

RESEARCH ARTICLE OPEN ACCESS

The Principles Of Maximum Inclusivity In Lexicographic Description Of Sociolinguistic Variation

Nasirov Davranbek Palvannazirovich

Senior Lecturer at Faculty of Philology of Alfraganus University, Uzbekistan

Received: 30 September 2025 Accepted: 23 October 2025 Published: 28 November 2025

ABSTRACT

The article examines the principles of maximum inclusivity in lexicographic description of sociolinguistic variation based on Uzbek and German language materials. A comparative analysis of 8903 sociolinguistic variants in the five-volume explanatory dictionary of the Uzbek language edited by A. Madvaliev and 72452 variants in the German universal Duden dictionary was conducted. The main sociolinguistic variables were identified: emotional-stylistic marking, functional-stylistic coloring, frequency of use, territorial and social limitations. It was established that professional variants predominate in Uzbek (34.68%), while dialectal differences dominate in German (49.24%). An improved system of lexicographic description principles was developed, including principles of maximum inclusivity, authentic recording, semantic equality, and user orientation. Recommendations for integrating corpus methods and artificial intelligence technologies into modern lexicographic practice are proposed.

Keywords: Sociolinguistic variation, lexicography, inclusivity, explanatory dictionaries, Uzbek language, German language, linguistic variants.

Introduction

Modern lexicography is undergoing a period of fundamental transformation driven by the need to adequately reflect linguistic diversity under conditions of globalization and intensified intercultural contacts. Traditional approaches to dictionary compilation, focused predominantly on the literary norm, are increasingly criticized for excluding or inadequately representing language varieties used by different social groups. In this context, the problem of developing principles of maximum inclusivity in the lexicographic description of sociolinguistic variation becomes particularly relevant.

Relevance of the study. The relevance of this study is determined by several key factors. First, the growing awareness of the social responsibility of lexicography as a discipline that shapes ideas about language norms and cultural values. Second, the revolutionary changes in the technological possibilities for collecting and analysing

linguistic data associated with the development of corpus linguistics and artificial intelligence. Third, the critical need to overcome discriminatory practices in lexicography, which for centuries have excluded the voices of marginalized communities from academic discourse.

Comparative study of the principles of lexicographic description in structurally different languages such as Uzbek and German, which represent different typological families and cultural-historical traditions, is of particular importance. The Uzbek language, belonging to the Turkic group and undergoing intensive change under the language policy of independent Uzbekistan, demonstrates specific forms of sociolinguistic variation. The German language, with its rich dialect tradition and well-developed system of functional styles, represents an alternative model for reflecting linguistic diversity in lexicographic practice.

Scientific novelty of the study. The scientific novelty of the

present research lies in a comprehensive analysis of principles of inclusivity in lexicography based on empirical material from two structurally different languages. For the first time, a quantitative analysis of sociolinguistic variables in explanatory dictionaries of Uzbek and German is carried out using statistical methods. Novelty is also found in the development of an improved system of principles for lexicographic description that integrates the achievements of modern sociolinguistics, corpus linguistics and digital technologies.

Object and subject of the study. The object of the study is the representation of sociolinguistic variation in explanatory dictionaries of the Uzbek and German languages. The subject of the study is the principles of lexicographic description of sociolinguistic variation and their implementation in specific lexicographic sources.

The main sources of empirical material were the five-volume “O‘zbek tilining izohli lug‘ati” (Explanatory Dictionary of the Uzbek Language) edited by A. Madvaliev (2006–2008) and “Deutsches Universalwörterbuch: Das umfassende Bedeutungswörterbuch der deutschen Gegenwartssprache” published by Duden (2019). The choice of these sources is due to their representativeness and authority within their respective linguistic traditions.

The aim of the study is to establish principles of maximum inclusivity in the lexicographic description of sociolinguistic variation and to identify their specific features in the Uzbek and German languages.

To achieve this aim, the following objectives are pursued:

1. Systematise the theoretical foundations of lexicographic interpretation of sociolinguistic variation;
2. Identify and classify the main sociolinguistic variables in the languages under study;
3. Carry out a comparative analysis of the principles of lexicographic description in explanatory dictionaries of the Uzbek and German languages;
4. Develop an improved system of principles of maximum inclusivity for contemporary lexicographic practice.

