

The MedDRA Paradox

Gary H. Merrill, Ph.D.
GlaxoSmithKline, Research Triangle Park, North Carolina

Abstract

MedDRA (the Medical Dictionary for Regulatory Activities Terminology) is a controlled vocabulary widely used as a medical coding scheme. However, MedDRA's characterization of its structural hierarchy exhibits some confusing and paradoxical features. The goal of this paper is to examine these features, determine whether there is a coherent view of the MedDRA hierarchy that emerges, and explore what lessons are to be learned from this for using MedDRA and similar terminologies in a broad medical informatics context that includes relations among multiple disparate terminologies, thesauri, and ontologies.

Introduction

MedDRA was developed to facilitate the coding of “regulatory data” in biopharmaceutical development, clinical trials, and the reporting of adverse events. Its subject matter comprises signs, symptoms, diseases, diagnoses, therapeutic indications, results of investigations, procedures, and medical/social/family histories ([1], p. 7).

Two or more medical terms may stand for the same medical concept. For example, “Prostate cancer”, “Cancer of prostate”, “Prostatic cancer”, and “Malignant neoplasm of prostate” are simply different names for the same medical entity (concept, in MedDRA’s phrasing). In addition, some terms are more general than others. Thus “Prostatic disorders” is more general than “Prostatic neoplasms and hypertrophy” which is in turn more general than “Prostate cancer”. Medical terms form a natural hierarchy of generality which is useful in categorizing and referring to medical concepts, and MedDRA represents such a hierarchy through its description of levels of generality into which its terms fall. This is the common sense view of MedDRA (and of medical terminologies more generally). It is also the view endorsed by standards for the construction and management of controlled vocabularies, as in [2].

The MedDRA hierarchy is described as consisting of five levels as illustrated in Figure 1. In general, LLTs (Lowest Level Terms) are characterized as different terms for the *same concept*, and a PT (Preferred Term) is characterized as the preferred way of referring to that (one) concept. Moreover, and in perfect

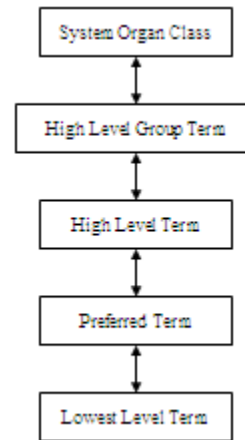


Figure 1: The MedDRA Structural Hierarchy

accord with this view, a PT is itself an LLT. Its distinction from the other LLTs to which it is linked is not a *conceptual* one (i.e., they all refer to the same concept), but is rather a pragmatic one relating to the *use* of terms in a controlled vocabulary.

Even though all of the LLTs linked to a PT are equivalent to one another – and this includes the PT itself – and even though they refer to the same medical concept, still section 3.1 of [1] informs us that LLTs are *subordinate* to their PT and occupy a distinct level in the MedDRA structural hierarchy. In a particularly confusing twist, it is further asserted that the LLT that is *identical* to the PT is also *subordinate* to it, that in that case the PT is the “parent” of the LLT, and that although the LLT and PT “have the same MedDRA code” (they are, after all, *identical* – they are the *same term*), they “appear at both levels”.¹ Thus it seems that a PT is subordinate to itself, which is not possible.

It might appear that the paradoxical nature of the PT/LLT relationship in MedDRA is simply an elementary error in failing to distinguish a term (as a lexical entity) from its properties (LLT or PT – which are not mutually exclusive), and certainly this does seem to be part of the difficulty. But the underlying problem is deeper, and it has consequences for how Med-

¹Note the use of the plural construction in talking about the single term that is at issue – as though in fact there are really two terms here that just “look the same”.

DRA should be represented and used in various informatics contexts because it encourages the view that in fact LLTs (being “subordinate” to PTs) are “less general” or “more specific” than PTs (rather than being truly *equivalent* to them), and that this distinction of generality between PTs and LLTs must be maintained and appear explicitly in any representation or use of them. This view then engenders a number of confusions in applying MedDRA in such areas as data mining, knowledge representation, and knowledge exploration – particularly when MedDRA is to be used in conjunction with other similar terminologies.

These confusions do not, of course, frequently make it into the published literature where they can be referenced. Rather, they often appear in online forums when attempts are made to understand and to apply MedDRA and similar schemes. However, one example of such a confusion appears in [3], where it is asked (a) whether LLTs represent “separate concepts from their PTs”, (b) whether LLTs “may be synonyms or lexical variants of their PTs or may be narrower”, and (c) whether “the relationship between LLTs and PTs is not specified by MedDRA”. And these questions are raised by an author (see [4]) with a substantial understanding of coding schemes.

