

Aggregation and Generalisation in SNOMED CT

Jon Patrick

Centre for Health Informatics Research & Development

School of Information Technologies

University of Sydney

Sydney Australia

jonpat@it.usyd.edu.au

Abstract

This paper discusses the manner in which SNOMED CT (SCT) has compounded the two hierarchies of aggregation and generalisation in a way that can lead to significant processing problems, and has introduced greater than necessary complexity in the model. The historical reasons for this strategy are analysed and a linguistic explanation is presented. The importance of differentiating between the two systems for language usage at the point of care is described.

1. Introduction

In this work we view SNOMED CT (SCT) for its functional purpose as an ontology to use for computation at the point of care. In this context it can be useful:

- a. to check that data from external sources are conformant and complete for the processing purposes intended,
- b. to assist in automatically correcting data when it doesn't conform ontological knowledge,
- c. to supply ontological information to fill in a wider picture around data collected at the point of care,
- d. To allow discussion, description and analysis of collected data at a language level of greater abstraction to that at which the data was collected.

Ontologies to be useful for computation need to be precise in two ways; firstly, they need to match as closely as possible our understanding of the natural world we wish to deal in. Ontologies need to be accurate and to closely follow human understanding otherwise they create confusion in their design and their use. This means the ontology needs to be constructed with very close attention to the meanings of the terms used in it. Hence the names used in the ontology need to represent an intuitive and agreed understanding of the real world.

Secondly the variety of linguistic usage of those terms needs to be explored to uncover the diversity of semantic roles of the terminology and ensure that only those roles that are useful are included in the ontological modelling, and to remove ambiguity in the use of those roles. Precision of definition is particularly important in establishing the relationships between elements and identifying the

fundamental atomic elements and how they combine together systematically to make more complex semantic concepts.

Ontologies also consist of abstractions with each level of the ontology being more abstract as one moves up the hierarchy. This structure allows us to talk about the world at the different levels of abstraction in the ontology. In using an ontology for computation there are two basic forms of abstraction available, generalisation and aggregation. Generalisation hierarchies are used throughout SNOMED CT (SCT) as the basic mechanisms for relating content. Aggregation hierarchies on the other hand have not been used but rather transposed so that they appear like generalisation hierarchies using a data structure called SEP. We investigate the adequacy of SEP as a replacement for a proper aggregation hierarchy and argue that its needless introduction is based on a misunderstanding of the both the linguistic use of terminology at the point of care and the logic arguments developed for its justification.

2. The Meaning of Aggregation and Generalisation.

The discussion of Aggregation and Generalisation is commenced with a view drawn from the database literature so as to give priority to their computational aspects. Smith & Smith make the basic statement "aggregation and generalization are independent activities" (1977, p116)¹. Although they are two fundamental means of organizing the knowledge about the relationships between concepts or entities, they are orthogonal to each other, that is they specify knowledge that is independent of each other. Generalisation creates a hierarchy of classes and subclasses in a subtype relationship, usually referenced as IS_A. Aggregation creates a hierarchy of linked components sometimes called a "Part of" or "part-whole" relationship. Generalisation and Aggregation are known as *relationship types* as there are many instances of different hierarchies that have the behaviour of these *types* in natural language. The similarities and differences between these two types of relationships are fundamental to understanding how ontologies are organised and the limitations to their processing.

Generalisation is a hierarchy of classes and subclasses where each subclass has a transitive

relationship between parents and children, that is, properties are carried from subclass to subclass down the hierarchy, but also the properties introduced into a subclass, separate to the inherited properties are *not inherited* back up the hierarchy.

In aggregation the individual types of parts are brought together in a hierarchy to make a whole, such as in the assembly of a motor car. Hence, a sparkplug is part of a motor, is part of a car. Importantly there is a transitive relationship between the parts, that is, the sparkplug is part of the car. However, there are no inherited properties, so red wheels do not make the car a red car, nor does a red car make the wheels red. Furthermore, the relation is *irreflexive*, that is, an entity cannot have a part-of relationship with itself, so a car is not part of a car.

Algebraically the transitive property can be expressed as:

If aRb, bRc, then aRc, where R is a defined relationship such as IS_A_SUBCLASS_OF, or IS_PART_OF.

