Comparative Analysis of Medical Adjectives in Croatian General Dictionaries

Martina Pavić[1], Daša Farkaš[2]

[1] Institute of Croatian Language, Republike Austrije 16, 10000 Zagreb

[2] Faculty of Humanities and Social Sciences, Ivana Lučića 3, 10000 Zagreb

E-mail: mpavic@ihjj.hr, dfarkas@ffzg.hr

Abstract

The representation of medical adjectives in Croatian general dictionaries reveals significant inconsistencies, reflected in uneven lemma inclusion, ambigous or absent domain labels, and limited definitional precision. This paper analyzes the 80 most frequent adjectives, based on corpus data from the *Croatian Medical Corpus* (CMC) (Kocijan et al., 2020), in the three major Croatian general dictionaries: *Veliki rječnik hrvatskoga standardnog jezika* (2015), *Hrvatski enciklopedijski rječnik* (2002), and *Rječnik hrvatskoga jezika* (2000). The analysis focuses on lemma status, the presence of domain labels, and the accuracy of definitions. To contextualize the Croatian practice, the study includes a brief comparison with *Merriam-Webster Dictionary* (2025), which demonstrates better lemma coverage and more terminologically informed definitions, but also exhibits inconsistencies that reflect the broader challenges of systematically representing medical adjectives in general lexicography. The paper's findings reveal inconsistencies in Croatian lexicographic practice and highlight the need for more conceptually grounded, corpus-based approaches that integrate terminological precision with lexicographic usability.

Keywords: medical adjectives; Croatian general dictionaries; *Croatian Medical Corpus*; *Merriam-Webster Dictionary*

1. Introduction

The integration of specialized terms into general dictionaries reflects both scientific advancement and the democratization of knowledge (Rondeau, 1984, as cited in Salgado et al., 2022). However, their lexicographic treatment is often inadequate, marked by incomplete definitions, unclear inclusion criteria, and inconsistent domain labeling. These shortcomings highlight the ongoing challenge of balancing linguistic accessibility with conceptual precision.

This issue is especially evident with adjectives, which in domains like medicine are not merely optional modifiers but central elements that specify conditions (e.g., *kroničnі* 'chronic'), anatomical structures (*srčani* 'cardiac'), or pathological processes (*infektivni* 'infectious'). Despite their high frequency and semantic importance, medical adjectives are often underrepresented or ambiguously treated in general dictionaries, where their specialized usage is rarely marked or explained.

This tendency is not unique to Croatian lexicography, but reflects a more general challenge in dictionaries regarding the consistent representation of specialized usages. Definitions tend to be brief and general, domain labels are often absent, and selection criteria remain opaque, blurring the line between general and specialized meanings.

This study addresses that gap by examining the treatment of medical adjectives in the three major Croatian general dictionaries: *Veliki rječnik hrvatskoga standardnog jezika* [Great Dictionary of the Croatian Standard Language] (2015), *Hrvatski enciklopedijski rječnik* [Croatian Encyclopedic Dictionary] (2002), and *Rječnik hrvatskoga jezika* [Dictionary of the Croatian Language] (2000). Based on the 80 most frequent medical adjectives extracted from the *Croatian Medical Corpus* (CMC) (Kocijan, Kurolt & Mijić, 2020), the analysis investigates their inclusion as lemmas, the precision of their definitions, and the presence of domain labels.[1] In addition to identifying inconsistencies across dictionaries, the study includes a contrastive component based on the same set of 80 adjectives from *Merriam-Webster Dictionary* (2025).

Although some research has addressed the role of adjectives in multi-word terms (e.g., Grčić Simeunović, 2021, 2020; Durán-Muñoz, 2019; Pitkänen-Heikkilä, 2015; Alonso Campos & Torner Castells, 2010), this study represents one of the first systematic analysis of their status as haedwords in general-purpose dictionaries.

The paper is structured as follows: Section 2 outlines the theoretical framework for the research. Section 3 presents the methodology, including the selection of 80 high-frequency medical adjectives from the *CMC* and the criteria for their analysis. Section 4 presents the results of the analysis regarding lemma status, definitions, and domain labeling, and includes a contrastive overview based on entries from *Merriam-Webster Dictionary* (2025). Section 5 reflects on the findings and proposes directions for future research.

2. Specialized Adjectives between Terminology and Lexicography

2.1 Theoretical Foundations and Challenges

The treatment of specialized lexical units (SLUs) in general-purpose dictionaries has attracted increasing scholarly attention, particularly in light of newer approaches to terminology that emphasize contextual and communicative factors. Traditional terminological frameworks, most notably the General Theory of Terminology (Wüster, 1979), restrict the status of terms to nouns, considering only nominal elements as capable of designating discrete and delineated concepts.

However, such a restrictive view has been widely challenged. The Communicative Theory of Terminology (Cabré, 1999, 2003) reconceptualizes terminology as a dynamic interplay of linguistic, cognitive, and social elements. In this model, adjectives—and

[1] The study examines only those medical adjectives that function independently as terms—such as *kirurški* ('surgical') or *bakterijski* ('bacterial')—as opposed to those whose terminological relevance emerges only in specialized lexical units (e.g. *dubok* 'deep' in *duboka venska tromboza* 'deep vein thrombosis').

even other predicative forms—may acquire terminological status through recurrent use in specialized contexts. Rather than being viewed as subordinate modifiers, these elements are acknowledged as potential carriers of specialized meaning. As Cabré (1999) emphasizes, lexicographic work must transcend purely linguistic description by incorporating contextual and encyclopedic knowledge to capture the full complexity of specialized communication.