Another example occurs in the context of [5] – and particularly in [6] where an NLM Biomedical Terminologies Specialist tries to reconcile the MedDRA documentation with the representation of MedDRA within the UMLS Metathesaurus. The result, within a single paragraph, is the combined assertions that MedDRA has five levels and that it has only four. A valiant attempt is made there to distinguish between a *hierarchy of term types* and a *hierarchy of MedDRA content*, but it devolves into confusion – partly because this distinction does not appear in MedDRA itself (i.e., in [1] or [7]), and partly because in the end it appears to suggest that the *same term* (“same name and code”) is really *different* terms, depending on what “role” it plays (they play?).

This is a significant issue. If we take seriously the PT/LLT distinction described in [1] and follow it closely, then the representation of MedDRA in the UMLS Metathesaurus is *incorrect* because it fails to preserve the PT/LLT relation as this appears in [1]. We will return to a consideration of the correctness of the Metathesaurus representation of MedDRA in a later section, but first we must look more closely into MedDRA’s description of the PT/LLT relation and the principles on which this is based.

Analysis

We begin by looking even more closely at the PT/LLT distinction as this is characterized in section 3.1 of [1] and asking how (and indeed whether) that account can support the view that PTs and LLTs should

be thought of as occupying different “levels” in the MedDRA structural hierarchy. It does this by distinguishing five senses (or conditions) in which terms may enter the PT/LLT relation (which recall is being characterized as a relationship of *equivalence*). I will consider each of these in turn, with the goal of determining which of them may be seen as supporting the view that a PT should be construed as being superordinate to its (non-PT) LLTs and what, in that case, “superordinate” means. For reasons that will become clear, I will not consider these in the order they appear in [1], but rather in the descending order of their “strength”.

The first, and most direct, way in which a PT may be related to an LLT is by being *identical* to it (i.e., they are not two terms but exactly the same term). In this case, it is difficult to see what could be meant by saying (as [1] does) that there could be a relation of subordination or superordination between the two since there are not two terms to enter such a relation.

The second way in which a PT can be related to an LLT is when the two are *lexical variants* of one another. In MedDRA’s sense, lexical variants include such differences as spelling differences, morphological differences, differences in how the sub-terms are ordered, and one term’s being an abbreviation of the other. In none of these cases can the one term be seen as being more general than the other in any sense, nor is there any coherent sense in which one of the terms could be thought of as being at a “higher level” than the other. Indeed, [1] appears to concede this in saying that the terms comprise “Different terms for the same expression.”² But it then immediately goes on in an example to assert that the LLT is subordinate to its PT. In this case, we must simply regard this assertion as devoid of any sense – or at least as beginning to reveal the bizarre sense in which “subordinate” and “superordinate” must be interpreted in this context.

The third sense in which a PT can be related to an LLT is by means of being *synonymous* with that LLT. Here, there are in fact two distinct terms (such as ‘Arthritis’ and ‘Joint inflammation’), but they *mean* the same thing. In MedDRA’s characterization, they stand for the *same concept*. Because of this, there is clearly no difference in specificity between the two terms, and so again in this case the PT/LLT relation could not be one of superordination.

The fourth basis for the PT/LLT relationship may be that in which one term is a *quasi-synonym* of the other. The relation of quasi-synonymy is not well characterized in [1], but it is essentially a relation of synonymy constrained to a particular context (“in a given terminology”). The distinction between syn-

²Note that this in itself is erroneous. They are in fact different expressions related by lexical transformations.

onymy and quasi-synonymy appears to be that synonymy holds in the broader context of natural language generally (or at least within biomedical language generally), while quasi-synonymy holds only within a particular controlled vocabulary or terminology (such as MedDRA). However, this *is* a relation of *synonymy* (however constrained), and so it falls under the same considerations (within MedDRA) regarding the subordination/superordination relation as does the stronger synonymy relation. And consequently, it too cannot be thought of as supporting a relation of superordination between a PT and its LLTs.

This leaves us with the final sense in which a PT may be related to an LLT: that of *sub-element*. Only the briefest attempt is made in [1] to characterize this relation where it is said that an LLT is a sub-element of its PT in the case that the LLT expresses “more detailed information such as anatomic specificity”.

