



## LINGUISTIC FEATURES OF COMPOUND WORDS IN ENGLISH

**Author:** Nasriddinova Surayyo Faxriddin kizi <sup>1</sup>

**Affiliation:** Tashkent University for Applied Sciences, master's student <sup>1</sup>

**DOI:** <https://doi.org/10.5281/zenodo.19640908>

### ABSTRACT

This article examines the linguistic features of compound words in English, with comparative reference to Uzbek. Word formation — understood as the systematic creation of new lexical units from existing linguistic material — is among the most dynamic processes in any living language, continuously enriching the lexicon. Among the various word-formation strategies, compounding occupies a particularly prominent position due to its productivity, structural transparency, and cross-linguistic prevalence. The article investigates how compound words are formed in English, reviews the major theoretical frameworks proposed by linguists, and discusses both structural and semantic dimensions of compounding. Special attention is given to the synchronic and diachronic perspectives on English word formation, as well as the growing influence of English compounding on other languages, including Uzbek.

**Keywords:** compound words, word formation, linguistic features, morphology, lexical units, endocentric compounds, exocentric compounds, English lexicology, Uzbek linguistics.

### INTRODUCTION

Language is the primary instrument of human communication, serving as the medium through which ideas, emotions, and social relationships are expressed and negotiated. As a structured system, language operates across multiple levels — phonological, morphological, syntactic, semantic, and pragmatic — each contributing to the overall capacity of a language to represent meaning. Among these levels, morphology occupies a central position, governing the internal structure of words and the rules by which new words are formed from existing elements.

The capacity of a language to generate new vocabulary is directly related to the vitality of its word-formation processes. Word formation creates new lexical units by exploiting the material already available within the linguistic system, thereby adapting the lexicon to the communicative demands of its speakers. As language communities grow, diversify, and engage with new knowledge domains, the lexicon must expand accordingly. Morphological processes — including affixation, conversion, blending, clipping, and compounding — are the mechanisms by which this expansion occurs.

Of these processes, compounding — the combination of two or more free morphemes to produce a single new lexical item — is widely acknowledged as one of the most productive in English. Compound words such as cell phone, cheeseburger, and playboy illustrate the ease with which English forms new lexemes from existing words, and the frequency with which such forms are borrowed into other languages reflects the global reach of English as a source of neologisms. The present article investigates the structural and semantic properties of English

compound words, drawing on a range of theoretical perspectives and focusing on both endocentric and exocentric compound types.

### **Theoretical Background: Defining the Compound Word**

The definition of the compound word has been a subject of sustained scholarly debate. Across the literature, however, a broad consensus exists: a compound is a lexical unit formed by the combination of two or more bases that can function independently as words [4, 11]<sup>1</sup>. Several defining properties distinguish compound words from phrases: their components are inseparable, no additional element can be inserted between them, and their internal order is fixed [1, 17]<sup>2</sup>.

Carstairs-McCarthy [7]<sup>3</sup> further specifies that compounds are produced by combining roots with phrasal words — items that possess the internal structure of a phrase but function syntactically as single words. Matthews [14]<sup>4</sup> defines the formation of a compound as the construction of a compound lexeme from two or more simple lexemes. Ullmann [19]<sup>5</sup> notes an important semantic dimension: the constituent elements of some compounds are semantically transparent, while in others the meaning is conventionalized to the point of opacity, with no recoverable connection between the parts and the whole.

Ginzburg et al. [8]<sup>6</sup> regard compounding as one of the most effective word-formation processes in modern English, emphasizing that compound words consist of exactly two immediate constituents and function as integral lexical units. Meshkov [15]<sup>7</sup> similarly describes compound words as indivisible lexical units formed by the union of two or more bases, dividing them into structurally motivated compounds — whose meaning derives compositionally from the parts — and structurally unmotivated compounds, whose meaning is not derivable from the parts.