Sager (1990) also recognizes the lexicographic challenge of documenting specialized adjectives, noting their frequent marginalization in both general and terminological resources. Despite their high frequency and relevance in domain-specific discourse, adjectives are often excluded from lemma status or given only cursory definitions. This underrepresentation is particularly problematic in fields such as medicine, where adjectives serve a key role in expressing attributes, classifications, and diagnostic categories.

Recent empirical work reinforces this shift toward a contextual view of terminology. Grčić Simeunović (2015) stresses that the terminological value of adjectives is not inherent but emerges through their specialized use in scientific discourse. Similarly, Alonso Campos and Torner Castells (2010) show that adjectives originating in general language can become terminologically marked when stabilized in domain-specific contexts. Although their study focuses on environmental science, the same mechanisms apply to medical discourse, where relational adjectives (e.g., *cardiac*, *pulmonary*) and qualifying adjectives (e.g., *chronic*, *malignant*) are integral to professional communication.

However, when adjectives express terminological meaning only as part of fixed collocations, they are often excluded as lemmas in general-purpose dictionaries. This omission poses a significant lexicographic challenge, as the conceptual importance of such adjectives may be overlooked despite their terminological relevance. These issues lie at the intersection of lexicography and terminology, two fields that share methodological foundations but diverge in their primary objectives: general lexicography tends to prioritize descriptive documentation, while terminography focuses on conceptual clarity and prescriptive standardization (Alberts, 2001; Sager, 1990). Given this divergence, empirical and corpus-based methods have become essential for identifying adjectives that function as terms within specialized domains such as medicine. Moreover, the distinction between lexicography and terminography has been analyzed from multiple perspectives, including functional (Sager, 1990), socio-cognitive (Costa, 2013; Temmerman, 2000), and user-oriented (Bergenholtz & Tarp, 1995) frameworks, highlighting the complexity of their relationship and the need for integrated approaches in lexicographic practice.

Salgado and Costa (2024) further emphasize that a precise definition of terms requires integrating both linguistic forms and conceptual content. They argue for a dual-definition approach—providing both a general-language and a domain-specific

definition—to better serve diverse users while maintaining terminological precision. Their critique of general dictionaries includes vague definitions, inconsistent domain labeling, and omission of equivalents, all of which diminish the clarity and utility of specialized entries. They also note the need to treat not only nouns but also adjectives, modal verbs, collocations, and visual elements as part of terminological resources, reflecting the evolving nature of user expectations and communicative needs.

Fontenelle (2021) had previously raised related concerns, pointing to the neglect of existing terminological databases in general lexicography. He questioned the inclusion of specialized vocabulary in comprehensive dictionaries without concept-based approaches or curated resources like IATE (Interactive Terminology for Europe), highlighting a broader institutional divide between lexicography and terminology that undermines the methodological coherence of specialized entries.

Together, these perspectives support a more integrated approach to the lexicographic treatment of specialized adjectives. The recognition of adjectives as terminological units —especially in highly codified fields such as medicine—requires both theoretical adjustment and practical innovation. As this study demonstrates, the current representation of medical adjectives in Croatian general dictionaries remains inconsistent, often failing to reflect their communicative and conceptual relevance. A more systematic, concept-oriented, and corpus-informed methodology is therefore needed to align lexicographic practice with contemporary terminological standards.

2.2 Lexicographic vs. Terminological Definitions

Lexicographers generally prioritize creating user-oriented definitions that enhance comprehension, focusing on facilitating the user's understanding of a lexical item. In contrast, terminologists aim to preserve conceptual coherence within specialized knowledge systems by situating terms within a structured conceptual network (Landau, 2001: 154). While both disciplines rely on intensional definitions, their objectives differ: terminological definitions specify the concept that a term designates and establish its relations to other concepts, whereas lexicographic definitions describe the meaning(s) conveyed by a lexical unit (Salgado, 2021).

Sager (1990) distinguishes between two fundamentally different types of definitions: the linguistic definition, which provides a descriptive account of a concept based on a list of features that convey its meaning, and the terminological definition, which identifies a concept exclusively within the framework of a conceptual system, positioning it accordingly. This distinction is further developed by Mel'čuk and Polguère (2018), who differentiate between lexicographic definitions—designed to capture the usage-based meaning of a lexical unit—and terminological definitions, which aim to represent the conceptual content of a term within a structured knowledge domain. Lexicographic definitions are typically extensional and paraphrastic, focusing on observable usage, whereas terminological definitions are intensional, emphasizing essential features and

hierarchical classification. Both perspectives underline the importance of anchoring specialized terms not merely in linguistic usage, but in the underlying conceptual systems that give them disciplinary meaning.

This theoretical distinction proves particularly useful for analyzing the inconsistent treatment of medical adjectives in general dictionaries. For example, the adjective *infektivni* ('infectious') may have a varying number of definitions across general dictionaries, as will be shown in our analysis. Depending on the context, it can denote a cause (*infektivni uzročnik* 'infectious pathogen'), a condition (*infektivna bolest* 'infectious disease'), or a process (*infektivna faza* 'infectious phase'). In contrast, a terminological approach would treat these as distinct conceptual relations embedded within a structured medical knowledge system. Such analysis reinforces the need for dictionary definitions that go beyond surface-level generalization and reflect the internal structure and logic of the domain in question.

