Philology Matters / ISSN: 1994 – 4233
Uzbekistan State World Languages University

2026/D2
LINGUISTICS
DOI: 10.36078/987655525

Received: April 11, 2026
Accepted: May 23, 2026
Available online: May 25, 2026

Mukhtasarkhon Azimova

PhD Student
Uzbekistan State World Languages University
Tashkent, Uzbekistan
E-mail: mkamoldinova@gmail.com
ORCID iD: 0009-0005-4855-9238

Muxtasarxon Azimova

Tayanch doktorant
O‘zbekiston davlat jahon tillari universiteti
Toshkent, O‘zbekiston

CLASSIFICATION AND FUNCTIONAL SIGNIFICANCE OF KEY LEXEMES

ABSTRACT

The primary aim of this study is to examine the role of core lexemes as system-forming elements in lexical organization and to develop principled criteria for their identification and classification. The research seeks to determine how core lexemes function within semantic fields and how they generate derivational, combinatorial, and discourse-level units.

To achieve these objectives, the main tasks of the study include defining diagnostic parameters of lexical dominance, formulating a structured classification model, and demonstrating its applicability through systemic and cognitive-linguistic analysis.

Furthermore, the methodology integrates structural-semantic analysis, semantic field modeling, prototype theory, and corpus-informed observation. The proposed Core Lexeme–Unit Matrix (CLUM) serves as the analytical framework, combining semantic centrality, morphological productivity, collocational stability, and discourse recurrence as identifying criteria. Comparative insights from semantic-field research on idiomatic expressions further support the gradient nature of lexical units.

The results indicate that core lexemes are not merely high-frequency words but function as organizing nodes within lexical systems. Their structure reflects a hierarchical core–periphery continuum, extending from basic embodied meanings to abstract conceptual domains (inflectional paradigms, derivational families, phraseological units, and discourse formulae). The findings demonstrate that lexical centrality is multidimensional and cognitively motivated. The

TAYANCH LEKSEMALARNING TASNIFI VA FUNKSIONAL AHAMIYATI

ANNOTATSIYA

Mazkur tadqiqotning asosiy maqsadi leksik tizimda yadro leksemalarning tizim hosil qiluvchi element sifatidagi rolini o‘rganish hamda ularni aniqlash va tasniflashning asoslangan mezonlarini ishlab chiqishdan iborat. Tadqiqot yadro leksemalarning semantik maydonlar doirasida qanday faoliyat ko‘rsatishini va ularning so‘z yasash, kombinatorlik hamda diskurs darajasidagi birliklarni qanday hosil qilishini aniqlashga qaratilgan.

Tadqiqotning asosiy vazifalari leksik dominantlikning diagnostik parametrlarini belgilash, strukturaviy tasnif modelini ishlab chiqish hamda uning tizimli va kognitiv-lingvistik tahlil asosida qo‘llanish imkoniyatini ko‘rsatishdan iborat.

Metodologiya struktur-semantik tahlil, semantik maydon modellashtirish, prototip nazariyasi hamda korpusga asoslangan kuzatuv usullarini o‘z ichiga oladi. Taklif etilgan Core Lexeme–Unit Matrix (CLUM) modeli yadro leksemalarni aniqlashda semantik markaziylik, morfologik mahsuldorlik, kollokatsion barqarorlik va diskursiv takrorlanish mezonlarini birlashtiradi. Idiomatik birliklarning semantik maydoniga oid qiyosiy kuzatuvlar leksik birliklarning gradiyent xususiyatini tasdiqlaydi.

Natijalar shuni ko‘rsatadiki, yadro leksemalar faqat yuqori chastotali birliklar emas, balki leksik tizimning tashkil etuvchi markaziy tugunlaridir. Ularning tuzilishi asosiy mujassam ma’nolardan abstrakt konseptual darajalargacha (flektiv paradigmalar, derivatsion oilalar, frazeologik birliklar va diskurs formulalar) kengayuvchi iyerarxik yadro-periferiya kontinuumini aks

E-ISSN: 2181 – 1237

paper's contribution is an analytical model that makes the author's position explicit: core lexemes are best identified by the convergence of semantic centrality and patterned usage.

Finally, the discussion highlights the theoretical implications for systemic lexicology, cognitive linguistics, and corpus-based research. The study concludes that core lexemes constitute dynamic conceptual anchors within the mental lexicon and communicative practice, offering a replicable framework for lexical analysis across languages. An important theoretical contribution of this paper is therefore a strengthened interface claim: core lexemes are a privileged observation point for lexicon–grammar–discourse interaction.

Key words: core lexeme, lexical system, semantic field, prototype theory, cognitive linguistics, derivational productivity, collocation, lexical centrality, systemic lexicology, core–periphery model.

ettiradi. Olingan natijalar leksik markaziylikning ko‘p o‘lchovli va kognitiv motivatsiyalangan hodisa ekanini ko‘rsatadi. Mazkur maqolaning asosiy hissasi – muallif pozitsiyasini aniq ifodalovchi tahliliy modelni taklif etishdir: yadro leksemalar semantik markaziylik va qoliplangan qo‘llanishning kesishish nuqtasida aniqlanadi.

Munozara qismida natijalarning tizimli leksikologiya, kognitiv lingvistika va korpus tadqiqotlari uchun nazariy ahamiyati yoritiladi. Xulosa sifatida yadro leksemalar mental leksikonda va nutq amaliyotida dinamik konseptual tayanch vazifasini bajarishi hamda turli tillarda leksik tahlil uchun takrorlanuvchi modelni taqdim etishi asoslanadi. Mazkur maqolaning muhim nazariy jihatlaridan biri kuchaytirilgan interfeys da'vosini ilgari surishdan iboratdir, ya'ni yadro leksemalar leksikon–grammatika–diskurs o‘zaro ta'sirini kuzatish uchun imtiyozli nazariy nuqta sifatida xizmat qiladi.

