

# Grouping of Nouns in The Uzbek Language into Semantic Fields

Tirkashev Jasurjon Uktam ugli

Senior Specialist at the Agency for the Assessment of Knowledge and Skills, Uzbekistan

**Received:** 12 February 2026; **Accepted:** 09 March 2026; **Published:** 31 March 2026

**Abstract:** This article analyzes the issue of grouping nouns in the Uzbek language into semantic fields. The main objective of the research is to comprehensively group Uzbek nouns thematically and semantically, and to conduct a comparative analysis of this classification with existing semantic systems in English, Russian, and Turkish. The study utilized materials from the electronic corpus of the Uzbek language and the explanatory dictionary. As a result, 10,000 Uzbek nouns were categorized into the following six main semantic fields: Nature, Human Activity, Material Objects, Abstract Concepts, Place and Space, and Food. Each semantic field was further divided into smaller subgroups, whose proportions were statistically analyzed and represented using diagrams. The findings were compared with other languages, and commonalities and differences were examined. The research results can be applied in compiling semantic dictionaries, automatic semantic tagging, and developing multilingual ideographic dictionaries.

**Keywords:** Semantic field, Uzbek language, nouns, semantic classification, dictionary, hyponymy, hyperonymy, abstract concepts, electronic corpus, thematic groups, ideographic dictionary, automatic tagging, comparative analysis, corpus linguistics.

**Introduction:** Classification of lexical units by semantic fields is a method of studying vocabulary by dividing it into content domains. The concept of the semantic field was introduced into linguistics in the 20th century by the German scientist Jost Trier, who emphasized the need to consider semantically related words of a language as a single system [1]. According to Trier, a semantic or conceptual field is a range of concepts that includes certain lexical units connected through semantic relationships. A change in the meaning of one word within a field affects other words [6]. Later, scientists such as L. Weisgerber developed this theory and divided semantic fields into single-level and multi-level types. For example, Weisgerber cites the fact that in one dialect of the German language, paternal uncles and maternal uncles are expressed by separate words, while in standard German both are expressed identically by the word "onkel" [1]. This situation is

interpreted as a level differentiation in the language's "kinship" field.

## METHODS

Currently, the method of dividing lexicon into semantic groups in different languages is of crucial importance in the fields of dictionary compilation, translation, and computational linguistics. For example, in the English language, the WordNet system divides all nouns into 26 semantic classes (person, place, event, quantity, etc.), and each synset, or group of words with the same meaning, is considered to belong to the same semantic domain [3]. In Russian linguistics, there is a tradition of ideographic dictionaries and thematic classifications. In Yu. N. Karaulov's work "General and Russian Ideography," works in this direction in English, German, French, Russian, and Czech are analyzed in detail [1]. He demonstrates that in these works, various scholars attempted to group vocabulary by topics based on the

principle of the "Linguistic Picture of the World." In particular, German scientists R. Hallig and W. Wartburg, over 20 years of research, developed a comprehensive conceptual scheme - the "Linguistic Picture of the World" [1]. Their "Universe" classification included four major divisions: Celestial bodies, Earth, Plant world, and Animal world [1]. This approach has served as a model for many subsequent studies [6]. In Turkic languages, however, work on dividing the dictionary into semantic fields has been conducted less frequently.

In recent years, several studies have been conducted in Uzbek linguistics to examine lexicon from the perspective of semantic fields. Notably, Sh. Rakhmatullayev's "Study of Uzbek language lexicon as a semantic field (personal microfield)," R. Safarova's "Hyponymy in the Uzbek language (based on frequently used zoonyms)," and G. Nematova's "The system and artistic usage of plant name lexicon in the Uzbek language" have shed light on the systematic semantic properties of the language's vocabulary. These works analyzed semantic relationships of nouns - such as hyponymy (genus-species relationship), partonymy (whole-part relationship), and graduonymy (gradation series) - within specific microfields. However, the comprehensive classification of nouns in the Uzbek language as a system of broad semantic fields remains a topic that has not yet been fully explored.