The integration of terminology into general-purpose dictionaries is guided by international standards, particularly ISO 704:2009 and ISO 1087:2019. These standards emphasize the need for systematic conceptual analysis, domain-specific precision, and interdisciplinary coherence in the treatment of specialized language. In practice, this means that definitions should be grounded in a clear conceptual hierarchy, and accompanied by domain labels, contextual examples, and, if possible, cross-references to related terms.

Applying these principles to the lexicographic treatment of adjectives requires a shift in perspective: from viewing adjectives as mere modifiers to recognizing their function as term-forming elements in specialized domains. This is especially relevant in medicine, where adjectives such as *kroničnі* ('chronic') or *maligni* ('malignant') encapsulate crucial diagnostic and prognostic meanings. Therefore, a lexicographic approach aligned with terminological standards and supported by conceptual clarity is essential to improve the representation of medical adjectives in general-purpose dictionaries.

A key component of this approach is the use of domain labels, which serve as a vital bridge between specialized and general language. As Salgado, Costa, and Tasovac (2022) emphasize, domain labeling enhances both the readability and disambiguation of definitions in general dictionaries. It signals to the user that a word carries a specialized meaning and should be interpreted within the appropriate conceptual frame. Moreover, hierarchical domain structures—where labels are organized from broad categories (e.g., MED) to more specific subdomains (e.g., PAT, ANAT)—enable a more systematic representation of knowledge and promote terminological consistency across entries.

Despite their theoretical importance, domain labels are often inconsistently applied in Croatian general dictionaries. Some adjectives may be marked with general labels such as *med*, others more specifically with labels such as *anat*, while a number of entries remain unmarked, even when predominantly used in medical discourse. This inconsistency impedes both comprehension and accurate lexical categorization,

underscoring the need for a more principled and terminologically informed approach to domain labeling.

3. Methodology

This study employs a qualitative and comparative lexicographic approach to examine the treatment of medical adjectives in the three major general-purpose Croatian dictionaries: *Veliki rječnik hrvatskoga standardnog jezika* (VRH) (2015), *Hrvatski enciklopedijski rječnik* (2002), and *Rječnik hrvatskoga jezika* (2000).[2]

The primary objective of the paper is to analyze the lemma status, definitions, and domain labeling of medical adjectives in these dictionaries, and to contrast these findings with those from an English-language general dictionary, specifically *MWB* (2025). The analysis seeks to identify systematic patterns—or their absence—thereby contributing to a deeper understanding of how to reconcile broad general language coverage with terminological precision.

3.1 Selection of Dictionaries

The three monolingual Croatian general dictionaries were selected based on their comprehensiveness, scope, and recency of publication. These dictionaries were chosen under the assumption that they represent the most authoritative and extensive Croatian general-language lexicographic resources published in the last few decades. Although the Croatian Web Dictionary – Mrežnik represents the most contemporary and innovative effort in Croatian lexicography, it could not be included in this study because it is still in development and does not yet provide the full range of adjectives required for the systematic analysis.

In *Rječnik hrvatskoga jezika* (2000), which contains approximately 64,000 entries, the editorial principles explicitly state that multi-word terms consisting of an adjective and a noun will not be treated as separate lemmas. Instead, such terms are defined within the section on terminology, either under the adjective or the noun, depending on which component is considered lexicographically more prominent.

In contrast, *Veliki rječnik hrvatskoga standardnog jezika* (2015), the most comprehensive Croatian dictionary to date, contains more than 120,000 lemmas and subentries. The selection of entries in this dictionary was informed by a broad linguistic database that includes authoritative lexicographic sources, language corpora such as *Hrvatska jezična riznica* [Croatian Language Repository], *Hrvatski nacionalni korpus* [Croatian National Corpus], and *hrWaC 2.0* (Croatian Web Corpus), as well as other linguistic handbooks and specialized dictionaries, the most significant of which is *Hrvatski jezični savjetnik* (1999). The dictionary emphasizes its goal of capturing the

[2] In the analysis and tables, we list them using the following abbreviations: *Veliki rječnik hrvatskoga standardnog jezika* = VRH; *Hrvatski enciklopedijski rječnik* = ENCIKL.; *Rječnik hrvatskoga jezika* = RHJ.

contemporary state of language use and explicitly states that each lexical entry includes terminological indicators of usage.

Furthermore, *Hrvatski enciklopedijski rječnik* (2002) combines two approaches to the dictionary, encyclopedic and linguistic, and these two components are intertwined in the definitions. The dictionary contains 110,000 processed lemmas and their derivatives. The sources for this dictionary were *Rječnik hrvatskoga jezika* [Dictionary of the Croatian Language] by Vladimir Anić, *Rječnik stranih riječi* [Dictionary of Foreign Words] by Anić and Goldstein, and the Croatian National Corpus as an electronic source. The concept of this dictionary is extremely descriptive.

For comparison with English-language practice, *Merriam-Webster Dictionary* (2025) was selected due to its accessibility, frequency of use among both experts and the general public, and its established reputation for comprehensive treatment of both general and specialized meanings.

3.2 Corpus-Based Selection of Adjectives

The analysis is based on the *Croatian Medical Corpus* (CMC) (Kocijan et al., 2020), which consists of a collection of medical texts known as MedCoreA, forming the core textual material of the corpus. These texts include drug information leaflets intended for healthcare professionals, selected for their representativeness of contemporary Croatian medical discourse. The corpus was linguistically annotated using the NooJ platform, which enabled detailed morphosyntactic tagging and the extraction of lexical and grammatical patterns. The integration of MedCoreA texts and NooJ-based annotation results in a linguistically enriched resource that supports the systematic study of Croatian medical terminology, particularly the behavior and function of medical adjectives within specialized contexts. As emphasized by the authors, making these resources available to the broader scientific community via *Sketch Engine* (Kilgariff et al., 2014) is expected to facilitate further research in medical linguistics, including the development of algorithms for medical document classification, information retrieval, and machine translation—tasks that require high levels of terminological accuracy, reliability, and sensitivity to domain-specific variation.

