CZU: 81‘322:81‘276.6=61 https://doi.org/10.59295/teia2025.17

DIGITAL DATABASES AND MEDICAL TRANSLATION: OPTIMIZING TERMINOLOGICAL ACCURACY IN THE HEALTH FIELD

Inga STOIANOVA

Free International University of Moldova

ORCID: 0000-0003-2413-3468

In today's digital environment, medical translators take advantage of digital tools and specialized terminology databases. The study offers an overview of e-resources like SNOMED CT, MeSH, UMLS, TermiumPlus®, and other medical terminology databases. It discusses the benefits of these tools in making terminology more consistent, their impact on the translation quality and the way translators can use both their language and cultural knowledge along with digital tools to produce accurate translations.

Key terms: *digital medical databases, accuracy, standardization, hierarchical organization of medical terminology, interoperability, granular complexity, medical translation*

BAZE DE DATE DIGITALE ŞI TRADUCEREA MEDICALĂ: OPTIMIZAREA ACURATEŢEI TERMINOLOGICE ÎN DOMENIUL SĂNĂTĂŢII

În mediul digital de astăzi, traducătorii în domeniul medicinii profită de instrumentele digitale și de bazele de date terminologice specializate. Studiul oferă o prezentare a resurselor digitale precum SNOMED CT, MeSH, UMLS, TermiumPlus® și alte baze de date de terminologie medicală. De asemenea, sunt evidențiate beneficiile acestor instrumente în uniformizarea terminologiei, impactul lor asupra calității traducerii și modul în care traducătorii pot aplica cunoștințele lingvistice, și culturale, precum și instrumentele digitale în traducerea medicală de calitate.

Cuvinte-cheie*: baze de date medicale digitale, acurateţe, standardizare, organizare ierarhică a terminologiei medicale, interoperabilitate, complexitate granulară, traducere medicală*

The medical translation field is at the nexus of language, medical science, and patient safety. Properly translated terms are essential to all facets of health care delivery, including public health communications, regulatory documentation, clinical trial operations, patient educational materials, etc. As such, it is incredibly important when it comes to the proper transfer of information between languages and making sure that any medical information is correctly translated. If medical terms are translated incorrectly, it can create confusion, lead to incorrect treatment, result in problems with regulation, a breakdown in trust between the patient and the health care provider [1, p. 394]. With the globalization of medicine, the increasing number of multilingual clinical trials and explosive growth in medical knowledge, there continues to be a dramatic increase in the amount and complexity of terminology that requires translation [2].

This article examines how digital databases support the use of terminology in medical translation. The main objectives are to review digital resources, identify the main linguistic challenges in medical translation, and assess how databases help reduce errors and improve translation accuracy. The study draws on theories from text translation, lexical and phraseological studies, and computer systems for medical information [3].

Medical translation is a significant field, and digital databases play a major role; used together optimal results may be achieved. As healthcare moves toward a more global economy and large collections of medical data continue to grow in number and language, medical terminological resources have emerged including MeSH, SNOMED CT, UMLS, ICD, Termium Plus® etc., becoming essential in identifying precise medical terms via the medical translator's profession.

Medical vocabulary forms the foundation of communication in healthcare, acting as a detailed, well organized, and specialized language that allows for accurate information sharing among clinicians, researchers, and, more recently, artificial intelligence systems. In contrast to everyday language, medical terminology is defined by its need for accuracy, strict standardization, and its ability to adapt across multiple languages to cater to global healthcare requirements. Yet, the same features that make them identical also create major difficulties, especially when medical data are communicated through different languages or when technology is involved. Mistakes in the use of medical terms can lead to severe consequences, hitting both the patients' and the public's health, causing misdiagnoses, putting patients' lives at risk, and making people sceptical about health care services. In this context, medical terminology arises not only as a set of specialized words but as an organized framework, that supports clinical communication, research, and health information management. To understand its role in today's medicine, we have to examine the main characteristics of medical terminology, such as its standardization, structured hierarchy, and semantic accuracy.

Standardization. The standardization of medical terminology is a global acceptance that guarantees interoperability, quick knowledge transfer, and patient safety. The coding systems like SNOMED CT and the UMLS provide internationally recognized codes or ontologies that are considered structured vocabularies. These vocabularies enable health care facilities and practitioners globally to communicate using the same terms, regardless of country, language, and institution. Standardized terminology provides a foundation for how to map symptom terms (for example, *myocardial infarction*) to the same object, regardless of the digital or written language of the country where the term is used. Furthermore, standardized terminology serves as the basis for constructing computational models and clinical decision support systems. For example, Schulz and Juric used SNOMED CT to create large datasets for the purpose of conducting semantic similarity analyses of medical terms [4, p. 4-6], while Yang and Huang described several of the reasons why standardization is essential as a component for facilitating the practice of machine learning within healthcare [5, p. 58].

