Textual and logical definitions in ontologies

Selja Seppälä
Department of Philosophy
University at Buffalo
Buffalo, NY, 14260, USA
Email: seljamar@buffalo.edu

Yonatan Schreiber
CUBRC
4455 Genesee St.
Buffalo, NY, 14225, USA
Email: yonatan.schreiber@cubrc.org

Alan Ruttenberg
School of Dental Medicine
University at Buffalo
Buffalo, NY, 14215, USA
Email: alanruttenberg@gmail.com

***Abstract*—We discuss the structure and functions of definitions and axioms in ontologies from the perspective of a terminologist and logician respectively. By working through a few examples of the correspondence between parts of the textual definitions and the axioms, we show how to compare and contrast each and how each perspective reveals areas for improvement. Having established a correspondence between the textual and logical parts of ontology term definitions, we discuss the possibility of developing tools that help developers improve their ontologies. Such tools could be used to check both the textual definitions against the asserted axioms and vice versa. In addition, we propose a few other ways of checking the contents of textual definitions.**

Keywords*—*textual definitions, natural language definitions, logical definitions, OWL axioms, checking definition contents, problems in definitions, functions of definitions in ontologies, recommendations for definitions in ontologies, ontologies, terminology

I. Introduction

Ontologies have on the one hand axioms that form parts of the logical definition of terms, and on the other hand natural language definitions and other documentation of those terms.

However, the ontological world does not seem to have a theory of what the functions of textual as opposed to logical definitions are. The result of that is authoring practices that vary widely. There are nevertheless correspondences (to a certain extent) between phrases in the textual parts and the logical parts. We can use an expectation of correspondences between the textual and logical parts to build tools that help developers improve their ontologies and provide guidelines for identifying issues in axioms and definitions. Aspects we can exploit are:

- Leverage logic to help establish correspondences between the textual definition and the axioms.
- Leverage principles of organizing terminological entities (definitions, notes,...) to characterize the logical parts.
- Measure some part of the quality of an ontology in terms of these correspondences.

Thus, it may be feasible to bring automated methods used in the terminological world to bear on both establishing the correspondences and identifying quality issues in the textual part that could be mapped to quality issues in the logical form.

In this communication, we show examples of varying definition practices in ontologies to support our first thesis and describe issues in definition practices. We discuss the structure and functions of definitions and logical parts from the perspective of a terminologist and a logician respectively. By working through a few examples of the correspondence between parts of the textual definitions and logical parts, we show how to compare and contrast each and how each perspective reveals areas for improvement.

We suggest that it is possible to write tools that analyze textual definitions with the goal of offering places for improvement. We discuss how such tools could be leveraged to check the contents of both textual and logical definitions for terms in ontologies. Our recommendations could also contribute to supplementing the specifications of the OBO Foundry principles on textual definitions.[1]

II. Textual Definitions

In an ontology, a textual definition is, ideally, a short sentence found as the object of an annotation property designated for that purpose. This kind of natural language definition is also found in specialized terminological dictionaries. The account we give in the present communication is thus based on the more developed account of terminological definitions in [1], [2].

A good definition conveys the intended meaning of an ontology term — we will come back to this later — by describing the type of thing to which the term refers. For example, the Cell Type Ontology (CL) contains the following definition for the term *leukocyte*:

(a) *An achromatic cell of the myeloid or lymphoid lineages capable of ameboid movement, found in blood or other tissue.*

This example shows that the term *leukocyte* refers to those things that are of the type achromatic cell and that are distinguished from other achromatic cells in virtue of being: of the myeloid or lymphoid lineages; capable of ameboid movement; found in blood or other tissue.

As we can see, a definition normally states the type of thing to which the instances of the defined term belong, and distinguishes these instances from the type and from other things falling under the same type by listing one or more of the characteristics of the instances of the term.