Kalit so‘zlar: yadro leksema, leksik tizim, semantik maydon, prototip nazariyasi, kognitiv lingvistika, so‘z yasash mahsuldorligi, kollokatsiya, leksik markaziylik, tizimli leksikologiya, yadro-periferiya modeli.

INTRODUCTION

The research into lexical studies revisits the ostensibly simple entity, the word, again and again. But much of modern theory and practice insists that the word is not a sufficient scientific primitive until it is narrowed down to the concept of a lexeme, i.e., meaning-bearing lexical unit abstracted over grammatical forms and realized through a variety of word forms. The lexeme is considered to be the basic and central unit of language that has lexical meaning and is central to the lexical system [Abdirazakova, 2024]. Uzbek linguistic scholarship explicitly views this shift as viewing the lexeme as the unit that constitutes lexical meaning, and thus distinguishes it as such, as opposed to merely grammatical units and to mere intuitions of orthographic words. Such re-specification by means of lexemes is non-terminological decoration: it allows systematic classification, lexicographic representation, and cross-level description [Rahmatullayev, 2006]. Meanwhile, the lexicon cannot be treated as an undifferentiated inventory. The lexical system is not a mechanical system of words but a system where the lexemes are related components of a single linguistic system [Begimova, 2022]. Russian/CIS structural traditions, stress that lexis is a system... the systematicity of lexical inventories is difficult to exhibit systematically compared with grammar, in that lexical inventories are large, open, and heterogeneous. This is exactly the reason why the designation of core lexemes is a methodologically appealing choice: it limits the analysis to a manageable subset that is postulated to bear a disproportionate semantic and functional burden. Of vital importance, though, the

challenge of the lexical system as something visible does not justify the downgrading of core to raw frequency but necessitates principled standards by which lexical units are linked to systemic relations and patterned use [Reformatsky, 1996]. The issue of defining subdomains (lexicology vs. grammar vs. word-formation) has traditionally been viewed as a theoretically consequential issue in Uzbek and wider Turko-logical scholarship. Word-formation research, such as that by recurring, reports that the result of derivation is itself a unit of lexicon (lexeme) and thus incorporates lexicological structure, despite its possible morphological production. This point extends beyond disciplinary politics: this is an indication that the lexeme units do not consist merely of the individual lemmas, but of the families and networks formed by the morpho-semantic processes which tend to focus on the central meanings [Hojiyev, 1989]. A further complication is brought about by international corpus linguistics: meaning can be spread over lengthy units, and not limited to individual words. Word-based corpus-driven approach can start with words, but soon finds recurring, semi-fixed structures which act as meaning-bearing units in their own right. Such understanding inspires an analytic approach whereby core lexemes are not only assessed as a single lexical unit; they are also assessed as the center of stable phraseological and collocational structures [Sinclair, 1996]. The paper is therefore strongly authoritative: core lexemes should be most appropriately viewed as system-forming nodes the centrality of which is both semantic (field-structural) and functional (discourse-operational). In other words, core lexemes are not just some common words; they are structurers of lexical domains, focal points of lexical bonds, and creators of repetitive patterns of lexical use [Apresjan, 2009]. The gap in the research to be filled here is that numerous studies either (i) define core vocabulary in terms of quantitative prominence, or (ii) consider lexical fields and relations without incorporating corpus-attested multiword patterning. The lexical system has been modelled in detail with the help of Uzbek systemic lexicology and Russian systematic lexicography, whereas the usage-based units have been modelled with the help of international corpus-oriented work; but these threads are too seldom combined into a single identification and classification process of the so-called core lexemes in languages [Nurmonov, 2002; Sinclair, 1991]. The aims are threefold: first, to establish and operationalize a cross-linguistically valid concept of core lexeme; second, to suggest a typology of the units, which are related to core lexemes (derivational, phraseological, and discourse formulaic); and third, to demonstrate why identification of core lexemes has functional importance to linguistic theory, corpus research, lexicography, and discourse [Lyons, 1977; Apresjan, 2009].

Literature review

The Uzbek lexicology offers a descriptively vocabulary based on lexemes. An example of such a textbook level account is that by Sh.Rahmatullayev where lexeme is the carrier of lexical meaning and lexemes are placed in larger units of the dictionary, clearly differentiating between lexical unit and phraseological unit as a part of the lexicon. Phraseological units are not related to lexemes in terms of structure and semantic composition; however, they are directly related to the lexical system because they form based on the lexical meaning [Abdirazakova, 2024]. This is