From the *CMC*, the 80 most frequent medical adjectives were extracted. These adjectives were manually verified for their medical relevance to ensure that only domain-specific adjectives were included in the analysis. Examples include: *infektivni* (‘infectious’), *maligni* (‘malignant’), *akutni* (‘acute’), *kronični* (‘chronic’), *sistemski* (‘systemic’), among others. Eponymous adjectives are not included in the database, as they appear exclusively in fixed multi-word terminological units (e.g., *Crohnova bolest* 'Crohn's disease', *Turnerov sindrom* 'Turner syndrome', *Aspergerov sindrom* 'Asperger syndrome') (Campos & Torner Castells, 2010).

3.3 Analytical Procedure

Each of the 80 adjectives was examined across the three Croatian general dictionaries.[3] The analysis provides a cross-dictionary comparison focusing on the following variables:

- **lemma status**: whether the adjective is included as a lemma or only mentioned within other entries
- **definition structure**: number of definitions provided, clarity and consistency of definitions, semantic distinction between senses
- **domain labeling**: whether domain labels are present, type (e.g., MED, ANAT), and consistency of domain labels
- **collocations**: whether they are included and to what extent they reflect specialized usage.
- **comparison to *Merriam-Webster Dictionary*** (2025): highlights differences in the treatment of medical adjectives between Croatian and an English general-purpose dictionary.

The analysis of these five variables evaluates possible omissions, simplifications, or overlaps. The results are analyzed to identify strengths and gaps in the current lexicographic treatment of medical adjectives, leading to suggestions for improvements in Croatian lexicographic resources.

Data from each dictionary entry were recorded and organized as a base, with comparative data arranged in separate columns to enable systematic cross-dictionary analysis.

4. Results

The analysis revealed significant variation in the treatment of medical adjectives across the dictionaries. As for the existence of the lemma itself in the dictionaries, 20 adjectives have no lemma at all in any of the three dictionaries. We consider 25% of the total number to be significant, especially since these adjectives include, for example, *refluksni* 'reflux', N, *ulcerozni* 'ulcerative' or *urinarni* 'urinary', which, according to the data from the corpus, form multi-word terms (*urinarna inkontinencija* 'urinary incontinence', *urinarna retencija* 'urinary retention', *ulcerozni kolitis* 'ulcerative colitis', *ulcerosni stomatitis* 'ulcerative stomatitis', *refluksna bolest* 'reflux disease', *refluksni ezofagitis* 'reflux esophagitis'). In Table 1, the numbers of adjective lemmas per dictionary are shown.

[3] To support transparency and reusability, the complete dataset is included as an appendix to this paper. It contains lemma status, definitions, domain labels, and collocations, where applicable. The dataset is formatted to facilitate further processing and integration into terminological and lexicographic resources.

corpus	number of adj. lemmas
VRH	50
RHJ	39
ENCIKL.	51

Table 1: Number of lemmas in each dictionary.

Regarding the listing of lemmas either in the indefinite or definite form, the dictionaries show a lack of consistency for adjectives that have exclusively medical meanings. For example, the adjective *abdominalan* ('abdominal') appears in the indefinite form in the VRH and ENCIKL dictionaries, but in the definite form (*abdominalni*) in RHJ. We agree with the approach taken in RHJ, as we believe that adjectives with terminological meaning should be listed in the definite form in general dictionaries.

Further inconsistencies can be seen in the treatment of other medical adjectives. For instance, *renalan* ('renal') is listed in the indefinite form in ENCIKL, in the definite form (*renalni*) in VRH, and is not included at all in RHJ. Similarly, *kardiovaskularan* ('cardiovascular') is listed in the indefinite form in ENCIKL, but an example in the entry includes the definite form in the collocation *kardiovaskularni kirurg* ('cardiovascular surgeon'). The other two dictionaries do not include this lemma.

The dictionaries also exhibit inconsistency in the listing of lemmas for adjectives that have both general and specialized meanings. For example, in VRH, the adjective *živčan* is treated as a single lemma with two meanings:

1. *koji se lako i brzo uzruja* 'easily and quickly upset'
2. *(živčani) koji se odnosi na živce* 'relating to nerves'

Here, the specialized meaning is immediately indicated by the definite form in parentheses. In contrast, the RHJ dictionary separates the general and specialized meanings into two distinct lemmas, which we consider the better approach as it signals two different senses—one general and one specialized. In general language, *živčan* means *koji se lako, brzo uzrujava, koji pokazuje nemir, koji se živcira* 'easily or quickly upset, showing restlessness, nervous', while in specialized usage *živčani* means *koji se odnosi na srce* 'relating to nerves.'

A similar situation occurs with the adjective *srčan/srčani*. VRH lists it as a single lemma with two meanings:

1. *koji je pun hrabrosti ili volje za što* 'full of courage or will for something'
2. *(srčani) koji se odnosi na srce* 'relating to the heart'.

Additionally, in collocations, the anatomical label ANAT is given (e.g., *srčani mišić*, *srčane komore*). ENCIKL also lists *srčan* as a single lemma with two meanings:

1. *hrabar, odvažan* 'brave, courageous'

2. *(odr.) koji se odnosi na srce* 'relating to the heart'.