Consequently, the hierarchical structure of medical terminology enables clinicians or researchers to use terms that are appropriate for a given level of specificity, while maintaining the logical relationship between one medical term and another (e.g., *parent to child* or *part to whole*). To illustrate the idea, SNOMED CT parent–child (*is–a*)

hierarchical relationship is provided: *pneumonia* is a subtype of *infectious disease*, and *bacterial pneumonia* represents a further specialization of *pneumonia*, where each child concept inherits the clinical meaning of its parent, while adding greater specificity. A *part–whole relationship* exists between anatomical concepts, such as the *left ventricle*, which is part of the *heart*, which in turn is part of the *cardiovascular system*. All of them allow SNOMED CT to model complex clinical knowledge accurately to facilitate precise documentation, semantic querying, and reliable data aggregation across healthcare systems. The structured arrangement of medical terminology is essential for data management and knowledge representation. It enables precise coding, effective information retrieval, and significant data aggregation for epidemiological research, health metrics, and policy development. Additionally, hierarchical organization supports interoperability among terminological systems and health information technologies, facilitating uniform understanding of clinical data across institutions, languages, and healthcare environments.

Accuracy. Medical terminology should contain accurate definition and eliminate any ambiguity. The intent of a term should be clear, whether that is an illness, a process, or a part of the body. Schulz and Juric emphasize that a physician must possess a robust vocabulary of medical terminology and a mature understanding of its semantic relationships, such as synonymous, hierarchical, and associative [4, p. 2]. In the context of computational systems, the accuracy requirement is heightened because AI systems rely on the accurate representation of terms to perform functionally similar operations on clinical data. The accuracy that distinguishes medical terminology also makes it difficult for non-experts to interpret. Accurately acquiring knowledge from medical dialogue is a challenge for automated systems, for example, as they must be able to distinguish medically appropriate language from expressions that are not. If an automated model cannot correctly recognize a medical term's meaning, any output it produces will likely be inaccurate, misleading, and could even be harmful [6, p. 1].

The healthcare industry is a global one, and healthcare providers need to use medical terminology that is adaptable to many languages and cultures. Being able to speak multiple languages means that there will be standardized terms for medical concepts available to healthcare professionals, patients, and research scientists all over the world. This will enhance their ability to collaborate internationally and provide medical care to diverse groups of people. For example, SNOMED CT is referred to as the "most comprehensive, multilingual clinical healthcare terminology in the world" [4, p. 4]. Nevertheless, true multilingual equivalence is very difficult to achieve, and direct term-for-term translations are often insufficient, as medical concepts may not have exact equivalents in other languages or may not have the same connotations or cultural meaning. This is further complicated in low-resource languages, where medical vocabularies have not been established, and there are no computational resources to support the translation of healthcare communication [7].

Digital medical databases are structured and electronically accessible repositories that systematically store, manage, and disseminate health-related information to support healthcare functions such as clinical decision-making, research, education, and administration. The main purpose of digital medical databases is to guarantee semantic

interoperability: the capability of different healthcare systems to share, understand, and apply clinical data without losing significance. Standardized terminology databases like SNOMED CT and ICD serve as the linguistic foundation for electronic health records, decision support systems, and AI models, guaranteeing that a diagnosis, procedure, or finding documented in one system is clearly comprehended by another, irrespective of language or context [4, p. 7].

The typology of digital medical databases can be broadly classified as follows:

- *Terminological databases* are used to store, manage, and retrieve standard medical terms and their relationships, which provide the linguistic and conceptual basis for accurate clinical documentation, coding, billing, and interoperability between healthcare systems (SNOMED CT, ICD, TERMIUM Plus® etc).
- *Bibliographic databases* index literature of the scientific type like journal articles, conference papers, and clinical guidelines. They are subsequently used for literature reviews, evidence-based practice, and knowledge synthesis such as in the case of PubMed and MEDLINE.
- *Clinical and Patient Databases* consist of patient-level data that includes EHRs, medical images, lab test results, and treatment histories, used for patient care, epidemiological research, and AI-assisted diagnostics [8, p.129].
- *Multilingual and Translation Databases*. As healthcare becomes more global, databases with multilingual terminologies and translation, such as TERMIUM Plus® and translation engines are crucial for cross-border communication and inclusivity [7].

The divisions between these categories are not well-defined, and many current systems combine characteristics from several typologies (e.g., clinical documentation, research, public health surveillance, and medical translation support).

The most influential systems are SNOMED CT, ICD, UMLS, Termium Plus®, each with a different purpose, but all contributing to semantic interoperability across healthcare domains.

SNOMED CT (***Systematized Nomenclature of Medicine – Clinical Terms***) is a multilingual most comprehensive clinical terminology in the world, designed to capture detailed clinical information at the point of care. It covers over 349,000 medical concepts, a broad spectrum of concepts related to diseases, clinical findings, procedures, body structures, microorganisms, and substances [9]. One of the main linguistic benefits of SNOMED CT is that language-neutral Concept IDs identify its concepts. Each SNOMED CT concept has a unique, language-independent identifier. This identifier acts like a permanent, universal label for the underlying clinical meaning, regardless of how clinicians describe it in different languages or vocabularies. For example, the term *fracture* is captured in terminology by a single CT ID (such as 125605004).