In this example the relation IS_A_SUBCLASS_OF is defined to be of the relationship type Generalisation, while the relation IS_PART_OF is of relationship type Aggregation. So whilst both relations are transitive and we can apply that algebraic function to their hierarchies, the meaning content of the relationship names are different and the relationship types are different, the former having inheritance and the latter not.

3. Linguistic explanation (meronyms and metonyms)

The Generalisation-Specialization relationship is known as a *hypernym-hyponym* relationship in linguistics and it is thought of as having the same features as in logic, that is, transitive with one-way inheritance. Aggregation is known as the *holonym-meronym* relationship and likewise behaves the same way with an hierarchical structure but without inheritance. A further linguistic concept, *metonym*, is needed to complete the explanation of the problems with the SCT structures. A *metonym* is a word used in place of the correct word because it is associated with it in some way, so the word “dish” can be used in place of “fish” when talking about the evening meal. This is a very common linguistic strategy and in one of its extreme forms it is called a *euphemism*. In keeping with consistent terminological usage we will use generalisation as a synonym for hypernym-hyponym, and aggregation for the holonym-meronym relationship.

4. Processing with Ontologies

In SNOMED CT there are 19 generalisation hierarchies such as Organism, Body Structure, Clinical Finding, etc. The concept “part-of” is defined in the Attribute generalisation hierarchy. It is a means of representing part-whole concepts by using SEP structures based on the ideas created in

Schulz, Romaker & Hahn² which in turn claims its heritage in the GALEN project (Rector, Bechofer, Goble, Horrocks, Nowlan & Solomon)³. These papers present the problem of part-whole reasoning as a problem of differentiating it from reasoning for subsumption. They point out that some classification systems, MESH for example, treat reasoning for generalisation relations and part-whole relations the same way producing inconsistencies such as *blood* being a hypernym for both *foetal blood* (a type of blood) and *blood plasma* (a part of blood). They point out that a different logic is needed to compute over a part-whole hierarchy which consists of the features *transitivity* and *part-whole specialisation*.

5. Transitivity

A key issue for Schulz et al is the notion of whether the meronymic relationship (IS_PART_OF) is transitive as in the relationship IS_SUBCLASS_OF (or IS_A). The example provided is

appendix IS_PART_OF colon

colon IS_PART_OF intestine

as this is manifestly true for common sense knowledge then it must be true that

appendix IS_PART_OF intestine

hence transitivity is demonstrated.

However Schulz et al do not deal with the issue of inheritance of properties in either a modelling or a linguistic sense for a Generalisation hierarchy. Importantly later in their paper they assume that given the transitivity of both relationship types (that is generalisation and aggregation) that any property is also inherited between types. This is manifestly untrue. Take the example of a motor car and its wheels. A blue motor car does not necessarily have blue wheels and a car with a set of racing wheels is not necessarily a racing car. Hence the characteristics of each part are not necessarily inherited either up or down a holonymic hierarchy, in fact that is the point of differentiation between the two types of relationships, hyponyms have inheritance of characteristics and meronyms don't have inheritance. Furthermore blue wheels on a blue car are not instances of the same blue characteristic, they are merely coincidental facts about the car, in the same way a broken arm and broken leg are different. While it is possible to assign algebraic symbols for all these phenomena and then to perform computational manipulation that doesn't mean that any computation is sensible. Moderation of what is sensible to compute must lie with the meaning intention of the symbolic representation.

The misconception of sharing inheritance due to the equivalence of transitivity between the two relationship types needs to be investigated from both a linguistic perspective and an historical perspective to understand how their interpretation arose.

6. The Usefulness of the “Part-Whole Specialisation”

Shulz et al take *specialisation* from Horrocks et al⁴ and call it *part-whole specialisation*. This phrase creates a warning sign of future difficulties. Specialisation is the inverse of generalisation, and generalisation-specialisation are distinctly different to the part-whole relationship, so the term *part-whole specialisation* is a challenging idea.

Schulz et al produce a holonymic example with “shaft of femur” is PART_OF “femur” (Fig 1), in a parallel structure they produce the pair “fracture of shaft of femur” IS_A “fracture of femur” as a specialisation. They then point out that not only are the structures related by the relationship “FRACTURE_OF” (relations R1 and R3) but the hyponym “fracture of shaft of femur” is related to the holonym “femur” by the same relation, R2, that is “fracture of shaft of femur” is a FRACTURE_OF “femur”.