The indication that it is a definite form signals a specialized meaning. On the other hand, RHJ separates this adjective into two lemmas: *srčan* in general language means *koji ima srca, hrabrosti, volje, žara u kakvu poslu* 'full of heart, courage, will, passion for some activity', while *srčani* means *koji se odnosi na srce* 'relating to the heart'.

Research has shown that the dictionaries are not consistent in the presentation and definition of adjective lemmas. Definitions are frequently tautological or circular (e.g., *hematološki* = 'koji se odnosi na hematologiju', *krvni* = 'koji se odnosi na krv'), lacking conceptual context. In some cases, the examined dictionaries do not list medical adjectives with clear definitions and appropriate labels.

Such practices contrast with recommendations from terminological literature (Sager, 1990; ISO 704; Salgado, Costa & Tasovac, 2022), which emphasize that derivative forms like adjectives should be defined concerning the concepts expressed by their base nouns. In terminological systems, adjectives such as *pulmonološki* are not autonomous but function relationally, indicating a connection to a medical specialty (*pulmonologija* 'pulmonology').

Adjectives are commonly listed without domain identification and syntagmatic context. Table 2 contains information on many types of definitions and domain labels in which dictionary.

	VRH	RHJ	ENCIKL.	MERRIAM W.
basic def.	47	38	46	64
med. domain label	3	3	8	
domain labels	20	8	21	

Table 2: Types of definitions in the dictionaries.

Table 3 shows the number of adjectives according to how the basic definitions and domain labels are combined in their description. In all these dictionaries, most adjectives have one or more domain labels in addition to the basic definition.

dictionary	**def +** **med +**	**def −** **med +**	**def +** **dom. lab +**	**def −** **dom. lab +**
VRH	3	0	19	1
RHJ	1	2	8	0
ENCIKL.	7	1	17	5

Table 3: Combinations of basic definition and domain labels through the dictionaries.

In RHJ dictionary, only one label is used – **anat**. There are nine different labels in the encyclopedic dictionary, while there are eight in the VRH dictionary. Chart 1 shows the distribution of domain labels. Only the label **anat** appears in all three corpora and is the most common. In addition to it, there are also frequent **pat** and **fiziol** labels, as subfields of medicine.

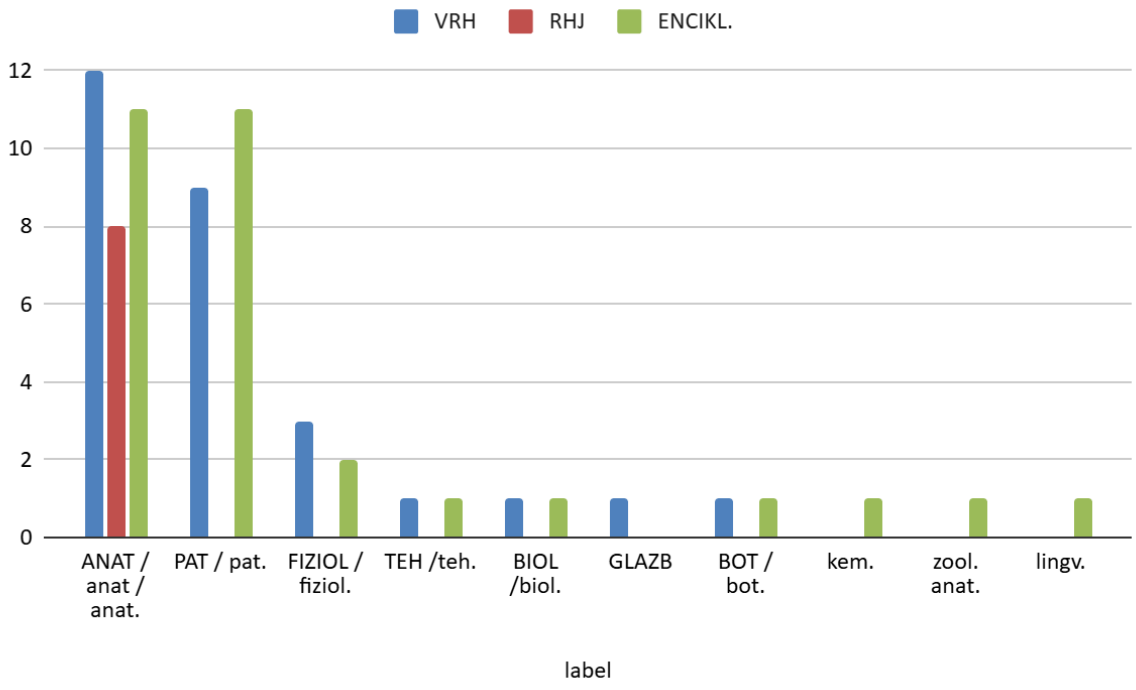

Chart 1: Distribution of labels through the dictionaries.

Although all adjectives are frequent in the medical corpus, only some of them have the additional **MED** (**med** or **med.**) label. In total, in all three dictionaries, only three adjectives have an exclusively medical definition:

- *koronaran* ('coronary') – **med/vet** koji opskrbljuju srce krvlju 'which supply the heart with blood' (RHJ)

- *maligan* ('malignant') – **med/vet** zloćudni, opaki, pogubni 'malignant, wicked, disastrous' (RHJ)

- *kardiovaskularan* ('cardiovascular') – **med**. koji se odnosi na srce i krvožilni sustav 'relating to the heart and circulatory system' (ENCIKL).