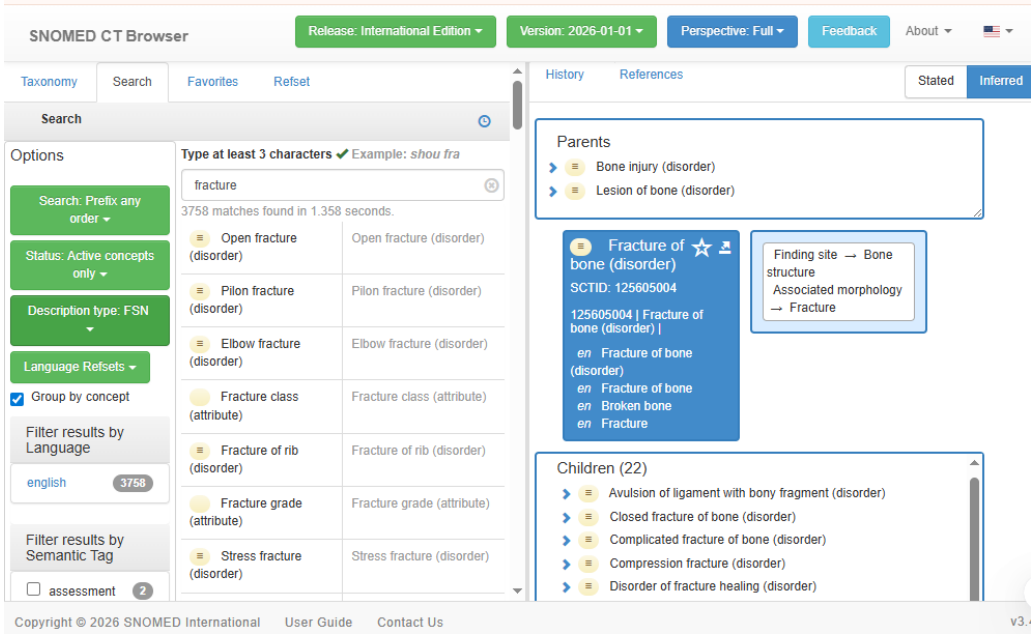


Fig. 1. Screenshot illustrating the SNOMED CT term *Fracture* [Accessed January 2, 2026].

Since Concept IDs are not only non-semantic but are also very stable, they act as interlingual pivots that enable a translator to find the equivalent term in the target language, even if the two languages do not share any lexical similarity at the level of the original text. For instance, a term can be found corresponding to different expressions in different languages, e.g., *accident vascular cerebral* in Romanian or *инсульт* in Russian for the English word *stroke* (SNOMED CT ID 230690007-IDs may vary by edition). Thus, by linking all these terms together with a single SNOMED CT concept, the translators can maintain the conceptual equivalence despite the variation in the lexical forms. The proposed way of working, thus making literal translation errors and false equivalences, is virally impossible.

From the point of view of language and translation interplay, SNOMED CT is an excellent example of how medical meaning can be standardized, restructured, and transferred across languages. Conventional terminologies, which have the tendency to mix up forms of words with the content of concepts, are contrasted here by SNOMED CT which actually insists on keeping a clear division between concepts (units of meaning) and descriptions (linguistic realizations). Moreover, the translators should be instructed that at times, the reason for not grasping a term or concept might not be their ignorance but rather a confusion in the core terminology (international release) that needs to be clarified. This difference matters a lot when it comes to the translation of medical documents, terminology management, and achieving interoperability among languages.

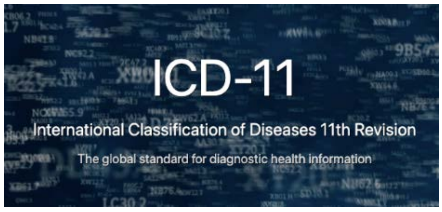
On the level of description, SNOMED CT offers an option to incorporate both multilingual variation and synonymy via its classification of description types. Preferred

terms are said to be those that most accurately represent the usages of the culturally and institutionally normalized languages. Synonyms are basically alternative terms that could be used in different contexts of the same language and have different language registers [8, p. 130]. As an illustration, the general concept of *myocardial infarction* (SNOMED CT Concept ID: 22298006) may be represented in the following way: the preferred clinical term goes hand in hand with the synonyms such as *infarction of heart, MI - myocardial infarction, myocardial infarct, cardiac infarction,* while *heart attack* is most frequently used in communication with patients. Consequently, the translators can make a decision whether the terms they want to use fit the communicative context or not, and at the same time, they remain in conceptual agreement with the standardized clinical meaning.

Nevertheless, the system introduces some difficulties in the translation area as well, especially if FSNs (Fully specified names) are considered. FSNs were not really made with linguistic naturalness in mind, but rather they are meant to be as terminologically precise as possible; therefore, they normally come with parenthetical semantic tags (e.g., *chronic kidney disease* (disorder), c*oronary artery bypass* graft (procedure)*).* The direct translation of FSNs is likely to produce, on the one hand, expressions that are linguistically unnatural, but, on the other hand, they might also be too explicit, e.g., FSN *appendectomy* (procedure) ID 80146002 and its synonyms *appendix removal and appendix excision*. Translating the FSN literally might yield *excizie de apendice (procedură)* in Romanian, which is technically accurate but awkward; the preferred term *apendicectomie* is natural in clinical use.

