

ENHANCING THEMATIC GROUPING IN ENGLISH AND UZBEK IDEOGRAPHIC DICTIONARIES

Bekmurodova Maftuna Jumanazar kizi

Teacher at the Department of English Teaching Methodology

№2, Uzbekistan State World Language University

e-mail: maftunabekmurodova52@gmail.com

+998931041993

Introduction

Thematic ideographic dictionaries organize lexical items by meaning rather than alphabet. Unlike alphabetical dictionaries, ideographic dictionaries “organize words and expressions into thematic groups based on semantic proximity” This thematic organization helps learners and users uncover the semantic relationships among words and understand vocabulary in context we emphasize that ideographic dictionaries reveal internal semantic connections, showing “how lexical units are interconnected on a conceptual and functional level”. Such dictionaries also serve important pedagogical and cognitive functions, providing translators and language learners with intuitive conceptual maps of word meaning. In short, thematic grouping can accelerate vocabulary acquisition, support semantic analysis, and highlight conceptual fields (e.g. human activities, emotion, nature, social life) in a language.

Despite these benefits, thematic classification faces challenges. Identifying clear semantic domains is not always straightforward. Conceptual boundaries between themes can be fuzzy, and words often belong to multiple fields. Alisherova and Sodikova note that lexicographers must balance semantic precision and cultural relevance, as thematic grouping requires reliable criteria for word inclusion. In practice, ensuring semantic accuracy - correctly assigning words to themes and cultural fit - capturing culture-specific concepts can be difficult. For example, some domain names or idiomatic concepts may not map neatly across languages. Maintaining coherence in a thematic taxonomy often demands expert judgment, extensive lexicographic analysis, and sometimes compromise or broad contextual definitions.

These structural differences reflect linguistic and cultural factors. For instance, Alisherova and Sodikova found that English thematic dictionaries typically employ a hierarchical classification system, while Uzbek dictionaries use a contextual grouping approach due to less standardized terminology and evolving technical vocabularies. English has a wealth of idiomatic expressions with no direct Uzbek equivalents, thus Uzbek entries often include extended definitions or explanatory notes to bridge gaps. Conversely, Uzbek culture-specific terms like *mahalla* “traditional neighborhood” or *osh* “communal rice dish” may lack precise English counterparts. These examples highlight how translation challenges arise: culture-bound concepts resist simple mapping between languages. Alisherova and Sodikova list “difficulties in translating culture-bound concepts” and the “lack of standardized Uzbek lexicographic resources” among the primary challenges of bilingual thematic dictionary compilation.

Linguistic structure also impacts thematic grouping. English is an analytic language with relatively fixed word order, whereas Uzbek is agglutinative with rich morphological formation. Uzbek's suffix-based derivations allow flexible word formation, which means Uzbek lexemes can be more granular different affixed forms for related ideas. As a result, thematic categories in Uzbek may require multiple entry forms or cross-references to cover affixal variants. In practice, Uzbek thematic dictionaries often include synonyms, antonyms, and extended phrase explanations to ensure comprehensive coverage (e.g. providing both *tadbirkor* and *biznes yuritish* for “entrepreneur” concepts).

Despite these challenges, English and Uzbek thematic systems share similarities. Both languages can organize fundamental semantic fields such as emotion, family, nature, and social relations. Sadikova notes that both English and Uzbek “categorize words based on shared meanings or themes,” for example clustering terms for family or emotions. Moreover, certain universal human domains (e.g. happiness, sadness, anger) are recognized across languages. However, cultural nuances differ: for instance, Uzbek has a richer set of kinship terms (e.g. *ota* “father” vs *bobo* “grandfather”) reflecting the importance of extended family, while English has comparatively broader terms with fewer relational distinctions. Cultural priorities also shape vocabulary: English lexical fields often emphasize individual achievement and concepts like *success*, whereas Uzbek may emphasize collective identity and hospitality in its thematic fields. These differences underscore that thematic dictionaries must be attuned to social values and cognitive patterns of each culture.

To improve thematic grouping quality, lexicographers can leverage modern theoretical and technological tools. A semantic-field approach provides a theoretical foundation: lexical items are grouped into fields of related meaning rather than by surface form as Sadikova explains, semantic fields consist of “groups of words that share a common theme or concept,” enabling a nuanced mapping of meanings. By systematically applying semantic-field theory, compilers can define clear macro- and micro-topics (e.g. broad fields like emotion subdivided into joy, sadness, etc.) and ensure coverage of paradigmatic relations. In practice, thematic blocks or domains (e.g. *human body*, *nature*, *social life*) serve as organizing units, and semantic subfields (synonym and antonym clusters) structure each block. Thus, explicitly incorporating semantic-field theory can guide consistent, conceptually meaningful categorization.

Corpus linguistics offers practical support. Analyzing large text corpora of each language helps identify natural word groupings and term usage. Lexicographers can use corpora to gather relevant terms for each theme and to see how words co-occur in real contexts. For example, Alisherova and Sodikova recommend using corpora to enhance term selection: “collections of written or spoken texts... provide insights into real-world language usage”. Corpus analysis can reveal common collocations and topical clusters, helping to define the boundaries of a semantic field. Corpus-derived examples can also illustrate word usage within themes, making definitions clearer. In fact, the study by Alisherova and Sodikova found that newer digital thematic dictionaries increasingly employ corpus-based analysis and AI tools to improve consistency and coverage.