Theoretical foundations of the study. The conceptual

foundations for the study of sociolinguistic variation were laid in the pioneering works of William Labov, who in his study “Language in the Inner City” (Labov, 1972) demonstrated the systemic nature of linguistic variation depending on social factors. Labov established that variants of linguistic units are not random deviations from the norm but constitute an ordered system that correlates with the social stratification of society.

The development of sociolinguistic theory in the works of J. K. Chambers and P. Trudgill expanded the understanding of factors influencing linguistic variation. In their fundamental work “Dialectology” (Chambers & Trudgill, 1998), the authors systematised methods for analysing dialect differences and their social distribution. Of particular importance for the development of the theory of variation is the concept of the “sociolinguistic variable”, which is defined as a linguistic unit that varies depending on the social characteristics of speakers.

Within the framework of the theory of sociolinguistic variation, two main types of variation are distinguished: stratification and situational variation. Stratification variation reflects stable differences in the speech of representatives of different social groups and is associated with such factors as socio-economic status, education, age, gender, and ethnic affiliation. Situational variation is characterised by changes in linguistic behaviour depending on the communicative situation, including the formality of the setting, the type of discourse, and the relationships between the participants in communication.

Traditional lexicography for a long time was oriented toward a prescriptive model aimed at the codification and standardisation of language. Classical explanatory dictionaries, beginning with Samuel Johnson’s Dictionary of the English Language (1755), established language standards and often excluded variants considered incorrect or inappropriate. This approach reflected the social hierarchies and cultural prejudices of an era in which access to education and cultural capital was limited to privileged strata of society.

The shift to descriptive lexicography in the twentieth century marked a fundamental change in the understanding of the tasks of lexicographic work. The descriptive approach, theoretically grounded in the works of structuralists, presupposes the description of language “as it is” rather than “as it ought to be.” However, the practical implementation of descriptive principles faces objective

constraints: the necessity of material selection, problems of source representativeness, and the cultural stereotypes of lexicographers.

Contemporary critical lexicography, developing under the influence of postcolonial studies and feminist theory, questions the neutrality of traditional lexicographic practices. As Russell (2012) notes in his analysis of feminist dictionaries, “dictionaries do not simply reflect language, but actively construct representations of reality by including and excluding particular voices and experiences” (Russell, 2012, p. 15).

Principles of inclusive lexicography. The concept of inclusive lexicography is formed as a response to criticism of traditional approaches for excluding marginalised groups. Inclusivity in the lexicographic context means the aspiration to provide the most complete and fair representation of linguistic diversity without discrimination on social, cultural, or ideological grounds.

Key principles of inclusive lexicography include: The principle of equal representation presupposes proportional representation of different sociolinguistic groups in the dictionary. This does not imply mechanical equality in the number of dictionary entries, but requires taking into account the actual distribution of linguistic variants in the speech practice of the community.

The principle of neutrality of description is aimed at avoiding value judgements and stigmatising characteristics when presenting linguistic variants. Instead of labels such as “incorrect” or “distorted,” descriptive markers are used that indicate the sphere of usage and the social conditions of use.

The principle of multiple sources requires drawing on data from various sociolinguistic communities and discursive practices. This is especially important for overcoming the traditional dominance of written sources and for including the oral speech of different social groups.

The principle of community participation presupposes the involvement of representatives of different language communities in the process of dictionary creation. Crowdsourcing platforms and collaborative projects such as Wiktionary demonstrate the potential for democratising the lexicographic process.

The methodological basis of the study is a comprehensive

approach that integrates quantitative and qualitative methods of analysis. The main research methods are:

Content analysis of dictionary entries, aimed at identifying the ways in which sociolinguistic information is presented in the dictionaries under study. The structure of dictionary entries, the system of usage labels and comments, and the methods of illustrating the use of linguistic variants are analysed.

Statistical analysis is used to quantitatively assess the representation of different types of sociolinguistic variants. The frequency of various types of labels is counted, as well as the distribution of variants across sociolinguistic categories and their comparative representation in the two languages studied.

The comparative method is used to identify common and specific features of the lexicographic description of sociolinguistic variation in Uzbek and German. Comparison is carried out at the level of principles of material selection, classification systems and methods of presenting information.

Discourse analysis of lexicographic texts is aimed at revealing hidden ideological attitudes and cultural biases in dictionary descriptions. The metalanguage of lexicographic descriptions and the ways of constructing social identities through linguistic characteristics are analysed.

