



Formalizability As A Criterion: Linguistic Paradigms And The Annotation Of Simple Sentences In Typologically Contrasting Languages

Khamdamova Dilnoza

Doctoral student, Karshi State University, Uzbekistan

ABSTRACT

The paper compares four contemporary linguistic paradigms generative, functional, cognitive, and typological as applied to the analysis of the simple sentence, and evaluates them against the criterion of formalizability for corpus annotation. Drawing on parallel English and Uzbek examples, it is argued that no single paradigm provides a comprehensive account of sentence structure and that an integrative framework offers the most productive basis for the syntactic annotation of typologically contrasting languages.

KEYWORDS: Simple sentence, syntactic annotation, formalizability, generative grammar, functional linguistics, cognitive linguistics, linguistic typology, Universal Dependencies, English, Uzbek.

INTRODUCTION

Over the past century, approaches to the analysis of the simple sentence the fundamental syntactic unit have evolved from structural and generative models to functional, cognitive, and typological perspectives, each highlighting different aspects of sentence organization. Recent advances in corpus linguistics and natural language processing have further transformed the field: simple sentences are now examined not only as objects of theoretical description but as units requiring formal representation for annotation, dependency parsing, and multilingual language resources.

Despite extensive research within individual frameworks, their systematic comparison from the perspective of modern syntactic analysis remains underexplored. The present paper compares the principal contemporary paradigms employed in simple sentence analysis and evaluates their contributions to modern syntactic research. It is argued that no single paradigm offers a comprehensive account of sentence structure, and that an integrative approach combining structural, functional, cognitive, and typological insights evaluated against the criterion of formalizability for annotation purposes provides a more effective foundation for both theoretical description and annotation of typologically diverse languages.

Contemporary Linguistic Paradigms in Simple Sentence Analysis

Contemporary research on sentence structure is characterized by the coexistence of several complementary paradigms, each highlighting distinct properties of syntactic organization formal, semantic, communicative, cognitive, and typological [1].

The generative paradigm, established by N. Chomsky, views the sentence as a hierarchically organized system governed by abstract grammatical principles [2]. The simple sentence is regarded as the realization of an underlying structure generated by universal rules; key notions

include constituent hierarchy, recursion, and the distinction between deep and surface representations. Thus, The architect designed the bridge is analyzed not as a linear sequence but as a phrase structure in which the NP (the architect) functions as subject and the VP (designed the bridge) as predicate, revealing dependencies inexplicable by word order alone [1]. These concepts remain a theoretical foundation for many computational models of language processing. However, the generative approach has been criticized for its limited attention to communicative context and authentic language use [6], which motivated the development of functional, cognitive, and typological alternatives.

Functional Perspective on the Simple Sentence

In contrast to the generative paradigm, the functional approach treats language as a communicative system whose grammatical organization is shaped by speakers' intentions and discourse needs [6]. According to M. A. K. Halliday, every clause simultaneously realizes three metafunctions: the ideational (representing experience), the interpersonal (reflecting speaker-listener interaction), and the textual (organizing elements into coherent discourse). Sentence analysis thus extends beyond identifying constituents to explaining how structures contribute to meaning construction.

Compare:

- (1) The teacher explained the rule carefully.
- (2) Carefully, the teacher explained the rule.

Generative syntax assigns both sentences essentially the same underlying structure, whereas functional linguistics interprets the fronting in (2) as an information-structuring strategy assigning discourse prominence to the manner of the action [6]. Uzbek exhibits a comparable mechanism realized through constituent order:

- (3) O'qituvchi qoidani batafsil tushuntirdi.
- (4) Qoidani o'qituvchi batafsil tushuntirdi.

While both sentences share the same propositional content, the initial position of qoidani in (4) topicalizes the object, altering the information structure of the utterance.

The functional approach is particularly valuable for comparative research, as it explains why speakers choose specific constructions; it has substantially influenced corpus linguistics and multilingual annotation, where information structure increasingly complements grammatical description [8]. However, communicative functions depend on discourse context and pragmatic interpretation and are therefore difficult to formalize; large multilingual annotation projects employ functional concepts as complementary rather than primary information [8].

Cognitive Perspective on the Analysis of Simple Sentences

Cognitive linguistics, unlike generative grammar with its view of syntax as an autonomous component, assumes that grammatical patterns arise from general cognitive abilities and reflect the ways speakers perceive and categorize reality [7]. A simple sentence is thus a conceptual representation of an event: R. Langacker treats constructions as symbolic units linking semantic structures with their expression, syntactic relations emerging from conceptual rather than purely formal dependencies [7].

Compare:

- (5) The boy opened the door.
- (6) The door opened.

(7) Bola eshikni ochdi.

(8) Eshik ochildi.

Sentences (5) and (7) construe the event as agentive, whereas (6) and (8) foreground the change of state without identifying its cause a difference in conceptual construal rather than a merely syntactic alternation [7]. Notably, the two languages encode this identical conceptual contrast by different grammatical means: English employs a labile verb (open in both transitive and intransitive use), while Uzbek marks the shift morphologically through the voice suffix (och- → och-il-). Cognitive principles thus prove valuable for explaining why typologically different languages select distinct strategies to represent similar conceptual situations [7].