essential to core-lexeme studies, since a core lexeme can be central in the lexicon, and can also be involved in phraseological stratification; and similarly, functional cores can be phraseological (multiword) but not single-lexeme [Rahmatullayev, 2006]. The thoughts of A.Hojiyev on the disciplinarity of the word-formation are highly supportive of the lexeme-centred perspective: the derived product is a lexeme, and as such, is a part of the lexicon as a system, even in case its creation is mediated by the morphological operations. In the case of core-lexeme identification, this means lexemic core status is able to project into derivational families, and derivational productivity and semantic transparency are diagnostically relevant. That is, a core lexeme is not just a node in a list, but it frequently serves as a node in a derivational and semantic net [Hojiyev, 1989]. The historiographic synthesis by A.Nurmonov is especially useful as it records the shift in the lexicology of the Uzbeks toward the systemic analysis of the lexicology. Lexemes are examined as complexes of semes in that transition and the lexical-semantic grouping is represented using an explicit centre-periphery system. Of interest to this paper, the central lexemes are analytically visible in the account of A.Nurmonov: in the lexical-semantic groups, central items form the relations, including the gradonymy, hyponymy, part-whole relations, and hierarchical structuring. This is in effect a Uzbekian theoretical foundation of tayanch leksemalar being comprehended as field-cores [Nurmonov, 2002]. The role of A.Hojiyev in Uzbek lexicographic practice is also described as system-building work: descriptive and explanatory dictionary work, synonym lexicography, terminological lexicography are all discussed as in need of principled structure of lexical meaning and usage [Hojiyev, 1974]. Though peripheral, this kind of accounts is relevant in that it demonstrates how the choices made in lexicographic design operationalise assumptions about lexical centrality and active vocabulary. The system perspective is further extended by Russian/CIS theory. The tradition of textbooks by A.Reformatsky asserts that lexicon is systemic, but the systematicity of lexicon is more difficult to establish than grammatical systematicity since lexical inventories are infinite and unlimited. Here is the very place where core lexemes come in: they give a relatively small set with the help of which relations between systems can be rebuilt and experimented [Reformatsky, 1996]. This system-commitment is made operational through the system-commitment of Y.Apresjan. His description specifically suggests that the lexical meaning is organized systematically and that the lexical system must be represented with the elements of classification and operations, i.e., the associations between lexical meanings and rules regulating the interaction of meanings in text. This bi-faceted view fits well into core-lexeme studies: a lexeme can be considered core since it is the nucleus of classification (a field nucleus) or it can be operationally central (it is the vehicle of semantic rules and patterns of combinability) [Apresjan, 2009]. Although it is easy to relate the works of Y.S. Kubryakova to the broader field of cognitive linguistics, it is the classification and boundary issues that were the bane of core-lexeme identification. Her argument about categories with fuzzy edges and prototypical structure provides a methodological caution: both categories of text and lexicon are not associated with sets of necessary and sufficient features that are

clean, and researchers should be prepared to see gradient membership and prototype effects in the categorization of lexical items. This makes the case that core-lexeme identification needs converging criteria as opposed to a threshold [Kubryakova, 2001]. Complementary foundations are received in international scholarship. J.Lyons describes linguistic semantics as the study of linguistic meaning to the extent that it is systematized within vocabulary and grammar; a definition that justifies the study of core lexemes as an intermediary between lexical meaning and grammatical/discourse encoding. The importance of core lexemes, according to this perspective, is that they are common points where encoding is exposed and that lexical meaning is subject to grammatical patterning [Lyons, 1995]. The corpus-driven semantics of Sinclair insists that meaning is described by words but cannot end there as meaningful relations are discovered in regular co-text and repetitive structures. The implication to core-lexeme studies is simple: in case a lexeme is core, it is likely to have stable phraseological settings that can serve as units of extended meaning. This means that lexical core analysis should be both lexeme and unit-of-meaning-based [Sinclair, 1996]. The relevance of the prototype theory and lexical relationships by D.Cruse highlights the fact that lexical sense is strongly influenced by relational patterns, such as taxonomical and partwhole structure, and the possibility that membership of a category can be graded instead of categorical. This is significant in that a core lexeme is not just a common label, but usually a focal point in lexical relations (e.g., basic-level categories, salient superordinates or relational focal points in synonomy and contrast) [Cruse, 1986]. Lastly, the Natural Semantic Metalanguage program by A.Wierzbicka offers another, but also highly applicable, concept of lexical core: a collection of minimal, cross-linguistically shared semantic primitives that serve as a universal conceptual framework. With or without the acceptance of the NSM at its entirety, such is relevant to the paper under discussion because it (i) isolates conceptual universality and (ii) lexical frequency that is specific to a language, and thus inspires a stratified concept of core lexemes. The assertion by A.Wierzbicka that the set stabilized at 65 primes provides an explicit candidate of an operational core layer of conceptual nature, which may be superimposed onto Uzbek and Russian field cores and corpus-attested operational cores [Wierzbicka, 1972; Wierzbicka 2021]. These traditions are followed by a critical synthesis. The field-structural intuition of centre-periphery and seme analysis is provided by Uzbek systemic lexicology; explicit modelling of lexicon as a system of semantic rules and lexical types is provided by Russian systematic lexicography; corpus linguistics provides empirical diagnostics of extended meaning units and co-textual regularities. This drawback of each tradition separately is unsurprising: semantic-field and systemic approaches may fail to capture evidence of usage, whereas corpus approaches may over-privilege frequency without capturing semantic centrality. These constraints can only be scientifically fruitful when they are not considered competing paradigms as the case of core-lexeme research [Apresjan, 1966; Nurmonov, 2002; Sinclair, 1996].

Besides, research in structural semantics has long established that lexical meaning is organized relationally within semantic fields, where words derive their

value from their position in a conceptual network [Trier, 1931; Lehrer, 1974; Lyons, 1977; Cruse, 1986]. Extending this perspective to idiomatic expressions, later studies argue that idioms form thematic semantic networks and can be modeled as radial, prototype-based categories with graded core–periphery structure rather than binary membership [Lakoff, 1987; Rosch, 1975; Geeraerts, 2010; Grant, 2022].