The sub-element relation in MedDRA is highly suggestive of the relation of *generic posting* described in section 8.2.4 of [2]. In generic posting, an explicitly hierarchical relationship is “treated as” an equivalence relationship by employing a broader class name as a preferred term (of the narrower terms subsumed by that class). In this case, *if* we are able to interpret the PT as being in some sense “more generic” than its non-PT LLTs, then *this* sense of the PT/LLT relation would provide a coherent way in which a PT should be regarded as superordinate to its LLTs. We will reconsider this point shortly, but first let us turn to a consideration of the sort of confusions engendered by MedDRA’s accounts of the PT/LLT relation and claims about superordination between PTs and their LLTs.

Can the Paradox be Explained Away?

One approach to explaining away the appearance of paradox is hinted at by Fung in [6]. This requires seeing Figure 1 as a hierarchy of *term types* rather than a hierarchy of medical concepts. Only in this way can a PT be seen as superordinate to its LLTs, and can “superordinate” have a uniform sense when applied to PTs and LLTs as well as to SOCs, HGLTs, and HLTs. But in this sense, “superordinate” does not mean “more general”. Rather, it means only “occurs previously in the hierarchy”. It acquires a purely positional sense in the description of a terminology, and the “hierarchy” no longer represents a genuine hierarchy of generality – it becomes just a way of partitioning *terms* into different (lexical) classes.

This avoids the paradox (narrowly) by sacrificing any meaningful and principled relation of the hierarchy to medical concepts, it results in a purely formal and sterile view of MedDRA and its hierarchy, and even then it is not obviously compatible with passages in [1], section 3, which describe relations of terms to concepts, symptoms, signs, diseases, etc., and the rela-

tive “specificity” of terms. In the end, then, such an interpretation is untenable and would require us to abandon the common sense view of MedDRA which is otherwise compatible with [2] and with much of what is said in [1].

An alternative attempt at avoiding the paradox may start with a very close reading of the ANSI/NISO standard [2]. In sections 8.1 and 8.2 of that work, great care is taken in characterizing the set of equivalence relationships among terms, and an equivalence relationship is said to be one in which “each term is regarded as referring to the same concept” (p. 43).

However, if we look closely at the description of equivalence relationships in section 8.2 of [2], it appears that these are not in fact genuine equivalence relationships – or at least they are not always so.³ For example, the relation USE FOR is not symmetric nor is it transitive. Its subject must be a preferred term and its object is intended to be a non-preferred term so that if “ τ USE FOR η ” is true, then “ η USE FOR τ ” *cannot* be true.

Thus USE FOR is *asymmetric* and there is an asymmetry between preferred terms and their non-preferred “equivalents”.⁴ If we acknowledge this asymmetry and abandon the attempt to see the relation among a PT and its LLTs as an equivalence relation, then this may serve as the ground for holding that a PT occupies a different “level” in the structural hierarchy of MedDRA. However, a general case cannot be made for such a view since among the five basic relations that a PT may have to its LLTs only one (the sub-element relation) may exhibit such asymmetry. The others (identity, lexical variance, synonymy, and quasi-synonymy) are all strong equivalence relations, and in all those cases it would be incorrect to say that the PT is at a higher level than its LLTs. Moreover, MedDRA does not formally represent the basic relations on which the PT/LLT relation is said to rest; and given this lack of information, it would be unjustified to take the general view that a PT is superordinate to its LLTs. In addition, the treatment of generic posting in [2] strongly suggests that this should be treated as a genuine equivalence relation and *not* as a hierarchical relation, saying that it “places limits on the specificity of the controlled vocabulary”.

An Ontological Alternative⁵

Rather than attempting to explain away the paradox by splitting terminological hairs or attempting to jus-

³The concept of equivalence relation is well established in mathematics and formal logic as that of a relation which is reflexive, symmetric, and transitive. See Ch. 10 of [8] for details.

⁴The issue here is seriously muddled in [2] by describing the relation between USE FOR and USE as “asymmetric” when in fact one is the *converse* of the other. See [8], p. 226.

⁵A precise and correct distinction between terminologies and ontologies is drawn in [10].

tify MedDRA’s confusing attempt at characterizing the PT/LLT relation, I propose an alternative approach illustrated by Figure 2. This may, on first inspection, be thought of as merely adding some detail to Figure 1, but in fact it is quite different in that it carefully separates the hierarchical relation of subordination (among the SOC/HGLT/HLT/PT levels) from the equivalence relation that holds among LLTs (including PTs).