The first part, the head of the definition is called the *genus*; a distinguishing part, *differentia*. Thus, a definition has a structure where each part is related to the defined term's instances by some type of relation:

[1] http://obofoundry.org/wiki/index.php/FP_006_textual_definitions

- In the classical Aristotelian form, the genus (implicitly) expresses an *is_a* relation, as in example (a) above, which we read as: *a leukocyte is an achromatic cell*.
- The differentia may express any kind of relation relevant for describing and distinguishing the kinds of things to which the defined term refers. In example (a) above, the relations expressed in the definition of *leukocyte* are respectively develops_from (*of the myeloid or lymphoid lineages*), capable_of (*capable of ameboid movement*), and located_in (*found in blood or other tissue*).

A textual definition also has a logical form that derives from the relationship between its *intension* (that which is said about the referent) and its *extension* (the set of instances that fall under the intension). We can distinguish three main logical forms:[2]

Classical definition A definition where the intension holds for *all* instances of the type that is defined, as in *Every instance of X is a Y and all instances of X Z. . . .* In this case, the characteristics expressed by Y and Z are necessary and, in the ideal case, they are jointly sufficient for including all instances of X and distinguishing them from other instances of Y. The ideal case corresponds to the Aristotelian definition by necessary and sufficient conditions. A standard example of classical definition is that of triangle: *A rectilinear figure that has three sides.* (All triangles are rectilinear and have three sides.)

Typical or prototypical definition A definition where the intension holds for *most* of the instances of the type that is defined, as in *Every instance of X is a Y and most instances of X Z. . . .* An example of prototypical definition for a swan would be *An aquatic bird with a long neck, usually having white plumage.* (Most swans are white.)

Instance definition A definition where the intension holds for *only a single* instance, as in *X is the only Y that Z. . . .* These correspond to proper definite descriptions. This kind of definition would apply, for example, to ontologies that include what may be considered as proper names, such as the *Large Hadron Collider* (LHC) in an ontology of nuclear physics. In this case, the relevant kind of differentiae would probably inform us about the geographical location of the LHC and specify that it is (or was until some point in time) "the world's largest and most powerful particle accelerator."[3] The definition could be even more specific and tell us about the length of the ring and the number of magnets that compose it.

Normally, ontologies contain classical definitions because their function is to disambiguate terms. This is not to say that the other forms cannot appear in the textual definitions, but this would not be ideal with respect to the function they are meant to fulfill in this context; without necessary and sufficient conditions it becomes possible to interpret terms in a manner that deviates from their intended use.

Indeed, the main function of textual definitions in ontologies is to specify the intended meaning of the ontology terms in order to avoid ambiguities and errors when, for example, annotating biomedical research texts or importing terms into other ontologies. Of course, this is also the function of the axioms, as we will see in the next section. However, the latter can be somewhat obscure to non-ontologists who may need more detailed and explicit information about the term and its referent.

Therefore, there is a cognitive advantage in including textual definitions in ontologies. As argued in [1, section 1.3], dictionary-type definitions are meant to adjust users' lexical competence [3] by modifying (or confirming) their knowledge about the use of terms. In ontologies, definitions allow users to make their use of a term converge toward that of the rest of the users of the ontology. Both the genus and the differentia contribute to the cognitive adjustment: the genus is meant to provide a sort of cognitive anchor by stating a term that should be familiar to the user of the definition; the differentiae are meant to tell the user how the defined thing differs from the thing that is expected to be already known.

III. Axioms in Ontologies

Axioms in ontologies restrict the intended meaning of a term by asserting necessary conditions for its use. They thus function in a manner analogous to the necessary conditions previously discussed under **Classical definition** in section II. In OWL, it is rarely possible to provide sufficient conditions, so axioms do not on their own constitute full definitions. We distinguish three primary functions of ontology axioms: disambiguation, taxonomic schematization, and fact-modeling.

The function of axioms in the disambiguation of terms is analogous to the function that textual definitions play in disambiguation. Every axiom represents a necessary condition for entities in the terms extension. Axioms thus help to determine the extension of a term by restricting it to those entities meeting the asserted condition. Each additional axiom restricts the extension further, though it is usually not possible to restrict the term to only its intended extension by providing conditions that are jointly sufficient. The most common type of axiom asserts an *is_a* relation that relates the defined term to a parent class by means of the subClassOf relation. For the most part, the relatum of such an axiom should correspond directly to the genus in the textual definition.