Theoretical framework

The theoretical approach taken here is that a lexeme is a lexical-meaning-bearing unit underlying grammatically inflected wordforms; thus, it is the lexeme rather than the orthographic word that is the correct unit of marking out cores. A lexeme is a complicated linguistic component that is defined by semantic, morphological, phonetic, and syntactic properties and acts as a central organizer in the structure of linguistic systems [Nizomova, 2025]. This promise is based directly on Uzbek linguistic terminology making lexeme the carrier of lexical meaning and isolating it of grammatical units [Rahmatullayev, 2006]. It is on the basis of this lexeme that we adopt a systemic perspective of lexis. In the wake of A.Reformatsky saying that lexicon is systemic (though less easily captured), and Y.Apresjan openly saying that lexical meanings are organized and that lexical system has both a classificatory and a rule-central aspect, we take core status as a property that can be instantiated in two different ways (i) field-centrality (classification core) and (ii) rule-centrality [Reformatsky, 1996; Apresjan, 2009]. In order to put the classification core into operation, we use the seme-based modelling of Uzbek systemic lexicology. A.Nurmonov records a move to the study of lexical meaning as a framework of semes, both integrative (unifying) and differential (contrastive) semes, and explains lexical-semantic groups with centre-periphery structure. A core lexeme in such a model is often a centre lexeme: it bears semes inherent to it which bind a group together and is often involved in several kinds of relationships (gradonymic, hyponymic, partonomic). This gives a theoretical explanation as to why treating core lexeme is a structural concept and not a purely statistical concept [Nurmonov, 2002]. To operationalize the core of operations, we refer to the corpus-based unit-of-meaning orientation of Sinclair: the lexical meaning is formed in repetitive co-textual patterns, and the unit, the word, is a preliminary but incomplete unit since patterned sequences are internalized by the language users. A lexeme may be core in the current model in that it reiteratively roots extended configurations that perform important semantic and discoursal labor (e.g., stance, evaluation, coordination, institutional formulae) [Sinclair, 1996]. This is supplemented by a cognitive constraint to avoid over-formalization. The discussion of categories with fuzzy boundaries and prototypical organization by Y.S. Kubryakova justifies the treatment of lexical categories, such as core, as graded, not strictly delimited. The prototype-based lexical relations proposed by D.Cruse supports the idea that lexical categorization does not necessarily admit necessary and sufficient conditions and tends to be graded about prototypical centres. In this case core-lexeme identification is supposed to produce strong cores, weak cores and borderline cases instead of a sharp edge [Kubryakova, 2001; Cruse, 1994]. Lastly, we add a conceptual-universality layer as a hypothetical upper limit: A.Wierzbicka suggests a collection of semantic

primes and claims that this collection has reached 65 by 2014. This nucleus is not a list of frequencies but a statement of semantic indefinability and cross-linguistic existence. In my system, NSM primes serve as a comparative baseline: they determine candidate meanings which, when lexically central in Uzbek/Russian/English, are likely to engage in core behaviour; when non-lexically central, nonetheless serve as semantic decomposition targets when synthesising field structure and extended units [Wierzbicka, 2021]. These commitments have a net result which we refer to as the Core Lexeme–Unit Matrix (CLUM). The matrix differentiates (a) types of cores (conceptual, field-core, operational core) and (b) unit realizations (inflectional paradigm, derivational family, phraseological unit, discourse formula). The CLUM is specifically integrative: it is postulated that other languages can converge at the conceptual level and diverge at the field and operational level, due to the difference in lexicalization patterns and discourse routines [Apresjan, 2009; Sinclair, 1991].

The gradient organization of lexical structure has also been observed in idiomatic semantic fields, where expressions radiate from embodied core meanings toward metaphorical abstraction [Azimova, 2025]. The gradient structuring of lexical systems is further supported by evidence from semantic-field analysis of idiomatic expressions. In a recent study on laughter and crying idioms in English, a five-tiered model was proposed, distinguishing Core, Near Periphery, Mid Periphery, Far Periphery, and Ambiguous/Mixed categories. The analysis demonstrated a systematic progression from direct embodied emotional expression (e.g., burst out laughing, cry your eyes out) to metaphorical abstraction (e.g., laugh all the way to the bank, cry over spilt milk) [Azimova, 2025].

This continuum mirrors the structural dynamics proposed in the Core Lexeme–Unit Matrix (CLUM). Just as emotional idioms extend from physiological experience to abstract conceptual domains, core lexemes generate layered realizations that radiate outward into derivational, collocational, and discourse-level units. The structural parallel between idiomatic semantic peripheries and lexemic operational layers reinforces the claim that lexical organization is inherently hierarchical and cognitively motivated [cf. Goffman, 1959].

The existence of an "Ambiguous/Mixed" category in idiomatic structuring further supports the view that lexical systems are not binary but spectrum-based. This aligns with prototype theory [Rosch, 1975] and conceptual metaphor theory [cf. Lakoff & Johnson, 1980], confirming that lexical cores function as dynamic centers rather than rigid categorical boundaries.

METHODS

Its study design is qualitative-corpus-oriented-comparative, yet conceptual and methodological integration, and not comprehensive quantitative modelling, is its main novelty. The interest is to show how core lexemes can be found and categorized through a synthesis between systemic-lexicological features and corpus-based unit-of-meaning diagnostics, resulting in a process of analysis that is analytically explicit and replicable and scalable [Sinclair, 1991; Apresjan, 2009]. The number of sources

