Using SNOMED CT as a Semantic Model for Controlled Natural Language Guided Capture of Clinical Data

Kristian Kankainen[1], Toomas Klementi[1], Gunnar Piho[2] and Peeter Ross[1]

[1]*Department of Health Technologies, Tallinn University of Technology, Ehitajate tee 5, 19086 Tallinn, Estonia*
[2]*Department of Software Science, Tallinn University of Technology, Ehitajate tee 5, 19086 Tallinn, Estonia*

Abstract
Capturing clinical data in written textual form is by far the most popular and accepted method in health care. The workshop paper presents the author's preliminary results from creating a controlled natural language based on the SNOMED Clinical Terminology. This language can be used as a user interface that allows bidirectional translation between lexical expressions in a natural language and SNOMED post-coordinated expressions. E.g., natural language can be used for capturing machine-readable SNOMED data. Also, in the opposite direction, captured SNOMED CT terms can be displayed as a sentence in a natural language, shown perhaps with different wordings or linguistic styles based on the presentation context. This controlled language is shown to be useful for reporting clinical situations, which has long been placed in the gray area known as the boundary problem between information and terminology models. The paper concludes that clinical situations can be recorded within the terminology model if an appropriate user interface is used. The paper's contribution is to propose and evaluate a methodology that can be used for recording clinical data in a formal, machine-readable form and validate the data correctness already during data capture at the point of care.

Keywords
controlled natural language, clinical data capture, clinical terminology, boundary problem, user interface

1. Introduction

Todays clinical terminologies are undoubtedly expressive. The Systematized Nomenclature of Medicine – Clinical Terms (SNOMED CT) contains over one third of a million pre-coordinated terms which can be further composed into a nearly endless amount of combinations (that is, post-coordinated term expressions). Nonetheless, a recent scoping literature review [1] concluded that this expressivity is rarely used and that "*there is no easy solution for mapping free text to this terminology and to perform automatic post-coordination*".

We propose a methodology for building an innovative user interface. The user interface consists of a controlled natural language (CNL) with which the user writes free text – composing sentences with this CNL composes machine-readable term expressions in the background. The CNL employs SNOMED CT for its semantic model, that is, whatever SNOMED CT allows to express, the CNL can express. But perhaps more importantly, the word combinations of the

The International Health Data Workshop (HEDA-2022) June 19-24, 2022, Bergen, Norway
kristian.kankainen@taltech.ee (K. Kankainen)
0000-0002-0551-927Xwork (K. Kankainen)

CNL is not able express anything more than what SNOMED CT is able to. An early proof of concept prototype was positively validated within obstetrics domain with an experiment and survey with 12 midwives in [2]. Here we present a new generalized model.

Kuhn [3] defines controlled natural language being "*a constructed language that is based on a certain natural language, being more restrictive concerning lexicon, syntax, and/or semantics, while preserving most of its natural properties*". For the concerns of this paper, the semantics of our CNL is restricted by the Systematized Nomenclature of Medicine Clinical Terminology concept model, to which we add lexicon and syntax for the Estonian language. Infact, our proposed controlled natural language is quite a trivial idea, as (pre-coordinated) terminologies are a kind of controlled vocabularies, so it follows rather intuitively, that post-coordinated terminologies are already a kind of controlled languages. The only thing we add is the "natural".

Terminological languages, or more broadly information languages, can be pre-coordinated or post-coordinated [4]. The distinction is whether the term's meaning i.e. its position in the taxonomy is know before or after its inception. SNOMED CT is an advanced terminological language which is able to use both. For example the single (pre-coordinated) term *161077003 |Father smokes (situation)|* is defined by its position in the taxonomy (it is subsumbed by *443877004 |Family history of smoking (situation)|*). But the term is also fully defined being logically equal to a post-coordinated term expression which is a situation with explicit context where the subject relationship context is specified as *father of subject*, the associated finding is specifed as *smoker*, the finding context is *known present*, and the temporal context is *current* (see fig. 1). There is no such pre-coordinated concept as *"Uncle smokes"* – but a post-coordinated term can be composed expressing this meaning. In fact, anyone of the 462 terms denoting persons in SNOMED CT can be defined to be a smoker (or not to be, or have been, etc).

