

 Research Article

Reflection Issues of Geographical Terms in The Uzbek Language Educational Corpus

Submission Date: April 12, 2025, **Accepted Date:** May 08, 2025,

Published Date: June 10, 2025

Crossref doi: <https://doi.org/10.37547/ijasr-05-06-03>

Journal Website:
<http://sciencebring.com/index.php/ijasr>

Copyright: Original content from this work may be used under the terms of the creative commons attributes 4.0 licence.

Ikrom Xushboqovich Islomov
Renaissance Education University, Uzbekistan

ABSTRACT

The article discusses the issue of designing a database on the semantic interpretation of geographical terms, which occupies a special place in the educational corpus, in particular, the educational corpus of the Uzbek language and its related database. The information on the structure, composition, corpus material, search potentials of the educational corpus are also provided. Recommendations for semantic tagging of geographical terms, creation of their database, selection of lexicographic ground have been developed for Uzbek language educational corps.

KEYWORDS

Educational corpus, educational corpus of the Uzbek language, geographical terms, semantic tagging, linguistic support, concordance, educational dictionary, electronic dictionary.

INTRODUCTION

Although the types of corpora have not been extensively studied in Uzbek computational linguistics, some monographic studies on specific linguistic corpora have already emerged. The issue of developing linguistic support for Uzbek language

corpora is also attracting the attention of specialists. Among these studies, the dissertation by U. Kholiyorov entitled “Linguistic Foundations of Creating an Educational Corpus of the Uzbek Language” is particularly significant in the context of our research.

The researcher defines an educational corpus as follows:

“An educational corpus is a language corpus with a linguo-didactic nature, whose materials are aimed at language teaching. The educational corpus of the Uzbek language is a corpus in Uzbek, designed to teach the possibilities of the Uzbek language, which includes electronic texts with a linguo-didactic function and operates as a dedicated website. It is a specific language corpus that consists of a large volume of texts, a simple/advanced search system, and functions for searching a particular unit within texts and Uzbek language learner dictionaries” [2:13].

From this definition, it is clear that such a corpus differs from a national corpus due to its linguo-didactic character. Additionally, its search system incorporates Uzbek language learner dictionaries.

The educational corpus differs from other corpora in its interface, texts, and lexicographic products. As emphasized by U. Kholiyorov, the main goal of the educational corpus is to present language material in accordance with the age and worldview of the learner. Currently, a team of specialists from the Tashkent State University of Uzbek Language and Literature is working on a practical project titled “Creation of an Educational Corpus of the Uzbek Language”. As a result of this project, the Educational Corpus of the Uzbek Language has been launched on a special website [6].

This corpus consists of two main blocks:

1. Concordance search for words in the Uzbek language.

2. Electronic library of Uzbek language learner dictionaries.

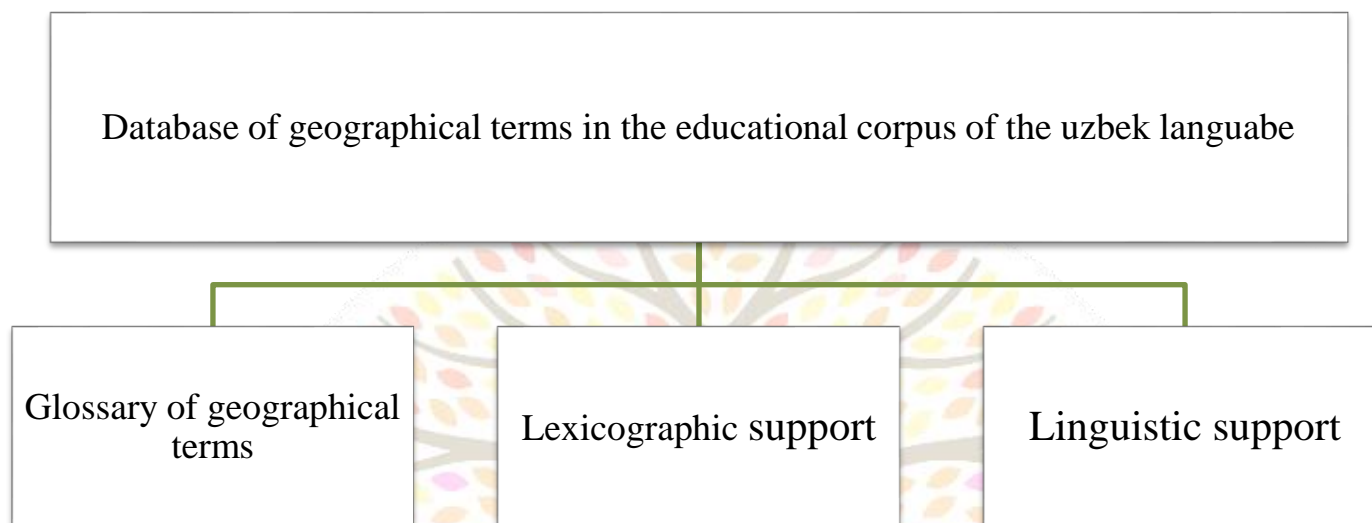
Both blocks can be used as corpus material. The searches are interconnected. This educational corpus provides the possibility to search for words, word forms, and bigrams, and has the functionality to create their concordances. It also features a function that allows searching for a word across various (existing in the corpus database) dictionaries. However, this corpus lacks a semantic search function — a problem that has not yet been fully resolved even in global corpus linguistics. In order to implement semantic search in such corpora, a semantically annotated base of language units or a semantic annotation tool is required.