of data is three times. We are initially going to take explanatory and terminological dictionaries as repositories of lexeme differences and systematised descriptions of meanings; this is by the standards of Uzbek lexicographic tradition, in which the description of lexemes and the typology of dictionaries are explicitly debated in pedagogical and scholarly texts. Second, we refer to phraseological and collocational material as the evidence to the fact that multiword units belong to the organization of lexicon (the Uzbek scholarship clearly identifies lexical units and phraseological. Third, corpus-inspired reasoning, frequency/dispersion, and repeated co-text are used to detect operational cores and longer units, which has been stressed by Sinclair that the meaning can be created in patterned use. [Rahmatullayev, 2006; Sinclair, 1996]. Since Uzbek scholarship focuses on the lexeme as a carrier of lexical meaning, the first level is lemmatization at the lexeme level: inflected wordforms are traced to lexemes. This is essential in Uzbek and Russian, where rich inflection has the effect of swelling the number of tokens and obfuscating centrality at the lexeme level. The seme-based modelling used in systemic lexicology also corresponds to the lexeme level unit, as a lexeme is assumed to be a carrier of sememe structure [Rahmatullayev, 2006; Nurmonov, 2002; 186]. The second stage involves the creation of candidate core sets by convergent criteria. We have three criteria families which are based on one of the theoretical strands: 1. Conceptual criterion: whether the lexeme lexicalizes a meaning, which is plausibly compatible with a semantic primitive or near-primitive (conceptual core). The rationale behind this criterion is the fact that NSM specifically focuses on minimal meanings, which are cross-linguistically shared [Wierzbicka, 2021]. 2. Field-centrality criterion: the lexeme is a centre item of a lexical-semantic group, and integral semes bind the group membership, and there are more than two relational connections (gradonymic, hyponymic, partonomic). This criterion is related to the centre-periphery modelling of Uzbek systemic lexicology [Nurmonov, 2002]. 3. Operational criterion: whether the lexeme anchors recurring, extended co-textual patterns that act as meaning-bearing units, i.e., collocational/phraseological regularities which are sufficiently stable to act as a unit-of-meaning. This criterion adheres to semantics by corpus [Sinclair, 1996]. Unit identification is the third step: the units are identified in which the core-ness is realised of each candidate core lexeme. In this case I explicitly consider the units as multi-layered category. The Uzbek word-formation school focuses on the fact that the derived forms give rise to new lexemes and that the derived lexemes become the objects of lexicology; hence, derivational families are taken as the basic constructs when a core lexeme is derivationally productive [Hojiyev, 1989]. Simultaneously, Uzbek lexicology separates the lexical and phraseological units within the lexicon. Therefore, phraseological patterns are considered units that may be core due to (i) being anchored by a core lexeme (lexeme-core = unit expansion) or because (ii) they are multiword functional core discourse formulae [Rahmatullayev, 2006]. In the case of Russian, the systematic lexicography by Y.Apresjan offers a methodologically transparent reason as to why rules of interaction of meaning in a text should be regarded as part of lexical system. Practically, it involves recording selectional tendencies, permitted alternations and paraphrastic equivalences as

phenomena of lexicographic-grammar interface that tend to be focused around core predicates and high-level relational lexemes. [Apresjan, 2009]. In order to prevent an entirely frequency-based approach, we explicitly treat frequency and dispersion as necessary but not sufficient diagnostics. The fact that lexicon is systemic but hard to capture, as warned by A.Reformatsky, means that a simple measure of numeric prominence will not be able to tell information about system relations without the provisions of semantic structure and relational networks [Reformatsky, 1996]. Cross-linguistic comparison is made at the level of functional-semantic correspondence, but not at the level of one-to-one correspondence of lexical relations. This is in line with a general definition of linguistic semantics as meaning coded into words and syntax: which is that the same semantic task can be coded by different lexemes and lexical units as well as constructions across languages, and therefore core analysis would need to permit cross-lexical, cross-lexical unit and cross-construction mapping [Lyons, 1995; Sinclair, 1996]. Lastly, we deal with the category of core as a gradient. The fact that Y.S. Kubryakova focuses on indistinct delimiting and D.Cruse is a prototype-oriented methodology indicates that borderline cases must not be subjected to binary classification but rather the methodology produces a ranked or tiered core set, which is theoretically more faithful and empirically more robust [Kubryakova, 2001; Cruse, 1994].

RESULTS AND DISCUSSION

The results of the analysis are a three-layer taxonomy of core lexemes. We do not offer the layers as a list of exhaustive constructs but as an analysis, since the identification logic and the unit typology are the main contribution and not a closed inventory. This is in line with the focus of systemic lexicography on modelling lexical structure and rules and with the focus of corpus-based semantics on units that arise out of use [Apresjan, 2009; Sinclair, 1991]. The conceptual core is the first layer, which is characterized by lexemes that lexicalize very general meanings which are consistent with semantic primitives or near-primitives (e.g., PERSON, DO, SAY, KNOW, GOOD/BAD, TIME-like relations). The account offered by Wierzbicka is important in this respect as it considers semantic primes as common conceptual atoms and demonstrates that the offered inventory became stable at 65. The conceptual core layer does not purport that these meanings are the same lexemes in different languages; it merely purports that when a language lexicalises such meanings, they will tend to exhibit core-like behaviour since they are involved in innumerable semantic compositions and discourse functions [Wierzbicka, 2021]. Lexemes encoding highly general human and action concepts (i.e., *odam* - ‘person’), *qilmoq* - ‘to do’, *bilmoq* - ‘to know’, *demoq* - ‘to say’) are supposed to be core candidates of concepts in Uzbek. Similar candidates in Russian are: человек, делать, знать, сказать. In English, there are such analogs as person/people, do, know, say. The theoretical point is not lexical equivalence but functional prediction, but in prediction that these lexemes will produce large operational signatures, i.e., high combinability, large scope of semantic domain, and high rates of occurrence in generalized constructions. The prediction is in line