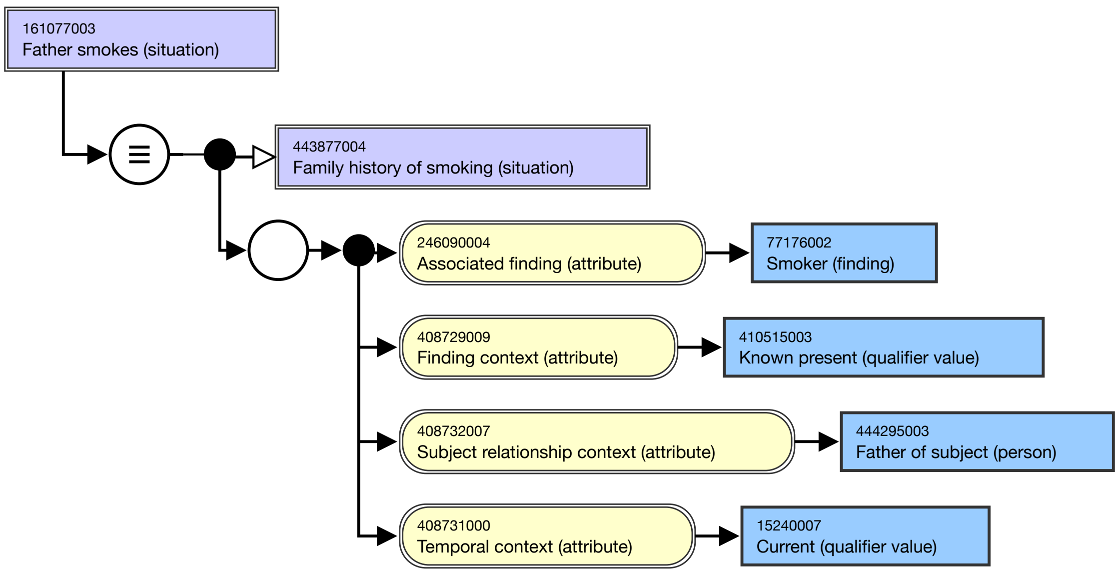

Figure 1: SNOMED CT pre-coordinated term *161077003 |Father smokes (situation)|* is fully defined with a post-coordinated expression denoting a family history of situation. Illustration by IHTSDO, using SNOMED International diagramming guidelines [5].

The rules for how post-coordinated expressions can be composed are defined in the SNOMED CT Compositional Grammar rules [6], but they allow all combinations, also non-sensical combinations. It is instead the Concept Model [7] that constrains what values are permitted in different positions of the expressions. E.g., that only persons can be the subject of a clinical situation or that the finding site of clinical findings can only be an anatomical or acquired body structure. The Concept Model constraint rules eliminate many, but not all non-sensical combinations.

1.1. Semantic interoperability and the boundary problem

The motivation behind our work is to enable the recording of more machine-readable and semantically interoperable data without burdening data entry.

The SNOMED International, the non-profit organisation behind SNOMED CT, states in [8, p. 3] that overall semantic interoperability is "*achieved through the combined functioning of the information architecture of the application and the terminology that populates it*".

The split between an information architecture (i.e., information model) and terminology is referred to as the boundary problem [9, p. 24]. Philosophically, the boundary problem is between the two types of knowledge: ontological knowledge, i.e., what is known to exist, and epistemological knowledge, i.e., how is it known to exist. Practically, it is the decision of how much data is recorded using terms from a terminology model (e.g., ontology model) and how much using an information model. The practice of binding or aligning the two models is known as terminology binding [10].

Markwell et al reports an agreed consensus position in [11] that terminology models are well suited for *What*, *How*, and *Why* whereas information models are better suited for *Who*, *When*, and *Where*. But there exists a grey area where the preference is unclear or dependent on the use case. Most of the grey area concerns the representation of contextual information related to instances of clinical situations (for example family history, presence/absence, certainty, goals, past/current, procedure planned/done/not-done).

The SNOMED CT can express context for both procedures and clinical findings by wrapping these terms in a *situation with explicit context* term expression. Although, in practice, simple terms are used instead and much context is assumed by default, e.g., procedures has actually occurred (rather than being planned or cancelled) or that findings are subject of the patient and are actually present (rather being ruled out or considered) [8, p. 4].

This practice of default contexts results in the peculiarity that positive findings belong to the clinical finding or procedure hierarchy whereas other, non-positive findings belong to the situation with explicit context hierarchy (compare *717234006 |Allergy to animal protein (*finding*)|* and *716220001 |No known animal allergy (*situation*)|*)

Rector & Brandt argued in 2008 [12] for a unified representation of findings, observables and procedures in SNOMED CT as situations that include any required context and would deal with negation explicitly and formally.