The inconsistency in listing domain labels can be seen from Table 4, which shows how adjectives referring to organs are defined in the VRH dictionary. Although they all refer to human organs, they all have a different way of defining them: *želučani* has only a circular definition without a domain label, *bubrežni* is defined exclusively with a domain label, and *moždani* has both a basic definition and a domain label. It is also interesting how *bubrežni* and *moždani* have the **ANAT** label listed differently. In *bubrežni,* it is a circular definition, and in *moždani* it is a contextual example, respectively example of collocation. We notice the same inconsistency in RHJ dictionary, while the encyclopedic dictionary lists adjectives more consistently, but also lists domain labels inconsistently (*bubrežni* does not have **anat** label, although other definitions of organs have it).

	lemma	**basic def.**	**domain label**
VRH	**želudčani (*gastric*)**	koji se odnosi na želudac 'relating to the stomach'	-
	bubrežni (*renal*)	-	***ANAT*** koji se odnosi na bubrege 'relating to kidneys'
	moždani (*cerebral*)	koji se odnosi na mozak 'relating to brain'	***ANAT*** moždane ovojnice 'meninges'; ***PAT*** moždani udar, moždana kap 'stroke, cerebral infarction'
RHJ	**želudčani (*gastric*)**	koji se odnosi na želudac 'relating to stomach'	***anat*** želučana sluznica, želučane žlijezde 'gastric mucosa, gastric glands'
	bubrežni (*renal*)	koji se odnosi na bubreg 'relating to kidney'	-

	moždani (*cerebral*)	koji se odnosi na mozak ‘relating to brain’	***ANAT*** moždane ovojnice ‘meninges’; ***PAT*** moždani udar, moždana kap ‘stroke, cerebral infarction’
ENCIKL	**želudčani (*gastric*)**	koji se odnosi na želudac ‘relating to stomach’	***fiziol.*** želučana kiselina ‘gastric acid’; ***anat.*** želučane žlijezde ‘gastric glands’
	bubrežni (*renal*)	koji se odnosi na bubrege ‘relating to kidney’	***pat.*** bubrežni edem, bubrežni kamenci ‘kidney edema, kidney stones’
	moždani (*cerebral*)	koji se odnosi na mozak ‘relating to brain’	***anat.*** moždana ovojnica ‘meninges’; ***pat.*** moždani udar ‘stroke’

Table 4: Definitions of adjectives relating to organs in VRH.

The absence of a label may imply general usage, although the adjective is specialized. This problem has been extensively discussed in recent terminological studies (e.g., Costa, 2019; Salgado et al., 2022), which argue for a structured and hierarchical approach to domain representation. The Croatian general dictionaries examined here rarely incorporate such a structure, leading to a flattening of conceptual distinctions.

In our corpus, 79 adjectives are forming nominal collocations, and just one adjective is exclusively predicative (*indiciran* 'indicate, V'). This adjective has a lemma only in the encyclopedic dictionary, but it has no basic definition, only a given example of a collocation (*indicirana snaga stroja* 'indicated machine power'). This collocation is not related to the field of medicine, but is from the technical field. The adjective *reumatoidni* 'rheumatoid' has only one basic definition in all three dictionaries and no domain designation or usage example. This definition is circular, *koji se odnosi na reumatoid* ‘relating to rheumatoid’, and without a definition of *rheumatoid* in the same dictionary (VRH). This example is very interesting because in the corpus it is used in only one type of collocation with the noun *artritis* ‘arthritis’. Reumatoidni *artritis* is a well-established medical term, and this adjective has no general meaning outside the term. We believe that such examples should be listed with a collocational relation and

labeled with MED. The situation is similar with the adjective *duodenalni* 'duodenal', where there is only one domain label, **pat,** and an example of collocation is *duodenalni ulkus*. Corpus data confirms that this adjective has only one collocational relationship and that it is a strong medical term, *duodenalni ulkus* 'duodenal ulcer'.

Some adjectives, such as *protuupalni* 'anti-inflammatory', *probavni* 'digestive', *gastrointestinalan* 'gastrointestinal', do not have any listed examples of contextual use or syntagmatic connections at all. Only 24 adjectives include examples of contextual use in at least one of the three dictionaries, and only three adjectives have such examples in all three dictionaries; they are presented in Table 5. The data reveal that the dictionary usage examples in the dictionaries do not correspond with the actual findings from the corpus.

adjective	VRH	RHJ	ENCIKL	CORPUS DATA
mokraćni (*urinary*)	mokraćna cijev 'urethra'; mokraćni kamenci 'urinary stones'	mokraćni mjehur, mokraćni organi 'urinary bladder, urinary organs'	mokraćna cijev, mjehur, organi, mokraćna kiselina, mokraćni kamenci 'urethra, bladder, organs, uric acid, urinary stones'	mjehur, sustav, dizurij, put, kiselina 'bladder, system, dysuria, path, acid'
cerebralni (*cerebral*)	cerebralni sustav 'cerebral system'	cerebralni sustav 'cerebral system'	cerebralni glasovi, cerebralna afekcija 'cerebral voices, cerebral affection'	paraliza, angiografija, tromboza, infarkt, arterija 'paralysis, angiography, thrombosis, infarction, artery'
limfni (*lymphatic*)	limfne žile, limfni čvorovi 'lymphatic vessels, lymph nodes'	limfne žile, čvorovi, organi 'lymphatic vessels, nodes, organs'	limfne žile, čvorovi 'lymphatic vessels, nodes'	središte, staza, sustav 'center, path, system'

Table 5: Contextual use in the three dictionaries and in the corpus.