with the operation conception of lexical system by Y.Apresjan, which is the meaning interaction rules in the text [Apresjan, 2009; Wierzbicka, 2021]. The second layer is the systemic field-core which are lexemes that take up the centre position in lexical-semantic groups. Uzbek systemic lexicology defines lexical meaning seme-structured and focuses on integral and differential semes in grouping. In this type of model, field-core lexeme is characterized by (i) bearing the inherent seme that determines a group and (ii) becoming the focus of semantic oppositions and relational connections. A.Nurmonov directly provides a framing of lexical-semantic group membership and modelling centre-periphery relations of such groups. [Nurmonov, 2002]. One definite example of what Nurmonov is discussing is the seme-based breakdown of lexemes based on kinship (his example of *aka* elder brother is broken down into semes which include PERSON/MALE/RELATIVE/OLDER). The core-lexeme inquiry is methodological, in that, by being a lexeme with seme structure containing integrative features to bind a group and being involved in multiple systematic opposition, it becomes field-core even when not the most often occurring object. That is, field-core is not merely a relational architectural property, but a property of relational counts [Nurmonov, 2002, p.186]. In projecting field-core behaviour in cross-linguistic projection, it is evident that field-cores tend to occur at basic-level or cognitively salient categories (e.g., common kin terms, basic motion predicates, core evaluation adjectives). This prediction is in line with prototype theory: categories are arranged around prototypical centres, and prototypically-centred lexemes in general structure lexical relations, like hyponymy or meronymy [Cruse, 1994].

The third layer is the discourse-operational core which is defined as lexemes whose core-ness is mostly apparent in patterned use: anchoring stable or semi-stable multiword units, discourse formulae, and recurrent patterns which carry out high-frequency text functions. This is because Sinclair asserts that the word is the place of departure to describing meaning, but that meaning are in relations with other words and in recurrent arrangements, which has a direct stimulus to the claim that operational cores are lexemes with robust extended-unit signatures [Sinclair, 1996]. The operational cores tend to be small lexemes with a high functional value: high-level verbs (DO, BE, KNOW), discourse-organising conjunctions and connectives, evaluative adjectives, and evidential/stance-related elements. Notably, operational core does not reduce to the presence of functional words alone; instead, it also contains the content lexemes, which serve as patterns of repetitive discourse moves (e.g., X is important, it turns out that X, it is possible to X). This aligns with Y.Apresjan concept which holds that lexis possesses rules of operation in the meaning interaction within text [Apresjan, 2009]. An important analytic finding is that every core layer is matched with unique unit realizations. We conclude with these types of units as four so-called lexeme units, all based on the Uzbek and Russian systemic perspective and corpus-based unit-of-meaning theory. Unit Type A: Inflectional paradigms (lexeme 2 wordform sets). Uzbek lexicological description actually encourages the lexeme-level analysis specifically due to the grammatically determined realization of wordforms that lexical content is represented by the lexeme as an abstract unit.

Thus, inflectional paradigms are not noises; they are the organized representations of core lexemes that occur in grammatical situations. [Rahmatullayev, 2006]. Unit Type B: Derivational families (core lexeme as derivational hub). The discussion on word-formation by A.Hojiyev is categorical here: derivative products are lexemes, and therefore, derivation extends lexicon by creating lexemes. When a lexeme is a core hub, the derivational family can focus semantic space on core meanings forming unit clusters that are structurally and functionally important (e.g., action-agent-instrument families) [Hojiyev, 1989]. Type C: Phraseological and collocational units (extended meaning units). Lexical units and phraseological units are explicitly separated in Uzbek lexicological treatment, allowing a two-track analysis: a core lexeme can be central as a lexical unit and at the same time occupy phraseological strata, and in other cases phraseological units themselves can be functional cores whether singly central lexically or not. The unit-of-meaning approach of Sinclair gives the usage-based reason as to why such extended units should be regarded as primary objects of description of meaning [Rahmatullayev, 2006; Sinclair, 1996]. unit D: Discourse formulas and pragmatic structures (operational templates). The study of text types by Y.S. Kubryakova accentuates pragmatic orientation, intentionality, functional organisation of the discourse objects; it justifies the treatment of recurrent discourse formulae as the unit whose contribution cannot be reduced to the lexical meaning of the single elements. Such templates are then anchored by operational cores [Kubryakova, 2001]. The CLUM model is known to predict and the cross-linguistic comparison to assure at the level of analytic plausibility that the same lexeme may occupy several layers of the core. As an example, DO-verbs (qil-, делать, do) are conceptual cores (basic action concept), field-cores in the domain of action predicates (central relational hubs) and operational cores as they anchor many light-verb and constructional templates. This stratified core-ness is what is exactly lost in defining core vocabulary as a frequency list [Wierzbicka, 2021; Sinclair, 1996]. A concise diagram is in the table below, which shows the alignment between core layers, unit types, and functional signatures. The objects are set as analytic exemplars not as a comprehensive empirical list; the idea is to illustrate the rationale of classification as well as the place of units. Exemplar Core Lexeme-Unit Matrix between Uzbek, Russian, and English. The conceptual/ pragmatic anchoring and unit types are based on the systemic and unit-of-meaning theory as opposed to one frequency threshold [Apresjan, 2009; Nurmonov, 2002; Sinclair, 1996].