The aim of this research is to categorize Uzbek nouns into large semantic fields based on their meaning and to conduct a comparative analysis of this classification with established semantic systems in English, Russian, and Turkish. Within the scope of the study, firstly, nouns occurring in the Uzbek language will be thematically collected and their ideographic classification will be compiled. Subsequently, this classification will be compared with categories in other languages (for example, in English WordNet or Russian ideographic dictionaries). The outcomes of this research will contribute to a comprehensive understanding of the Uzbek language vocabulary as a semantic structure and create a knowledge base that can be applied in translation, lexicography, and automated text analysis. This study was conducted based on the analysis of data corpora and lexical sources. Initially, a list of the most frequently used

nouns was compiled using materials from the explanatory dictionary of the Uzbek language and the national corpus. These nouns were then grouped into thematic categories according to their semantic proximity. The grouping adhered to the onomasiological principle of language units, meaning that objects and phenomena were systematized under general categories based on their semantic content. In determining lexical-semantic groups, the hyponymy-hyperonymy and species-genus relationships of words were used as the foundation. For instance, names such as crow, nightingale, and eagle were grouped under the general hyperonym "birds" [6].

The integration of approaches from local and foreign scientific sources played an important role in forming the classification. In particular, the "Linguistic Picture of the World" scheme created by Hallig and Wartburg and its categories were considered [2]. Their classification includes comprehensive sections such as nature, human, and material objects, and similar broad fields were identified in our study. Additionally, 26 categories (noun.person, noun.animal, noun.artifact, etc.) from the English-language WordNet system, known as lexicographer files, were used comparatively [3]. This experience of dividing English nouns into broad classes was adapted to the Uzbek language. Dictionaries compiled on the ideographic principle in the Russian language and the theoretical works cited by Karaulov (1976) were used as a methodological basis [1].

To ensure the completeness and empirical validity of the classification, data from the electronic corpus of the Uzbek language were utilized. Several hundred nouns were collected from the corpus for each semantic field, and their correspondence to the respective group was verified. When necessary, certain borderline cases, such as whether the word "oltin" (gold) is a material object or a natural resource, were resolved through expert assessment. Using statistical analysis methods, the proportion of parts of speech in each field was calculated and depicted graphically. For instance, it was determined what percentage of nouns belong to the "nature" field and what percentage belongs to the "material objects" field.

For continuous verification of intermediate results and classification obtained during the research process, broad categories were first identified, then subdivided into smaller levels based on semantic differences. As a

result, nouns in the Uzbek language were placed in a semantic network with a multi-level hierarchical classification. This approach is theoretically based on the field theory of Trier and Weisgerber, and practically on the principles of creating lexical databases such as WordNet.

## **RESULT AND DISCUSSION**

Nouns in the Uzbek language are semantically divided into the following large semantic fields:

1. Nouns related to nature - A group encompassing natural and environmental phenomena, elements of animate and inanimate nature. This field includes:

1.1 Landforms and relief: geographical landscape names such as mountain, hill, valley, desert, peninsula.

1.2 Water bodies: sea, river, lake, stream, spring and the like.

1.3 Weather and Climate Phenomena: sun, rain, snow, hail, fog, wind, storm, thunder, etc. (For example, the word "wind" itself has several synonyms according to its strength and characteristics: shabada - light breeze, shamol - ordinary wind, bo'ron - strong storm, to'fon - hurricane, etc. [12]).

1.4 Plant world: all names related to the plant world.

Trees: plane, oak, birch, juniper, willow and others.

Shrubs: rosehip and other shrub plants.

Herbs: grass, camel thorn, poppy (flowering herbs) and others.

Water plants: lotus, algae and other plants that grow in water.

Cultivated plants: wheat, grains, cotton, rice, corn, and similar cultivated crops.

Flowers: tulip, rose etc.

Mosses and riparian plants: moss, reed, sedge and other plants growing in moist areas.

1.5 Animal kingdom: terms related to the animal world.

Wild animals: lion, wolf, fox, bear, tiger and other predators and desert-dwelling wild animals.