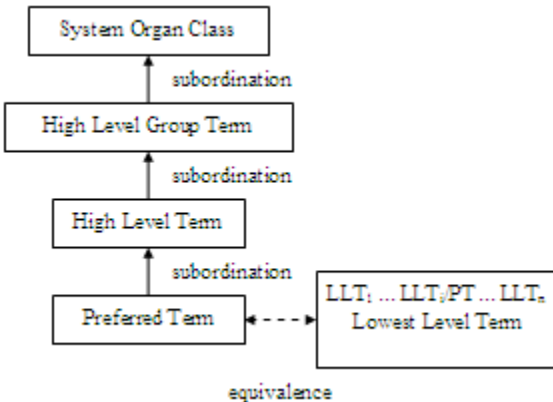


Figure 2: A Coherent MedDRA Hierarchy

This makes it possible to interpret the subordination relation (semantically or ontologically) as a genuine *is-a* (inclusion, subsumption) relation in a fully uniform and principled manner without becoming entangled in the non-hierarchical equivalence relation among LLTs and PTs. It also allows us to treat the relation between a PT and its LLTs as a semantic equivalence relation without treating it as a pragmatic equivalence relation (which would require USE and USE FOR to be symmetric in the usual sense).

A consequence of this is that only the top four categories of Figure 1 represent genuine hierarchical *levels* while the category of Lowest Level Term represents what may be thought of as a *pseudo-level* (if it needs to be thought of in this way at all). The (now genuine) four levels represent “levels of generality” in conformance with the common sense view, and there is no confusion as to whether a PT occurs at two distinct levels. A PT occurs only at the lowest of the four levels, but it bears the equivalence relation to its associated set of LLTs (including itself). No longer is a PT superordinate to its LLTs (and perhaps even to itself), and so the paradox is escaped in a clear and principled manner. The description of [1], section 3 can easily be rewritten to conform to this representation, and such phrases as “its subordinate LLT” can be replaced with “its associated LLT” with a resulting increase in clarity, no loss of content, and compatibility with the ANSI/NISO standard [2].

MedDRA in the UMLS Metathesaurus

An additional advantage of the ontological alternative is that it is fully compatible with the representation of MedDRA in the UMLS Metathesaurus. There, for example, the term ‘Arthritis’ has both an LT representation (atom A0026574) and a PT representation (A2849292) in the MRCONSO.RRF file which represents “Concept Names and Sources”. And in addition, the term ‘Joint inflammation’ – an LLT of ‘Arthritis’ in MedDRA – is also represented as an LT (A0726451). But of these, only the PT representation (A2849292) appears in the “Computable hierarchies” file (MRHIER.RRF) that represents the MedDRA hierarchy within the Metathesaurus.

Thus from the point view of the Metathesaurus, the MedDRA (non-PT) LLTs do not appear in the MedDRA hierarchy, and the equivalence relation among PTs and their LLTs is handled by representing it in the “Related Concepts” file (MRREL.RRF) through term relations such as RQ (*related and possibly synonymous*) and implicitly in the MRCONSO.RRF file by means of the LT and PT term types. In short, from the perspective of the Metathesaurus, there is a clear sense given to “level”; and according to that, MedDRA has only four levels rather than five.

The Metathesaurus representation of MedDRA indicates that UMLS is taking the sort of ontological view of that source that I have recommended, while at the same time it is taking care to clearly distinguish term relations (such as equivalence) from more properly semantic or ontological relations (such as subordination) as these appear in MedDRA. The result is a coherent and highly useable representation of MedDRA terms and the MedDRA hierarchy – but only on the basis of the ontological view I have urged in the previous section. Users (particularly novice users) of the UMLS MedDRA source are sometimes confused by the fact that the UMLS MedDRA hierarchy does not contain “lowest level terms” (and thus that the Metathesaurus appears not to represent MedDRA correctly), but we can now see why it does not, and why it should not.

Generalizing this example of MedDRA and its Metathesaurus representation suggests that, more broadly, the Metathesaurus should be conceived as a *meta-ontology* in addition to being viewed as a *metathesaurus*. This means that the Metathesaurus should be seen not merely as representing relations among terms, but rather as providing a *semantics* that relates terms to the “things” that they denote (medical entities, drugs, etc.). This semantics is not clearly and explicitly represented in the Metathesaurus, but it is possible to see it lurking in the shadows in terms of how atoms and AUIs appear and are related to UMLS concepts and CUIs.

