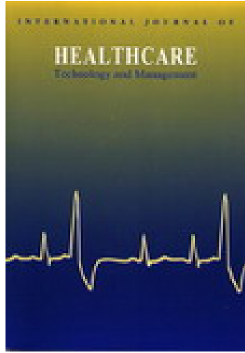
INDERSCIENCE PUBLISHERS
Linking academia, business and industry through research



International Journal of Healthcare Technology and Management

ISSN online: 1741-5144 - ISSN print: 1368-2156
https://www.inderscience.com/ijhtm

AI and linguistics give a new meaning to the concept of semantic interoperability in electronic medical records systems

Leslie Klieb, Marko Niinimaki, Raghid El-Yafouri, Pinnaree Tea-Makorn, Thanakij Wanavit, Samuel Sallee, Atamfon Udofia

DOI: 10.1504/IJHTM.2026.10079160

Article History:

Received:	23 October 2025
Last revised:	27 April 2026
Accepted:	09 May 2026
Published online:	18 June 2026

Copyright © 2024 Inderscience Enterprises Ltd.

AI and linguistics give a new meaning to the concept of semantic interoperability in electronic medical records systems

Leslie Klieb

Tenxor Inc.,
584 Castro St #2639, San Francisco,
CA 94114-2512, USA

and

Webster University Tashkent,
Navoi Avenue 11A, 100011, Tashkent, Uzbekistan
Email: leslieklieb@tenxor.sh

Marko Niinimaki*

Chulalongkorn School of Integrated Innovation,
Chamchuri 10, Wang Mai, Pathum Wan, 10330, Bangkok
Email: marko.n@chula.ac.th
*Corresponding author

Raghid El-Yafouri

Fanooce LLC,
2665 Aster Dr, Palm Harbor, FL 34684, USA
Email: raghid.el-yafouri@fanooce.com

Pinnaree Tea-Makorn

Sasin School of Management,
Soi Chulalongkorn 12, Wang Mai,
Pathum Wan, 10330, Bangkok
Email: pinnaree.tea-makorn@sasin.edu

Thanakij Wanavit and Samuel Sallee

Tenxor Inc.,
584 Castro St #2639, San Francisco,
CA 94114-2512, USA
Email: nic@tenxor.sh
Email: sam@tenxor.sh

Copyright © The Author(s) 2024. Published by Inderscience Publishers Ltd. This is an Open Access Article distributed under the CC BY license. (http://creativecommons.org/licenses/by/4.0/)

Atamfon Udofia

Chulalongkorn School of Integrated Innovation,
Chamchuri 10, Wang Mai, Pathum Wan, 10330, Bangkok
Email: 6558104256@student.chula.ac.th

Abstract: Interoperability, the exchange of information from a user of one electronic medical record system (EMR) to a user on another system, is of practical importance and is theoretically an unsolved problem. Most modelling has used ontologies and assumed that meaning of texts and information can be preserved. This elusive goal was named semantic interoperability and was never realised in practice. This work shows, using deep arguments from linguistics, that fundamentally this is a problem stemming from what "meaning" in human communication is. Certain ambiguities are studied in linguistics and cannot be avoided. The problems are especially pertinent if the sender and receiver come from a different culture, have a different medical background, or the receiver is a patient with limited medical knowledge. Also discussed is to what degree large language models can improve on these fundamental linguistic problems and on ontologies.

Keywords: EMR; electronic medical record systems; semantic interoperability; large language models; automatic translation of medical records; linguistic approaches to ambiguities.

Reference to this paper should be made as follows: Klieb, L., Niinimaki, M., El-Yafouri, R., Tea-Makorn, P., Wanavit, T., Sallee, S. and Udofia, A. (2024) 'AI and linguistics give a new meaning to the concept of semantic interoperability in electronic medical records systems', *Int. J. Healthcare Technology and Management*, Vol. 21, No. 5, pp.1–27.

Biographical notes: Leslie Klieb earned his PhD in Theoretical Physics from the State University of Groningen, the Netherlands. After some post-doctoral work in physics at the Racah Institute of Physics of the Hebrew University, Jerusalem, Israel, and elsewhere, he taught computer science in the USA. His interest in statistical analysis led him to quantitative business research. He has been a supervisor of DBA candidates at the Grenoble School of Management, France, and taught online complexity science and statistics in the PhD Doctoral Program for a business degree of the University of Liverpool in Association with Laureate, Inc. He is currently an assistant professor at Webster University in Tashkent, Uzbekistan. His research interests are in consumer behaviour, healthcare and linguistics.

Marko Niinimaki gained his PhD in Computer Science at the University of Tampere, Finland. He has worked as a researcher at CERN, software specialist in private companies in Japan and Thailand and lectured computer science in many universities in Europe and Asia. His special interests are artificial intelligence, databases and data management. He currently works at Chulalongkorn University School of Integrated Innovation.

Raghid El-Yafouri, DBA, is a Principal Consultant, Practitioner-Researcher, and Adjunct Faculty specialising in operations, performance optimisation, and technology-enabled transformation. He helps organisations improve outcomes by aligning people, processes, and systems to support effective execution. With more than 25 years of experience across healthcare and enterprise environments, his work spans product strategy, system integration, and compliance-driven operations. He holds a Doctorate in Innovation Management

from Grenoble École de Management and has published peer-reviewed research on digital transformation, health IT adoption, and the human factors influencing technology acceptance.

Pinnaree Tea-Makorn is a Lecturer and AI Strategist at Sasin School of Management and a founding member of Sasin's Artificial Intelligence and Data Analytics Center. Her expertise is in computational psychology, artificial intelligence, and data science. She teaches AI & Business to Sasin's MBA and Executive MBA students, as well as various executive education and corporate programs. She received her PhD, MS, and BS in Electrical Engineering from Stanford University.

Thanakij Wanavit is a Technologist and Entrepreneur specialising in artificial intelligence, deep learning, blockchain, and software engineering. He is the Founder of Tenxor Inc, Co-Founder of Capyverse and Head of Software Engineering at Meticuly Co., Ltd., where he leads AI and computer vision initiatives for medical applications. He holds a Bachelor's degree in Integrated Engineering (UK, 2012) and an MBA (USA, 2021). His work spans AI research, blockchain innovation, manufacturing technology, and venture investment, with publications and award-winning research in neural computation and deep learning.

Samuel Sallee served in the military from 2013 to 2018 before pursuing a degree in Computer Science at Webster University Thailand, where he graduated in May 2022. Since graduating, he has worked in software development and artificial intelligence, including as an AI engineer and researcher, where he contributed to projects involving medical image segmentation and AI recruiting. His interests include hobby programming, AI and machine learning, and most importantly, his family.

Atamfon Udofia is a graduate of the Chulalongkorn School of Integrated Innovation with experience in scientific research, innovation, social impact, and technology development. Her work focuses on innovation-driven solutions across education, sustainable development, and digital systems, with particular interests in African development, entrepreneurship, and inclusive economic growth. She has contributed to research projects, consulting initiatives, and technology solutions across Southeast Asia and Africa.

1 Introduction

The interoperability of electronic medical record (EMR) systems has long been a critical challenge in healthcare, with implications for patient care, research, and global health initiatives. While much attention has been focused on the technical aspects of data exchange, the complexities of medical communication extend far beyond code conversion. In this work, the interactions are the transfer of medical information that was entered into an EMR by a user to a user at another EMR. The two users may possibly speak a different language, have a different cultural background, different specialty (specialist, GP, even patient or caretaker), and different type of medical education. The two users may also differ much less in background and speak the same native language. Every change that the system must make so that the report is more understandable for the receiver is in linguistic terms a translation. Linguistics knows that it is impossible to

make boundaries between languages, dialects, sociolects (the language of social groups), jargon (professional language) and all possible other variations in communication. That is also not necessary for this work.

A summary of this work is as follows. In semantic interoperability, one of several levels of interoperability, meaning is supposed to be preserved. Examples (Adebesin et al., 2013; Heiler, 1995; Tolk et al., 2007, 2020).

It is assumed implicitly in semantic operability that the meaning of words or text is clear and unambiguous and conversion can be done at the level of systems in interaction with other systems. By considering the extreme cases of speakers of different native languages who are different in training and culture, it will be seen that meaning cannot be always unambiguously determined and that the standard way of using ontologies for medical terminology will fall short. Linguistics and translation science have heuristic ways of defining meanings that accommodate ambiguities but will not be able to always transmit the perfect meaning that the input user had in mind. This presents fundamental limitations to what semantic interoperability can accomplish. In order to present the issues in the most persuasive way, this paper uses translations from one natural language to another. But the same problems show in any human fallible communication.

The second part of this work will argue that linguistic ways of pinning down meanings are also leading to more useful transmissions of meaning in medical translations. Apart from promising higher accuracy, they are closer to how Large Language Models (LLMs) embed tokens, so that those can perform reasonable quality translations. Unfortunately, this will introduce ambiguities in the translation of the natural language and the medical terminology. This is empirically shown using a publicly available custom app (Niinimaki). Techniques like Retrieval-Augmented Generation (RAG) (Gao et al., 2024), API lookup of databases, and MoE (Mixture of Experts in LLM) (Cai et al., 2025) can help with mitigating those problems, but some ambiguities will never be able to get resolved, according to the linguistic part of this work.