Domestic animals: horse, cow, sheep, chicken, dog, cat and other domesticated animals

Birds: eagle, swallow, quail, swan, ostrich etc.

Aquatic animals: fish, whale, dolphin and others.

Insects: bee, grasshopper, butterfly, fly, beetle and others.

Reptiles: snake, turtle, lizard, crocodile and so on.

Imaginary or mythical animals: dragon, Simurgh and similar creatures.

1.6 Celestial and space objects: concepts related to celestial bodies and the universe.

Stars and galaxies: Sun, Star, Venus (Zuhra), Milky Way etc.

Planets and their satellites: Earth, Mars, Moon and others.

Solar system objects: asteroid, comet, meteor, planet and other astronomical terms.

Atmospheric phenomena: aurora, sky and other phenomena.

Cosmic objects: galaxy, universe, black hole, quasar and other astrophysical concepts.

Space and cosmic concepts: orbit, gravity, rocket, space exploration and other terms related to the cosmos.

1.7 Natural resources: minerals and natural wealth - gold, oil, gas, coal, iron, timber, water resources and others. This group also includes names of underground resources and other raw materials.

2. Nouns related to humans and their activities - concepts pertaining to humans, their place in society, activities, relationships, and products.

2.1 Kinship relations: father, mother, older brother, younger brother, older sister, younger sister, paternal aunt, maternal aunt, paternal uncle, maternal uncle, nephew/niece, grandchild, and other kinship terms. Note: The Uzbek language has a very rich system of kinship terms, including words that distinguish between relatives on the father's and mother's sides. This feature is also observed in Turkish (for example, in Turkish "amca" - paternal uncle, "dayı" - maternal uncle), while in English or Russian there is no such distinction (in English "uncle," in Russian "дядя" is used for both) [7].

2.2 Professions and positions: teacher, doctor, engineer, farmer, president, minister, officer, professor, academician and other occupational and official titles.

2.3 Social concepts: people, nation, state, society, family, group, friend, enemy and other nouns denoting

social categories and relationships.

2.4 Personal qualities and traits: intelligence, wit, courage, bravery, cowardice, generosity, stinginess and other terms characterizing a person.

2.5 Human activities and actions: work, war, travel, play, celebration, ceremony, education (in the sense of a process) and other names of actions and processes performed by humans.

2.6 Legal and Political Concepts: terms related to state and law, such as law, right, obligation, freedom, election, court, crime, punishment, constitution, democracy, politics.

2.7 Spirituality and Faith: spiritual concepts and religious terms such as religion, faith, worship, prayer, sin, virtue, spirit, conscience, morality, philosophy.

2.8 Culture and Art: literature, music, theater, cinema, dance, song, image, culture, customs and the like.

2.9 Education and Science: nouns associated with teaching and scientific activities, such as school, university, lesson, science, knowledge, discovery, mathematics, history, biology.

2.10 Sports and Physical Activities: football, wrestling, chess, running, Olympics, champion, gymnastics, athlete and so on.

3. Material Objects and Things - names of all material objects and items created or used by humans. This field has the largest branching and is divided into 19 subgroups:

3.1 Household Items: furniture, table, chair, bed, carpet, sheet, lamp and other items used in the home.

3.2 Clothing and Accessories: clothing and jewelry such as shirt, pants, trousers, skullcap, hat, scarf, ring, bracelet, glasses.

3.3 Vehicles: car, bus, train, airplane, bicycle, ship, cart and the like.

3.4 Technology and Electronics: radio, TV, computer, phone, tablet, refrigerator, washing machine, motor, engine and such.

3.5 Construction and architectural objects: house, building, bridge, road, minaret, wall, edifice, palace, mosque and others.

3.6 Tools and equipment: hammer, saw, shovel, lock, key, grinding wheel, workbench, machine tool.

3.7 Toys and sports equipment: doll, ball, chess set,

bicycle, ice skates, skis, ball, racket and other children's toys and sports items.