Post-coordination additionally hinders translation work. Here, translators are faced with the challenge of having to find the right equivalent conceptual compositions in their languages when the source-language clinical ideas are not a single concept. They are rather a complicated combination of the latter with various attributes [8, p. 131]. A post-coordinated phrase such as *fracture of femur* with *left* and *open* attributes requires quite a syntactic and semantic reorganization if the target language involved is one that has grammatical agreement or different word order rules and the like (English *left femour open fracture* vs Romanian *fractură deschisă a femurului stâng* vs Russian *открытый перелом левой бедренной кости*).

The International Classification of Diseases (ICD) is a coding framework created and overseen by the World Health Organization (WHO), mainly utilized for classifying diseases, epidemiological studies, and health data analytics [10]. In contrast to SNOMED CT, which is intended for intricate clinical expression, ICD is tailored for the aggregation, reporting, and comparability of health data internationally and throughout time. ICD codes serve purposes such as morbidity and mortality statistics, health policy formulation, reimbursement processes, and monitoring of public health. The shift to ICD-11 (launched in 2018 and approved by WHO in 2019) has brought about more detailed coding, enhanced digital access, and better connections with other nomenclature systems like SNOMED CT.



ICD is organized hierarchically to enable systematic recording and reporting of health concerns. At the most fundamental level, chapters classify wide categories of diseases. For instance, Chapter IX – Disorders of the circulatory system in ICD-10 comprises all heart and vascular ailments, whereas Chapter VI – Mental, behavioural or neurodevelopmental disorders in ICD-11 addresses issues like depression and anxiety. In chapters, diseases are arranged in sections that categorize related conditions. In the circulatory system chapter, the section I10– I15 *Hypertensive diseases* encompasses all disorders caused by high blood pressure. Every block is additionally segmented into categories, denoted by three-letter abbreviations. For example, *I10 – Essential (primary) hypertension* denotes the primary disease category without associated complications. To offer additional clinical data, categories are divided into subcategories. For instance, *I10.0 – Hypertensive heart disease with (congestive) heart failure,* indicating a causal link where uncontrolled high blood pressure leads to heart failure, often involving structural changes like hypertrophy. ICD-11 implements extension codes to document extra features like severity or laterality; for instance, XM0PW can signify mild severity, whereas *BD11.Z - left ventricular failure, unspecified* might indicate that the left side is involved. Each code is assigned a specific term and definition, ensuring consistency across clinical documentation, research, and statistical analysis. Another illustration comes from mental health: *6A70.3 - Single episode depressive disorder, severe, without psychotic symptoms; 6A71 - recurrent depressive disorder; 6A71.3 -recurrent depressive disorder, current episode severe, without psychotic symptoms; 6A71.1-recurrent depressive disorder, current episode moderate, without psychotic symptoms* providing very precise clinical details.

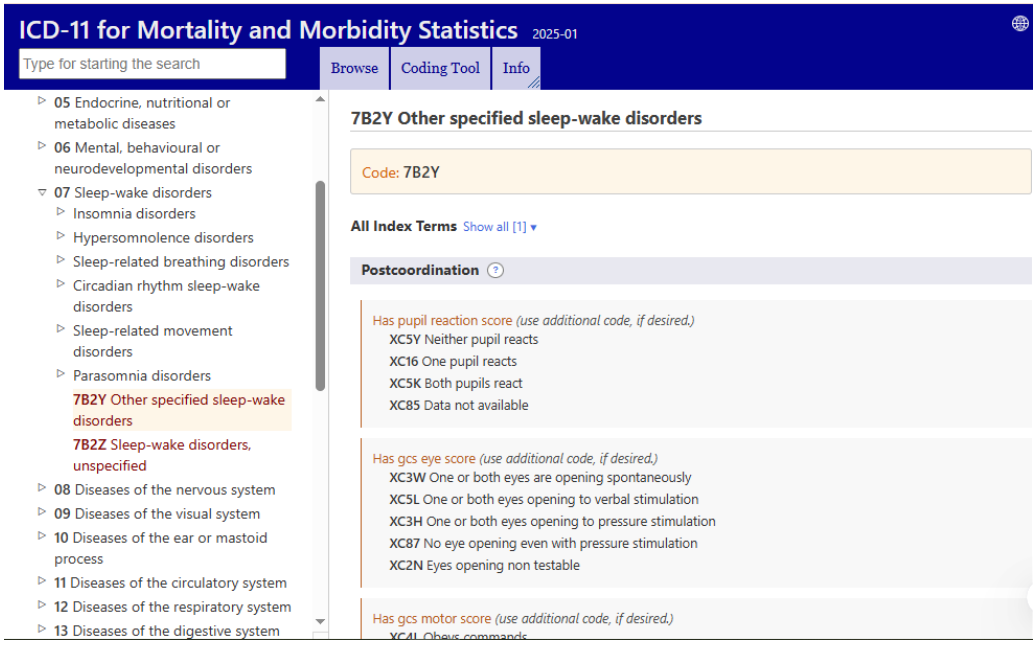


Fig. 2. The concept of *Sleep-Wake Disorders* in ICD -11 (screenshot) [Accessed January 2, 2026].