The advantages to such an ontological view of the-

sauri (and other “sources” or “vocabularies”) within the Metathesaurus should now be clear on the basis of the MedDRA example we have considered in detail, and in contrast to the otherwise bewildering paradoxes that a purely term-oriented approach invites.

Conclusions

Our examination of MedDRA’s paradoxical treatment of PTs and LLTs has provided some valuable insights into the relations of controlled vocabularies and ontologies. In conjunction with this, a consideration of the ANSI/NISO standard for the construction and management of controlled vocabularies has shown that these insights extend beyond MedDRA and its potential uses to the more general use of vocabularies, thesauri, and ontologies – and to the relations among these.

If a terminology system such as MedDRA is considered or used in isolation (and particularly for the specific purpose for which it was designed), its paradoxical features may have no untoward consequences and thus may be tolerated – or perhaps they may go totally unnoticed. But if we desire to make use of it in a broader informatics context and in conjunction with other vocabularies, thesauri, or ontologies, then we must examine these features with a more critical eye, eliminate inconsistencies in a principled manner, determine which features of the original structure are essential and must be preserved, and provide a semantic or ontological interpretation that supports its integration with other – potentially disparate and perhaps incommensurate – structures. In the case of MedDRA, we have seen that this requires abandoning the five-level view of its hierarchy in favor of a four-level view that rests on a more careful characterization of the relations between preferred and non-preferred (lowest level) terms. And we have seen this approach illustrated in the integration of MedDRA within the UMLS Metathesaurus.

Finally, we should reflect on a more general lesson that has been illustrated here; and this involves the relation between a vocabulary or terminology (i.e., a system of terms) and an ontology (a system of categories, classes, or other abstract entities). Part of this lesson is to remember that, historically, biomedical terminologies have each been developed for specific purposes and from a particular terminological perspective (represented, for example, by [1] and [2]). If we need to make use of such terminologies for new or different purposes (such as coordinating them with one another or using one as a “reference terminology” to represent disparately coded data sources), then we must be sensitive to “parochial” features of them that may not transition to their new roles. In particular, we must beware of an inclination to preserve terminology relations in

true ontologies and of inferring semantic or ontological relations from terminological relations – even when an explicit appeal to such relations has been made in characterizing the original terminology. Failure to follow such cautions will continue to lead us down the path to confusion and paradox.

References

- [1] MedDRA Maintenance and Support Services Organization. Introductory Guide, MedDRA Version 10.1. International Federation of Pharmaceutical Manufacturers and Associations. Reston, VA.; 2007 Sep.
- [2] National Information Standards Organization (NISO). Guidelines for the construction, format, and management of monolingual controlled vocabularies. NISO Press. Bethesda, Md.; 2005. ANSI/NISO publication Z39.19-2005. Available at: <http://www.techstreet.com/cgi-bin/pdf/free/455225/Z39-19-2005.pdf>.
- [3] Lipow S. MedDRA 9.0 Information Page. 2006 Jul 26. Available from: http://gforge.nci.nih.gov/docman/view.php/53/2278/MedDRA_Source_Information.html.
- [4] Hole T, Carlsen B, Tuttle M, Srinivasan S, Lipow S, et al. Achieving “Source Transparency” in the UMLS Metathesaurus. MEDINFO. 2004 Sep;11(Pt 1):371-375.
- [5] Sa’adon R. Re: AUI and SUI. In: UMLSUSERS-L [Internet]. Bethesda (MD): National Library of Medicine (US); 2008 Jan 15, 10:48 PM.
- [6] Fung K. Re: AUI and SUI. In: UMLSUSERS-L [Internet]. Bethesda (MD): National Library of Medicine (US); 2008 Jan 16, 2:55 PM.
- [7] MedDRA Maintenance and Support Services Organization. MedDRa ASCII and Consecutive Files Documentation, MedDRA Version 10.1. International Federation of Pharmaceutical Manufacturers and Associations. Reston, VA.; 2007 Sep.
- [8] Suppes P. Introduction To Logic. Dover Publications; 1999. 312 p.
- [9] Unified Medical Language System. Bethesda(MD): U.S. National Library of Medicine; 1999 Jan 1 [updated 2008 Mar 5]. Available from: <http://www.nlm.nih.gov/research/umls/>.
- [10] Gruber T. Ontology. 2007 Sep. Available from: <http://tomgruber.org/writing/ontology-definition-2007.htm>.