The empirical base of the study consists of data obtained by continuous sampling from explanatory dictionaries of the Uzbek and German languages. In the five-volume “O‘zbek tilining izohli lug‘ati” edited by A. Madvaliev, 8,903 dictionary units with sociolinguistic labels and comments were identified. In the “Deutsches Universalwörterbuch” published by Duden, 72,452 units with sociolinguistic marking were recorded.

The significant quantitative difference in the volume of material is explained by differences in lexicographic traditions and principles of material selection. The German lexicographic tradition is characterised by a more detailed consideration of dialectal and regional variants, which is due to the historical features of the formation of the German standard language and the federal structure of German-speaking countries.

On the basis of the analysis of empirical material, the

following main types of sociolinguistic variables were distinguished:

1. Emotional-stylistic marking. This category includes variants with emotional or expressive colouring. In Uzbek, 246 units were identified (2.76% of the total), and in German – 6,374 units (8.8%). Examples from Uzbek: онажон (affectionate), болакай (diminutive-affectionate), яшшамагур (abusive), ваъдабоз (pejorative). German examples: abschleppen (colloquial, humorous), Kuckuck (iron.), glatt wie ein Aal sein (disparaging).

The predominance of negatively marked units in both languages reflects a tendency toward more detailed recording of pejorative vocabulary, which may be related to the preventive function of dictionaries – warning users against inappropriate use of stylistically marked words.

2. Functional-stylistic colouring. Functional-stylistic variants are characterised by their belonging to a particular functional style or register of speech. In Uzbek, there are 2,355 such units (26.45%), and in German – 7,218 (9.96%). Uzbek examples: бемаъно (elevated), ширу шакар (literary). German examples: renitent (elevated), Gischt (technical language). The higher representation of functional-stylistic variants in Uzbek can be explained by the specific features of the linguistic situation in Uzbekistan, where different lexical layers coexist: original Turkic, Arabo-Persian and Russian–international.

3. Frequency of use. Labels indicating frequency of usage are only slightly represented in both languages: 34 units (0.38%) in Uzbek and 1,807 units (2.5%) in German. Examples: бўлимли (rarely used) in Uzbek, Futter (selten “rarely”) in German. The low representation of this category is connected with the objective difficulties of determining frequency without corpus data, which were not available to the compilers of the dictionaries under study.

4. Territorial limitation. Dialectal and regional variants amount to 1,172 units (13.16%) in Uzbek and 35,672 units (49.24%) in German. Examples: юрум (dial.) in Uzbek, abkratzen (mundartl. “dialectal”) in German. The drastic difference in the representation of dialect variants reflects the differences in the dialect situation of the two languages. German is characterised by extremely rich dialect diversity with 29 major dialect groups, whereas Uzbek has a relatively unified dialect structure with four

main groups of dialects.

5. Social limitation. Socially marked variants are subdivided into several subgroups:

By gender and age: 44 units (0.49%) in Uzbek, 698 units (0.96%) in German. Examples: гиргиттон (female), пўм (child.) in Uzbek; Abi (Schülerspr. “school slang”), abmurksen (Studentenspr. “student slang”) in German.

By professional sphere: 3,087 units (34.68%) in Uzbek, 19,134 units (26.41%) in German. Examples: куштомоқ (anat.), пештоқ (archit.) in Uzbek; Ozonschicht (Meteorol.), Bocksbart (Bot.) in German.

The dominance of professional vocabulary in both lexicographic traditions reflects the importance of the terminological function of explanatory dictionaries. However, the differences in percentage ratios point to different priorities: Uzbek lexicography pays relatively greater attention to professional differentiation, whereas German lexicography prioritises territorial differentiation.

6. Historical and national-cultural significance.

Variants with special cultural significance amount to 281 units (3.16%) in Uzbek and 96 units (0.13%) in German. An example is ҳазрат (relig.) in Uzbek.

The higher representation of this category in Uzbek is associated with particular attention to religious and ethnocultural vocabulary during the period of forming the national identity of independent Uzbekistan.

7. Chronological marking.

Obsolete and historical variants are represented by 1,684 units (18.92%) in Uzbek and 1,453 units (2%) in German. Examples: нечукким (obsolete), аббосийлар (hist.) in Uzbek; Bein (landsch., südd., österr.) in German. The significant representation of chronologically marked vocabulary in the Uzbek dictionary reflects the aspiration to preserve the linguistic heritage and to document historical changes in the language.