This hierarchical arrangement (from chapters to blocks, categories, subcategories, and extension codes) enables the ICD to support comprehensive epidemiological research and accurate patient classification. The effective implementation of ICD heavily depends on accurate translation into different languages. Since healthcare is inherently multilingual in various sectors, translating ICD codes and terminology is essential for consistent clinical documentation, research, epidemiology, and worldwide reporting [10].

Languages frequently use different terms to denote (the same) illnesses, and occasionally, classifications might differ slightly (common usage, abbreviations, or medical standards) based on regional medical practices. For instance, *type 2 diabetes mellitus* is translated as *diabet zaharat tip 2* in Romanian and *сахарный диабет 2 типа* in Russian, emphasizing the significance of consistent terminology across languages. Besides, languages vary in their construction of medical terminology. English frequently employs compound nouns or analytic structures, whereas Romance or Slavic languages might utilize phrases or morphologically intricate forms. For example, the English term *heart failure with reduced ejection fraction* becomes *insuficiență cardiacă cu fracție de ejecție redusă* in Romanian, which reorganizes the components and uses a prepositional construction. In Russian, it is *сердечная недостаточность с пониженной фракцией выброс,* where the word order and modifiers differ. These structural disparities can hinder automated translation, indexing, and integration with electronic health records.

Medical Subject Headings (MeSH) is a standardized word and phrases (controlled vocabulary) of biomedical terms produced by the U.S. National Library of Medicine (NLM) to help index, categorize, and search for articles. The idea behind this is that the same types of articles can be retrieved consistently using the same or different terminology. The structure of MeSH is hierarchical in nature, where several methods are available to search for medical literature, and the search can be refined to the level of specificity. The terms are presented as descriptors, which include qualifiers, as well as by Supplementary Concept Records (SCGs) for all compounds/drugs/chemicals, including emerging terms. Therefore, a descriptor could exist in more than one branch of the hierarchy, indicating the multifaceted aspects of biomedical knowledge. An example of this would be *diabetes mellitus* appearing in both endocrine system and metabolic disorders branches of the hierarchy, which provides flexible and context-sensitive searching. This polyhierarchical structure provides MeSH with an important support structure as it reflects the overlap of concepts in medicine.

MeSH is not designed to be used for clinical documentation; rather, it was designed to organize knowledge and information to facilitate biomedical research, systematic reviews, and evidence-based medicine [11].

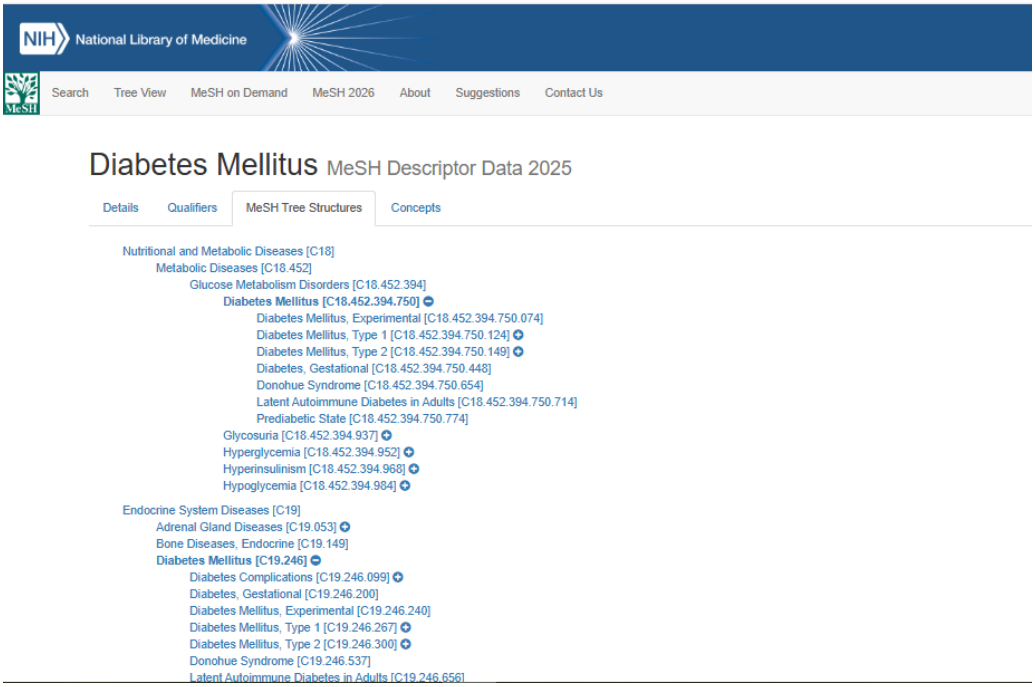


Fig. 3. Polyhierarchical representation of *Diabetes Mellitus* in MeSH (screenshot) [Accessed January 2, 2026].

Linguistically, MeSH is focused on controlling terminology and managing synonyms. A descriptor has its synonyms (which the indexer does not use), lexical equivalents, abbreviations, and other variations. Users can apply the synonyms to locate the best term, thus reducing confusion and ambiguity in the terminology. This is particularly helpful for medical translators, who can find the best term and alternative forms in the literature, and determine how to equate non-preferred terms to preferred terms for each text. MeSH also has a multilingual aspect, which is very helpful for translation and cross-linguistic studies. There are floating translations of the MeSH into French, Spanish, Portuguese, German, and others, and the floating translations keep the same concept as the best English term for the concept without being literal translations, especially because of the large number of composite nouns in English that are often replaced with flexional or prepositional forms in the Romance languages.