3.8 Dishes and kitchen utensils: saucer, knife, fork, spoon, cauldron, basin, teapot, tea bowl, platter, stove, gas stove and so on.

3.9 Writing equipment and books: pen, pencil, paper, notebook, book, encyclopedia, dictionary, atlas, pencil case, eraser and so on.

3.10 Jewelry and ornaments: ring, necklace, bracelet, chain, pendant and similar adornments.

3.11 Artworks and musical instruments: painting, sculpture, violin, doira, piano, dombra, drum and others.

3.12 Military supplies and weapons: weapon, rifle, pistol, sword, spear, armor, tank, shell, bomb, shield and so on.

3.13 Packaging and containers: box, crate, sack, pouch, container, bottle, envelope, barrel, vessel.

3.14 Stationery: stamp, seal, folder, file, glue and other office supplies.

3.15 Horticultural and agricultural equipment: hoe, shovel, plow and other tools used in gardens and fields.

3.15 Medical equipment: syringe, thermometer, stethoscope, ultrasound machine, X-ray machine, bandage, scalpel, microscope and so forth.

3.16 Travel and camping equipment: tent, flashlight, compass, map, suitcase, thermos and others.

3.17 Industrial and manufacturing equipment: machine tool, conveyor, crane, motor, turbine, drilling rig, generator and the like.

3.18 Items for pets: saddle, bridle, horse cart, dog chain, bird cage, aquarium and other objects related to keeping animals.

3.19 Collection and antique items: coins, stamps, souvenirs, artifacts.

This "Material Objects" field is the most comprehensive, and the number of material nouns in the Uzbek language is very large. The analysis showed that about 30-40% of the dictionary belongs to this group. The reason for this is the diversity of objects created by humans during their activities and the abundance of products of cultural life. For example, there are dozens of names in the kitchen utensils group alone. At the same time, there is a hierarchy of

meanings within this group: for example, "weapon" is a broad concept, under which types such as rifle, pistol, sword stand as hyponyms.

4. Abstract concepts - names of concepts and phenomena that do not exist directly in material form. This group includes the following subfields:

4.1 Emotions and experiences: love, hatred, fear, joy, sadness, excitement, trust, enmity and other inner experiences and feelings of a person.

4.2 Concepts related to mind and thinking: thought, mind, memory, imagination, knowledge, perception, concept, experience, fantasy, intention and the like.

4.3 Moral and Spiritual Concepts: Categories related to morality and belief, such as goodness, evil, justice, truth, conscience, faith, sin, virtue, perspective, repentance.

4.4 Concepts of Time and Period: Time units and concepts of periodicity, such as time, era, century, minute, night, day, season, spring, summer, year, epoch, history.

4.5 States and Processes: Nouns denoting dynamic phenomena and states such as sleep, death, illness, peace, war, silence.

4.6 Nouns Denoting Signs and Properties: Abstract nouns denoting the quality or properties of objects and phenomena, such as speed, weight, color, beauty, light, hardness,.

4.5 Legal and Social Concepts: Social order and legal abstract concepts such as legality, freedom, equality, obligation, property, responsibility, rank, reputation, liberty, democracy.

5. Place and Spatial Nouns - Place names and spatial words of various sizes and categories, expressing the concept of space.

5.1 Geographic Locations: Major natural geographic features such as continents, oceans, seas, islands, deserts, mountains.

5.2 Territorial Units: Administrative-territorial units such as country, state, province, district, city, village, neighborhood.

5.3 Buildings and Structures: Human-built places and structures, such as houses, schools, hospitals, mosques, bridges, parks, stadiums, museums.

5.4 Rooms and Interior Spaces: Indoor spaces such as

rooms, halls, kitchens, bedrooms, bathrooms, corridors, attics.

5.5 Natural landscapes: Place names associated with nature, such as garden, forest, field, pasture, park, and reserve.

5.6 Road and transport infrastructure: street, boulevard, highway, motorway, railway, station, terminal, airport, port and the like.