Artificial intelligence and distributional semantics further expand possibilities. Modern NLP techniques (such as word embeddings) can automatically discover semantic groupings. For instance, Alvarado demonstrates that vector-space embeddings (e.g. Word2Vec, FastText) can cluster words into coherent semantic domains based on usage patterns. This data-driven clustering offers a “usage-based classification” that complements expert-designed fields. In other words, embedding models

learn semantic relations from corpora and can propose fields or subfields without manual predefinition. Such AI-supported lexicon structuring could help generate candidate thematic clusters, refine existing categories, and even reveal underrepresented areas. In practice, combining AI clustering with expert review yields more robust classification.

Moreover, existing semantic databases can be leveraged. For English, WordNet already provides a rich semantic network connecting words by synsets and relations. Uzbek is developing similar resources. Integrating these lexical-semantic databases ensures that thematic groupings align with recognized synonym/antonym relations. We mention UzWordNet among emerging tools in Uzbek lexicography. In summary, applying corpus methods and AI along with semantic theories can greatly enhance thematic dictionary design by making it systematic, data-informed, and sensitive to usage.

Practical Methodologies for Refining Thematic Classification

In practical terms, refining thematic grouping involves a structured compilation process. Lexicographers typically start by selecting core domains (macro-themes) such as physical world, social life, thought and emotion, etc. According to us, an ideographic dictionary often progresses “from large semantic clusters macro-themes to smaller thematic subdivisions (micro-themes)”. Within each macro-theme, terms are grouped into semantic fields (synonym sets, antonyms, hypernyms) and cognitive networks cause-effect, actions. For example, in a “nature” domain one might have micro-themes like *weather*, *plants*, *animals*, each containing relevant lexemes. Such hierarchical organization makes the dictionary both comprehensive and user-friendly.

Refinement also relies on semantic relations. Bkmurodova and Maxkamova detail core components: thematic blocks - domains, semantic fields - related word groups, and cognitive structures - relations like hypernymy. For instance, under a “movement” topic one could cluster synonyms (run, jog, sprint), arrange hyponyms/hypernyms (animal → bird → crow), and link antonyms (e.g. *freeze* vs *melt* in the context of temperature changes). Using these relations ensures that words are classified by meaning relations rather than arbitrary labels. In practice, lexicographers often build or draw on thesauri and semantic networks like WordNet or multilingual concept lists to identify such relations.

Empirical methods complement these semantic strategies. As noted, large corpora and digital tools are essential. The compilation process can be augmented by automated queries and concordancing. For each theme, lexicographers can run corpus searches to find frequent collocates and expressions related to the domain. These corpus findings help confirm which words naturally belong together. Alisherova and Sodikova specifically recommend using digital tools and corpora to systematize term collection. For example, a concordance for “hunger” could reveal related verbs and adjectives, suggesting items for a *food* or *nutrition* category. Similarly, lexical databases and NLP software can cluster terms via semantic similarity or topic modeling, providing suggestions for category membership.

Another practical method is involving user feedback. Surveys or usability studies often done in pedagogical lexicography can indicate which groupings make sense to learners. For instance, testing thematic groupings on language students can reveal if certain words are expected to co-occur or if some categories are confusing. Alisherova and Sodikova mention that user-oriented classification and clear organization of entries aid learners. Although not explicit in their study, best practice suggests

iteratively refining themes: initial hypotheses of semantic fields are adjusted based on corpus evidence and expert evaluation.

Finally, accuracy in definitions is crucial. Once words are grouped, each entry must explain the meaning in context. Definitions should be concise yet cover the thematic sense; where meanings overlap categories, cross-references may be provided. Alisherova and Sodikova (2025) note that precise definitions and illustrative examples help users grasp thematic usage. In theme-based dictionaries, example sentences drawn from corpora are particularly valuable, as they show the word in action within its semantic domain. By combining structured classification (thematic blocks, fields, relations) with corpus-verified usage examples and clear definitions, lexicographers can refine thematic classifications that are accurate, consistent, and pedagogically effective.

Conclusion

Thematic grouping is a powerful approach in lexicography that highlights semantic relationships and supports language learning. English lexicography has long harnessed this method (as in Roget's Thesaurus or WordNet), while Uzbek lexicography is developing similar tools. Both traditions face analogous challenges: organizing vocabulary into coherent semantic fields and accommodating cultural nuances. Comparative studies show that English tends toward systematic hierarchical schemes, whereas Uzbek approaches often emphasize contextual grouping to bridge terminology gaps.

To enhance thematic dictionaries in both languages, linguists should apply semantic-field theory to define robust domains, and adopt corpus-based and AI techniques to ground classifications in actual usage. Modern resources corpora, WordNet-like databases, embedding models can uncover patterns of semantic similarity and suggest optimal groupings. Practical methodology involves starting with broad topics, then iteratively refining categories through semantic relations (synonymy, hyponymy) and usage evidence. By blending theoretical insights, empirical data, and digital tools, lexicographers can improve thematic accuracy and usability. Such enriched thematic dictionaries will better reflect cultural contexts and aid learners of English and Uzbek in understanding the conceptual tapestry of each language.

References

1. Alvarado, J. (2023). *Embedding based multilingual atlas of semantic fields*. CEUR Workshop Proceedings, 3516, 1–6.
2. Alisherova, S. A., & Sodikova, S. A. (2025). The lexicographic principles for creating thematic dictionaries in English and Uzbek languages. *International Journal of Science and Technology*, 2(5), 133–135.
3. Bekturodova, M. J., & Maxkamova, S. M. (2025). Structure of ideographic dictionaries. *Luchshie Intellektualnye Issledovaniya*, 43(2), 476–494.
4. Sadikova, S. A. (2024). Semantic fields of ideographic dictionaries in English and Uzbek. *Periodica Journal of Modern Philosophy, Social Sciences and Humanities*, 37, 44–46.