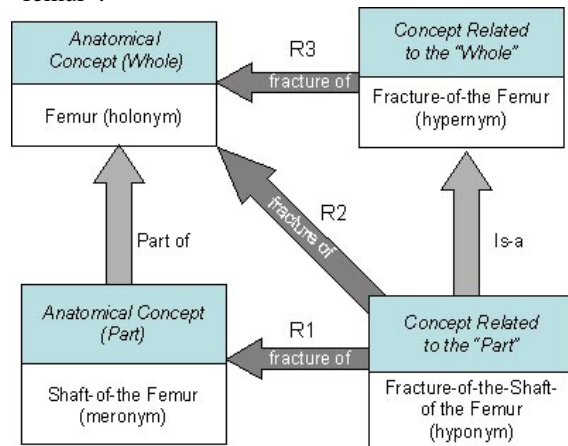


Fig 1. Part-Whole Specialisation as represented by Shulz et al. R1, R2, R3 are the relationship “fracture of”. R2 is a relationship from the hyponym to the holonym.

There are a number of difficulties with the configuration in Fig 1 and the line of argument presented in Shulz et al. The logical deduction follows these lines:

Premiss: the “femur” and “shaft of femur” have a holonym-meronym relationship,

Deduction:

If the hyponym to meronym have a relationship, in this case FRACTURE_OF,

then the hyponym to holonym have the same relationship (R2),

and thereby the holonym is also a generalisation of the meronym,

furthermore by asserting the premiss that the hypernym has the same relationship to the holonym(R3) as the hyponym has with the meronym (R1), that a hypernym-hyponym relationship is thereby verified.

A general form of this argument is embodied in the SEP structure as used to model an ANATOMICAL-PART-OF relation for the relationship between physical parts of an organism (*op cit*). It consists of three nodes: S, E and P nodes. S is a structure node which subsumes the E-node and the P-node and there is an ANATOMICAL-PART-OF relation from the P-node to the E-node (fig2).

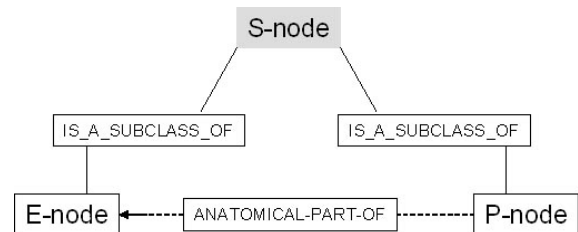


Fig 2. Structure of S-node for modelling Part-Whole Specialisation (after Schulz et al).

The E-node denotes the *whole* anatomical part being modelled and the P-node denotes any part of the E-node entity. The P-node is then the parent of the same structure repeated down a generalisation hierarchy, so:

the intestine-structure (S-node) has an intestine (E-node) and intestine-part (P-node) which has a colon-structure (S-node) which has a colon (E-node) and colon-part (P-node) which has a caecum-structure (S-node) which has a caecum (E-node) and caecum-part(P-node) which has an appendix-structure(S-node) which has an appendix (E-node).

Schulz et al justify this structure in saying “by introducing a special data structure for part-whole encoding we build up specifically structured IS-A hierarchies which support the emulation of inferences typical of transitive PART-OF relationships. The same formalism allows for conditioned part-whole specialization”. They have in effect created a data structure that masks the PART-OF relationship between anatomical elements on the basis that they don’t have a transitive relationship (whereas they are transitive) so as to allow for inheritance (which is not permissible).

However they concede that the inheritance of the specialisation relationship down the holonymic hierarchy is not always found to be consistent with medical knowledge. The cited contrary example being:

While

Perforation of Appendix IS_A_SUBCLASS_OF Intestinal Perforation is true,

it is NOT true that ,

Inflammation of Appendix (Appendicitis) IS_A_SUBCLASS_OF Intestinal inflammation (Enteritis).

This should be a warning that there is some trouble with this type of modelling. Their arguments for introducing specialisation into the Aggregation hierarchy are not convincing and do not account for a number of matters and introduce the possibility of significant errors of reasoning.