We call the second function we identify 'taxonomic schematization'. When employed in this capacity, an axiom asserted for a class provides a schema or template for the axioms of any subclasses. This provides, in our view, robust, principled taxonomic relations between parent, child, and sibling classes. A class's axioms are inherited by all of its subclasses. This makes it possible to use axioms to suggest differentiae for its child classes, in other words to use these axioms as templates for the axioms of the subclasses. This can be done by asserting a relational axiom for the parent class relating it to some other kind of entity (e.g. by writing an axiom for a class X asserting that any X is 'part_of some Y'). For every subclass of this related kind, a subclass of the parent can then be distinguished. For example, the axioms specifying the term *infection* in the Infectious Disease Ontology (IDO) can be used to generate the subclass axioms of its child terms, such as *amebiasis* (see the axiom under *SubClass Of (Anonymous Ancestor)* in Figure 1; see also the discussion of this example in section IV-C below).

[2] X, traditionally called the *definiendum*, stands for the defined term's referent; Y for the genus; Z for a differentia; Y and Z together for the definition itself, traditionally called the *definiens*.

[3] http://home.web.cern.ch/topics/large-hadron-collider

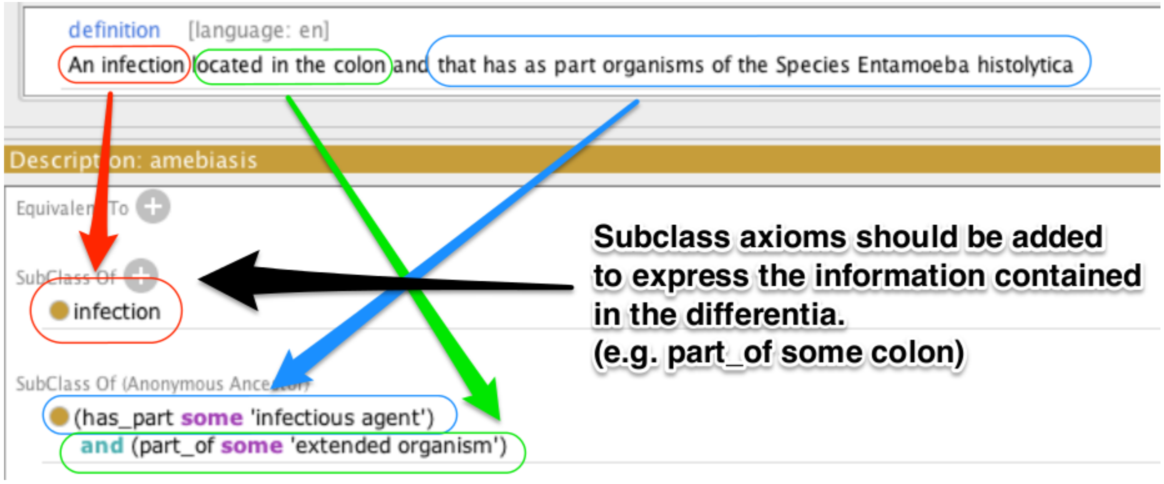

Fig. 1. Correspondences in the parts of the textual definition and the axioms of the IDO term *amebiasis*.

Lastly, we distinguish a fact-modeling function of axioms. An ontology can be considered a specification of a controlled vocabulary for expressing facts in a given domain. Such a vocabulary is much sparser than the vocabulary that would be used to express these facts in natural language, that is, there is a one-many correspondence between ontology terms and words in domain-relevant portions of natural language. This means that the syntax for expressing facts (i.e., assertions between instances) using ontology terms necessarily diverges from the syntax used for expressing the same facts in natural language. The RDF-schema regularizes this syntax substantially, but it is still generally the case that RDF syntax plus the list of terms in the ontology underdetermine how any given fact should be translated from natural language into an expression using the ontologys controlled vocabulary. An important function of axioms in ontologies is to provide a schematic suggestion of how this should be done. Thus, axioms complement textual definitions in contributing cognitively towards regularizing users' employment of terms. For example, the axiom 'is_about some document' in one of the axioms specifying the term *abstract* in the Information Artifact Ontology (IAO) tells us that the relation expressed by the verb *to summarize* in natural language is expressed at the logical level by the *is_about* relation that is part of the controlled vocabulary of the ontology (see annotations in blue in Figure 2).