The conclusion is that the expectations and solutions for semantic interoperability with ontologies are overreaching but a satisfactory level of automatic translation is probably possible. However, errors will possibly always creep in. It will depend on the use case, what kind of users will be using the system, and their tolerance for errors if the benefits outweigh the disadvantages.

The point of departure of this work is that medical information transfer across different healthcare systems, languages, and user types presents unique challenges that traditional approaches to semantic interoperability have struggled to overcome (Dolin et al., 2001). The language of medicine is not merely a set of standardised terms and codes; it is a rich, context-dependent communication system deeply rooted in the cultural practices of various medical specialties and healthcare systems worldwide (Ethier et al., 2018; Rosenbloom et al., 2011).

By focusing on these aspects, this paper aims to contribute to a more nuanced understanding of semantic interoperability in healthcare. To make the arguments less abstract, we show a list of possible errors in a conversion of medical texts to another language, specialty, cultural beliefs, etc., that any system should try to avoid or at least make transparent. We propose that generative LLMs can solve some of those issues but at the costs of introducing different kinds of mistakes. This work proposes a more holistic, AI-driven solution that respects the cultural and contextual richness of medical communication (Cimino, 1998).

LLMs can generate remarkably well human-like texts in many areas (Brown et al., 2020). Their ability to capture contextual nuances and adapt to different writing styles makes them particularly promising for handling the complexities of medical communication (Esteva et al., 2019).

Unlike traditional rule-based systems or earlier machine learning approaches, LLMs can potentially grasp the subtle cultural and contextual cues embedded in medical texts, making them well-suited for this challenging task. But a naïve implementation will suffer from hallucinations and lack of understanding of the context and so, while gains come up in one area, new problems will arise in different areas.

The research questions of this work can be formulated as “Is perfect semantic interoperability possible?” (which is answered as ‘no’) and “What possibilities, difficulties and fundamental limitations are there for automatic transformation of medical patient data to a different EMR system in the most general way, using a different natural language and code system?”, with sub-questions: “What linguistic issues cause problems to do this?” and “How can large language models and connected techniques be leveraged to mitigate these problems?”

This work assumes that a safe and secure way to access global EMR systems has been developed if needed in the best interests of the patient. This is a difficult problem with mostly legal aspects, especially when the patient is not able to give consent in emergency situations or from impairments and no legal representative is known. The discussions here assume free unlimited access to patient records.

Accordingly, this paper is intended as a conceptual and explanatory contribution that examines why semantic interoperability remains difficult in practice, rather than proposing a new technical standard or implementation framework.

2 Literature overview

A selection of the relevant literature is discussed in this chapter.

2.1 Levels of interoperability

The topic of the interoperability of two disparate electronic medical record systems (EMRs) has a long history of being discussed in the scientific literature. The final goal of interoperability is clear. Physicians in any part of the world should be able to look up a patient record in a system anywhere else as easily as on their local systems. Reasons are e.g., emergencies (Potin et al., 2024), continuity of care for travellers, migrants (Chongthawonsatid, 2015) and medical tourism, medical research, and minimise redundancy in the administration of medical tests and save hospital time. However, medical professionals (physicians, physical therapists, etc.) in different countries may have had a different education, have different national and organisational protocols to follow, or have different cultural backgrounds. It seems therefore wise to first leave out the users and discuss only the transmission of medical information from one system to another system.

Different authors have published proposals to classify the levels of interoperability in different ways. This work uses the HIMMS classification as a useful handle, but the reasoning applies to every kind of classification as long as there is an equivalent semantic interoperability. HIMMS defines interoperability as “exchange, integrate and

cooperatively use data in a coordinated manner, within and across organisational, regional and national boundaries, to provide timely and seamless portability of information and optimise the health of individuals and populations globally" (HIMSS-Interoperability, 2020). It also adds "Health data exchange architectures, application interfaces and standards enable data to be accessed and shared appropriately and securely across the complete spectrum of care, within all applicable settings and with relevant stakeholders, including the individual". This definition covers everything from entering the data up to its transport to users of the other system and explicitly includes the patient as well. It not only considers the flow from machine to machine, but also includes the asynchronous interaction between the creator of medical information on one system and the user of that medical information on a different system.

The HIMSS classification provides four levels for interoperability (Literal citations are quoted, some remarks are added).

- "*Foundational* (*Level 1*): Establishes the inter-connectivity requirements needed for one system or application to securely communicate data to and receive data from another" (HIMSS-Interoperability, 2020).
  - This is satisfactorily addressed by TCP/IP and similar protocols.
- "*Structural* (*Level 2*): Defines the format, syntax and organisation of data exchange including at the data field level for interpretation" (HIMSS-Interoperability, 2020).
  - Many authors call this level syntactic. Examples of such structures are XML (with delimiters '<' and '>') and HL7 (which uses the pipe character as a delimiter for data segments (Benson and Grieve, 2021)) and SQL. This only indicates how to parse the message. HL7 uses segments, groupings of fields. Content is not prescribed (it is for instance a patient name) but the kind of fields and segments (with names like MSH – message header -, etc.) are defined by standards.
- "*Semantic* (*Level 3*): Provides for common underlying models and codification of the data including the use of data elements with standardised definitions from publicly available value sets and coding vocabularies, providing shared understanding and meaning to the user" (HIMSS-Interoperability, 2020).
  - "Semantics is the study of linguistic meaning" (Saeed, 2015) Most authors state that at the semantic interoperability level, meaning is preserved during exchange of messages. Examples: (Adebesin et al., 2013; Heiler, 1995; Tolk et al., 2020).
  - The HIMSS classification does not give a definition of what semantic interoperability is, but what tools should be used to accomplish it. The word 'semantic' is only in the definition here as that the elements in the transfer should be standardised enough so that the user can understand the contents of the message. Meaning has become an afterthought.
  - The implication seems to be that enough of the meaning will emerge by itself if the tools are precise enough. However, as will be clear in discussions below, they will not satisfactorily determine meaning in the messy world in which healthcare operates. Linguists have shown that meaning cannot be separated from context (Cummings, 2015). LLMs can solve some of these real-life problems but not all of them.

  - Unstructured text like written medical reports will have problems with standardised coding vocabularies.

- "*Organisational* (*Level 4*): Includes governance, policy, social, legal and organisational considerations to facilitate the secure, seamless and timely communication and use of data both within and between organisations, entities and individuals. These components enable shared consent, trust and integrated end-user processes and workflows" (HIMSS-Interoperability, 2020).

  - Level 4 was not included in the original HIMSS level classification of 2013 but the current definition includes it.

  - The HIMSS definition of interoperability contains "use data in a coordinated manner, within and across organisational, regional and national boundaries" (HIMSS-Interoperability, 2020). Across national boundaries at this level would imply automatic translation of the source language and automatic adjustments for users (physicians and patients) who use other coding schemes, have had other educational background, follow other policies, and necessary clarifications for regions where the pharmaceutical market is different. This would need adjustments by an expert or use of automatic language translation tools. Developing dedicated translation systems would be prohibitively expensive and time-consuming. In this paper we try to understand how well LLMs can be used.

  - This work will not address workflow. Workflow interoperability is important in military applications (Tolk et al., 2007). There is no formal or informal language to describe workflows, so 'translating workflows' will be very difficult. Most workflows are implied and not explicitly written down.

Across the interoperability levels, standards such as HL7 FHIR, SNOMED-CT, and LOINC are commonly presented as mechanisms for achieving semantic interoperability, particularly at Level 3. However, their role is primarily to formalise structure, terminology, and exchange constraints, rather than to guarantee preservation of meaning across heterogeneous contexts. A scoping review of the electronic health record (EHR) interoperability literature shows that semantic inconsistencies persist even in settings where these standards are widely adopted, due to heterogeneous information models, partial or inconsistent mappings, and context-dependent interpretation by users and systems (El-Yafouri and Klieb, 2025). In this light, standards are treated in the present work as necessary infrastructural components of interoperability, but not as sufficient solutions for meaning preservation.

2.2 Ontologies and meaning

Meaning is not fixed. It is different for the sender, the user who inputs a medical record, and for the receiver, the user who needs the message as part of patient care or being a patient. This section focuses on the shortcomings of ontologies seen in the light of translation science (Langeveld, 2013) and pragmatics (Cummings, 2015). For the relationship between those two fields of linguistics, see for instance (Allan, 2007). Both provide a useful point of view and lead to the same conclusions about ontologies. Interoperability is about human communication and analysing it needs the science of human communication, i.e., linguistics.

It is not realistic to expect free-text medical reports to adhere to only using standardised terminology in order to fix meaning and bring down the ambiguity of human communication. Older research seems to have implicitly assumed that users of EMR systems are homogenous qua education (medical professionals), and qua cultural and linguistic background (speakers of English, usually in the USA), and would always understand each other completely. Terminology lists like SNOMED – CT (Systematised Nomenclature of Medicine – Clinical Terms (SNOMED-CT) provide a hierarchical structure (viral pneumonia is a descendent of pneumonia). SNOMED – CT is the most often used ontology, but the discussion here relates to other ontologies too. The background of the user of those tools is implicitly assumed to be a USA 'middle of the road' physician.