Table 6 also shows that the contexts listed in the dictionaries do not match the usage in the medical corpus. Although this could mean that those listed in the dictionaries have a meaning in general language contexts, this is not the case because the collocations listed in the dictionaries have a medical meaning.

adjective	VRH	ENCIKL	CORPUS DATA
koronaran (*coronary*)	koronarna insuficijencija 'coronary insufficiency'	koronarna insuficijenciija, koronarne žile, koronarna jedinica 'coronary insufficiency, coronary vessels, coronary unit'	arterija, arteriografija, arteriospazam, kateterizacija, angiografija 'artery, arteriography, arteriospasm, catheterization, angiography'
patološki (*pathological*)	patološki prijelom 'pathological fracture'	patološki prijelom 'pathological fracture'	promjena, stanje, tkivo 'change, condition, tissue'
endokrini (*endocrine*)	endokrini sustav, endokrine žlijezde 'endocrine system, endocrine glands'	endokrine stanice, endokrine žlijezde 'endocrine cells, endocrine glands'	parametar, stanica, poremećaj 'parameter, cell, disorder'

Table 6: Contextual use in VRH and ENCIKL dictionaries and in the corpus.

To provide contextual insight and an external benchmark, a contrastive analysis was conducted using entries from *Merriam-Webster Dictionary* (2025), focusing on English equivalents of selected Croatian medical adjectives. The analysis showed the following findings:

- 75 adjectives have listed lemmas in the dictionary, which is much better coverage than in the Croatian dictionaries.
- 14 adjectives of these 75 adjectives are listed as lemmas but do not have definitions, and they are hierarchically listed and conceptually related to the noun that carries the meaning (e.g., *neurological* – *neurology*; *ischemic* – *ischemia*). We consider this a great way to avoid giving circular definitions and to show a hierarchical structure. Furthermore, such listing of adjectives is not implemented systematically and consistently in *MWB* (e.g., *epidermal* = of, relating to, or arising from the epidermis), but it can serve as an example of good practice for the Croatian dictionaries.

- Definitions relating to body parts, organs, and medical disciplines are also listed circularly, as in the Croatian dictionaries (e.g., *gastric* = of or relating to the stomach).

- There are no domain labels (e.g., *medical*, *pathological*, *clinical*) although it is clear from the definitions and examples of usage that these adjectives have a medical meaning (e.g., *pediatric* = 1. of, relating to, or specializing in pediatrics or its practice [pediatric *dermatology*, *a pediatric nurse*]; 2. of, relating to, affecting, or being an infant, child, or adolescent [*pediatric patients*, *pediatric cancer*s, *renal disease*, *renal failure*]).

5. Conclusion

The lexicographic representation of medical adjectives remains inconsistent and insufficiently grounded in a coherent methodological framework. This study examined how Croatian general-purpose dictionaries reflect specialized usage, focusing on the presence of lemmas, definitions, domain labels, and collocations. The analysis of 80 high-frequency medical adjectives, drawn from the Croatian Medical Corpus, revealed several notable shortcomings:

- inconsistent lexicographic inclusion of medical adjectives as lemmas
- a lack of clear or precise definitions for adjectives with exclusively medical meanings (e.g., *bubrežni*, *srčani*, *jetreni*)
- inconsistent or absent domain labels (e.g., *med*, *anat*, *pat*)
- irregular treatment of lemma status, particularly regarding definite vs. indefinite forms (e.g., *abdominalan* vs. *abdominalni*)
- weak or missing links to related noun terms and underlying medical concepts.

These inconsistencies diminish the usability of dictionary entries for both expert and non-expert users. For example, non-specialists (such as translators, students, or patients) may be misled by imprecise definitions or the absence of medical labels, while specialists may find it difficult to rely on such dictionaries for domain-specific reference.

We therefore recommend a more integrated and systematic approach to the lexicographic treatment of medical adjectives, grounded in terminographic principles and international standards. Specifically, we propose:

- **systematic application of domain labels** following ISO 704 and ISO 1087 to indicate medical subdomains (e.g., *anat*, *pat*)
- **definitions that are conceptually anchored**, preferably referencing the base noun term or conceptual unit to which the adjective belongs

- **clear lexicographic distinction between general-language and specialized meanings in cases where two distinct senses exist**, ideally through the use of separate lemmas (as seen in RHJ for *živčan* vs. *živčani*)

- **alignment with curated terminological resources**, such as medical thesauri, terminological databases (e.g., IATE), and national terminology infrastructures, particularly the Croatian terminology database Struna.

Such improvements would enhance not only the descriptive adequacy of general-purpose dictionaries but also their practical relevance in fields like health communication, public information, translation, and education. Users would benefit from clearer guidance on when a term has a specialized meaning, what domain it belongs to, and how it relates to other terms in the conceptual system of medicine.

In support of this research, a database of the 80 most frequent adjectives with medical meaning was created based on corpus data. As noted in earlier studies (e.g., Pitkänen-Heikkilä, 2015), adjectival terms are particularly abundant in domains such as medicine, botany, and chemistry. Therefore, the database was designed with a modular and expandable structure, enabling the inclusion of new lemmas and the extension into other specialized domains. This resource is intended to support future work in terminology management, dictionary compilation, and digital lexicography, and can serve as a basis for enriching e-terminology databases and e-dictionaries.

It is planned that the dataset will be published and made publicly available through the CLARIN-HR repository after final validation and updates. By providing structured, corpus-informed lexical data with potential for integration into broader language technology applications, this study contributes to the advancing terminological standardization and improving treatment of medical language in general lexicographic practice.