IV. Correspondences Between Textual and Logical Definitions

As we have seen, axioms and textual definitions have overlapping and complementary functions. Hereafter, we examine how they contribute to conveying the intended meaning of terms. We compare and discuss some examples in the biomedical domain to show how these different forms relate. The examples will show what kinds of issues or inconsistencies can be identified by these comparisons; they reveal at least five types of correspondences. We also give some recommendations as to how to improve both the textual definitions and the related axioms. For sake of readability, we will illustrate the cases with screenshots of the ontology editor Protégé.

A. General recommendation

Based on the identified functions for textual definitions and axioms, we make the following general recommendation: textual definitions should contain content analogous to what is expressed in the axioms, i.e., descriptive content that motivates the logical axioms. The expressions used in natural language may however be more idiomatic than the ontology vocabulary (e.g., the expression *inheres_in* is not very natural). Any complementary information that is deemed useful for understanding the intended meaning of the term but which cannot be included in the axioms should be systematically asserted using other annotation properties.

B. Exact correspondence

Figure 3 shows that the parts of the textual definition of *dead-end host* in IDO correspond exactly to the logical definition by necessary and sufficient conditions. The only difference is in the natural language expression (*bearing*) used for the *has_role* ontological relation — perhaps to avoid the seemingly redundant use of 'role twice. Here, the logical part is useful to fix the intended meaning of the natural language expression.

C. Structural correspondence but more specific content in textual definitions than in axioms

Figure 1 shows that both differentia of the textual definition of the IDO term *amebiasis* contain information of the type expressed in the subclass axioms inherited from the parent class *infection* (see annotations in blue). However, the content conveyed by the parts of the textual definition of *amebiasis* are more specific than the properties and classes expressed in the axiom; they are subproperties of the relations and subclasses of the relata in the axiom.

If these inherited parts are relevant for distinguishing all the subclasses, then all textual definitions at that subclass level should include that kind of information with the specific content that actually distinguishes each entity at that level. If the comparison reveals a match of logical and textual parts at the level of inherited logical parts, this might be a sign that

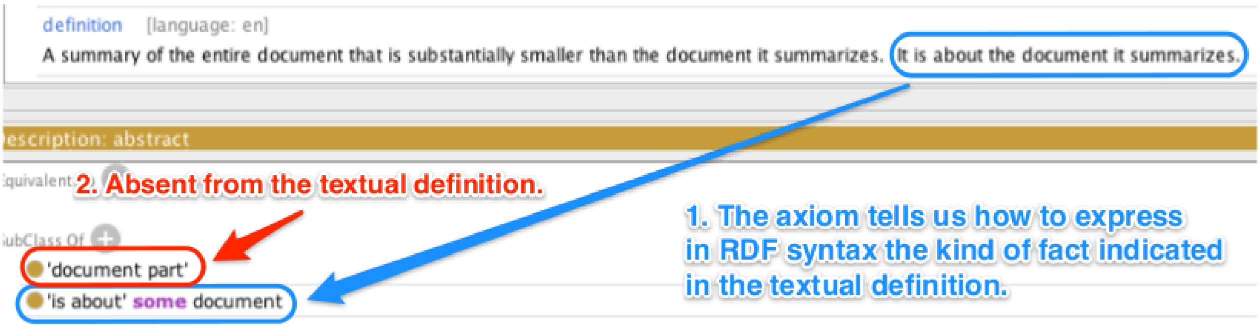

Fig. 2. Correspondences in the parts of the textual definition and the axioms of the IAO term *abstract*.