The analysis of the structure of dictionary entries in the sources under study reveals both common and specific features in the organisation of sociolinguistic information.

The principle of hierarchical organisation of information is implemented in both dictionaries through the sequential

placement of sociolinguistic labels after grammatical characteristics but before the definition of meaning. This emphasises the importance of sociolinguistic information for the adequate understanding and use of a lexical unit.

The principle of systematisation of labels is manifested in the use of standardised abbreviations and conventional symbols. However, the labelling systems in the two languages differ significantly in the degree of detail and in their classification principles.

The principle of contextualisation is realised through the provision of illustrative examples demonstrating the use of sociolinguistic variants in natural contexts. In the German dictionary, the examples are more diverse and include both written and oral sources.

The Uzbek lexicographic tradition is characterised by several specific features:

Priority of the literary norm is expressed in the careful marking of deviations from the standard language. Sociolinguistic variants are often accompanied by recommendations on the use of literary equivalents.

Ethnocultural orientation is reflected in the detailed description of vocabulary related to the traditional culture, religion and history of the Uzbek people. This reflects the function of the dictionary as a repository of cultural memory.

A conservative stance towards borrowings is manifested in a critical attitude to Russianisms and internationalisms, which are often marked as undesirable or to be replaced by native equivalents.

The German lexicographic tradition demonstrates the following characteristic features:

Descriptive orientation is expressed in the aspiration to provide an objective description of language usage without explicit prescriptive guidelines. Sociolinguistic variants are presented as equally valid forms of linguistic expression.

Detailed territorial differentiation reflects the federal structure of the German-speaking area and the traditional respect for dialect diversity.

Professional specialisation is manifested in the thorough recording of terminology from various fields of activity,

which corresponds to the high level of professional differentiation in contemporary German society.

On the basis of the analysis carried out, a system of principles of maximum inclusivity for contemporary lexicographic practice has been developed:

1. Principle of equal representation. This principle requires ensuring proportional representation of all sociolinguistic groups in accordance with their actual linguistic activity. This does not mean mechanical equality, but implies overcoming traditional imbalances in favour of privileged social strata.

Practical implementation of the principle includes:

- Systematic collection of data from different social environments
- Use of stratified samples in corpus-based research
- Involvement of speakers from different social groups in the dictionary-making process
- Regular monitoring of the representation of different sociolects

2. Principle of authentic recording. The principle of authentic recording requires documenting linguistic variants in their natural form without normalisation or “correction.” This is especially important for spoken language and informal registers, which have traditionally been subject to stylistic editing.

Implementation of the principle presupposes:

- Use of audio and video recordings of natural speech
- Preservation of phonetic features in transcription
- Recording grammatical variants without normalization
- Documenting contexts of natural usage

3. Principle of semantic equality. Semantic equality means the equal treatment of all linguistic variants without hierarchisation according to the “correct/incorrect” principle. All recorded variants are regarded as legitimate ways of linguistic expression.

Practical aspects include:

- Avoidance of evaluative characteristics such as “distorted,” “incorrect”
- Use of descriptive labels indicating conditions of use
- Equal treatment of the meanings of all variants
- Presentation of sociolinguistic information as an additional characteristic rather than a restriction

4. Principle of dynamic updating. The sociolinguistic situation is subject to constant change, which requires regular updating of dictionary information. The traditional model of the dictionary as a finished product must give way to a dynamic model of a continuously updated resource.

Mechanisms of implementation:

- Integration with corpus databases in real-time mode
- Crowdsourcing platforms for collecting new data
- Automatic systems for monitoring language change
- Regular revision of existing dictionary entries

5. Principle of user orientation. Modern lexicography must take into account the diverse needs of users: language learners, translators, researchers, representatives of various professions. This requires flexible organisation of information and multiple ways of accessing data.

Practical implementation:

- Multi-level structure of dictionary entries
- Adaptive interfaces for different categories of users
- Integration of multimedia materials
- A feedback system for users

Technological innovations in inclusive lexicography. Modern corpus technologies open up new possibilities for inclusive lexicography. Balanced corpora that include texts from various social groups make it possible to obtain an objective picture of linguistic variation.