### **Structural Classification of Compound Words**

The classification of compound words on structural grounds has produced several competing taxonomies in the literature. One of the most influential is that of Bauer [5]<sup>8</sup>, who identifies two primary features of English compounding: first, the majority of English compound words follow the N+N pattern; and second, the semantic and syntactic relations between constituents are diverse, necessitating a multi-group classification. Bauer distinguishes four structurally-functional groups.

The first group comprises endocentric compounds, in which the second (head) element is grammatically and semantically dominant, while the first element functions as a modifier or determiner. Examples include beehive and armchair, where the head noun determines both the grammatical category and the core meaning of the compound.

The second group consists of exocentric compounds — referred to in Sanskrit morphological tradition by the term bahuvrihi. In these cases, neither component serves as the grammatical or semantic head; instead, the whole refers to something only metonymically or metaphorically related to the parts. The word skinheads

<sup>1</sup>Arnold, I. V. The English word (1986); Katamba, F. Morphology (1993), p. 291

<sup>2</sup>Adams, W. An introduction to modern English word-formation (1973); Smirnitsky, A. I. Lexicology of the English language (1956)

<sup>3</sup>Carstairs-McCarthy, A. An introduction to English morphology (2002), p. 59

<sup>4</sup>Matthews, P. H. Morphology (1974), p. 82

<sup>5</sup>Ullmann, S. Semantics (1972), p. 81

<sup>6</sup>Ginzburg, R. S. et al. A course in modern English lexicology (1979)

<sup>7</sup>Meshkov, O. D. Word formation in modern English (1985)

<sup>8</sup>Bauer, L. English word-formation (1983)

exemplifies this type: the compound does not denote heads covered in skin, but a subculture identified by a specific appearance.

The third group treats the components of the compound as syntactically coordinate, resembling sentence-level constructions. The word maidservant belongs here, where both elements contribute equally to the compound's reference. The fourth group — the dvandva compound in Sanskrit terminology — includes cases where it is impossible to determine which component is grammatically or semantically dominant, as in poet-translator.

Arnold [4]<sup>9</sup> provides an additional layer of classification, distinguishing between asyntactic and syntactic compounds, and between endocentric and exocentric types. He identifies transparency, connectedness, and idiomatic meaning as central problems in defining and classifying compound words, synthesizing the criteria proposed by scholars including Nida, Bloomfield, Paul, Quirk, Bloch, Trager, and Marchand.

### Syntactic and Semantic Relations in English Compounds

A major contribution to the analysis of compound words comes from Lees [12]<sup>10</sup> and Marchand [13]<sup>11</sup>, who argue that compound formation derives from underlying syntactic structures. Marchand [13]<sup>12</sup> specifically proposes that compounds are formed on the basis of a determinative–definite relation, representing the morphological fusion of a syntactic construction. Based on this framework, eight types of grammatical connection underlie English compound formation:

- (a) Subject–Predicate: fighter plane (the plane is a fighter)
- (b) Subject–Middle Object: marrow bone (the bone has marrow)
- (c) Subject–Verb: wading bird (the bird wades)
- (d) Subject–Object: police dog (a dog used by the police)
- (e) Verb–Object: pushbutton (pushes the button)
- (f) Subject–Prepositional Object: coffee cream (cream for coffee)
- (g) Verb–Prepositional Object: grindstone (grinds on the stone)
- (h) Object–Prepositional Object: school grammar (grammar taught in school)

Greenbaum [9]<sup>13</sup> situates compounding among the four fundamental word-formation processes of modern English — alongside prefixation, suffixation, and conversion — and similarly applies grammatical terminology to describe internal compound relationships. He notes that compound words appear across all lexical categories, though nouns and adjectives predominate among newly formed items. His taxonomy of compound noun patterns includes: Subject + predicate (bee sting); Predicate + object (chewing gum); Object + predicate (air-conditioner); Subject + object (cable car); Predicate + place (dance hall); and Predicate + time (closing time).