Irregardless of this, the SNOMED International education instead advocates for terminology binding and to use the information architecture for much of context. An example being to specify a list of allergies in a health record only using values for substances from the terminology model, instead of recording the full contextual meaning (e.g., *762952008 |Peanut (substance)|*

instead of *91935009 |Allergy to peanut (finding)|*) [13], [14].

Our work relates to the boundary problem as our controlled natural language can be seen as a user interface that effectively enables recording context using the terminology. We see the current practice of using terminology binding between the information model and terminology model as part of what creates semantic heterogeneity between data sources. Terminology binding is a kind of external knowledge, and as every data source's terminology binding schema is to some extent unique, semantic heterogeneity inevitably arises.

1.2. Related work

A recent scoping literature review [1] covering the years 2002–2019 on the use of SNOMED CT for processing free text in health care highlights that 1) most work is done for information retrieval and analysis purposes and few works have mapping to SNOMED CT as their primary goal 2) very rare usage of post-coordination capabilities 3) most often rule-based approaches used 4) the number of publications on the subject has decreasing trend since year 2012.

The review concludes that "*the need for formal semantic representation of free text in health care is high, and automatic encoding into a compositional ontology could be a solution*".

We will now show related work and argue that this *automatic encoding* can be implemented in many ways, including a bidirectional controlled natural language.

All works brought out by the previous review use Natural Language Processing (NLP) techniques and go in the analysis direction, going from text to (partial) meaning representation. We are aware of little work going in the opposite synthesis direction. Natural Language Generation (NLG) techniques go from a computable meaning representation to a natural language representation. One example is ontology verbalization employed in [15] for translating SNOMED CT post-coordinated term expressions into natural language paragraphs that they state is helpful both for quality assurance of terminology authoring and helping users understand complex post-coordinated expressinos. A similar application has been done with the GALEN medical terminology in [16] that generated unambiguous definitions to the French national classification of surgical procedures, and has been noted as "one of the major applications of Galen technology" in [17, p. 445].

Kuhn argues in [3] that many user interface approaches such as verbalization is similar to controlled natural languages in that they inevitably restrict the expressiveness i.e., the universe of meaning. There seems to be a research gap on using controlled natural language in medical informatics. A search on PubMed (on 10.06.2022) for the phrase "*controlled natural language*" retrieves only three related publications, each representing a certain capability of a controlled natural language. [18] presents a human-friendly search query user interface that could simplify querying across linked biomedical semantic web resources; [19] limits vagueness of written recommendation statements in clinical practice guidelines; and [20] discusses the opportunities of a prototype for rapid learning in precision oncology 3.0 due to closed loop feedback which would be accomplished by using a controlled natural language as the main unit of information and thus ridding the representational distinction between data capture and data publication information.

It is unclear to us whether the prototype in the last mentioned work has evolved further, but the presented CNL, called Biomedical Controlled English, is described as a separate tool used by a specially trained scribe to capture case summaries at tumor board meetings. The main difference we see is that our language would be used in everyday documentation by healthcare workers. This allows an even more closed feedback loop, as information is recorded also before a case is summarized. A fundamental similarity is the indistinction between the representation data is captured in and the representation tha already captured data is represented in. This similiraty is enabled not by identical representation but by a isomorphic mapping, that is, bidirectional translation.

To the best of our understanding, no work has been done on bidirectional translation between SNOMED CT representation and free text.

Bidirectional CNLs can be found in other domains, such as question answering over biomedical linked data has been reported in [21]. And work on a multilingual, multimodal speech-to-speech translation application for maternal health care is described in [22]. Both works employ the Grammatical Framework, but not the SNOMED CT.

Grammatical Framework (GF) is a programming language for writing type-theoretical grammars of natural languages [23].

A report on the state of the art of controlled natural languages for ontology authoring [24] states that the Grammatical Framework has the potential to become the de-facto open source general framework for developing resources for engineering multilingual controlled natural languages.