6. Acknowledgements

This work was created as part of the project Semantic Frames in the Croatian Language (SEF), funded by the European Union – NextGenerationEU, and within HR-CLARIN.

7. References

Alberts, M. (2001). Lexicography versus Terminograpy. *Lexikos*, 11, pp. 71–84.

Alonso Campos, A. & Torner Castells, S. (2010). Adjectives and collocations in specialized texts: lexicographic implications, In A. Dykstra & T. Schoonheim (eds.) *Proceedings of the XIV EURALEX International Congress*, Leeuwarden: Fyrske Akademy – Afûk, pp. 872–881.

Bergenholtz, H., & Tarp, S. (eds.) (1995). *Manual of Specialised Lexicography: The preparation of specialized dictionaries.* Amsterdam/Philadelphia: John Benjamins.

Cabré, M. T. (1999). *Terminology: Theory, methods, and applications*. Amsterdam/Philadelphia: John Benjamins Publishing Company.

Costa, R. (2013). Terminology and Specialised Lexicography: two complementary domains. *Lexicographica*, 29(1), pp. 29–42.

Cabré Castellvi, M. T. (2003). Theories of terminology: Their description, prescription and explanation. *Terminology*, 9 (2), pp. 163–199.

Durán-Muñoz, I. (2019). Adjectives and their keyness: a corpus-based analysis of tourism discourse in English. *Corpora*, 14 (3), pp. 351–378.

Fontenelle, T. (2016). From Lexicography to Terminology: a Cline, not a Dichotomy. In A. Abel; C. Vettori & N. Ralli (eds.) *Proceedings of the XVI EURALEX International Congress: The User in Focus*, Bolzano: Institute for Specialised Communication and Multilingualism, pp. 25–45.

Grčić Simeunović, L. (2015). Prilog metodologiji opisa sintagmi u stručnom diskursu. *Rasprave: Časopis Instituta za hrvatski jezik i jezikoslovlje*, 41(1), pp. 29–47.

Grčić Simeunović, L.; Stepišnik, U. & Vintar, Š. (2020). Klasifikacijska uloga pridjeva u domeni geomorfologije krša. *Rasprave: Časopis Instituta za hrvatski jezik i jezikoslovlje*, 46(2), pp. 619–633.

Grčić, Simeunović, L. (2021). *Terminološki opis u službi stručnoga prevođenja. Dinamično modeliranje specijaliziranoga znanja*. Zadar/Zagreb: Sveučilište u Zadru/Institut za hrvatski jezik i jezikoslovlje.

ISO 704:2009. *Terminology work – Principles and methods*. Geneva: International Organization for Standardization.

ISO 1087:2019. *Terminology Work – Vocabulary – Part 1: Theory and Application*. Geneva: International Organization for Standardization.

Jojić, Lj. & Matasović, R. (2000). *Hrvatski enciklopedijski rječnik* [Croatian Encyclopedic Dictionary]. Zagreb: Novi liber.

Jojić, Lj. (2015). *Veliki rječnik hrvatskoga standardnog jezika* [Great Dictionary of the Croatian Standard]. Zagreb: Školska knjiga.

Kocijan, K.; Kurolt, S. & Mijić, L. (2020). Building the Croatian medical dictionary from medical corpus. *Rasprave: Časopis Instituta za hrvatski jezik i jezikoslovlje*, 46(2), pp. 765–782.

Kilgarriff, A. et al. (2014). *The Sketch Engine: Ten years on*. Lexicography, 1(1), pp. 7–36.

Landau, S. I. (2001). *Dictionaries. The art and craft of lexicography*. Cambridge: Cambridge University Press.

Mel'čuk, I., & Polguère, A. (2018). Theory and practice of lexicographic definition. *Journal of Cognitive Science,* 19(4), 417–470.

Merriam-Webster. Accessed at: https://www.merriam-webster.com. (1-8 July 2025)

Pitkänen-Heikkilä, K. (2015). Adjectives as terms. *Terminology*, 21 (1), pp. 76–101.

Sager, J. C. (1990). *A practical course in terminology processing*. Amsterdam/Philadelphia: John Benjamins Publishing Company.

Salgado, A. (2021). *Terminological Methods in Lexicography: Conceptualising, Organising and Encoding Terms in General Language Dictionaries* (Doctoral

dissertation). Universidade Nova de Lisboa, Faculdade de Ciências Sociais e Humanas, Lisboa.

Salgado, A.; Costa, R. & Tasovac, T. (2022). Applying terminological methods to lexicographic work: terms and their domains. In A. Klosa-Kückelhaus et al. (eds.) *Dictionaries and Society. Proceedings of the XX EURALEX International Congress*, Mannheim: IDS-Verlag, pp. 181–195.

Salgado, A. & Costa, R. (2024). Enhancing Lexicographic Work with Terminological Methods. In Margalitadze, T. (ed.) *Proceedings of the I International Conference Lexicography in the XXI century: Lexicography – by combining traditional methods and modern technologies*, Tbilisi: Centre for Lexicography and Language Technologies, Ilia State University. pp. 15–26.

Šonje, J. (2000). *Rječnik hrvatskoga jezika* [Dictionary of the Croatian Language]. Zagreb: Leksikografski zavod Miroslav Krleža.

Temmerman, R. (2000). *Towards New Ways of Terminology Description. The Sociocognitive Approach.* Amsterdam/Philadelphia: John Benjamins Publishing Company.

Wüster, E. (1979). *Einführung in Die Allgemeine Terminologielehre Und Terminologische Lexikographie.* Beč/New York: Springer International Publishing.