The indication of Core lexemes and their units

Core Layer	Uzbek Exemplar	Russian Exemplar	English Exemplar	Dominant Unit Types	Typical Functional Significance
Conceptual Core	odam 'person' bil- 'know'	человек знать	person / people know	A, C, D	• Conceptual composition • Stance / evidence • Reference tracking
Systemic Field-Core	aka (kin domain centre)	брат / kin nuclei	brother / kin nuclei	A, B	• Field organisation • Oppositions • Semantic grouping
Systemic Field-Core	aka (kin domain centre)		brother / kin nuclei	A, B	• Formulaic discourse moves • Pragmatic routines • Cohesion
Discourse-Operational Core	qil-, de- templates		делать, сказать templates	C, D	A Frequent lexical items B Domain-specific terms C Collocational patterns D Discourse templates
	qil-, de- templates		do, say templates	C, D	

A Frequent lexical items B Domain-specific terms C Collocational patterns D Discourse templates

Lastly, the results indicate that "core lexeme" is not an inherently evident natural category; rather, it is a theory-mediated classification whose validity is contingent upon diagnostics. In systemic lexicology, diagnostics encompass seme structure and field relations; in systematic lexicography, they comprise semantic rules and lexical types; in corpus-driven semantics, diagnostics involve stable co-textual regularities. The main result of these diagnostics coming together is that core lexemes are those that stay central when looked at from different angles [Nurmonov, 2002; Apresjan, 2009; Sinclair, 1996].

DISCUSSION

The layered taxonomy clarifies why simplistic operationalizations of "core vocabulary" are theoretically underpowered. If core status is equated with frequency alone, then lexemes that are field-central but not maximally frequent will be misclassified as non-core; conversely, high-frequency items that lack field-centrality may be misclassified as core despite having limited organising power in lexical relations. A.Reformatsky's observation that lexical system is difficult to establish is instructive here: frequency is not a system model; it is a surface indicator that must be interpreted through theory [Reformatsky, 1996].

The systemic lexicology tradition in Uzbek scholarship provides a direct solution by shifting attention to seme-based modelling and centre-periphery field architecture. When a lexical-semantic group is modelled as a structured whole, "central lexemes" become empirically and theoretically justifiable. These central

lexemes are not "central" because they are frequent; they are central because they unify and differentiate semantic space via integral and differential semes, and because they anchor relational networks (gradonymy, hyponymy, partonymy). This is precisely the field-core component of the CLUM model [Nurmonov, 2002].

Apresjan's distinction between classification and operational aspects of lexis strengthens the same point from a different angle. A lexeme can be system-core because it sits at a classificatory centre (semantic class membership), or because it participates in rules of meaning interaction in text. This helps explain why some lexemes "feel" core in discourse even when not obvious field-centres: they may be operational pivots, especially among predicates and relational lexemes that drive combinability and semantic alternations. [Apresjan, 2009].

Corpus linguistics, as represented by J.Sinclair's unit-of-meaning position, provides the empirical mechanism that makes operational core detectable. International corpus linguistics adds a further complication: meaning is often distributed across extended units rather than confined to single words. All linguistic units, including lexemes, morphemes, and syntactic structures, realize their functional roles within sentence structure, where they operate as syntactic-semantic elements [Mamurov & Tilakova, 2022]. This insight motivates an analytic strategy. If meaning and function are distributed across recurring patterns, then the lexeme's core-ness is partly a property of its *habitual environments* and the extended configurations in which it participates. This perspective corrects a persistent weakness in purely field-based approaches: lexical relations alone do not predict the pragmatics of real discourse unless linked to usage [Sinclair, 1996].

The cognitive-prototype perspective [Kubryakova, 2001; Cruse, 1994] adds an essential meta-theoretical caution: "core" is likely to be a graded category, not a crisp set. If categories like "text" or lexical categories are prototypically organized with fuzzy boundaries, core lexemes will also show degrees of membership, with prototypical cores and peripheral candidates. This has a direct methodological implication: researchers should replace single-threshold classification with tiering (strong core vs. weak core) and should explicitly justify borderline decisions with converging diagnostics rather than with ad hoc judgment [Kubryakova, 2001; Cruse, 1994].

A.Wierzbicka's NSM framework reframes the debate by separating conceptual core from usage prominence. Even if a meaning is "primitive" and cross-linguistically shared, its lexicalization and operational prominence can vary across languages and registers; conversely, some operational cores may be language-specific discourse routines. This supports the CLUM claim that core-lexeme research must distinguish at least three layers: conceptual core (shared atoms), field core (system organization), and operational core (discourse routines) [Wierzbicka, 2021].

An important theoretical contribution of this paper is therefore a strengthened interface claim: *core lexemes are a privileged observation point for lexicon–grammar–discourse interaction*. Sinclair explicitly notes that the word aligns grammar and vocabulary, while the sentence aligns grammar and discourse; this alignment framing

can be extended: core lexemes are those lexical items that most visibly mediate between the grammatical system and discourse-level meaning work through their patterned co-textual behaviour [Sinclair, 1996].

Uzbek word-formation scholarship, exemplified by A.Hojiyev, further supports the interface claim by showing that word-formation cannot be confined to morphology because the outputs are lexemes and thus re-enter the lexicological system. Core lexemes, from this perspective, can (i) structure derivational families and (ii) provide templates for lexical innovation, thereby linking lexical centrality to productivity and change. This is methodologically relevant because it suggests that a core lexeme's units are not only collocations but also *derivational clusters* that may be culturally and historically salient [Hojiyev et.al., 2001].

Lexicography is an obvious applied domain. Uzbek scholarship on A.Hojiyev's lexicographic work highlights that dictionary design depends on principled structuring of lexical meaning, including the identification of active vocabulary, the modelling of synonymic relations, and the presentation of illustrative examples. The CLUM model can be used as a meta-design principle: dictionary entries for core lexemes should prioritize unit-rich descriptions (collocations, phraseology, discourse patterns) and system relations (field links, oppositional structure) because those properties constitute core status.

From a discourse-analytic perspective, the operational core layer has a particularly direct payoff. If operational cores are lexemes that anchor recurrent discourse formulae, then identifying them helps explain discourse cohesion, stance marking, and pragmatic routine across genres. Y.S. Kubryakova's insistence on intentionality and pragmatic orientation in defining textual categories generalizes: discourse units are not reducible to single lexemes, and core analysis must therefore treat multiword units as first-class research objects [Kubryakova, 2001].