First and foremost a fracture to the “shaft of femur” is a change of state of the “shaft of a femur” and not a type of “shaft of a femur”, that is, it is not a subclass of either a “femur” or “shaft of femur” and to create a specialisation hierarchy is to contradict the intrinsic definition of an aggregation hierarchy. To make an analogy, a broken car is not a broken wheel and a broken wheel is only a broken car through a metonymic substitution where the correct word, “wheel” is substituted by a word associated with it, “car” in this case. If the “shaft of the femur” is broken then it is not the “femur” that is broken but rather the “shaft” and on the basis of non-inheritance between holonym and meronym the one is not the other. The key confusion occurs because the term “femur” which is a holonym in the hierarchy is used as well as a metonym (that is, as a substitute for the correct word by virtue of association with it).

Turning to the logic of the argument for the part-whole specialisation, there is no credible basis for any steps in the deduction. The argument hinges on the notion that “specialisation” exists because a hypernym has the same relationship to each of the members of the holonym-meronym pair. However the argument relies on deducing that the relationship between hyponym and holonym is created by virtue of the relationship of hyponym to meronym, then meronym to holonym –this is a circular argument.

A further weakness in the argument lies in the nature of the modelling of the physical world. A fracture to anything is not an entity of itself but a CHANGE OF STATE of an entity and is better modelled that way. It would be possible to model it as a property but then the unfractured femur would be the unmarked case. No doubt clinicians at the point of care speak about “a fracture to the shaft of the femur” but attention is not given to the linguistic structure. If you invert the possessive case you get the “femur’s fracture” and you can then see the possessor of the phenomena as the femur. The “fracture to” is really serving as a descriptor of the entity and acts far more in the role of an epithet than an entity as in “a fractured shaft of the femur” where the participle “fractured” is clearly serving in the epithet role.

This analysis questions the applicability of the SEP model to a holonymic structure and demonstrates that it is not applicable at least in the scope of the examples used so far. The explanation for the justification used by Schulz et al is shown to be flawed because they shift from defining a term in a holonymic role and then change the role of the holonym to a metonym.

Schulz et al use another example to discuss the generalisation of their SEP model where their process is also manifest. They assert that if INTESTINE-STRUCTURE is an S-node then a PERFORATION-OF-APPENDIX is also a PERFORATION-OF-INTESTINE as Appendix IS-A-PART-OF Intestine. This is the same faulty reasoning. It is true that if the holonymic tree is only defined as far as “intestine” then there is no such thing as a PERFORATION-OF-APPENDIX and any perforation of anywhere in the Intestine is a PERFORATION-OF-INTESTINE. However if the holonymic tree is further described in all of its parts then the word “intestine” can only be used in a metonymic role and as such does not have any of the characteristics of the real perforated part, that is, it is acting as a surrogate. Taking the example of a car-motor-spark_plug meronymy, if the spark plug is cracked then the motor won’t work and the car won’t work, but neither the motor nor the car is cracked. Both the motor and the car have emergent properties, that is properties they have by virtue of being a “whole” that none of their parts have, e.g. a motor runs and a car runs or works or can be driven (in some contexts). In English we have words that differentiate between a cracked spark_plug and a motor that doesn’t work and car that’s broken down that help create the separation of metonymy from meronymy - we would not (normally) say “My car is cracked” or “My motor is cracked” (excluding the British propensity to contract “motor car” to “motor” while most others use “car”).

This argument does not discount the legitimate use of the PERFORATION-OF-INTESTINE when the meronymic component is unknown as might well happen at point of care. But that situation is as above where the word is being used in the role of a metonym and not as a meronym. This issue leads into the further example used by Schulz et al where they assert that “enteritis” has the relation INFLAMMATION-OF with “intestine”, and “appendicitis” has the relation INFLAMMATION-OF with “appendix”, yet “appendicitis” cannot be classified as a type of “enteritis”, even though the “appendix” is a PART_OF the “intestine”. Importantly they argue that each of these are attached to E-nodes and thereby do not automatically have a subsumption relationship as is true in real life, whereas they claim the Perforation example is a true subsumption and is attached to the P-nodes. This is a failure to understand the difference between metonymic usage, very common at the point of care, and meronymic usage, fundamental for correct ontological rendition.