Before stating the main contribution of our work, let us return back to mainstream NLP analysis approaches. Work on automating the transformation of free text into SNOMED CT post-coordinated expressions has been presented in [25]. The authors state their effort is primarily motivated by the advantage to downstream analytics, which is a reasonable objective for allowing statistical errors. Their work extends previous approaches on relation identification by deep learning methods. Relation identification is the act of analysing a phrase (e.g., *severe asthma*) and connecting the identified elaboration term (e.g., *severe*) to the identified focus term (e.g., *asthma*) by guessing the correct relation attribute (e.g., *severity*) in order to correctly compose the post-coordinated term *195967001|Asthma (disorder)| : 246112005|Severity (attribute)| = 24484000|Severe (severity modifier)|*. Instead of guessing this kind of relations, our method presupposes the SNOMED CT concept model [7] rules for how relations can be composed and we map in the opposite direction, that disorders can have a severity modifier which can be expressed by severity attributes of which one is *severe*. The opposite direction guarantees full accuracy of our approach. Although the authors explain their candidate relations are pruned according to the concept model, their main example analysis is a post-coordinated expression that does not conform to it. They refine a morphological abnormality rather than a disease, something which is not allowed by the concept model.

The main contribution of our work to the problem of automatical encoding or mapping between free text and SNOMED clinical terminology are 1) the use of controlled natural language and 2) support for bidirectional translation. The impact of our approach that addresses these existing limitations is the use of text as a user interface for data capture at the point of care.

2. Method

The gist of our method is to use the SNOMED CT concept model as a semantic model and add a a linguistic grammatical representation to. That is, create a controlled natural language based on the concept model. Our method's direction of synthesis has been inspired by the Meaning–Text Theory [26] set forth by Žolkovskij and Meľčuk in the late 1960's, where language is seen as a vehicle to express meanings (direction of synthesis, from meaning to form) instead of language being some form that conveys some meaning (direction of analysis, from form to meaning).

The SNOMED CT concept model is a set of rules that specifies which attributes can be applied to refine a concept and what attribute values are permitted. To illustrate, clinical findings can be refined with 21 different attributes, e.g., finding site, finding informer, associated morphology, etc.

The editorial guide [8] is a human readable version of the concept model, but parts of the model is released in computable form as the Machine Readable Concept Model (MRCM) [7]. The MRCM uses Expression Constraint Language [27] for representing the permitted attribute values either intensionally or extensionally.

A *situation with explicit context* is a concept that specifically defines the context of a clinical finding or procedure. The MRCM describes it as two sub-domains, *finding with explicit context* and *procedure with explicit context.* See table 1 for the two domains' attributes.

Table 1
Attributes specified by the SNOMED CT Machine Readable Concept Model for the two domains *finding with explicit context* and *procedure with explicit context*

Finding with explicit context	Procedure with explicit context
Associated finding	Associated procedure
Finding context	Procedure context
Subject relationship context	Subject relationship context
Temporal context	Temporal context

The attributes are constrained for possible values, i.e., the value of *|Subject relationship context|* can only be a term from the *|Person|* hierarchy and the value of the *|Associated procedure|* can only be any term from either the *|Procedure|* hierarchy or the *|Observable entity|* hierarchy.

We convert these constraint rules into an abstract grammar using Grammatical Framework (GF), which is a special purpose programming language for writing type-theoretical grammars of natural languages [23]. In Grammatical Framework, abstract grammars are rules that govern how abstract syntax trees can be constructed. The abstract syntax represents semantically relevant combinations of structure, e.g., the universe of meaning.

Thus, the universe of meanings for a *situation with explicit context* consists of the following combinations:

$$U_{findingSituation} = ClinicalFinding \times FindingContext \times SubjectRelationContext \times TemporalContext,$$

and analogically for *procedures with explicit context* it is:

$$U_{procedureSituation} = Procedure \times ProcedureContext \times SubjectRelationContext \times TemporalContext.$$

To express the universe of meaning in text, we have written two concrete grammars. One concrete grammar for Estonian and one for SNOMED CT syntactic expressions. Concrete grammars are rule sets that relate how an abstract tree is represented textually, e.g., how it is linearized. The concrete grammar for post-coordinated expressions that adheres to the SNOMED CT compositional grammar [6] can be compiled automatically.

The Estonian expressions are built by hand using the Grammatical Framework's language-neutral Resource Grammar Library API (Application Programming Interface) [28]. Using the API makes it easier to add more languages later using a technique called example-based grammar writing [29].

For expressing a situation's temporal context, both grammatical and lexical means are used. E.g., grammatical tense is used to express the temporal contexts *|Current|* and *|Past|*, but a lexical phrase is used to express the temporal context *|Since last encounter|*.

