findings should not be converted into rigid rules. A collocational preference is not the same as a grammatical prohibition. An expression may be less typical in a specialized corpus without being ungrammatical or semantically impossible. Learners should therefore interpret corpus frequencies as evidence of typicality and contextual preference rather than absolute rules of correctness. For academic writing, corpus-based lexical instruction can help learners choose among the verbs according to intended conceptual emphasis: protection from harm, preservation of an existing or valued state, or conservation through careful and sustainable management.

7. LIMITATIONS AND FUTURE RESEARCH

The study has several limitations. First, while the specialized 10-million-word corpus offers rich domain-specific evidence, it may not fully represent all registers of contemporary English. Second, semantic categorization inherently involves qualitative interpretation, which could be expanded in future studies by employing intercoder reliability checks. Future research could also compare these environmental verbs across broader general corpora or examine their translation equivalents in parallel corpora involving languages such as Vietnamese.

8. CONCLUSION

This study investigated the collocational behavior and semantic preferences of protect, preserve, and conserve in contemporary environmental English. The analysis indicates that the three verbs share a broad semantic domain of environmental safeguarding but display different collocational tendencies. Protect is broadly associated with vulnerable people, species, habitats, ecosystems, and other entities requiring safeguarding. Preserve is more strongly associated with continuity, ecological or cultural integrity, biodiversity, and valued existing states. Conserve is particularly associated with resources and ecological assets requiring careful management, sustainable use, and future availability.

The findings support the view that near-synonyms should not be treated as fully interchangeable merely because their dictionary definitions overlap. Their recurrent collocational environments provide additional evidence for lexical meaning and discourse function. For Environmental English and ESP, vocabulary instruction should therefore incorporate authentic lexical patterns and semantic preferences rather than relying exclusively on lists of synonyms.

REFERENCES

1. Firth, J. R. (1957). A synopsis of linguistic theory 1930–1955. In *Studies in linguistic analysis* (pp. 1–32). Philological Society.
2. Hunston, S. (2007). Semantic prosody revisited. *International Journal of Corpus Linguistics*, 12(2), 249–268.
3. Louw, B. (1993). Irony in the text or insincerity in the writer? The diagnostic potential of semantic prosodies. In M. Baker, G. Francis, & E. Tognini-Bonelli (Eds.), *Text and technology* (pp. 157–176). John Benjamins.
4. McEnery, T., & Hardie, A. (2012). *Corpus linguistics: Method, theory and practice*. Cambridge University Press.
5. Partington, A., Duguid, A., & Taylor, C. (2013). *Patterns and meanings in discourse: Theory and practice in corpus-assisted discourse studies*. John Benjamins.
6. Sinclair, J. (1991). *Corpus, concordance, collocation*. Oxford University Press.
7. Sinclair, J. (2004). *Trust the text: Language, corpus and discourse*. Routledge.
8. Stubbs, M. (2001). *Words and phrases: Corpus studies of lexical semantics*. Blackwell.