5.7 State and administrative units: government, parliament

5.8 Institutions and organizations: university, factory, bank, company, theater, library, market, shop and other place-related organization names.

5.9 Religious and cultural sites: cemetery, shrine, mausoleum, theater, stadium (entertainment venues), gallery, temple, pilgrimage sites.

5.10 Tourist and historical places: Samarkand, Khiva, Registan, Kuhna Ark, Kyzylkum and other names of historical or notable places (as common nouns or toponyms).

6. Food and Dishes - names of all products and dishes related to human nutrition. Its designation as a separate semantic field has cultural significance, as the Uzbek language has a rich vocabulary of national dishes and food products.

6.1 Main dishes: plov, manti, shorva, norin, lagman, shashlik, kasan kebab and other national dishes.

6.2 Drinks: tea, coffee, juice, kvass, cola, water, milk, yogurt, wine, beer and other liquid beverages.

6.3 Fruits and vegetables: apple, grape, melon, watermelon, cucumber, tomato, potato, carrot, cherry, peach and so on.

6.4 Meat and dairy products: beef, mutton, fish, sausage, cheese, cream, butter, qurut (dried yogurt balls), eggs and so forth.

6.5 Bread and flour products: bread, flatbread, pasta, samsa, fried dough balls, small round bread, dough, flour, groats.

6.6 Sweets and desserts: halva, sugar, honey, cake, parvarda (a type of sweet), chocolate, etc.

6.7 Spices and additives: salt, pepper, cumin, cinnamon, mint, ginger, vinegar, mayonnaise, tomato paste and similar food additives.

The above classification forms a hierarchy of common

semantic fields. In total, nouns in the Uzbek language were divided into 6 macro-fields and 60 micro-fields. Each micro-field can be further divided into smaller groups. For example, if we take the micro-field of "birds" in the macro-field of the animal world, it can be divided into informal groups such as predatory birds (falcon, eagle, vulture), domestic birds (chicken, rooster, goose, duck), and ornamental birds (parrot, nightingale - kept by humans). However, our classification does not include such detailed subdivisions and is limited to general categories.

When examining 10,000 nouns from the corpus for each semantic field, it was found that the group of material objects contains the largest number of words (approximately 35%). The following places are occupied by the fields of nature (20%) and human activity (18%). The group of place names and abstract concepts is around 10%. The food group had the smallest share (7%) due to its relatively limited vocabulary. Of course, these numbers are relative, because although there are many abstract concepts in the dictionary, many of them can be expressed in other parts of speech, such as adjectives and verbs. Since we only considered the noun category, the proportion of the abstract field turned out to be lower than expected.

The results show that the vocabulary of the Uzbek language can be divided into broad categories in terms of content, and the boundaries between these categories are often relative. There were cases where some words belonged to several fields. For example, the word "gold" can be both a natural resource (underground wealth) and a jewelry item (jewelry material). In such cases, the main semantic range of the word was taken into account. "Gold" was included in the field of material objects because it was used more often as jewelry and metal, while in the field of natural resources, terms such as gold ore remained.

Another example is the word "qalam" (pen/pencil). This word belongs to the group of writing instruments. However, historically, since the qalam was a product derived from a plant (reed pen), it can sometimes be associated with the plant world. Nevertheless, modern language considers the qalam as a concrete object, therefore it is included in the field of material objects.

In general, when creating a classification, it should be understood that the boundaries of semantic fields are

variable, and shifts between fields can be observed through different meanings of a word. Our classification is based on basic, prototypical meanings. In this case, the most fundamental conceptual scope of each noun has been selected.

## **CONCLUSIONS**

As a result of the conducted research, nouns in the Uzbek language were comprehensively classified according to their semantic domains. A total of 6 major semantic fields were identified (Nature, Human, Material objects, Abstract concepts, Place-space, Food), which were further divided into smaller microfields. This classification allows for a systematic examination of the Uzbek language vocabulary from an ideographic perspective. As each semantic field represents a specific lexical set and their interrelationships, semantic connections and differences between words were identified.