Empirical analyses of ontology-based interoperability reinforce these conceptual limitations. Comparative studies examining SNOMED-CT–centric implementations report issues such as inconsistent inheritance structures, ambiguous concept boundaries, and unsatisfiable mappings when ontologies are applied across heterogeneous clinical information models (El-Yafouri and Klieb, 2025). These findings suggest that ontology alignment errors are not exceptional edge cases but recurring properties of real-world EHR integrations. As a result, ontology-driven semantic interoperability remains fragile when extended beyond narrowly controlled contexts, particularly when free text, translation, or cross-specialty interpretation is involved.

El-Sappagh et al. (2018) provides a rather scathing literature review of problems in ontologies, with a focus on SNOMED – CT. The most important one listed there is that it is "not compliant with any formal upper-level ontology, and it allows for multiple inheritances, which causes a messy situation in the classification of entities". The lack of a formal upper-level ontology is a weak point. However, it is not possible to define a top-level concept that is guaranteed not to change and is the same globally in every culture. Examples are 'patient name' (where extensions were needed to encompass entry of names in cases of gender dysphoria) and mRNA, which was just an obscure term in genetics but later has gotten very different contexts after the development of mRNA vaccines. The relationships are many to many and break a clean hierarchy. Infections can be placed in a hierarchy with the body part, or with the pathogen. The 'IS_A' relationship is overloaded and can be for instance a 'PART_OF' or an 'IS_AN_INSTANCE_OF' relationship. Free-form medical reports often hedge their bets, 'patient has probably A or B'. Fuzzy ontologies can handle that but are not part of SNOMED-CT (Adel et al., 2019).

El-Sappagh et al. (2018) tries to remedy this situation by a more formal set-theoretic approach. SNOMED-CT gets updated monthly in an attempt to cope with changing knowledge and contexts. However, this does not help to differentiate 'before COVID' and 'after COVID', or that 'gall-bladder surgery' shifted its meaning from 'open surgery' to 'laparoscopy'. It is too optimistic that an ontology, a snapshot in a fast-evolving field like medicine, can function with fixed meaning. This raises the question: is it necessary to use ontologies and vocabularies to make EMR systems interoperable at the semantic and organisational HIMSS level? Is it necessary for the conservation of meaning to use tools like ontologies? Is conservation of meaning actually possible? This work tries to find an answer to these questions from a linguistic analysis of meaning, including pragmatics (Cummings, 2015) and considers the use of LLMs to 'translate'

from one specialist to another one in a different field, from specialist to family physician of therapist, and even from medical professionals to patients. Ontologies might be useful for lookups but not for full translations.

2.3 Linguistics and meaning

To extend the concept of interoperability between EMR systems to interacting systems with users at the organisational HIMSS level, it is necessary that the meaning of text is well enough preserved. Unfortunately, it will be shown that ‘complete preservation of meaning’ is often not possible.

In information science, one of the goals of ontologies is to study classifying entities and describing the relationships between them, like SNOMED-CT is doing. Defining a meaning has never been the purpose of building ontologies and their limitations become visible especially when translations are involved.

Besides in philosophy, ‘meaning’ has been studied in depth in linguistics. This work focuses on helpful results for written communications from linguistic research.

The meaning of a written (also oral, but this is not relevant here) communication is derived from many sources for the recipient. It is picked up at word level, but also at the sub-word level (Example in English: ****iser, like in organiser). Meaning can also be expressed by terms consisting of more than one more word, lexemes; medical terminology often describes an illness or syndrome by a combination of words. Cruse (2001) defines a lexeme as “a set of related meanings associated with a set of related word forms”. Higher in the hierarchy of expressions of meaning are paragraphs, chapters, and longer windows. All levels contribute to what the user considers is the overall meaning of the communication. Authors have a meaning in mind and are influenced in their choice of words by previous experiences and education, but so are readers. Ambiguities and misunderstanding are part of human interaction and considered unavoidable when the context is informative about the meaning of the communication (Piantadosi et al., 2012). This will cause problems for all translations, including medical ones.

While there are different approaches in linguistics, the separation into referential and connotative meaning of lexemes is the most straightforward and useful here.

Referential (also called denominative) meaning refers to something in reality, like an object, action, property, characteristic, or abstraction. The lexeme is a pointer to the object or concept.

But even if the referents are the same, the objects might not be necessarily interchangeable. For instance, trivially, in medical systems, the units of a measurement can be different. That other aspect of meaning is covered by the concept of connotation. This is (Allan, 2007) defined as the community and personal attitude of the sender. A slightly different definition is that it is the attitude of the original (source) author towards the referent (Langeveld, 2013). The receiver will have an understanding of the meaning that is hopefully aligned sufficiently with what the author meant. Pragmatics is the subfield of linguistics that treats interactions from this point of view (Stalnaker, 2002). In medical communications, context, education, background, personality, place of work, and many other characteristics may be different between sender and receiver and their environment influences the connotation. The connotations of the source are only partially knowable by the receiver. Connotations should be clear in a factual report, like most medical reports are, but this is often less transparent than one would think. Multi-lingual

terminology servers help in combating ambiguities and differences in understanding but are not enough. Remaining ambiguities should be infrequent and harmless enough that the probability of retrieving a sufficient part of the overall meaning of the communication and not distorting it is high enough, but make communication more efficient when contexts are meaningful (Piantadosi et al., 2012).

Not always it is possible to find an exact referential meaning of a lexeme in a source language (which can be the same natural language but used by a different type of medical professional), and in that case the recipient needs to be provided with an approximation, according to the HIMSS definition of organisational interoperability. An example is the brand name of a medicine that is marketed in the source country, but is available as a generic or under a different brand name in the country of the recipient. If the equivalence is not mentioned to the receiver, then not only the referential meaning is not transmitted, but understanding the report is made unnecessarily difficult for the reader.

Language has – at least – six functions (Jacobson, 1960). One is referential. The second function is poetic/aesthetic. Medical reports will usually not be written in a poetic style but the expressivity and clarity of the writing plays an important role to persuade the recipient reader of its value. Aesthetics is heavily influenced by culture, in this case the indoctrination that happens during a medical education. In some more subjective medical fields, like psychiatry, the style of a report is an important issue. The third function is emotive. Most medical reports are emotionally flat and distant. This might need to be adapted for a patient, so that the remoteness is not so obvious and a certain sympathy and compassion is showing. Another function is conative, it pushes the recipient to an action. It could instruct the patient to adopt a certain lifestyle. The phatic function, checking if the communication channel is working, is not relevant for EMR systems. The metalanguage function, communication about the message itself, is also not relevant.

Connotations can point to a hierarchy. A textbook is a book, it will have a 'book' connotation. But hierarchies are not often needed or useful for understanding a meaning – people do not use hierarchies consciously to derive a meaning from a lexeme.

Humans learn in a concentric manner. It is not a problem that expansion of vocabulary takes place by using familiar words. In LLMs the connotations are indicated by numbers that point to other vectors.

Translators need to find lexemes in the foreign language with the most appropriate referent and connotations while preserving the function of the language. Often enough there will not be an exact equivalent in the target language and some approximation is required. Translators make choices with respect to understandability and precision for technical documents and try to preserve the poetic emotion for some literary texts.

2.4 Large language models and meaning

LLMs like ChatGPT parse a text into tokens identified by integer numbers. A window of the text is used to look for the tokens of lexemes that are frequently found together. Those are embedded close to each other. The GPT (generative pre-trained transformer) architecture finds patterns that amount to connotations at certain coordinates. This is already sufficient to generate texts that give people the impression of a meaningful document. It shows that the hierarchical structure of an ontology with limited relationships is not as necessary and effective for letting people see meaning in the text as the messy many-to-many relationships that the transformer architecture builds. That

structure might miss insights into hierarchy and also maybe some other intuitive notions about meaning that people have, but it is much better than using other ways like ontologies.

The multi-dimensional space is a vector space for the embedded vectors, arbitrary axes can be chosen by rotations as base vectors. In some rotations a cloud of points can have simple coordinates, but often these axes will represent a weighted sum in a kind of connotation that is not recognisable for people. This makes following the decision process of the large language model and debugging of mistakes more difficult. A vector is much larger than the unit on the axes, because the length is determined by the theorem of Pythagoras. With hundreds or thousands of non-zero connotations the length is the square root of a sum over hundreds or thousands of the squares of the size of a unit on the axes of connotations. A corollary is that changing the meaning of related words by changing only a few connotations does not make a big difference in where the vector points to, and also not in its length (Vylomova et al., 2016). A large number of connotations in a natural language helps to determine the meaning more precisely and a ‘tiny language’ (Eldan and Li, 2023) cannot be too tiny.

2.5 EMR systems as finite-state machines

EMR systems are deterministic finite-state machines (Figure 1 first step). An input or deletion changes the state of the record, retrieval keeps the state as it was. Agents have predefined rules to follow. Within the organisational rules, like privacy, agents are free to choose their actions. In general, medical input consists of structured information like objects with names of tests with measured values and usually a reference interval, linked to demographic information about a patient and with a date. Other information, such as medical reports, mail, fax and email messages, is unstructured.