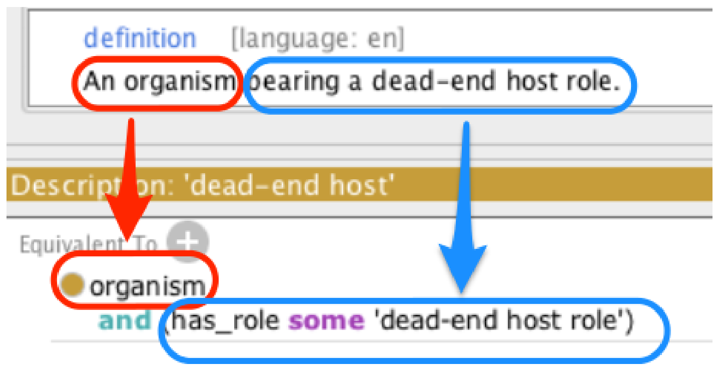

Fig. 3. Correspondences in the parts of the textual definition and the axioms of the IDO term *dead-end host*.

the entity lacks an available subclass axiom. If this is the case, the textual definition can be used as a basis for creating the missing axioms.

We thus recommend that more specific axioms be added whenever the ontology has the resources to include them, i.e., if the terms are defined elsewhere in the ontology. For example, the axioms specifying the IDO term *antiseptic role* in Figure 4 could be completed as follows:

```
subClassOf
realized_by only has_participant
        some (anatomical entity
              and part_of some organism)
```

D. Incomplete textual definitions

Figure 2 shows that the axiom specifying the term *abstract* in the IAO contains the information 'document part' which is absent from the textual definition.

We recommend that the textual definition be completed with this information.

E. Missing axioms

Figure 4 shows that the last part of the textual definition of the IDO term *antiseptic role* does not correspond to any logical part (see annotations in green). However, this more specific differentia serves to distinguish the defined term from (1) *antimicrobial disposition*, which has the same subclass axiom (in blue), and (2) the sibling term *disinfectant role* which is specified by exactly the same axioms. It would therefore be useful to have an axiom that allows these three terms to be logically distinguished.

Here again, we recommend that the axiom be added whenever the ontology has the resources to include the missing axiom.

F. Redundant parts of axioms or definitions

Logical parts may contain axioms specifying other terms.

Figure 4 shows that part of the axioms specified for *antiseptic role* in IDO correspond to:

- the subclass axioms specifying the term 'antimicrobial' — the 'material entity' (see annotations in red);
- the subclass axioms specifying the term 'antimicrobial disposition' (see annotations in blue).

This should not be a problem at the logical level, since the inferences that are made based on the logical expressions end up being the same.

We recommend nevertheless that the axiom be simplified by using the terms that are specified by those axioms. For example, in this example, the first part of the axiom

```
(inheres_in some
     ('material entity'
      and (has_disposition
           some 'antimicrobial disposition')))
```

can be replaced by the following simpler expression:

```
     inheres_in some 'antimicrobial'
```