From a clinical and translational standpoint, MeSH is somewhat limited because it is designed as a very broad indexing system for literature. For example, it lacks the necessary detailed descriptions of clinical aspects of disease processes and the ability to accurately represent the finer distinctions between individual diseases and conditions [11]. As such, it does not provide fine-grained diagnostic details (e.g., the level of granularity that would be required to be able to apply a diagnosis based on a specific patient's situation) or procedural data (i.e., detail on the procedure being performed on a specific patient). Compared to the above-mentioned databases, MeSH does not provide the same high level of detail. Additionally, MeSH does not have a short-term update cycle, and

thus, there can be delays in incorporating new diseases, technologies, or therapies as they emerge in public health settings where there are rapid changes.

The Unified Medical Language System (UMLS) is the USA meta-terminology source. It allows biomedical interoperability via a meta-terminology framework that offers a mechanism for mapping medical terminologies together through a series of overlays, as opposed to replacing one medical terminology with another (National Library of Medicine, 2023). It has a central role in extensive mining of text, biomedical informatics as a result of UMLS' comprehensive cross-database links and support of natural language processing applications. However, translators and the general public are unable to take advantage of UMLS' technical complexity and limited availability of languages to obtain direct access to UMLS' capabilities. UMLS is viewed primarily as an informatics tool and not a linguistic one, given that training is required for the majority of effective implementations of UMLS. The UMLS framework has three key elements: *Metathesaurus*, which maps similar concepts across a variety of terminologies; the *Semantic Network*, which groups similar concepts conceptually and describes their connections; and the *SPECIALIST Lexicon*, which provides tools to facilitate natural language processing. UMLS facilitates cross-terminology mapping and aligns semantics in biomedical research, retrieval of information from various databases, and multilingual applications [12].

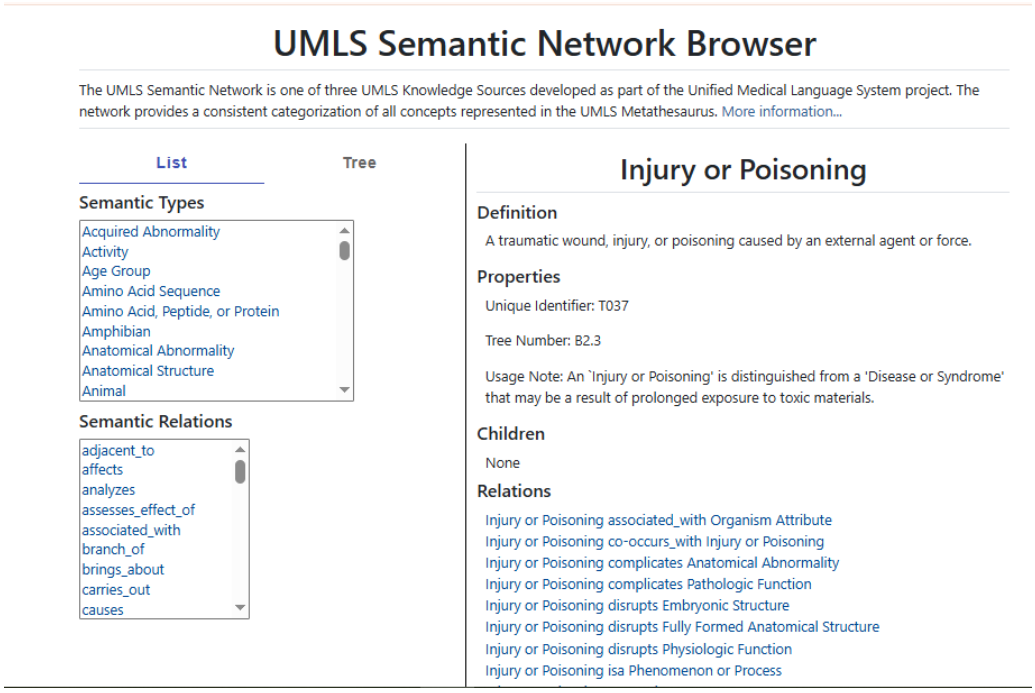


Fig.4. The UMLS Semantic Network for *Injury or Poisoning* (screenshot) [Accessed January 2, 2026].

Among the inherent complexities of the system are an extensive and layered set of hierarchies, along with a dense network of relationships that can complicate navigation, particularly for those without an adequate level of informatics training. This includes clinicians, linguists, and medical coders who may possess insufficient technical knowledge, skills, or computational tools necessary to use the system effectively. Additionally, the system has inherited inconsistencies from the source vocabularies used to build the UMLS because it incorporates different vocabularies generated for different purposes (clinical documentation, billing, indexing, regulatory reporting, etc.). As a result, some concepts were redundantly created, others have slightly different definitions or scopes, and some concepts are ambiguous or have multiple meanings, which is a significant issue when translating or cross-mapping [12].