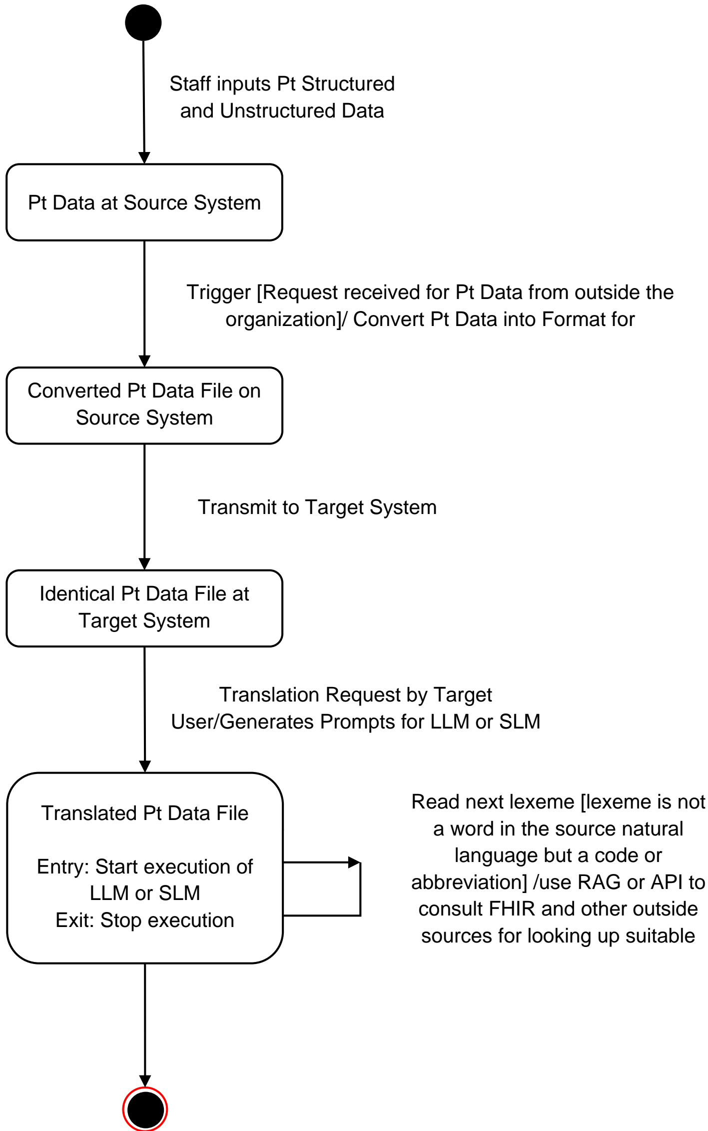
Retrieval may happen much later than shortly after input and in a different country with a different language. The value to the retriever is completely unrelated to the value to the one who entered the data. The connotative meaning of the text is different for the retriever, even if the referents are the same. As an aside, this makes that the person entering the data has no incentive to enter complete and accurate data for the benefit of other users, unless the user of the data is the same person as the inputter. Even in that case there might be a fear that a note or diagnosis is wrong and a reluctance to enter conjectures or self-incriminating divulgences about sub-optimal treatment. This contributes to very low user satisfaction with EMR systems (El-Yafouri et al., 2023) and a notion that some reports might be unreliable (Greiver et al., 2012).

Interacting EMR systems are, in a physical systems analysis, collaborative distributed multi-agent finite-state systems, comparable with a multi-player game. In such general systems, the states can differ between the two collaborative systems. The protocols for interacting finite-state machines are restricted (Brand and Zafiropulo, 1983) and certainly not Turing-complete. To generate arbitrary output on the target computer, other programs must be running on that machine besides a simple transformation using protocols. Simple transformations on ontologies are not sufficient.

For scaling reasons, it is impossible that the source system provides a translation for every system it might interact with. Therefore, the translation needs to be done on the target system. Translation means, as implied in the introduction, also to give an interpretation of the meaning that was intended by the sender in terms that the receiver can comprehend easier, faster, and without looking up unfamiliar expressions. Because a

request for patient records results in a one-way flow from source to target, there is in a logical system analysis only one state, one object, that simply jumps at one moment from one machine to another. Figure 1 describes such an UML state machine for the shared patient record object. If afterwards the source would like to see what the target has done, it will issue a request itself for the translated or modified result on the receiving system.

Figure 1 Logical description of the UML shared state machine of interoperating EMR systems (see online version for colours)



Medical codes, abbreviations and generic and brand names of medicine are usually not part of the vocabulary of a language translator, although ChatGPT-5 makes attempts. The codes might also be unfamiliar to caregivers in a different specialty from the sender. To

make the system more flexible, a language model is chosen as translator and not an application such as Google Translate. That language model can also 'translate' jargon into jargon in the same language but used by a different professional group. This is the final step in the UML state machine model. The architecture of the UML state machine mimics the role of human translators and their use of dictionaries.

Large language models are expensive to train and even transfer learning requires large resources. Eldan and Li (2023) shows that a natural ('tiny') language can be constructed with a small language model (SLM) on a small domain of knowledge, with a small vocabulary and a limited number of semantic fields. Those are "groups of words of which meanings are closely interrelated" (Gao and Xu, 2013) or "a set of lexemes which cover a certain conceptual area and which bear certain specifiable semantic relations to one another" (Faber and Usón, 1999). This can be used to generate texts in the tiny language. Also, the non-medical parts of the wording used by a certain group of physicians, like specialists in one field (neurosurgeons, general practitioners, etc.) are fairly uniform. Perplexity or LongPPL (Fang et al., 2025) should be generally low and this makes the lexemes fairly predictable. A smaller model could be used on its own or can be incorporated inside a larger model via the Mixture-of-Experts technology (Muennighoff et al., 2025), which would use the large model for general translation with an extensive vocabulary and a small medical model for specialised needs additionally to the RAG mechanism.

Medical codes, standardised names of medicines and treatments and dosages are easier to look up in already existing terminology servers like SNOMED-CT (SNOMED-CT) or ICD-11 (ICD-11). This allows the target system to get around the limitations of the kind of protocols that finite-state machines can use. This is symbolically indicated in the last step in the UML state model in Figure 1. It indicates that the translating Language Model is also able to use RAG, Retrieval-Augmented Generation (Gao et al., 2024) to obtain information from external sources, API calls (Patil et al., 2024; Liang et al., 2024), and dedicated programs and the MoE technology to use a dedicated trained language model. Figure 1 is not intended to specify an actual implementation.

2.6 Conclusions from the literature

A linguistic analysis shows that meaning of a lexeme is usually sufficiently described by labelling with a referent and by enough connotations. Alternatively, pragmatics can be used for the description. Ontologies fall short especially in handling the connotations and contexts.

The rest of this paper discusses the consequences of this model for the quality of the translated target paper. Section 3 discusses why it is impossible to change medical practice to avoid the need for translation. Section 4 draws directly from linguistic translation science and practice to show the structural similarities between natural language translation and medical translations. Section 5 analyses the pitfalls in translation. Section 6 presents the final conclusions.

3 Methodology for the empirical part of this work

The previous discussion implies that structured information can be generally translated in a straightforward fashion. Unstructured text is much more difficult. Unstructured text

contains ambiguities (Piantadosi et al., 2012; Boyarskaya, 2019). An important part of the work of a translator is to make choices for ambiguities, together with finding approximations for tone, context and other connotations in the original work. One could hope that the medical terminology is free of ambiguities and translation problems. Unfortunately, that is not the case. In the empirical part of this paper, it is explored if Large Language Models are able to provide interoperability at Level 3 and 4 of the HIMSS classification. The list of potential interoperability problems below is partly theoretical and partly empirical. It is not exhaustive. We used deidentified medical reports in American English with USA standards from (University_of_Chapel_Hill), a synthetic Autism Spectrum Disorder Diagnostic Assessment Report (PECS), a synthetic public example of the guidelines from the Psychiatric Association of the Netherlands in Dutch (Federatie Medische Specialisten), and a not publicly available Dutch technical medical note on cardiac surgery (permission for use granted by the owner). Translations were either done with an application (Niinimaki) built by one of the authors using ChatGPT version 3 or 4 via API, or directly using ChatGPT. The users of the application can upload a medical report and get a downloadable translation on screen. The application has options to translate into Dutch, French, and English. ChatGPT was instructed with the following prompts: “Translate this English text to Dutch for a medical specialist”, “Translate and summarise this English text to Dutch for a primary care physician” and “Translate and summarise this English text for a patient without medical knowledge or background”. For the Dutch to English translation text, the names of the languages were swapped.

The main focus of the experiments was on American English to Dutch translations. Dutch medical language has many slightly Dutchified calques adapted from and very close to English, but tricky exceptions exist and Dutchifying is sometimes done incorrectly by the LLMs. The app can translate and summarise at the level of specialists and family physicians, and write summaries that can be given to patients. The found problems can be used as a check-list in the quality control of automatic medical translations.

Using ChatGPT, we translated English also to Igbo, a language spoken in the South of Nigeria which is within the Benue-Congo branch of the Niger-Congo language family and has no connections with Western languages. ChatGPT uses the Western alphabet for Igbo and can show the pronunciation (tone) markings when prompted, although sometimes it seems necessary to prompt it more than one time. Tone markings distinguish between words in Igbo that have a different referential meaning and pronunciation, but are written in the same way. Medical terminology is not part of the Igbo language, because physicians in Nigeria use English (a mix of British and American English like all speakers there) to record patient information. Also, there is no standard code system. This makes complete semantic interoperability impossible. The quality of the translations by ChatGPT into Igbo was judged by one of the authors, an Igbo SL2 speaker, as moderate. This work restricts itself further to languages that seem to make semantic interoperability possible.