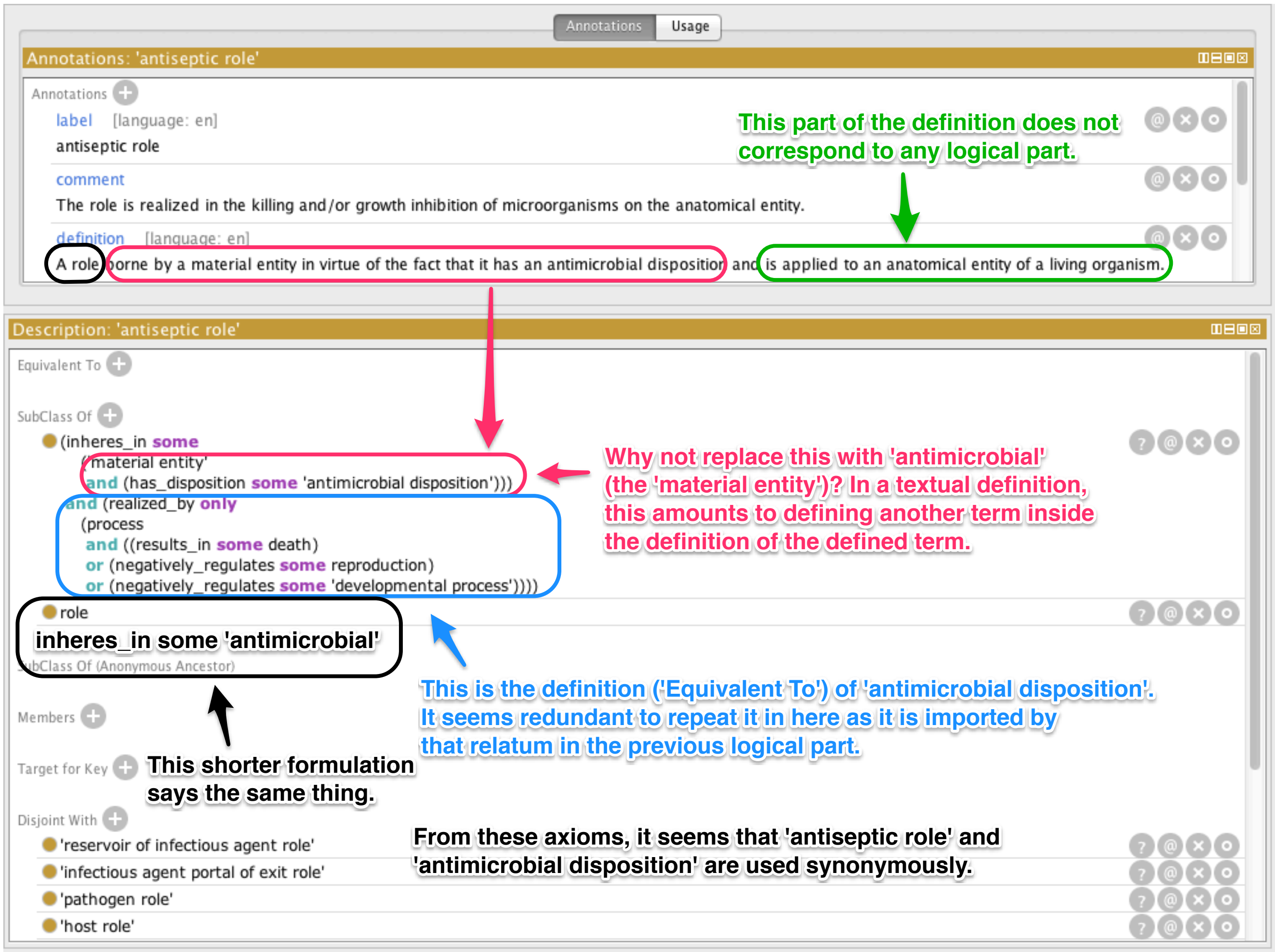

Fig. 4. Correspondences in the parts of the textual definition and the axioms of the IDO term *antiseptic role*.

In a textual definition, this amounts to defining another term within the definition of the defined term, as can be seen in the first differentia of the example (in red), which contains the definition of *antimicrobial*. This lacks conciseness and is generally considered bad practice (see for example [4, 28]). It unnecessarily overloads the contents of the definition — imagine if each term of a definition was replaced by its definition. More importantly, the reader might not recognize that it is the definition of another term and fail to link the defined term with that other one.

We thus recommend that whenever a textual definition contains the definition of another term from the same ontology or an imported ontology, this sub-definition be replaced by the corresponding term. In this example, the differentia *borne by a material entity in virtue of the fact that it has an antimicrobial disposition* should be replaced by *borne by an antimicrobial*. If the reader does not know the term used in the definition, she can (in principle) look it up in the ontology. A system of hyperlinks should also be provided for easier access, as it is done in electronic dictionaries and in the axioms.

V. Using the Correspondences to Help in Definition Checking

In ontologies that use semi-automated systems to create the logical and the corresponding textual definitions, such as *TermGenie*[4], both definition forms are expected to be reasonably consistent. However, when definitions are hand-crafted or imported from other sources, such as other ontologies or, for example, from Wikipedia, various kinds of errors or inconsistencies can creep in, as discussed above. Identifying these problems manually is less rigorous if no guidelines are provided.