Returning to the exemplar of the intestinal tract of Schulz et al. They have at the top of their example the S-node Intestine structure which descends down an IS_A hierarchy successively to colon, caecum, and appendix. In such a structure if one were to assign to the intestines a property of

8metres long we would have a colon, caecum and appendix all 8 metres and a very crowded pelvic zone. In terms of diseases in this diagram Enteritis is attached to the Intestine and Appendicitis to the Appendix, both attached by the relationship INFLAMMATION-OF. Apart from the convention that enteritis is a condition of the small intestine and colitis of the large intestine, the naming convention does not match any sensible characterisation of the real world. Furthermore the weakness in the SEP model is exposed. If the correct level of granularity is a match of the disease instance level to the anatomical location, that is Appendicitis IS_AN_INFLAMMATION_OF Appendix, then Appendicitis is a specialisation of the class INFLAMMATIONS_OF_THE_INTESTINAL TRACT, that is the correct level of generalisation of the terminology that must match the correct level of meronymy, but nevertheless importantly they are independent of each other. So to say a patient has “intestinal inflammation” is holonomically correct, but to say they have an “inflamed small intestine” is a metonymic use of “inflamed”, the meronymically correct expression being “enteritis of the small intestine”. Hence on Schulz et al’s diagram the correct entry for Enteritis should be Intestinal_inflammation a sub-class of inflammations that has members {enteritis, colitis, appendicitis} each of which are DISEASES_OF the anatomical components of the Intestine (small intestine, large intestine and appendix).

7. Origins of the processing fallacy

The aim in this section is to trace the history of the justification for this egregious shift and to see if something can be redeemed from the work.

Horrocks et al following from the work of Lenat & Guha⁵ make the claim:

“Specialisation allows the user to specify that a characteristic is inherited across relations other than is-a: stating that role R is specialisedBy role S leads to the inference that for any objects x, y and z, $xRy \wedge ySz \rightarrow xRz$.”

They then go on to apply this approach to a case of R(*haslocation*) and S (*isPartOf*) applied to relating bodily location with anatomical parts and thereby deduce that “fracture of femur” subsumes “Fracture of shaft of femur”.

However there are two matters that have been overlooked here. The Lenat & Guha text merely presents a mechanism, labelled *TransfersThro*, for such transitive assignment but leaves it entirely up to the user to define if the transformation is appropriate. Horrocks et al then restate this mechanism and make the claim “specialisation allows the user to specify...” thus relabelling the Lenat & Guha mechanism as a general principle and extending to themselves a process to be used automatically on all relationship pairings rather than

the original intention which was discretionary. Subsequently they have not considered that the *relationship types* are different and that the transformation rule will not produce sensible deductions if the relationship types of Aggregation and Generalisation are mixed together.

Secondly the assertion that this interpretation of specialisation is applicable to other relationship types is never tested, just assumed. It does not make sense to assert that the relationship type of aggregation can give the functionality of inheritance by the development of a processing system that can do it, or by describing a logic that asserts it. The Horrocks et al paper asserts this example:

“that the hasLocation role transfers through the isPartOf role so that any x (fracture) which hasLocation y (shaft of femur) which isPartOf z (femur) is classified as a kind of x (fracture) which hasLocation z (femur).” (example data not in original).

This is an example where paying attention to the meaning of language informs us about a weakness in the logical deduction. The *hasLocation* relation tells us that the fracture has occurred on the “shaft of the femur”, however the “shaft of the femur” has a meronym relationship with the holonym “femur” which therefore cannot inherit any of its properties or relationships. While the algebra may operate the semantic test fails which then puts into the question the usability of the algebra. While it might well be true that the algebra is useful across Generalisation relationship type it is not usable in any aggregation relationship type. The examples used by Horrocks et al have been accepted up to now and therefore the system of SEP used in SCT because it appears plausible and has a correspondence with habits of natural speech. However in ontological processing it is important to resolve description to the point of removing ambiguity and contradiction, otherwise processing systems will produce erroneous and nonsensical results.