Overall, the approach improves existing models by introducing a principled integration: instead of choosing between field theory, systematic lexicography, and corpus-driven units, it turns their tensions into constraints that jointly define core status. The result is not merely a new definition but a research programme: to identify core lexemes is to model the lexical system's centres, the operational rules of meaning in text, and the "units" through which this system becomes empirically observable [Apresjan, 2009; Nurmonov, 2002; Sinclair, 1996].

CONCLUSION

Core lexemes should be treated as system-forming lexical nodes rather than as mere members of a high-frequency list. Their theoretical value emerges when they are analysed as centres of lexical-semantic organisation and as anchors of recurrent multiword units that realise meaning in discourse.

The analysis proposed a three-layer taxonomy: conceptual core, systemic field-core, and discourse-operational core and demonstrated how each layer corresponds to distinct unit types: inflectional paradigms, derivational families, phraseological units, and discourse formulae. This layered approach explains why core status is gradient

and why robust identification requires converging diagnostics rather than single-threshold classification.

The paper's contribution is a unified framework (the Core Lexeme–Unit Matrix) that integrates Uzbek systemic lexicology, Russian systematic lexicography, and corpus-driven unit-of-meaning analysis into a replicable research logic. This framework clarifies how core lexemes mediate between lexicon, grammar, and discourse, and it provides a principled foundation for future corpus studies, lexicographic design, and comparative lexical research across Uzbek, Russian, and English.

REFERENCES

1. Abdirazakova, Z. K. (2024). Lexical and phraseological units, their meanings and usage features. *Pedagogs Journal, 59*(1), 51–53.
2. Apresjan, Y. D. (1966). *Ideas and methods of modern structural linguistics: A brief survey.* Moscow: Enlightenment Publishing House.
3. Apresjan, Y. D. (2009). *Studies in semantics and lexicography. Vol. 1: Paradigmatics.* Moscow: Languages of Slavic Cultures Press.
4. Azimova, M. (2025). Mapping the semantic field of laughter and crying idioms in English: A corpus-based analysis. *Philology Matters, 4*, 19–31.
5. Begimova, G. J. (2022). Interpretation of lexical-semantic level and its units. *Education and Innovative Research, 3*, 38–40.
6. Cruse, D. A. (1986). *Lexical semantics.* Cambridge: Cambridge University Press.
7. Cruse, D. A. (1994). Prototype theory and lexical relations. *Rivista di Linguistica, 6*(2), 167–188.
8. Geeraerts, D. (2010). *Theories of lexical semantics.* Oxford: Oxford University Press.
9. Goffman, E. (1959). *The presentation of self in everyday life.* New York, NY: Anchor Books.
10. Grant, L. E. (2022). *Idioms: A view from the web.* Amsterdam: John Benjamins Publishing Company.
11. Hojiyev, A. (1974). *Explanatory dictionary of Uzbek synonyms.* Tashkent: Science Publishing House.
12. Hojiyev, A. (1989). *Uzbek word formation.* Tashkent: Teacher Publishing House.
13. Hojiyev, A., Nurmonov, A., Zaynobiddinov, S., Kokren, K., Saidxonov, M., Sobirov, A., & Quronov, D. (2001). *Explanatory dictionary of active words in modern Uzbek.* Tashkent: Orient Publishing House.
14. Kubryakova, Y. S. (2001). On text and the criteria for its definition. In *Text: Structure and semantics* (Vol. 1, pp. 72–81). Moscow.
15. Lakoff, G. (1987). *Women, fire, and dangerous things: What categories reveal about the mind.* Chicago, IL: University of Chicago Press.
16. Lakoff, G., & Johnson, M. (1980). *Metaphors we live by.* Chicago, IL: University of Chicago Press.
17. Lehrer, A. (1974). *Semantic fields and lexical structure.* Amsterdam: North-Holland Publishing Company.
18. Lyons, J. (1977). *Semantics* (Vol. 1). Cambridge: Cambridge University Press.
19. Lyons, J. (1977). *Semantics* (Vols. 1–2). Cambridge: Cambridge University Press.
20. Lyons, J. (1995). *Linguistic semantics: An introduction.* Cambridge: Cambridge University Press.
21. Mamurov, Q., & Tilakova, N. (2022). Linguistic units of translation and directions for their study. *Zenodo.*
22. Nizomova, Z. M. (2025). The role of lexical typology and lemmatization in the formation of terminological systems. *Global Science Review, 4*(1), 326–332.
23. Nurmonov, A. (2002). *History of Uzbek linguistics.* Tashkent.
24. Rahmatullayev, Sh. (2006). *Modern literary Uzbek.* Tashkent: University Press.

25. Reformatsky, A. A. (1996). *Introduction to linguistics* (5th ed.). Moscow: Aspect Press.
26. Rosch, E. (1975). Cognitive representations of semantic categories. *Journal of Experimental Psychology: General, 104*(3), 192–233.
27. Sinclair, J. (1991). *Corpus, concordance, collocation.* Oxford: Oxford University Press.
28. Sinclair, J. (1996). The search for units of meaning. *Textus, 9*, 75–106.
29. Trier, J. (1931). *The German vocabulary in the semantic field of intellect: The history of a linguistic field.* Heidelberg: Winter.
30. Wierzbicka, A. (1972). *Semantic primitives.* Frankfurt am Main: Athenäum.
31. Wierzbicka, A. (2021). Semantic primitives, fifty years later. *Russian Journal of Linguistics, 25*(2), 317–342.