However, categories such as construal and conceptual prominence resist objective, reproducible definition. Contemporary annotation frameworks therefore incorporate cognitive insights as complementary interpretative information while relying primarily on explicit structural representations [4].

Typological Perspective on Simple Sentence Analysis

Linguistic typology investigates how different languages encode similar grammatical relations through diverse structural means, seeking recurrent cross-linguistic patterns and universals [3, 5]. Languages differ not only in surface organization but in the very strategies constituent order, case marking, agreement employed to express relations such as subject, object, and predicate. The English–Uzbek contrast illustrates this particularly well. English, an analytic SVO language, expresses grammatical relations predominantly through constituent order:

(9) The student wrote the report.

(10) The report wrote the student.

Reversing the noun phrases in (10) yields a semantically anomalous sentence, since syntactic functions in English are assigned positionally. Uzbek, an agglutinative SOV language, encodes relations morphologically:

(11) Talaba hisobotni yozdi.

(12) Hisobotni talaba yozdi.

Because the accusative suffix -ni explicitly marks the direct object, both sentences preserve identical grammatical relations; the reordering in (12) affects only information structure. Equivalent propositions are thus represented through different morphosyntactic strategies, which is why, as W. Croft argues, cross-linguistic comparison must rest on functional equivalence rather than language-specific categories [3].

This perspective has become central to multilingual corpus annotation: projects such as Universal Dependencies balance universal syntactic relations with language-specific phenomena, enabling consistent annotation across typologically diverse languages [4, 8]. Nevertheless, languages with rich morphology and flexible constituent order often require annotation decisions not fully captured by models developed for analytic languages a challenge that motivates combining typological insights with dependency-based frameworks.

Comparative Evaluation of Contemporary Linguistic Paradigms

The comparison demonstrates that generative, functional, cognitive, and typological approaches are complementary rather than competing models, each describing a distinct dimension of syntactic organization while exhibiting its own methodological limitations. Their principal contributions and constraints are summarized in Table 1.

Table 1. Comparative characteristics of contemporary linguistic paradigms in simple sentence analysis

Paradigm	Principal contribution	Limitation in terms of formalizability
Generative	Formal representation of hierarchical structure	High formalizability, but limited coverage of communicative context
Functional	Explanation of information structure and discourse functions	Difficult to formalize computationally
Cognitive	Interpretation of conceptual motivation and construal	Subjective analytical categories
Typological	Cross-linguistic comparison and identification of universals	Primarily descriptive rather than predictive

Since no individual paradigm adequately accounts for all structural, semantic, communicative, and typological aspects of the simple sentence, contemporary research increasingly integrates them: multilingual annotation frameworks combine formal dependency relations with functional, semantic, and typological information within unified models. An integrative analytical framework therefore appears to provide the most productive methodological basis for modern syntactic research particularly for the description and annotation of typologically contrasting languages such as English and Uzbek.

Conclusion

The comparative analysis demonstrates that no single linguistic paradigm provides a comprehensive account of simple sentence structure, and that the paradigms differ substantially in the degree to which their categories can be formalized for annotation: each framework explains particular dimensions of syntactic organization while exhibiting methodological limitations when applied independently. An integrative perspective combining structural, functional, cognitive, and typological principles is therefore both a theoretical necessity and a practical requirement particularly in corpus linguistics and NLP, where grammatical description must remain compatible with multilingual data and computational processing. Future research should accordingly focus on interdisciplinary methodologies capable of providing a linguistically consistent foundation for the syntactic annotation of typologically diverse languages, including English and Uzbek.

References

1. Carnie A. Syntax: A Generative Introduction. 4th ed. – Hoboken: Wiley-Blackwell, 2021. – 512 p.



2. Chomsky N. Syntactic Structures. – The Hague: Mouton, 1957. – 117 p.
3. Croft W. Typology and Universals. 2nd ed. – Cambridge: Cambridge University Press, 2003. – 341 p.
4. de Marneffe M.-C., Manning C. D., Nivre J., Zeman D. Universal Dependencies // Computational Linguistics. – 2021. – Vol. 47, No. 2. – P. 255–308.
5. Greenberg J. H. Some universals of grammar with particular reference to the order of meaningful elements // Universals of Language / Ed. J. H. Greenberg. – Cambridge, MA: MIT Press, 1963. – P. 73–113.
6. Halliday M. A. K., Matthiessen C. M. I. M. Halliday's Introduction to Functional Grammar. 4th ed. – London: Routledge, 2014. – 786 p.
7. Langacker R. W. Cognitive Grammar: A Basic Introduction. – Oxford: Oxford University Press, 2008. – 562 p.
8. Nivre J., de Marneffe M.-C., Ginter F., Hajič J., Manning C. D., Pyysalo S., Schuster S., Tyers F., Zeman D. Universal Dependencies v2: An Evergrowing Multilingual Treebank Collection // Proceedings of the 12th Language Resources and Evaluation Conference (LREC 2020). – Marseille: ELRA, 2020. – P. 4034–4043.