The SEP structure appears to be a contrivance to avoid the fundamental distinction defined by the difference between aggregation and generalisation, and more specifically to transform an aggregation hierarchy so that it has the inheritance characteristic of a generalisation hierarchy, despite the fact that they are intrinsically disparate and to substitute one by the other is a confusion of important distinctions much like confusing cats and dogs. Whilst one can compute transitivity on both of the hierarchies with the application of suitable logics, to use a device to merge the two hierarchies is meaningless at best and potentially damaging at worst. The logic for manipulating generalisation hierarchies are well established however a similar function for aggregation hierarchies can be first found in Padgham & Lambrix (1994)⁶. That paper discusses the similarity with a generalisation hierarchy but offers these words of warning “we have defined a

language for describing compositional concepts and a relationship which allows us to maintain compositional hierarchy, analogous to, but differing from the subsumption hierarchy.”

8. Meronymy and Metonymy at Point of Care

The difference between using words in holonymic and metonymic roles is important and context dependent. If a patient attends a clinic and reports “I have a broken leg” the clinician will say “we need to X-ray that and find out what is broken”. On the return of the X-ray the clinician will say to the patient “the femur is broken” however to an attending surgeon they will say “the shaft of the femur is broken” this being an important distinction to the “neck of femur being broken” as they have two entirely different treatment requirements. Once that definition is established they can readily revert to talking about the “fractured femur” without any risk of misunderstanding because they know they are really talking about the “shaft of the femur”. The expression “femur” is thereby being used as a metonym and not a holonym.

The question remains as to the level of granularity a meronymic hierarchy needs to be referenced in the point of care. The answer is resolved clearly by the detail at which the care regime is determined. If the difference in care is defined by differentiating between “shaft of femur” and “neck of femur” then the clinician(s) need to establish their reference terminology to that level, and use of the terminology at that level is literal and meronymic, however references to terms above that meronymic level are not holonymic but metonymic, and hence the metaphorical role of the metonym needs to be accounted for.

None of these arguments are meant to deny the need for clinicians to talk at various levels of abstraction for different listener communities. Rather the arguments are intended to direct the need for the ontology to catch all aspects of the language variations in systematic ways that exploit not only our knowledge of medicine but our knowledge of language and our tools of logic (and sometimes statistics).

9. Correcting the problem

We have argued that the SEP model creates contradictions to common sense knowledge. What has been provided as a computational function in Cyc has been taken as a generally applicable principle in GALEN and then borrowed without consideration of its semantic constraints.

The consequences of this analysis are:

The SEP model is not needed to model the transitivity relationship in a holonymic-meronymic hierarchy as it is intrinsically transitive.

The SEP model introduces the dangerous precedent of allowing attribute inheritance in a holonymic-

meronymic hierarchy which can readily lead to nonsensical assertions.

The identification of the appropriate level in the holonymic-meronymic hierarchy for clinical reference needs to be at the point at which clinical care is determined, and references to meronyms at that point are literal.

References above that point are to be considered metonymic unless otherwise resolved.

Other entities, such as diseases can be organised on generalisation hierarchies and inherit properties through that hierarchy and have relationships of any appropriate type to the components hierarchy however they do not inherit their properties.

¹ Smith, J.M. & Smith, D.C.P. Database Abstractions: Aggregation and Generalisation, ACM Transactions on Database Systems, Vol.2 , No. 2, pp105-133. 1977.

² Schulz, S., Romacker, M. & Hahn, U. Part-Whole Reasoning in Medical Ontologies Revisited – Introducing SEP Triplets into Classification-Based Description Logics. Proc AMIA. 1998.

³ Rector, A.L., Bechofer, S., Goble, C.a., Horrocks, I., Nowlan, W.A. & Solomon, W.D. The GRAIL concept modelling language for medical terminology. *Artificial Intelligence in Medicine*. 9:139-171, 1997.

⁴ Horrocks, I., Rector, A. & Goble, C. A description logic based schema for the classification of medical data. KRDB'96 – Proc of the 3rd Workshop on Knowledge Representation Meets Databases, pp 24-28, 1996.

⁵ Lenat, D.G. & Guha, R.V. Building Large knowledge-Based Systems. Addison-Wesley: Reading, 1983.

⁶ Padgham, L. & Lambrix, P. A framework for part-of hierarchies in terminological logics. IN J. Doyle, E. Sandewall and P. Tarasso (Eds.) Principles of Knowledge Representation and Reasoning: Proceedings of the 4th international Conference (KR'4), Bonn, Germany, pp485-496, 1994.