Source code examples are here shown for some of the Estonian concrete grammar rules needed for expressing the meaning shown in fig. 3.

- The substance radon is a noun phrase consisting of only one noun:

```
lin SCT72927002_Radon_substance = { s = mkNP (mkN "radoon") } ;
```

- The *Exposure to potentially harmful entity* is an event that can take a substance as its causative agent:

```
lin mkSCT418715001_Exposure_to_potentially_harmful_entity_event substance
    = { s = mkVP (mkA2 (invA "kokkupuutes")
                       (casePrep comitative)) substance.s } ;
```

- The brother of the subject has also a short variant specified *brother*:

```
lin SCT444303004_Brother_of_subject_person
    = { s = mkNP (mkCN (mkN2 (mkN "vend" "venna" "venda"))
                       (mkNP (mkN "patsient"))) }
    | { s = mkNP (mkN "vend" "venna" "venda") } ;
```

- For combining everything into a finding with explicit context:

```
lin findingsituation subject associatedfinding temporalcontext findingcontext
    = mkUtt ( mkS temporalcontext.tense temporalcontext.anteriority
                  findingcontext.polarity
                  ( mkCl subject.s associatedfinding.s ) ) ;
```

3. Results

We have created a Controlled Natural Language (CNL) for specifying SNOMED CT post-coordinated expressions. The semantic model of our controlled natural language is based on and restricted to SNOMED CT's concept model for *Situation with explicit context.*

We initially cover two languages: Estonian and SNOMED CT post-coordinated expression syntax. The post-coordinated expressions adhere to the syntactic rules of the SNOMED CT compositional grammar [6].

The Estonian linguistic expressions are built by hand using the Grammatical Framework's Resourge Grammar Library [28], which makes it easier to add more language later using a technique called example-based grammar writing [29]. As can be seen from the source code listings presented under the methodology section above, the code reflects syntactic constructions and can be compiled by a (computational) linguist together with a clinician informant who specifies the lexicon.

So far, our prototype Estonian concrete grammar expresses only a small part of the universe of meanings of the *Situations with explicit context* domain. We have 27 *|Clinical Finding|* and *|Event|* concepts for the Associated finding axis, and 5 *|Person|* concepts for the Subject relationship context. The temporal contexts implemented so far are *|Current|*, *|Past - time unspecified|*, and *|Since last encounter|*. Implemented finding contexts are *|Known present|*, *|Probably present|*, *|Known absent|*, and *|Refuted|*.

We can thus express $27 \times 4 \times 5 \times 3 = 1620$ different situations, e.g., combinations of findings with explicit contexts and translate these between Estonian and SNOMED CT post-coordinated expressions. See figure 3 for an illustrative alignment of attribute values and grammatical features.

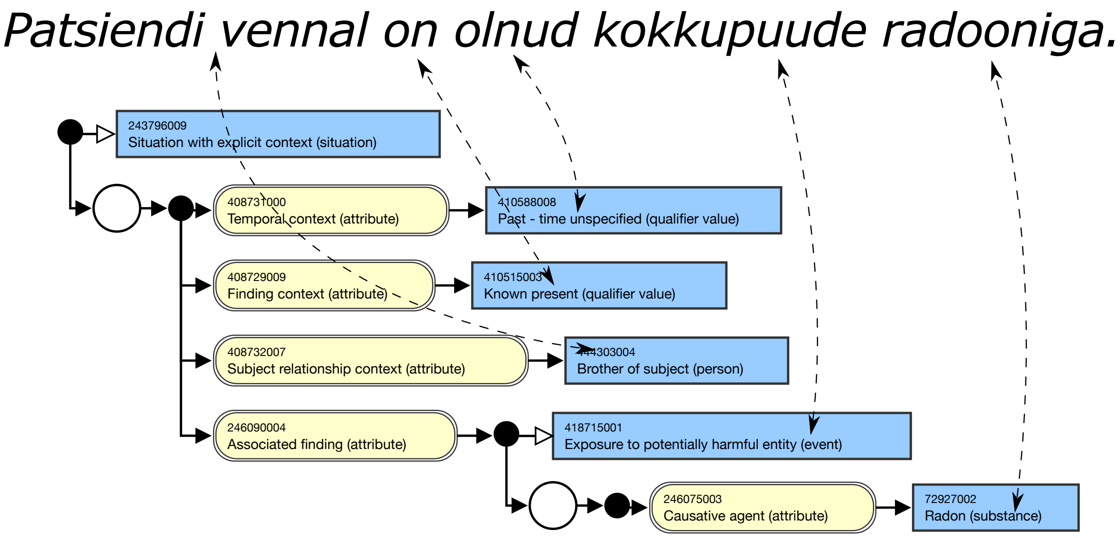

Figure 2: Estonian phrase (literally "the patient's brother has been exposed to radon") in the controlled natural language aligned with its SNOMED CT post-coordinated expression. Illustration by K. Kankainen, using SNOMED International diagramming guidelines [5].