4 Issues with medical codes and abbreviations

Concrete examples illustrate how semantically valid representations can still fail to preserve meaning. Consider the exchange of a laboratory observation such as serum

glucose between two EHR systems. In HL7 FHIR, glucose measurements are typically represented using an Observation resource with a LOINC code identifying the analyte and optional SNOMED-CT concepts describing clinical interpretation. Although this structure enables syntactic and partial semantic interoperability, ambiguity can arise when contextual qualifiers – such as fasting status, measurement units (mg/dL vs. mmol/L), or reference intervals – are not consistently modelled or exchanged. Prior analyses have shown that such gaps can lead to clinically misleading interpretations despite standards-compliant data exchange (El-Yafouri and Klieb, 2025). Serum glucose measured from venous blood is a standardised clinical observation whose interpretation depends on measurement protocol and clinical context. The venous blood determination of serum glucose measures the glycaemic status of red blood cells and is usually extremely reliable to diagnose diabetes, except in certain sometimes not recognised circumstances. Certain red blood cells disorders, hemoglobinopathies, thalassemia, chronic and end-stage kidney disease, and some forms of anaemia will affect the glycaemic status (The Pathologist). In such cases, identical numerical values can carry different clinical meaning. Formally resolving this requires linking the glucose observation ontology to multiple disease ontologies, which are themselves interconnected, quickly yielding dense and unbounded dependency structures. This illustrates how ontology-centric approaches encounter scalability limits when meaning depends on context and interaction effects. By contrast, linguistic approaches to meaning – and large language models that approximate them – handle such dependence through distributed associations rather than explicit hierarchical alignment.

Standardised codes can in principle be looked up and translated automatically into different languages on other systems. The concept of semantic interoperability suggests that there is always a standard and correct way to do this. That is not true. Consider a medical report written by specialist in USA English using ICD codes. The remote target user is a general practitioner in a non-English speaking country who uses the codes meant for primary care ICPC-2 (WONCA) or ICPC-3 (ICPC-3 Foundation). ICPC-2 is much less detailed than ICPC-3. The semantic fields of the two codes are not equal. An automatic system could provide a translation for a medical code in various ways: leave the code unmodified, maybe with a note about the name of the code system; it could translate the description of the code; and it could provide the equivalent code that is known to the user. For instance, general physicians will read an ICPC code in the translation, maybe together with a description. Or a terminology server attempts only a description without any code. A clarifying description that is not in the source text is the equivalent of a translator note (Buendia, 2013). It would be best if this is standardised.

Abbreviations for medical procedures or symptoms are much more difficult to handle automatically. First, they must be recognised as abbreviations. The language models used in our experiments were not always able to do that. If recognised, then not always has the target language/jargon a similar abbreviation. Both in the Netherlands and in the USA the abbreviation ‘RR’ stands in cardiology for taking blood pressure measurements with a manual blood pressure meter with a cuff, so no problem. However, many abbreviations differ between the source language and some target languages. The Dutch abbreviation for ventricular tachycardia, ‘VT’, an arrhythmia of the lower heart chambers, corresponds to the English abbreviation ‘V-tach’. The difference in abbreviation is not logical but needs to be resolved.

Context (connotations) can be important. The RR-interval in English is the distance between two R-waves in an ECG (electrocardiogram). This is a different referential use of RR and there is no relationship with the use of RR as blood pressure, so this needs to be recognised by the translator. USA English MS is in cardiology Mitral Stenosis, calcification of a heart valve, but to a neurologist it is Multiple Sclerosis. There is no current mechanism to look up this kind of abbreviations conveniently ((HL7 FIHR) is an attempt). Using the context (connotations) of a document is difficult to perform automatically in all circumstances.

5 A systematic analysis of common problems in interoperability

Interoperability issues at HIMSS level 3 and 4 can often be identified but cannot always be solved due to lack of context or differences in practices and cultures. The lists below show many pitfalls and ambiguities that need to be resolved for a workable interoperability. The methodology to discuss those is similar to Langeveld (2013). One can wonder if general large language models that were not specifically trained for medical translations, are already good enough to do most translations. It was found that they are often good enough at the patient level and adequate at the family physician level, but are not sufficient for the kind of medical notes entered in hospital systems by specialists. According to its own reports, ChatGPT-5 has made some improvements in that area.

Errors found in actual translations are noted in the comment section of the lists below.

5.1 Linguistic full text problems

Issues with linguistic full texts as found below (Table 1) probably will need to be solved by training or transfer learning of the language model. Training on only medical texts might be the best solution. L3 and L4 refer to the levels of the HIMSS level for which this is relevant. L3 corresponds to semantic interoperability.

This list is not meant as a full inventory of the problems in automatic language translation. For that see for instance (Lu et al., 2025) (conclusion: unclear if quality is high enough), (Yang et al., 2023) (conclusion: promise), (Meddeb et al., 2024) (“LLMs excelled in clarity, readability, and consistency with the original meaning but showed moderate medical terminology accuracy”), (Lundmark and Boglind, 2024) (conclusion: no perfect results, crucial in medical context).

Our results below agree with all those conclusions but go further in pointing out that the errors that LLMs make cannot be solved by ontologies and are a consequence of the characteristics of language as human communication.

5.2 Linguistic localised text problems

Issues in this subsection are local to a small part of the text, as shown in Table 2. That makes resolving them easier than full text issues (Table 1). Errors here will not always make the target text useless.

Table 1 Full text issues

Source	*Target*	*Correct?*	*Example or Comment*
Full medical text with only general everyday medical terms	Medical text with only general everyday medical terms	Yes	via Google Translate or language models or similar for most but not all languages (L3, L4)
Medical text with terms, codes, and similar	Part or all of text is translated into something not medical or wrong specialty	No, connotative error	Example (E → D): 'Chief Complaint' translated into a judicial term, 'aanklacht' (=indictment), instead of klacht (=complaint). (L3)
Ditto	Target language has no official medical terminology	Depends on how situation is solved	The South-Nigerian language Igbo and many other languages in Nigeria and other countries. (L3, L4)
Standard medical terminology and no codes or abbreviations	Standard terminology exists and is used	Yes	In Great Britain, 'chest infection' and 'upper-respiratory infection' is used synonymously. The NHS (National Health Service) runs a FHIR (Fast Healthcare Interoperability Resources) compliant terminology server. It contains SNOMED – CT (Systemized Nomenclature of Medicine – Clinical Terms) (SNOMED-CT) and ICD–10 (an earlier version of (ICD–11)) and some other lists. However, "complex mapping tables between the loaded code systems are NOT currently available via the Terminology Server. Future software developments should enable this capability." (NHS). RAG, MeO (Cai et al., 2025), and API calls can often solve such issues. A third approach is to use specific training for the language model. (L3)
Ditto	Standard terminology exists and is not used but a description is correct	Probably acceptable	Healthcare professionals will probably recognise the diagnosis or procedure. Possibly confusing or misleading for lay people (L3, L4)
Ditto	Standard terminology exists and is not used and alternative description is incorrect	No	(L3)
Identifiers like name, address in one orthography (e.g., Cyrillic)	Imperfect or standardised transliteration or original orthography	Depending	Preferably adjusted to target users and the provisions of their systems. (L3) Extreme risks for identity theft, billing errors, and other identity problems
All of the above	Inconsistent but correct use	Acceptable	Annoying (L3, L4)
Ditto	Inconsistent and incorrect use	No	(L3, L4)
Layout and style	Layout and style not changed to target system	Usually Acceptable	Varied styles might be an issue for meta-analysis, international research, and similar projects. (L4)

Table 1 Full text issues (continued)

Source	*Target*	*Correct?*	*Example or Comment*
Ditto	Layout and style adjusted	Yes, usually preferred	(L3, L4)
Tone of text is appropriate for source but inappropriate for target	Tone is unchanged	Might not be acceptable	Some ways of addressing medical issues in medical reports or in person are inappropriate for Muslim patients (Laird et al., 2007; de Visser, 2024). The same is probably true for other religious or cultural groups, like anti-vaxxers. (L4)
Ditto	Tone is adjusted	Might be preferable	The message might get distorted. There is no general 'best' solution. (L4)
Full text	Summary for patient with little formal education	Sliding scale from very good to very poor	Often too technical in the current implementations. During a chatbot session with ChatGPT–4 one of the authors was asked to compare two summaries for somebody with only primary school education to help with the development of new features. The newer version was much less technical and much better, but was never shown again by ChatGPT. (L3)
Full text of any type	Elements that are not recognised cause the chatbot to hallucinate	Unacceptable	Often not recognisable for the user. The architecture of LLMs makes it very difficult to solve this problem. Human correction might be necessary. (L3)
Full text	Elements that are not recognised cause chatbot to make up a translation	Sliding scale from very good to very poor	E → D Many medical terms in Dutch are calques from English. Many instances were found where the medical term calqued into Dutch was recognisable but not fully correct. (L3, L4)
Date in source region	Date in Target region	Automatic Conversion sometimes possible, but not always. Cannot be resolved without context like location of the source	E → D USA MM/DD/(YY)YY, Europe DD/MM/(YY)YY. For dates like 1/2/2025, no automatic conversion is possible without knowing the source region. Thailand uses a Buddhist Era year in some hospital systems. Year 2025 is Buddhist year 2568
Reports from non-Western medical traditions like Traditional Thai and Chinese Medicine	Target can contain accepted translations or made-up translations	Sliding scale from very good to very poor	Chinese → English. Readers schooled in Western Science do usually not understand the flow of energy (Qi) and meridians in Chinese Traditional Medicine. (L3, L4)