For compound adjectives, Greenbaum [9]<sup>14</sup> describes patterns such as subject + predicate (English-speaking), place/time + predicate (far-reaching), noun + adjective with a simile relation (dirt-cheap), and coordinated adjectives (bitter-sweet). These models underscore the rich variety of conceptual relationships that compounding can encode within a single lexical item. In endocentric compounds, the final component typically indicates the lexical class of the compound (travel

<sup>9</sup>Arnold, I. V. *The English word* (1986)

<sup>10</sup>Lees, R. B. *The grammar of English nominalizations* (1960)

<sup>11</sup>Marchand, H. *The categories and types of present-day English word-formation* (1965)

<sup>12</sup>Marchand, H. *The categories and types of present-day English word-formation* (1965)

<sup>13</sup>Greenbaum, S. *The Oxford English grammar* (1996), p. 461

<sup>14</sup>Greenbaum, S. *The Oxford English grammar* (1996)

guide is a type of guide; pop group is a type of group). In exocentric compounds, this head relationship is absent or obscured — a hotdog is not a type of dog, and a white lady is not a type of face.

### Compound Words in English and Uzbek: A Comparative Note

The influence of English compound formation on Uzbek is evident in a growing number of borrowed compound lexemes. Terms such as cell phone, cheeseburger, and playboy have entered Uzbek with minimal phonological adaptation, reflecting the prestige and global dominance of English as a source of neologisms in science, technology, popular culture, and commerce.

In the history of Uzbek linguistics, the study of word formation — including compounding — developed largely under the influence of Russian linguistic tradition, which treated word formation as a subsystem of morphology. This framework, derived from an inflectional language typology, did not always map cleanly onto the agglutinative structure of Uzbek. Consequently, the analytical categories of Uzbek word formation were not always addressed on their own terms, and a fully autonomous theoretical framework for Uzbek compounding has only more recently been developed.

Following Uzbekistan's independence, the study of the national language received renewed scholarly and institutional attention. Significant theoretical and descriptive work across all domains of Uzbek linguistics has been undertaken, with particular emphasis on language development and the role of Uzbek as a state language. This period has witnessed the emergence of more systematic approaches to compound word formation in Uzbek, no longer mediated exclusively through Russian linguistic doctrine.

### Computational and Corpus-Based Approaches to Compound Analysis

While the theoretical analysis of compound words has a long tradition in linguistics, the computational identification and structural disambiguation of compounds presents distinct challenges. Two tasks are central: identifying the constituent elements of a compound, and determining the dependency structure that organizes those elements.

Early computational approaches relied on handcrafted rule-based methods [16]<sup>15</sup> and probabilistic models such as Markov chains [18]<sup>16</sup>. Later work exploited corpus co-occurrence data [10]<sup>17</sup> and statistical association measures — including mutual information — to select the most probable structural analysis among competing candidates. The use of internet-scale corpora has helped address the data sparseness problem that limits smaller corpus studies, yielding more statistically reliable analyses of rare compound types.

The challenge of structural disambiguation is particularly acute for compounds with three or more elements, where multiple bracketings are possible — for example, [[nuclear power] plant] vs. [nuclear [power plant]]. Corpus and statistics-based approaches employing a deterministic process that progressively eliminates less probable structures have shown considerable promise in resolving such ambiguities. These computational advances complement traditional linguistic analysis and are increasingly central to natural language processing applications involving compound word recognition.

<sup>15</sup>Miyazaki, M. et al. Compound word analysis, COLING-84 (1984)

<sup>16</sup>Takeda, K. & Fujisaki, H. Segmentation of kanji compound words (1987)

<sup>17</sup>Han, Z. et al. Compound word segmentation using contextual information from a corpus (2001)

## CONCLUSION

This article has examined the linguistic features of compound words in English from structural, semantic, and comparative perspectives. Compounding is one of the most productive word-formation processes in English, generating new lexical items through the combination of free morphemes in patterns that range from transparent N+N constructions to semantically opaque exocentric formations.