In this context, we will focus on the semantic tagging of geographic terms in Uzbek educational corpora, the creation of a database for these terms, and the selection of appropriate lexicographic products. To achieve this, two main tasks can be outlined:

1. Creation of a database of geographic terms (its structure, lexicographic support, search capabilities).

2. Development of a search system for geographic terms in the educational corpus (issues of tagging units, establishing hyperlinks, and linking dictionary data).

When constructing the database, it is necessary to first define its architecture, structure, and content. The structure of the geographic terms database for the Uzbek educational corpus is presented in the diagram (See: Figure 1).



1-Figure. The structure of the geographic terms database in the Educational Corpus of the Uzbek Language

The geographic terms database of the Educational Corpus of the Uzbek Language is formed based on various sources such as the Explanatory Dictionary of the Uzbek Language, Children’s Encyclopedia, National Encyclopedia, Explanatory Dictionary of Uzbek Toponyms, Educational Dictionary of Geographic Terms, Russian-Uzbek Dictionary of Geographic Terms, and other reference materials.

The entries included in the attached database are selectively compiled from the sources listed above. Words from the Explanatory Dictionary of the Uzbek Language (EDUL) serve as the primary basis. However, if a term is not available in EDUL but is found in other sources, it is also included.

The lexicographic support of the geographic terms database for the Educational Corpus of the Uzbek Language serves as the foundation for entering data into the database.

The linguistic support of the geographic terms database consists of a set of rules for providing definitions of words within the data repository.

The database developed for the Educational Corpus of the Uzbek Language will be a part of the general database created for the National Corpus. The required data can be extracted from it via search queries.

The Explanatory Dictionary and the Explanatory Dictionary of Terms differ in the way they organize

the headword (dictionary entry) and the structure of the dictionary article. The differences between these types are presented in the table below:

Lexicographic source	Headword presentation	Structure of a dictionary article
<i>Explanatory dictionary</i>	Lexeme, fixed expression, free phrase, compound word	Heading word, semantics and explanations, original and borrowed meaning, lexicographic sign and explanation, index, reference, illustrative example
Terminological dictionary	Simple term (consisting of one word), compound term (consisting of two or more words)	Heading word and its explanation

Table 1. Comparative Table of Lexicographic Sources

As lexicographic support, P.N. G'ulomov's "Explanatory Dictionary of School Geography Terms and Concepts" was selected. This decision is supported by the following statement from the preface of the dictionary [3: 5]:

"In preparing the dictionary, the main criterion was to select terms and concepts presented in middle school curricula, textbooks, and educational maps.

In writing the explanatory texts for the terms and concepts, previously published dictionaries, encyclopedic dictionaries, and textbook definitions were used. Although the dictionary is intended as an explanatory dictionary of school geography terms and concepts, it also includes local geographical terms not covered in school curricula. It is necessary for our students to be familiar with

such local geographical terminology and concepts. When providing explanations for these local terms and concepts, the author also drew on information collected during scientific expeditions.”

These characteristics made the dictionary a suitable choice for lexicographic support. According to the dictionary’s annotation [3: 6], it contains terms and concepts related to:

physical geography, landscape studies, geomorphology, climatology, meteorology, hydrology, soil geography, biogeography, glaciology, economic geography, population and urban geography, industry, agriculture, transportation geography, and political geography.

The structure of dictionary entries differs from that of an explanatory dictionary. In the process of filling the database and assigning semantic tags, part of the information must be selected manually by specialists, since the dictionary mainly includes only headwords and their brief explanations. Therefore, information taken from this source is subject to manual processing before being added to the database.

To create a terminological semantic database, three components are required:

1. A list of terms,
2. Semantic explanations,
3. Semantic tags.

Words included in the lexicon of the Uzbek Geographic Terms Database are selected from the Explanatory Dictionary of the Uzbek Language [4],

the Russian-Uzbek Dictionary of Geographical Terms [5], and various encyclopedias. For adding geographic terms to the Educational Corpus of the Uzbek Language, the Children’s Encyclopedia [1] also serves, to some extent, as lexicographic support. A significant portion of the entries in this dictionary are proper nouns. Additionally, this dictionary is of an encyclopedic nature, covering information from various fields.

When providing definitions of geographic terms, the principle of combining data from a single source with supplementary data from others is followed.

The information included in the database consists of the following parameters:

1. Term
2. Explanation(s)
3. Marker indicating it is a geographic term
4. Semantic tag
5. Etymology
6. Category/type to which the term belongs

In general, building the lexicographic support for the geographic terms database within the Educational Corpus of the Uzbek Language involves:

- Adapting lexicographic explanations from domain-specific sources to suit the age characteristics of learners, and

- Complying with linguodidactic requirements, which are among the most important tasks in this process.

REFERENCES

1. Болалар энциклопедияси. – Toshkent: «O‘zbekiston milliy ensiklopediyasi» Davlat ilmiy nashriyoti, 2000. – 664 б.
2. Холиёров Ў. Ўзбек тили таълимий корпусини тузишнинг лингвистик асослари: Филол. фан. бўйича фалсафа доктори (PhD) дисс. – Термиз, 2021. – 147 б.
3. Ғуломов П.Н. Мактаб жўғрофия атамалари ва тушунчалари изоҳли луғати. – Тошкент: Ўқитувчи, 1994. – 143 б.
4. Ўзбек тилининг изоҳли луғати. V жилдли. 1-5-жилдлар. – Тошкент: “Ўзбекистон миллий энциклопедияси” Давлат илмий нашриёти, 2006-2008.
5. G‘ulomov P.N. Geografiyadan qisqacha ruscha-o‘zbekcha terminlar va tushunchalar lug‘ati. – Toshkent: “O‘zbekiston milliy ensiklopediyasi” Davlat ilmiy nashriyoti, 2013. – 80 b.
6. <http://uzschoolcorpara.uz/>