The coverage of multilingual terminology is still uneven from a linguistic viewpoint. In many cases, there are numerous non-English terms present; and there has been limited focus on certain important linguistic phenomena, including compounding, inflection, and syntactic variation, which makes it difficult to achieve an exact cross-lingual equivalent. The UMLS was created to serve as a meta-thesaurus and not a point-of-care terminology; as a result, the number of concepts may be considered too abstract or too many for clinicians to incorporate into their clinical documentation. The final issue identified pertains to maintenance and version alignment; updating the UMLS regularly presents a challenge, with the potential for discrepancies to occur when integrating with other systems.

The Canadian government's official terminology and linguistic repository, **TERMIUM Plus®**, is a multilingual terminology resource. It contains standardized terminology in French and English, as well as some Spanish and Portuguese terminology; in addition, it has definitions, context examples, use notes and grammatical information, etc. For translators, the goal of TERMIUM Plus® is to assist them in finding the best and most appropriate term(s) to use in a translation; it does so through means of its use of attested usage, preferred term(s), and use of proper register. Unlike other systems, such as SNOMED CT and UMLS, which may be organized and represented as formal ontologies based on concepts, the TERMIUM Plus® focuses on the terminology being used and how they relate to each other for the practical use of translation [13].

The structure of TERMIUM Plus® gives preference to those lexical units as opposed to abstract biomedical concepts. Each record may have preferred term(s) and accepted variants of use; the grammatical category; the gender, where applicable; domain labels (medicine, pharmacology, public health, etc.), examples of context reflecting the authoritative use of the word; and usage notes indicating any constraints of regional, institutional, and stylistic nature. The focus of the structure of TERMIUM is in supporting those who are translating and who make sure the terms used are both semantically and practically correct and stylistically appropriate.

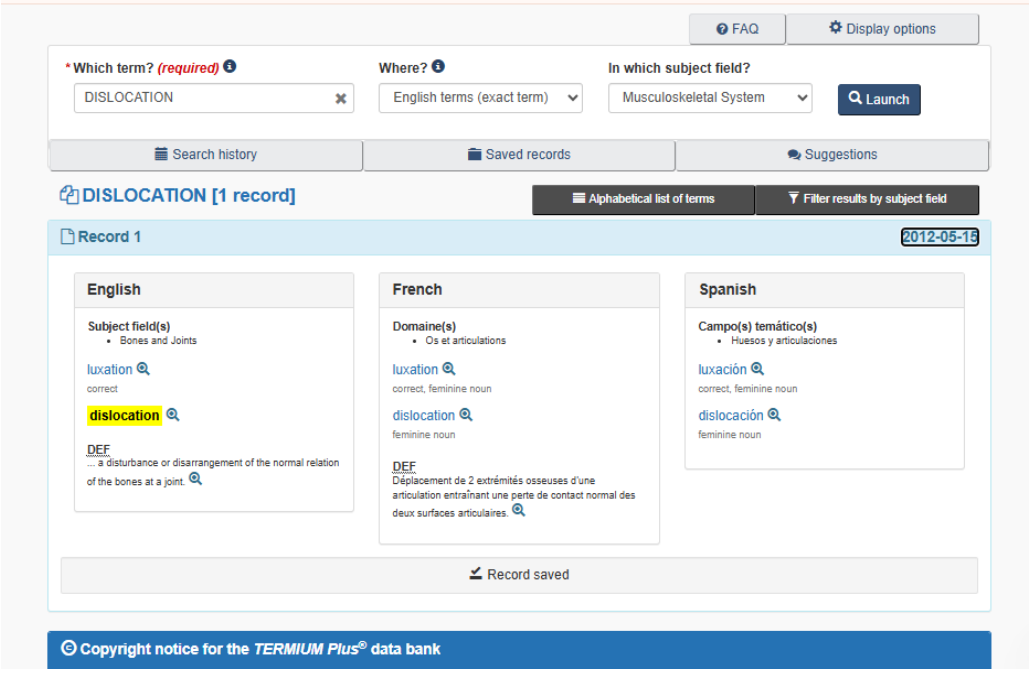


Fig.4. The concept *Dislocation* in TERMIUM Plus® (screenshot) [Accessed January 2, 2026].

One of the main benefits of using TERMIUM is its focus on the translation aspect of language. It allows for accurate cross-linguistic equivalencies across multiple languages, especially in areas where a literal translation could mislead. When creating bilingual medical documentation, this tool is invaluable for providing consistent terminology when resolving differences between British, American, and Canadian English; it helps to identify preferred versus deprecated terms that are crucial to consider when creating public health and regulation-related documentation. For instance, in English, the term *adverse drug reaction* translates to *réaction indésirable à un médicament* in French and keeps the same conventional phrasing and the same semantic meaning in both cases, allowing us to avoid the use of calques and non-idiomatic constructions. Another advantage of TERMIUM Plus® is its ability to handle morphosyntactic differences between working languages with ease, for instance, English compound nouns are often rendered in French as either analytic constructions. English compound nouns are typically expressed as either an analytic construction or a prepositional construction in French (e.g., disease *surveillance system* vs. *système de surveillance des maladies*). In addition to providing clear mappings of these terms for consistency in developing a translation memory, it also provides a system of recording the mappings, which permits building a consistent translation memory and the identification of similar terms or phrases across different languages.