Our controlled natural language supports translation in both directions, e.g., Estonian language can be used to specify SNOMED CT expressions. The other direction can be used for displaying already recorded SNOMED CT data in a human understandable way.

Additionally, the technology employed supports incremental parsing that can be used for guiding the writer *in situ* when documenting at the point of care.

4. Discussion

Our results has many implications for further work and usage scenarios for tools. Below we discuss using text as user interface, its implications for user experience, that can be made more interactive. Finally, we discuss limitations of our work and point towards future work.

4.1. Text as user interface

We think of text as a very general graphical user interface (remember, also written language is an invention, i.e., a technology). Other forms of user interfaces exist and we see the field of structured reporting [30] being very related to our work. But we see the relation on structuring a semantic basis, not structuring the form or shape of a reporting template.

Most of the cons of structured reporting stated in [30, p. 6] does not arise when using free text as the user interface. *"Prose-form dictation style is inherently flexible and customizable, and thus particularly preferred among older radiologists"* [30, p. 6] is not a problem at all, but rather a sign of maturity. Also, there is no risk of commoditization as the radiologists is able to use and modify their crafted text template in the same manner as they are used today. There is also no concern of errors caused by increased user interface interaction (e.g., clicking and scrolling) disrupting the user's existing search patterns and cognitive reasoning.

One concern that is not solved is the need to customize the structured templates to accommodate all needs of referring physicians or registries. Nevertheless, this problem is much different when considered from a structured semantics point of view rather than a mere structured template. We perceive both end parties of this communication to be competent in specifying their own needs – so, the writer might feel a certain sentence or phrase should be colored or trigger a certain search; or the receiving party might want to get more information – and what is most important, the common language for the specifications of these needs is SNOMED CT. Furthermore, development and implementation of these customizations does not change the main interface, which remains a text box.

Finally, no oversimplification of reporting can happen, as long as the text box accepts all text, also that which is not recognized by the controlled natural language to be inserted.

4.2. Interactive capture at point of care

As the translation to a formal representation is done *in situ* during the moment of data capture, it can have implications on the writer's user experience.

Incremental parsing can be used both to help and guide the writer while writing. Most of this can be thought of as auto-completion functionality that is context dependent – e.g., it doesn't suggest to insert *clinical findings* in subject relation contexts where *person* is instead expected.

As long as the writer allows herself to follow the suggested guidance, she can feel assured the computer understands what is inserted. Also colors can be used for marking the text, but even more feedback can be employed to create what we call dialogic data capture.

For example, when a clinical finding has been input that concerns a specific body part, an automatic search can show related data from the patient's EHR aside the text. Or, if necessary decision support rules exist and the inserted written text corresponds semantically to the rules, then the system could, for example, ask the writer whether a referral or request should be composed, or a submission to a registry or accounting be sent. The data is already available and the message can be prefilled.

4.3. Bidirectional translation

The previous example demonstrates how the bidirectionality of our CNL could be employed. Each SNOMED CT expression included in our CNL's universe of meanings can also be translated to Estonian sentence, e.g., be shown in human-readable form.

This has implications on many usage cases, because data that is captured in a machine-readable format is often illegible for humans to read. This can be applied to different linguistic representations of the data in generated summaries, patient portals, registry or accounting submissions, referrals and requests, etc.

4.4. Limitations and further work

A major limitation of confining the semantic model of our controlled natural language to the SNOMED CT concept model is due to the boundary problem (see section 1.1). We can only express ontological knowledge (e.g., clinical statements like *the patient's blood sugar is low*), but we can not express literal values of an information model (e.g., *the patient's blood sugar was measured to be 3.7 mmol/L*).

Although, recently, SNOMED International released a preview of added capability to express concrete domain values which initially contain numeric values for drug strengths of medicinal products [31]. Extending this capability to other concrete domains would solve our limitation, but we have no knowledge of current work in that direction.