The groups formed for the purpose of this study were compared with classifications from international experience. The comparison revealed that large categories of semantic fields are generally universal. Many languages have similar conceptual categories. However, the internal structure and lexical richness of each field are unique to each language and culture. In the Uzbek language, some areas are expressed in great detail, while other fields are lexically expressed mainly through loanwords. These features reflect the history of our language's vocabulary formation and cultural influences.

The results of this work are significant not only for theoretical linguistics but also for practical applications. In particular, this classification can serve as a fundamental basis for creating a semantic thesaurus of the Uzbek language. Presenting words in synonym groups is helpful for translators in choosing alternative words and for writers in enhancing the richness of their language. In computational linguistics, such semantic categorization can be useful for tasks like automatic text analysis, semantic disambiguation, and semantic data indexing. Currently, initial steps are being taken to create the Uzbek equivalent of the English WordNet - UzWordNet, where the semantic groups we have identified and their correspondences in other languages will certainly be valuable. The study recommends developing the following areas in the

future:

- Expansion of semantic fields to include other parts of speech;
- Creation and application of automatic semantic tagging algorithms;
- Development of multilingual ideographic dictionaries.

In conclusion, it can be said that the description of nouns in the Uzbek language, grouped into semantic fields, more clearly demonstrated the lexical systematic nature of the language. It was once again confirmed that each word is not isolated, but rather a part of a large system connected by meaning. Comparative analysis of differences and similarities between semantic fields helped to determine the place of our language in the general structure. The given classification, of course, is not the final word. It is an open system that can be supplemented with new data and observations in the future. Importantly, this research has created a solid scientific basis for understanding and organizing the rich and diverse vocabulary of the Uzbek language.

#### **REFERENCES**

1. Karaulov Yu. N. General and Russian Ideography. - Moscow: Nauka, 1976. - 240 p.
2. Hallig R., von Wartburg W. Begriffssystem als Grundlage für die Lexikographie (das Weltbild der Sprache). - Bonn, 1952.
3. Ciaramita M., Johnson M. Supersense Tagging of Unknown Nouns in WordNet // Proceedings of the Conference on Empirical Methods in Natural Language Processing (EMNLP) . - 2003. - P. 168-175.
4. Fazlyeva Z. The Semantic Field Theory as Applied to Phraseology // Mediterranean Journal of Social Sciences. - 2015. - Vol. 6, No. 3 S2. - P. 408-414.
5. Kholmanova Z. T. Semantic field and semantic network capabilities // Proceedings of the International Scientific-Practical Conference "Computer Linguistics: Problems, Solutions, Prospects". - Tashkent, 2022. - Vol. 1, pp. 37-43.
6. Sobirov A. Researching the lexical level of the Uzbek language based on the principle of "system of systems." - Andijan: Andijan State University named after Zahiriddin Muhammad Babur, 2019. - 320 p.
7. Rahmatullayev Sh. Studying Uzbek lexicon as a semantic field (personal microfield). Abstract of the doctoral dissertation in Philological Sciences. - Tashkent, 1990. - 45 p.
8. Safarova R. Hyponymy in the Uzbek language (based on actively used zoonyms). - Candidate's dissertation in Philological Sciences. - Tashkent, 2005.
9. Ne'matova G. Lexicon of plant names in the Uzbek language: system and artistic use. - Doctoral dissertation in Philological Sciences. - Tashkent, 2010.
10. Meliev K. Names of fruit trees in Turkic languages and lexical units derived from them // Research on the lexicon of Turkic languages of Central Asia and Kazakhstan. - Tashkent: Fan, 1990. - pp. 100-175.
11. Saporov M. Names of melon crops in Turkic languages // Research on the lexicon of Turkic languages of Central Asia and Kazakhstan. - Tashkent: Fan, 1990. - pp. 176-250.
12. Rustamov A. A word about the word. - Tashkent: Fan, 1970. - 112 p.
13. Dalaboyev Kh. From the history of astronomical terms // Uzbek terminology and prospects for its development. - Tashkent: Fan, 1986. - pp. 94-96.