From the standpoint of language, TERMIUM contains strategies for creating terms (compounding vs. analytic structures); differences in usage depending on the context of

language (technical, administrative, and common); rules of institutional language use (health policy, public administration etc.). For medical translators, the foundation allows a functional equivalent to be used instead of an identity for the terminology being translated; this can be especially useful when translating materials intended for patients, policy documents, or clinical guidelines, where clarity and acceptability may be more important than complete terminology [13].

There are some limitations regarding biomedical ontology and clinical terminology. Only English and French languages are included in the term base, meaning that it cannot be used in a multilingual setting. The term base does not provide formal concept identifiers or explicit hierarchical relations; therefore, it cannot provide the semantic reasoning capabilities to allow concepts to be normalized automatically, nor can it be used between different electronic health record systems. The content is selective rather than exhaustive, meaning that highly specialized clinical terms and/or rare diseases often have better representation in terms of covered content by SNOMED CT or UMLS. SNOMED CT provides machine-readable and clinical precision, with UMLS providing cross-terminology mapping, while TERMIUM provides access to practical terminology use.

These systems are complementary, not competing, although their scope and function are different: SNOMED CT, on the other hand, supports detailed forms of clinical documentation; ICD is used for standardizing reports or statistics between institutions, while UMLS helps create a common language between different systems of health information through its semantic framework.

All three (SNOMED CT, ICD & UMLS) assist in providing interoperable information systems for health care by allowing ways to share data between organizations and countries. They also support evidence-based medicine as well as growing use cases within health informatics, artificial intelligence and public health research requiring accurate terminology and concepts aligned to one another.

Therefore, numerous medical terminology databases have been developed since the rapid digitization of both biomedical research and healthcare to assist in standardizing, structuring, and disseminating specialized knowledge. These databases differ widely with regard to their scope, aims, granularity, and language compatibility; some focus on clinical documentation and interoperability, while others cover the areas of pharmacy, biomedical research, regulatory affairs, or multilingual communication. The diverse functions can lead to challenges associated with translating medical documents as well as harmonizing terminology across all languages and disciplines when analysing the description, strengths, and weaknesses of an individual database. A single medical terminology database cannot satisfy all the requirements for clinical, research, regulatory, and translational purposes. More user-friendly resources such as MeSH and ICD share much of the same linguistic/technical complexity, but grranular Systems, such as SNOMED CT and UMLS, excel at interoperability.

References:

1. CIMINO, J. Desiderata for controlled medical vocabularies in the twenty-first century. *Methods of Information in Medicine*. 1998. No. 37(4-5), pp. 394–403.
2. BOWKER, L., PEARSON, J. *Working with Specialized Language: A Practical Guide to Using Corpora*. London: Routledge, 2002. 256 p.

3. GOODMAN N. *Medical Writing: A Prescription for Clarity*. Cambridge: Cambridge Univ. Press, 2014. 382 p.
4. SCHULZ CL., JURIC D. *Can Embeddings Adequately Represent Medical Terminology? New Large-Scale Medical Term Similarity Datasets Have the Answer!*' [Retrieved 3 December, 2025]. Available at: https://arxiv.org/abs/2003.11082
5. YANG, J., LI, Y., LIU, Q., LI, L., et al. Brief introduction of medical database and data mining technology in big data era. In *Journal of Evidence-Based Medicine*. 2020. Nr.13, pp. 57 - 69.
6. TANG CH., ZHANG H., et al. *Terminology-aware Medical Dialogue Generation*. [Retrieved 20 December, 2025]. Available at: https://arxiv.org/pdf/2210.15551v2
7. DELFANI J., ORASAN C., HADEEL S. et al. *Google Translate Error Analysis for Mental Healthcare Information: Evaluating Accuracy, Comprehensibility, and Implications for Multilingual Healthcare Communication*. [Retrieved 15 December, 2025]. Available at: https://arxiv.org/pdf/2402.04023v1
8. BODENREIDER O., CORNET R. et al, Recent Developments in Clinical Terminologies — SNOMED CT, LOINC, and RxNorm. [Retrieved 23December, 2025]. Available at: https://pmc.ncbi.nlm.nih.gov/articles/PMC6115234/pdf/10-1055-s-0038-1667077.pdf
9. SNOMED International. *SNOMED CT International Edition.* 2022. [Retrieved 17 December 2025]. Available on: https://www.snomed.org/
10. INTERNATIONAL CLASSIFICATION OF DISEASES *11th Revision (ICD-11)*. [Retrieved 16 December 2025]. Available on: https://icd.who.int.
11. MEDICAL SUBJECT HEADINGS (MESH). National Library of Medicine. 2022. [Retrieved 15 December 2025]. Available on: https://meshb.nlm.nih.gov.
12. UNIFIED MEDICAL LANGUAGE SYSTEM *(UMLS).* Retrieved 20 December 2025]. Available on: https://www.nlm.nih.gov/research/umls.
13. *TERMIUM Plus.* [Retrieved 20 December 2025]. Available on: https://www.btb.termiumplus.gc.ca/