Table 2 Localised issues

Source	*Target*	*Correct?*	*Example or Comment*
Semantic field	Semantic field for target differs considerably	Sliding scale from very good to very poor	Context matters. 'Underfed child' could be child abuse or victim of famine. The names of addictive street drugs, especially designer drugs, can change from region to region. Zaza can be cannabis (in rap music), but also Tianeptine, an antidepressant sold as a street drug
Drug name	Drug not or no longer marketed	Generic name or source brand name is both correct	Target user preference or complete description

5.3 Medical abbreviations

Medical abbreviations are part of the medical language but not officially regulated or standardised. Empirically, they were found to be an important cause of confusion and errors. In most countries there are lists of abbreviations from hospitals or other sources. Google's translation service seems to try to locate a description for most of them. The lists in the Netherlands are often incomplete, even the list in the Dutch Wikipedia is missing many common abbreviations. RAG or API fuzzy lookup for such abbreviations would be preferable if good sources are available, which in Dutch is not always the case. These issues are shown in Table 3.

Table 3 Issues with abbreviations

Source	*Target*	*Correct?*	*Example or Comment*
Abbreviation	An abbreviation or standard terminology exists in the target's country with identical referential and connotative meaning. Can be an identical or different abbreviation or a description	Correct	E → D. English RR identical
Ditto	Code is identical to source abbreviation but incorrect in target, or not identical and incorrect	Incorrect	Context: E → D English PE means Pulmonary Embolism. Dutch PE is Pericard Effusie, (pericardial effusion) (Cardiologie Central Nederland, 2021). Left as PE in Dutch text by ChatGPT.$PE is also Physical Education in American schools
Ditto	No abbreviation in the target's country or its language but description is correct	Correct	E → D: English HTN (hypertension) correctly given Dutch description
Ditto	Not correctly translated. Sometimes the LLM seems to hallucinate	Unacceptable	E → D ChatGPT gave RR as Dutch abbreviation for English HTN
Ditto	Skipped in target text	Incorrect	

5.4 Medical codes

Medical Codes (see Table 4) can be looked up in external sources, because they are standardised and regulated. Usually, APIs can be used to search for them. RAG lookups will also work.

Table 4 Issues with medical codes

Source	*Target*	*Correct?*	*Example or Comment*
Standardized medical code like from SNOMED – CT (SNOMED-CT)	Same code exists in target's country. Meaning is the same	Yes	Developers of medical codes usually provide an API and/or a browser page to look up such codes. (EVS Explore; Rogers and Bodenreider, 2008; Göbel) (L3)
Ditto	Code is correctly described	Acceptable if this is user preference	The SNOMED Navigator (Göbel) can integrate SNOMED codes using ChatGPT with dedicated prompts. Example (E → E): *Chief Complaint:* *"Swelling of tongue (SNOMED: 267036007) and difficulty breathing (SNOMED: 267036007) and swallowing (Dysphagia, SNOMED: 162397003)."* --- *History of Present Illness:* *..... woman with a history of Coronary Artery Disease (CAD, SNOMED: 53741008), Diabetes Mellitus Type 2 (DM2, SNOMED: 44054006), Asthma (SNOMED: 195967001), and ....* Abbreviations are transcribed, the original abbreviation given if in the source text, plus the SNOMED code. (L3)
Ditto	Code is incorrectly described	Incorrect	(L3)

5.5 Units and related issues

Medical units are usually but not always standardised per country. This presents a problem with transmissions to other countries (see Table 5).

Table 5 Unit conversion

Source	*Target*	*Correct?*	*Example or Comment*
Lab value with unit, reference interval and no or negligible lab variability for that type of measurement	If needed, system converts the lab value and reference interval, using the unit of the target	Correct	(L3)
Ditto	System does not convert but reports source unit correctly	Correct	This is not an error but not helpful for target user. (L3)
Lab value without mentioning the unit and/or reference intervals. No or negligible variability between labs for that type of measurement	Depending on the measurement, the used unit can be deduced and the conversion done	Correct	Example: Blood glucose units can be mmol/L and mg/dL The values in mg/dL are about 18 times the values in mmol/L. No confusion possible. For Hb1Ac, the USA uses mostly %, but Europe mmol/mol. Factor is not a constant, but varies between 5 and 10. Conversion table is necessary. (L3)
Ditto	Depending on the measurement, used unit cannot be deduced from its value because plausible ranges overlap	Automatic conversion without context is not possible Unsolvable problem in general	Knowledge of standard practices of source country might make a unit determination possible. This could be automatic, with the possibility of errors, or entered by the target retriever. (L3)
Measurement processes may differ or lab or other variability is high	An automatic judgement of the measurement is not possible	Unsatisfactory	Variability is the inconsistency in results for the same test by the same lab (Laboratory Variability). Local determination needed. (L3)
Reports with diagrams or results as images and units are visible or implied	No OCR available	Automatic solution	Reusing the images is the only practical solution. (L3)

5.6 Granularity

Granularity, the level of detail of descriptions, is different between specialists, general practitioners, and patients. That is one of the drivers for the ICPC (WONCA, 2005) system for general practitioners. Language models can summarise medical texts for different audiences. The need for a different level of granularity can come from different

levels of education and knowledge, but also from different experiences and culture. Issues with granularity are listed here. The ICPC system is mapped to ICD-10, but ICD-10 is orientated to diagnostics while ICPC is more orientated to complaints. There is a large difference in granularity between ICPC-2 and ICPC-3. ICPC-3 is more patient-centred and aligns better with SNOMED-CT and ICD-10 and up. These issues are shown in Table 6.

Table 6 Granularity (level of detail)

Source	*Target*	*Correct?*	*Example or Comment*
Report is written by a specialist	Report is aimed at a primary care physician but at similar level of difficulty as source	Yes (assuming target physician is sufficiently educated)	(L3)
Ditto	Report's difficulty level lowered for reading by family physicians, nurses, physical therapists, and similar professionals. No code adjustments	Yes, but problematic when code is not adjusted	Example: USA Pulmonologists ICD-11 (ICD-11) code. CA40.Z, 'pneumonia, organism unspecified' Family physicians in certain countries will use ICPC-2 code R81 or ICPC 3 code RD09. Pneumonia. (L3)
Ditto	Report's difficulty level lowered nurses, physical therapists, and similar professionals. Codes are adjusted if there is a corresponding code	Correct	Some of the intended professionals will not know the codes. (L3, L4)
Ditto	Report's difficulty level lowered, introducing errors, for nurses, physical therapists, and similar professionals	Incorrect	(L3, L4)
General practitioner, physical therapist, etc.	Upward adjustment for specialists. Downward adjustment for patient	Upwards impossible because lack of information. Downwards, e.g., to patient, possible	(L3, L4)

There are no universal solutions for those problems. It depends on the reader of a medical record what is helpful. In a Thai pharmacy it was observed that staff always says about foreign medication 'we do not have that' to lay people and tell customers that 'this medicine works in the same way' even when it is the same medicine in a different box. Instead of explaining brand names, generic names, and substitutes to foreign lay customers, the pharmacy tells everybody that it is a substitute. Technically incorrect, but works in practice. Specialists in the same field probably would prefer the generic name with a mention of the original name that they can use in conversations with the foreign

patient. What about different protocols? What could be needed are settings ranging from 'give as much information about possible issues as possible' to 'give limited information', maybe with some generic hints how issues are solved under the various settings. The best solution will vary with language, background, specialty, local protocols, branding, and others. Looking at those use cases shows again the limited applicability of a developer stage like 'semantic interoperability'.

6 Conclusions

The development of interacting EMR systems at the level of 'Semantic interoperability', where 'meaning' is always and automatically preserved, will encounter fundamental barriers for EMR systems that exchange medical reports in an unstructured way. The non-medical part can already have errors in automatic translations; the medical parts of the text can and usually will contain additional errors. Some of the ambiguities cannot be resolved because of missing information and context. A general form of interoperability between users, independent of their physical location, culture, education, and medical background, can replace the concept of semantic interoperability. HIMMS level 3 and similar levels in other ways to classify interoperability, like in Tolk et al. (2020) cannot be perfect. It will, however, be practically useful in many circumstances for tourists, travellers, migrants, and international workers.

The problems are fundamental. Current large language models will need external retrieval to improve the quality of their work and cannot provide interoperability by themselves. That conclusion is in contradiction with (Grimm et al., 2024). Each issue needs to be handled separately for actual implementation; there is no general solution possible. This work provides a checklist of possible errors, some of them empirically seen in actual translations.

In light of these limitations, this work argues that further progress will require approaches that complement ontologies with mechanisms better suited to handling linguistic context, ambiguity, and interpretation – like AI and NLP.

Acknowledgements

The authors want to thank the physicians Nic Vreeswijk and T.H.N. Rinkes and the translator Arthur Langeveld for helpful discussions.

Conflicts of interest

All authors declare that they have no conflicts of interest.