Confining the abstract model too tightly with the target model is also against best practices brought out by a Grammatical Framework report, as it also makes forwards compatibility harder if the target model changes [32]. This is definitely part of future work.

Another strand of future work is with regard to a standardization work-flow on data capture. We have discovered that the main effort in our proposed methodology is not programming or creating the linguistic rules. What is hard and takes much time is instead the specification of what meanings need to be captured. The clear distinction between meaning and its expression that our methodology makes allows us to consider our controlled natural language from two points of views of data usage. 1) that of primary usage, i.e., what does the clinician write already in the free text that might be helpful for the clincian to make decisions; and 2) that of secondary usage, i.e., what fragment of the SNOMED CT termsinology is required to be captured by other data users. Including this extra terminology in the controlled natural language can make it easier for this data to be captured without burdening the clinician. More work is needed in settling both data usage's semantics.

5. Conclusion

Writing is natural and by far the most accepted user interface for reporting health data. We have created a Controlled Natural Language (CNL) that employs SNOMED CT's concept model as semantic model. The CNL's bidirectional translation capability allows using Estonian language to compose SNOMED CT post-coordinated expressions and vice versa, to express complex SNOMED CT term expressions in a human readable form.

Acknowledgments

The work of Mr Kankainen has been partially conducted in the project "ICT programme" which was supported by the European Union through the European Social Fund.

K.K. wrote the manuscript with support from T.K. All authors contributed to the final version of the manuscript. G.P. and P.R. supervised the project.

References

[1] C. Gaudet-Blavignac, V. Foufi, M. Bjelogrlic, C. Lovis, Use of the Systematized Nomenclature of Medicine Clinical Terms (SNOMED CT) for Processing Free Text in Health Care: Systematic Scoping Review, Journal of Medical Internet Research 23 (2021) e24594. URL: https://www.jmir.org/2021/1/e24594. doi:10.2196/24594, company: Journal of Medical Internet Research Distributor: Journal of Medical Internet Research Institution: Journal of Medical Internet Research Label: Journal of Medical Internet Research Publisher: JMIR Publications Inc., Toronto, Canada.

[2] L. Hamburg, Feasibility of maternity record standardization on the example of midwives' free text entries, Master's thesis, Tallinn University of Technology, Tallinn, 2021. URL: https://digikogu.taltech.ee/et/Item/b10c13a6-29f2-4967-9458-59226b3d6359.

[3] T. Kuhn, A Survey and Classification of Controlled Natural Languages, Computational Linguistics 40 (2013) 121–170. URL: https://doi.org/10.1162/COLI_a_00168. doi:10.1162/COLI_a_00168, publisher: MIT Press.

[4] E. Currás, Ontologies, Taxonomies and Thesauri in Systems Science and Systematics, Chandos Information Professional Series, 1 ed., Chandos Publishing, 2010. URL: http://gen.lib.rus.ec/book/index.php?md5=0be0412d588bea1dbcd5e0367d883d5d.

[5] SNOMED International, Diagramming Guideline, 2022. URL: http://snomed.org/diagram.

[6] SNOMED International, Compositional Grammar - Specification and Guide, 2022. URL: http://snomed.org/scg.

[7] SNOMED International, Machine Readable Concept Model, 2022. URL: http://snomed.org/mrcm.

[8] S. International, Editorial Guide, 2022. URL: http://snomed.org/eg.

[9] P. Kubben, M. Dumontier, A. Dekker (Eds.), Fundamentals of Clinical Data Science, Springer International Publishing, Cham, 2019. URL: http://link.springer.com/10.1007/978-3-319-99713-1. doi:10.1007/978-3-319-99713-1.

[10] A. L. Rector, R. Qamar, T. Marley, Binding ontologies and coding systems to electronic health records and messages, Applied Ontology 4 (2009) 51–69. URL: http://snomed.org/BindingOntologies. doi:10.3233/AO-2009-0063, publisher: IOS Press.