The theoretical frameworks surveyed — from Smirnitsky and Marchand to Bauer, Greenbaum, and Arnold — converge on several key points: compound words are integral lexical units with fixed constituent order; their meaning is not always compositionally derivable from their parts; and the structural relationship between constituents can be described in terms of endocentric versus exocentric organization. The syntactic-semantic typology developed by Lees [12]<sup>18</sup> and Greenbaum [9]<sup>19</sup> demonstrates that a wide range of conceptual relationships — agent-action, part-whole, purpose, resemblance — can be encoded within compound structures.

From a cross-linguistic perspective, English compounds exert significant influence on Uzbek and other languages through lexical borrowing, reflecting the global role of English in contemporary vocabulary expansion. The study of English compounding thus has both theoretical significances for morphological typology and practical relevance for lexicography, language pedagogy, and natural language processing.

## REFERENCES

1. Adams, W. (1973). *An introduction to modern English word-formation*. Longman.
2. Aitchison, J. (2008). *The articulate mammal: An introduction to psycholinguistics* (5th ed.). Routledge.
3. Antrushina, G. B., Afanasyeva, O. V., & Morozova, N. N. (1999). *English lexicology*. Drofa.
4. Arnold, I. V. (1986). *The English word*. Vysshaya Shkola.
5. Bauer, L. (1983). *English word-formation*. Cambridge University Press.
6. Campbell, L. (2013). *Historical linguistics: An introduction* (3rd ed.). Edinburgh University Press.
7. Carstairs-McCarthy, A. (2002). *An introduction to English morphology: Words and their structure*. Edinburgh University Press.
8. Ginzburg, R. S., Khidekel, S. S., Knyazeva, G. Y., & Sankin, A. A. (1979). *A course in modern English lexicology*. Vysshaya Shkola.
9. Greenbaum, S. (1996). *The Oxford English grammar*. Oxford University Press.
10. Han, Z., Guo, Y., & Sproat, R. (2001). Compound word segmentation using contextual information from a corpus. In *Proceedings of the 39th Annual Meeting of the ACL* (pp. 1–8). Association for Computational Linguistics.
11. Katamba, F. (1993). *Morphology*. Macmillan.
12. Lees, R. B. (1960). *The grammar of English nominalizations*. Indiana University Press.

<sup>18</sup>Lees, R. B. *The grammar of English nominalizations* (1960)

<sup>19</sup>Greenbaum, S. *The Oxford English grammar* (1996)

13. Marchand, H. (1965). The categories and types of present-day English word-formation (2nd ed.). C. H. Beck.

14. Matthews, P. H. (1974). Morphology: An introduction to the theory of word-structure. Cambridge University Press.

15. Meshkov, O. D. (1985). Word formation in modern English. Nauka.

16. Miyazaki, M., Ohyama, Y., & Kurematsu, A. (1984). Compound word analysis using a handcrafted rule-based method. In Proceedings of COLING-84 (pp. 142–145). ACL.

17. Smirnitsky, A. I. (1956). Lexicology of the English language. Moscow State University Press.

18. Takeda, K., & Fujisaki, H. (1987). Segmentation of kanji compound words using a Markov model. In Proceedings of the 25th Annual Meeting of the ACL (pp. 210–215). ACL.

19. Ullmann, S. (1972). Semantics: An introduction to the science of meaning. Blackwell.

20. Saxena, P., Paul, S. Sentiment Analysis of Idioms Using Neural Network Models // Proceedings of the 2022 Conference on Empirical Methods in Natural Language Processing. – Abu Dhabi, 2022. – P. 1142–1156.

21. Fadaee, M., Bisazza, A., Monz, C. Examining the Challenges of Neural Machine Translation for Idiomatic Expressions // Computational Linguistics. – 2018. – Vol. 44, No. 3. – P. 451–482.