Key advantages of corpus methods:

- Objectivity of data based on actual usage
- Possibility of quantitative analysis of sociolinguistic variables
- Automatic identification of new linguistic trends
- Statistical validity of conclusions about the frequency of variants

Digital technologies in modern lexicography

The development of digital technologies radically changes the principles of creating and using dictionaries. As Lew (2011) notes in his analysis of contemporary trends in dictionary-use research, “methodological standards are improving at a steady rate,” and interest in empirical studies of dictionary users is growing.

Applications of digital technologies in inclusive lexicography:

- Automatic detection of sociolinguistic patterns
- Machine learning for the classification of linguistic variants
- Digital platforms for collaborative dictionary creation
- Personalisation of content in accordance with users’ needs

Crowdsourcing platforms. Crowdsourcing is a powerful tool for the democratisation of the lexicographic process. The study by Sajous and Josselin-Leray (2022) identifies various models of crowdsourcing in lexicography: microtasks, collaborative models and user-generated content. Platforms such as Wiktionary demonstrate the potential of collective creation of lexicographic resources, providing rapid response to linguistic change and inclusion of marginalised language variants.

Advantages of crowdsourcing:

- Involvement of native speakers in the process of dictionary creation
- Rapid response to linguistic changes
- Inclusion of marginalised language variants

- Reduction of the cost of lexicographic projects

CONCLUSION

The conducted study has revealed significant differences in approaches to lexicographic description of sociolinguistic variation in Uzbek and German. Quantitative analysis of 8,903 sociolinguistic variants in Uzbek and 72,452 variants in German demonstrated differing priorities of the lexicographic traditions: an orientation toward professional differentiation in Uzbek (34.68% of the total) and territorial variation in German (49.24%).

The system of principles of maximum inclusivity developed in this study represents a comprehensive approach to modernising lexicographic practice. The key principles—equal representation, authentic recording, semantic equality, dynamic updating and user orientation—provide a theoretical foundation for creating more equitable and representative lexicographic resources.

Of particular importance is the integration of modern technologies—corpus linguistics, artificial intelligence and crowdsourcing platforms—into the lexicographic process. These technologies not only increase efficiency but also open fundamentally new possibilities for inclusive practice.

The comparative analysis has shown that the Uzbek lexicographic tradition is characterised by a conservative approach with an emphasis on the literary norm and cultural authenticity, whereas the German tradition demonstrates greater openness to linguistic diversity and a descriptive approach. Both traditions can mutually enrich one another: Uzbek lexicography through the adoption of more inclusive principles, and German lexicography through greater consideration of the cultural-historical context of linguistic variation.

Prospects for the development of inclusive lexicography are associated with the further democratisation of the lexicographic process, the strengthening of the role of language communities in dictionary creation, and the development of technologies for automatic analysis of linguistic data. A key direction is the elaboration of ethical standards for lexicographic activity that ensure fair representation of all language communities.

The present study contributes to the development of the theory and practice of modern lexicography by proposing

concrete solutions for increasing the inclusivity of dictionary-making. The principles developed here can be applied not only in Uzbek and German lexicography but also in other linguistic traditions striving for a more equitable representation of linguistic diversity.

REFERENCES

1. Chambers, J. K., & Trudgill, P. (1998). *Dialectology*. Cambridge University Press.
2. Deutsches Universalwörterbuch: Das umfassende Bedeutungswörterbuch der deutschen Gegenwartssprache. (2019). Duden.
3. Kramarae, C., & Treichler, P. A. (1985). *A Feminist Dictionary*. Pandora Press.
4. Labov, W. (1972). *Language in the Inner City: Studies in the Black English Vernacular*. University of Pennsylvania Press.
5. Lew, R. (2011). *Studies in Dictionary Use: Recent Developments*. *International Journal of Lexicography*, 24(1), 1-4. <https://doi.org/10.1093/ijl/ecq044>
6. Russell, L. R. (2012). *This is What a Dictionary Looks Like: The Lexicographical Contributions of Feminist Dictionaries*. *International Journal of Lexicography*, 25(1), 1-25.
7. Sajous, F., & Josselin-Leray, A. (2022). *Issues in Collaborative and Crowdsourced Lexicography*. HAL Science Archive.
8. Ўзбек тили изоҳли луғати [Толковый словарь узбекского языка]. (2006-2008). А. Мадвалиев (Ред.), Тома 1-5. Ўзбекистон миллий энциклопедияси.