References

Adebesin, F., Foster, R., Kotzé, P., Greunen, D., Adel, E., El-Sappagh, S., Barakat, S. and Elmogy, M. (2013) 'A review of interoperability standards in e-health and imperatives for their adoption in Africa : research article', *South African Computer Journal*, Vol. 50, No. 1, pp.55–72, [Online], DOI: 10.10520/EJC139542.

Adel, E., El-Sappagh, S., Barakat, S. and Elmogy, M. (2019) 'Chapter-14 A unified fuzzy ontology for distributed electronic health record semantic interoperability', in Dey, N., Ashour, A.S., Fong, S.J. and Borra, S. (Eds): *U-Healthcare Monitoring Systems, Advances in Ubiquitous Sensing Applications for Healthcare*, Academic Press, London.

Allan, K. (2007) 'The pragmatics of connotation', *Journal of Pragmatics*, Vol. 39, No. 6, pp.1047–1057, [Online], DOI: 10.1016/j. pragma.2006.08.004.

Benson, T. and Grieve, G. (2021) 'Why interoperability is hard', in Benson, T. and Grieve, G (Eds): *Principles of Health Interoperability: FHIR, HL7 and SNOMED CT*, Health Information Technology standards. Springer International Publishing, Cham, pp.21–40, [Online], DOI: 10.1007/978-3-030-56883-2_2 (Accessed 23 July, 2023).

Boyarskaya, E. (2019) 'Ambiguity matters in linguistics and translation', *Slovo.ru: Baltic accent*, Vol. 10, pp.81–93, [Online], DOI: 10.5922/2225-5346-2019-3-6.

Brand, D. and Zafiropulo, P. (1983) 'On communicating finite-state machines', *Journal of the ACM*, Vol. 30, No. 2, pp.323–342, [Online], DOI: 10.1145/322374.322380.

Brown, T.B., Mann, B., Ryder, N., Subbiah, M., Kaplan, J.D., Dhariwal, P., Neelakantan, A., Shyam, P., Sastry, G., Askell, A., Agarwal, S., Herbert-Voss, A., Krueger, G., Henighan, T., Child, R., Ramesh, A., Ziegler, D., Wu, J., Winter, C., Hesse, C., Chen, M., Sigler, E., Litwin, M., Gray, S., Chess, B., Clark, J., Berner, C., McCandlish, S., Radford, A., Sutskever, I. and Amodei, D. (2020) 'Language models are few-shot learners', *Advances in Neural Information Processing* Systems, Curran Associates, Inc., Vol. 33, pp.1877–1901, [Online], Available at. https://papers.nips.cc/paper/2020/hash/1457c0d6bfcb4967418bfb8ac142f64a-Abstract.html (Accessed 24 August, 2025).

Buendia, C.T. (2013) 'Listening to the voice of the translator: a description of translator's notes as paratextual elements', *Translation and Interpreting: The International Journal of Translation and Interpreting Research*, Vol. 5, No. 2, pp.149–162, [Online], DOI: 10.3316/informit.466106332038749.

Cai, W., Jiang, J., Wang, F., Tang, J., Kim, S. and Huang, J. (2025) 'A survey on mixture of experts in large language models', *IEEE Transactions on Knowledge and Data Engineering*, Vol. 37, No. 7, pp.3896–3915, [Online], DOI: 10.1109/TKDE.2025.3554028.

Cardiologie Central Nederland (2021) *Medische Afkortingen*. [Online], Cardiologie Centra Nederland, Available at. https://www.cardiologiecentra.nl/kennis/medische-afkortingen/ (Accessed 21 February, 2025).

Chongthawonsatid, S. (2015) 'Medical records management of foreign patients in Thai border hospitals', *Journal of Health Research*, Vol. 29, No. 6, pp.473–480.

Cimino, J.J. (1998) 'Desiderata for controlled medical vocabularies in the twenty-first century', *Methods of Information in Medicine*, Vol. 37, Nos. 4–5, pp.394–403, [Online], DOI: 10.1055/s-0038-1634558.

Cruse, D.A. (2001) '10 The Lexicon', in Aronoff, M. and Rees-Miller, J. (Eds.): *Handbook of Linguistics*, Blackwell publishers, Blackwell Oxford, pp.238.

Cummings, L. (2015) 'Chapter 8 pragmatics', in Braber, N., Cummings, L. and Morrish, L. (Eds.): *Exploring Language and Linguistics*, Cambridge University Press, Cambridge, [Online], DOI: 10.1017/CBO9781139548922 (Accessed 1 April, 2026).

de Visser, E. (2024) *Cultuurverschillen in de zorg: Nederlandse directheid leidt soms (onbedoeld) tot boosheid*, [Online]. de Volkskrant, Available at. https://www.volkskrant.nl/wetenschap/cultuurverschillen-in-de-zorg-nederlandse-direct-heid-leidt-soms-onbedoeld-tot-boosheid~b493aba4/ (Accessed 21 February, 2025).

Dolin, R.H., Alschuler, L., Beebe, C., Biron, P.V., Boyer, S.L., Essin, D., Kimber, E., Lincoln, T. and Mattison, J.E. (2001) 'The HL7 clinical document architecture', *Journal of the American Medical Informatics Association*, Vol. 8, No. 6, pp.552–569, [Online], DOI: 10.1136/jamia. 2001.0080552.

Eldan, R. and Li, Y. (2023) *TinyStories: How Small Can Language Models Be and Still Speak Coherent English?*, arXiv, [Online], DOI: 10.48550/arXiv.2305.07759 (Accessed 24 October, 2023).

El-Sappagh, S., Franda, F., Ali, F. and Kwak, K.S. (2018) 'SNOMED CT standard ontology based on the ontology for general medical science', *BMC Medical Informatics and Decision Making*, Vol. 18, No. 1, p.76, [Online], DOI: 10.1186/s12911-018-0651-5.

El-Yafouri, R. and Klieb, L. (2025) 'A scoping review of electronic health records interoperability levels, expectations, approaches, and problems', *Health Informatics Journal*, Vol. 31, No. 4, p.14604582251385986, [Online], DOI: 10.1177/14604582251385986.

El-Yafouri, R., Klieb, L. and Sabatier, V. (2023) 'The impact of office-related metrics on meeting physician expectations from electronic medical record systems', *Healthcare Analytics*, Vol. 4, p.100208, [Online], DOI: 10.1016/j.health.2023.100208.

Esteva, A., Robicquet, A., Ramsundar, B., Kuleshov, V., DePristo, M., Chou, K., Cui, C., Corrado, G., Thrun, S. and Dean, J. (2019) 'A guide to deep learning in healthcare', *Nature Medicine*, Vol. 25, No. 1, pp.24–29, [Online], DOI: 10.1038/s41591-018-0316-z.

Ethier, J-F., McGilchrist, M., Barton, A., Cloutier, A.M., Curcin, V., Delaney, B.C. and Burgun, A. (2018) 'The TRANSFoRm project: experience and lessons learned regarding functional and interoperability requirements to support primary care', *Learning Health Systems*, Vol. 2, No. 2, p.e10037, [Online], DOI: 10.1002/lrh2.10037.

Faber, P.B. and Usón, R.M. (1999) *Constructing a Lexicon of English Verbs*, Mouton de Gruyter, Berlin – New York.

Fang, L., Wang, Y., Liu, Z., Zhang, C., Jegelka, S., Gao, J., Ding, B. and Wang, Y. (2025) *What is Wrong with Perplexity for Long-Context Language Modeling?* ArXiv, [Online], DOI: 10.48550/arXiv.2410.23771 (Accessed 14 March, 2025).

Gao, C. and Xu, B. (2013) 'The application of semantic field theory to English vocabulary learning', *Theory and Practice in Language Studies*, Vol. 3, No. 11, pp.2030–2035, [Online], DOI: 10.4304/tpls.3.11.2030-2035.

Gao, Y., Xiong, Y., Gao, X., Jia, K., Pan, J., Bi, Y., Dai, Y., Sun, J., Wang, M. and Wang, H. (2024) *Retrieval-Augmented Generation for Large Language Models: A Survey*. arXiv, [Online], DOI: 10.48550/arXiv.2312.10997 (Accessed 28 February, 2025).

Greiver, M., Barnsley, J., Glazier, R.H., Harvey, B.J., Moineddin, R. and Grimm, D.R. (2012) 'Measuring data reliability for preventive services in electronic medical records', *BMC Health Services Research*, Vol. 12, No. 1, p.116, [Online], DOI: 10.1186/1472-6963-12-116.

Grimm, D.R., Lee, Y.J., Hu, K., Liu, L., Garcia, O., Balakrishnan, K., Ayoub, N.F. (2024) 'The utility of ChatGPT as a generative medical translator', *European Archives of Oto-Rhino-Laryngology*, Vol. 281, No. 11, pp.6161–6165, [Online], DOI: 10.1007/s00405-024-08708-8.

Heiler, S. (1995) 'Semantic interoperability', *CM Computmg Surveys*, Vol. 27, No. 2, June, p.199.

HIMSS-Interoperability (2020) *Interoperability in Healthcare | HIMSS*, [Online], Available at. https://nationalcapitalarea.himss.org/resources/interoperability-healthcare (Accessed 19 April, 2026).