[11] D. Markwell, L. Sato, E. Cheetham, Representing Clinical Information using SNOMED

Clinical Terms with Different Structural Information Models., volume 2008, 2008. Journal Abbreviation: In KR-MED Publication Title: In KR-MED.
[12] A. L. Rector, S. Brandt, Why Do It the Hard Way? The Case for an Expressive Description Logic for SNOMED, Journal of the American Medical Informatics Association 15 (2008) 744–751. URL: https://doi.org/10.1197/jamia.M2797. doi:10.1197/jamia.M2797.
[13] SNOMED International, SNOMED CT Implementation Course, 2021. URL: https://courses.ihtsdotools.org/product?catalog=ICE_home.
[14] J. Zammit, Re: [#33807] Problems understanding one specific question in Module B (Implementation course), 2021.
[15] S. Liang, D. Scott, R. Stevens, A. Rector, OntoVerbal: a Generic Tool and Practical Application to SNOMED CT, 2013. doi:10.14569/IJACSA.2013.040631, journal Abbreviation: International Journal of Advanced Computer Science and Applications Publication Title: International Journal of Advanced Computer Science and Applications Volume: 4.
[16] J. C. Wagner, J. E. Rogers, R. H. Baud, J. R. Scherrer, Natural language generation of surgical procedures, International Journal of Medical Informatics 53 (1999) 175–192. doi:10.1016/s1386-5056(98)00158-0.
[17] F. Baader, The description logic handbook : theory, implementation, and applications, 2nd ed., paperback ed. ed., Cambridge University Press, Cambridge [etc.], 2010.
[18] L. McCarthy, B. Vandervalk, M. Wilkinson, SPARQL assist language-neutral query composer, BMC bioinformatics 13 Suppl 1 (2012) S2. doi:10.1186/1471-2105-13-S1-S2.
[19] R. N. Shiffman, G. Michel, R. M. Rosenfeld, C. Davidson, Building better guidelines with BRIDGE-Wiz: development and evaluation of a software assistant to promote clarity, transparency, and implementability, Journal of the American Medical Informatics Association: JAMIA 19 (2012) 94–101. doi:10.1136/amiajnl-2011-000172.
[20] C. Sweetnam, S. Mocellin, M. Krauthammer, N. Knopf, R. Baertsch, J. Shrager, Prototyping a precision oncology 3.0 rapid learning platform, BMC bioinformatics 19 (2018) 341. doi:10.1186/s12859-018-2374-0.
[21] A. Marginean, GFMed: Question Answering over BioMedical Linked Data with Grammatical Framework., in: CLEF (Working Notes), 2014, pp. 1224–1235.
[22] L. Marais, J. A. Louw, J. Badenhorst, K. Calteaux, I. Wilken, N. van Niekerk, G. Stein, AwezaMed: A Multilingual, Multimodal Speech-To-Speech Translation Application for Maternal Health Care, in: 2020 IEEE 23rd International Conference on Information Fusion (FUSION), 2020, pp. 1–8. doi:10.23919/FUSION45008.2020.9190240.
[23] A. Ranta, Grammatical framework : programming with multilingual grammars, Studies in computational linguistics, CSLI Publications, 2011.
[24] H. Safwat, B. Davis, A brief state of the art of CNLs for ontology authoring, in: International Workshop on Controlled Natural Language, Springer, 2014, pp. 190–200.
[25] K. J. Peterson, H. Liu, Automating the Transformation of Free-Text Clinical Problems into SNOMED CT Expressions, AMIA Summits on Translational Science Proceedings 2020 (2020) 497–506. URL: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC7233039/.
[26] J. Milićević, A Short Guide to the Meaning-Text Linguistic Theory, in: Journal of Koralex, 8, 2006, pp. 187–233.
[27] SNOMED International, Expression Constraint Language - Specification and Guide, 2020. URL: http://snomed.org/ecl.

[28] A. Ranta, A. El Dada, J. Khegai, The GF resource grammar library, Linguistic Issues in Language Technology 2 (2009) 1–63.

[29] T. Hallgren, R. Enache, A. Ranta, A cloud-based editor for multilingual grammars, in: Proceedings of the Grammar Engineering Across Frameworks (GEAF) 2015 Workshop, 2015, pp. 41–48.

[30] O. Brook, W. H. Sommer, Radiology Structured Reporting Handbook: Disease-Specific Templates and Interpretation Pearls, 1 ed., Thieme, 2021.

[31] SNOMED International, SNOMED International delivers technical preview for enhancement of SNOMED CT with concrete domains in advance of July 2021, 2021. URL: https://www.snomed.org/news-and-events/articles/snomed-delivers-technical-preview-concrete-domains.

[32] A. Ranta, J. Camilleri, G. Détrez, R. Enache, T. Hallgren, GF Grammar Tool Manual and Best Practices, 2012. URL: http://www.molto-project.eu/sites/default/files/MOLTO_D2.3.pdf.