Jacobson (1960) 'Closing statements: linguistics and poetics', in Sebeok, T.A. (Eds.): *Style in Language*, MIT Press, Cambridge Massachusets, pp.350–377.

Laird, L.D., de Marrais, J. and Barnes, L.L. (2007) 'Portraying Islam and Muslims in MEDLINE: a content analysis', *Social Science and Medicine*, Vol. 65, No. 12, pp.2425–2439, [Online], DOI: 10.1016/j. socscimed.2007.07.029.

Langeveld, A (2013) *Vertalen Wat Er Staat*., 7th ed., Atlas Contact.

Liang, Y., Wu, C., Song, T., Wu, W., Xia, Y., Liu, Y., Ou, Y., Lu, S., Ji, L., Mao, S., Wang, Y., Shou, L., Gong, M. and Duan, N. (2024) 'TaskMatrix. AI: completing tasks by connecting foundation models with millions of APIs', *Intelligent Computing*, Vol. 3, p.0063, [Online], DOI: 10.34133/icomputing.0063.

Lu, S.C., Xu, C., Kaur, M., Edelen, M.O., Pusic, A. and Gibbons, C. (2025) 'Can machine translation match human expertise? quantifying the performance of large language models in the translation of patient-reported outcome measures (PROMs)', *Journal of Patient-Reported Outcomes*, Vol. 9, No. 1, p.94, [Online], DOI: 10.1186/s41687-025-00926-w.

Lundmark, A. and Boglind, F. (2024) *Evaluating ChatGPT for Machine Translation of Medical Terms*, [Online], Available at. https://urn.kb.se/resolve?urn=urn:nbn:se:su:diva-242698 (Accessed 17 April, 2026).

Meddeb, A,, Lüken, S., Busch, F., Adams,L., Ugga,L., Koltsakis, E., Antonios Tzortzakakis, A., Jelassi, S., Dkhil, I., Klontzas, M.E., Triantafyllou, M., Kocak, B., Sabahattin Yüzkan, S., Zhang, L., Hu, B., Andreychenko, A., Yurievich, E.A., Logunova, T., Morakote, W., Angkurawaranon, S., Makowski, M.R., Wattjes, M.P. , Cuocolo, R. and Bressem, K. (2024) 'Large language model ability to translate CT and MRI free-text radiology reports into multiple languages', *Radiology*, Vol. 313, No. 3, p.1, DOI: https://doi.org/10.1148/radiol.241736

Muennighoff, N., Soldaini, L., Groeneveld, Lo, K., Morrison, J., Min, S., Shi, W., Walsh, P., Tafjord, O., Lambert, N., Gu, Y., Arora, S., Bhagia, A., Schwenk, D., Wadden, D., Wettig, A., Hui, B., Dettmers, T., Kiela, D., Farhadi, A., Smith, N.A., Koh, P.W., Singh, A. and Hajishirzi, H. (2025) *OLMoE: Open Mixture-of-Experts Language Models*, arXiv, [Online], DOI: 10.48550/arXiv.2409.02060 (Accessed 25 July, 2025).

Patil, S.G., Zhang, T., Wang, X. and Gonzalez, J.E. (2024) *Gorilla: Large Language Model Connected with Massive APIs*. Vancouver, pp.1–22, [Online], Available at: https://proceedings.neurips.cc/paper_files/paper/2024/file/e4c61f578ff07830f5c37378dd3ecb0d-Paper-Conference.pdf (Accessed 14 March, 2025).

Piantadosi, S.T., Tily, H. and Gibson, E. (2012) 'The communicative function of ambiguity in language', *Cognition*, Vol. 122, No. 3, pp.280–291, [Online], DOI: 10.1016/j.cognition.2011.10.004.

Potin, M., Carron, P-N. and Genton, B. (2024) 'Injuries and medical emergencies among international travellers', *Journal of Travel Medicine*, Vol. 31, No. 1, p taad088, [Online], DOI: 10.1093/jtm/taad088.

Rogers, J. and Bodenreider, O (2008) SNOMED CT: browsing the browsers', in Cornet, R, Spackman, K.A. (Eds.): *Proceedings of the 3rd International Conference on Knowledge Representation in Medicine* (*KR-MED 2008*), Phoenix, AZ, USA, p.30.

Rosenbloom, S.T., Denny, J.C., Xu, H., Lorenzi, N., Stead, W.W. and Johnson, K.B. (2011) 'Data from clinical notes: a perspective on the tension between structure and flexible documentation', *Journal of the American Medical Informatics Association: JAMIA*, Vol. 18, No. 2, pp.181–186, [Online], DOI: 10.1136/jamia.2010.007237.

Saeed, J. (2015) 'Chapter 7: semantics', in Braber N, Cummings, L and Morrish, L (Eds): *Exploring Language and Linguistics*, Cambridge University Press, Cambridge, pp.168–193, [Online], DOI: 10.1017/CBO9781139548922 (Accessed 1 April, 2026).

Stalnaker, R. (2002) 'Common ground', *Linguistics and Philosophy*, Vol. 25, No. 5, pp.701–721, [Online], DOI: 10.1023/A:1020867916902.

Tolk, A., Diallo, S.Y. and Turnitsa, C.D. (2007) 'Applying the levels of conceptual interoperability model in support of integratability, interoperability, and composability for system-of-systems engineering', *Journal of Systemics, Cybernetics, and Informatics*, Vol. 5, No. 5, p.66.

Tolk, A., Katina, P. F., Rainville, A. and Selby, J. (2020) 'Towards a framework for enterprise-wide operational agility in complex environments', *Nt. J. System of Systems Engineering*, Vol. 10, No. 2, pp.91–107.

Vylomova, E., Rimell, L., Cohn, T. and Baldwin, T. (2016) *Take and Took, Gaggle and Goose, Book and Read: Evaluating the Utility of Vector Differences for Lexical Relation Learning*. arXiv, [Online], DOI: 10.48550/arXiv.1509.01692 (Accessed 28 September, 2025).

WONCA (2005) *ICPC – 2 – R International Classification of Primary Care Revised Second Edition*, Second Edition with Revisions, Oxford University Press, Oxford, [Online]. Available at: http://www.ph3c.org/PH3C/docs/27/000496/0000908.pdf

Yang, R., Tan, T.F., Lu, W., Thirunavukarasu, A.J., Ting, D.S.W. and Liu, N. (2023) 'Large language models in health care: development, applications, and challenges', *Health Care Science*, Vol. 2, No. 4, pp.255–263, [Online], DOI: 10.1002/hcs2.61.

Websites

EVS Explore. *EVS Explore*. [Online], Available at: https://evsexplore.semantics.cancer.gov/evsexplore/welcome (Accessed 21 April, 2026).

Federatie Medische Specialisten, *Voorbeeldverslagen status-mentalisonderzoek*, [Online]. richtlijnendatabase.nl, Available at: https://richtlijnendatabase.nl/index.php/gerelateerde_documenten/bijlage/10444/1/121/Voorbeeldverslagen%20status-mentalisonderzoek.html (Accessed 20 March, 2025).

HL7 FIHR Release 5 [Online], Available at: https://fhir.hl7.org/fhir/index.html

ICD-11, ICD-11 for Mortality and Morbidity Statistics, [Online], Available at: https://icd.who.int/browse/2025-01/mms/en (Accessed 22 April, 2026).

ICPC-3 Foundation, International Classification of Primary Care – 3rd Revision, [Online], Available at: https://www.icpc-3.info/ (Accessed 29 March, 2025).

Laboratory Variability, Laboratory Medicine Curriculum, [Online], Available at: https://webpath.med.utah.edu/EXAM/LabMedCurric/LabMed01_03.html (Accessed 6 March, 2025).

Niinimaki, M., ChatGPT Medical Report Translator Demo, [Online], Available at. http://medical.nya.pub/medical

NHS, The NHS England Terminology Server, [Online], NHS England Digital. Available at: https://digital.nhs.uk/services/terminology-server (Accessed 22 February, 2025).

Göbel, N., SNOMED Navigator GPTs features and functions, examples and prompts, [Online]. GPT Store, Available at: https://gptstore.ai/gpts/kMlgVU6X2p-snomed-navigator (Accessed 20 February, 2025).

PECS, PECS Autism Spectrum Disorder Diagnostic Assessment Report: John Example. Psychological and Educational Consultancy Services, [Online], Available at: https://www.registeredpsychologist.com.au/wp-content/uploads/2020/09/PECS-Example-Autism-Report.pdf (Accessed 7 October, 2024).

SNOMED-CT, What is SNOMED CT, [Online], SNOMED International, Available at: https://www.snomed.org/what-is-snomed-ct (Accessed 22 April, 2026).

The Pathologist. Bridging the Gap in Glucose Monitoring, [Online], The Pathologist, Available at: https://thepathologist.com/issues/2015/articles/sep/bridging-the-gap-in-glucose-monitoring (Accessed 13 April, 2026).

University_of_Chapel_Hill, Chapel Hill Clinical Notes, [Online], Available at: https://www.med.unc.edu/medclerk/education/grading/history-and-physical-examination-h-p-examples/.

WONCA, ICPC-2 – English International Classification of Primary Care – 2nd Edition Wonca International Classification Committee (WICC), [Online], Available at: https://orbi.uliege.be/bitstream/2268/223037/5/icpc_copydesk_en.pdf.