To increase reliability of definition-content checking, we propose a method that could be implemented in a computer program to assist ontology editors/curators in carrying out this task in a systematic way. This method can also be used as a guide to manual identification of issues in definitions.

The method consists in the following steps:

1) Determine whether any of the terms from either the ontology that is being checked or the imported ontologies appear in the textual definitions.

[4] *TermGenie* is used for creating definitions in the Gene Ontology (GO), (http://go.termgenie.org).

2) Get the taxonomic hierarchy of the matched terms to the top level.
3) Determine whether any of the terms in this hierarchy corresponds to one of the relata in the axioms.
4) If no correspondence is found between terms in the textual definition and terms in the axioms, look for a correspondence between the relations expressed in the differentiae of the textual definition and the object properties in the axioms. This can also be done by taking into account the hierarchy of object properties (if available).
5) If matches are found, tag the corresponding part of the textual definition with the corresponding relation–relatum pair (the tagging could supplement the textual definitions with hyperlinks to the entries of the terms and relations used in the definition).
6) If mismatches of this kind are identified, manually correct, modify or complete either the textual definition or the axioms, or both according to the recommendations put forward in this paper.

The proposed method may raise some implementation challenges. For example, the first and fourth steps require natural language processing (NLP) methods to correctly identify existing terms and relations in the textual definition. This involves using methods to find inexact matches, for example, plural forms of terms and partial matches, as when only the head of a complex term is used. Matching ontology relations to natural language expressions can also be challenging, as there can be several ways to express a single ontological relation. A solution for relation identification that also involves NLP methods would be analyzing large amounts of definitions in which each part is matched to the corresponding ontological relation to identify the different corresponding expressions. This solution might reveal domain-specific expressions for the more general ontological relations.

VI. Other Useful Ways of Checking the Contents of Textual definitions

In ontologies, definitions should include only necessary conditions that have the classical all-some form. Thus, they should avoid:

- Particularizing expressions such as *for example*, *especially*, *in particular*, *i.e.*, *such as*, ..., and punctuation signs such as parentheses and colons. Sometimes, differentia may contain hidden examples that should also be avoided, as in the definition of *leukocyte* above which states *found in blood or other tissue*. Here, the specification *blood* is superfluous since it is embedded in a conjunction of which the other conjunct is its superclass.
- Overly generalizing expressions such as *etc.*, *in general*, *normally*, ..., and disjunctions, as these are linguistic markers of conditions that are not necessary.

Although particularizing and generalizing expressions can be useful for a better understanding of the term (as in example (a) above). These kinds of information should be asserted using other annotation properties.

Futhermore, textual definitions should not contain definitions of other terms, as in the definition of *antiseptic role* examined above (Figure 4). Thus, they should avoid:

- Punctuation signs such as parentheses and colons which are also a sign of new definitions.
- Expressions introducing new information such as *i.e.*, *that is*, ...

The content-related issues presented in this section can be automatically checked with a simple rule-based program that uses, for example, lexico-syntactic patterns. This kind of program can also be used for checking the conformity of the surface form of the definitions to the editorial line of the ontology (if any) [5].

In addition to these ontology-specific recommendations, terminological manuals and guidelines state a number of other general principles and recommendations relating to definition writing [4], [6]–[8].[5]

VII. Conclusion

In this communication, we showed through examples that the defining practices in the ontology world lack systematic principles and theory. To fill this gap, we presented some background on textual definitions and axioms in ontologies from the terminologist's and logician's viewpoint, emphasizing their overlapping and complementary functions.

Based on a discussion of various kinds of correspondences between the parts of textual definitions and axioms, we put forward two primary recommendations to improve the contents of both textual definitions and axioms:

- Textual definitions and axioms should, whenever possible, represent the same content. As we hope our examples have indicated, it is frequently possible to do this with the resources of the ontology.
- Neither textual definitions nor axioms should include content that defines another term in the same ontology.

Finally, we proposed an implementable procedure to help systematize content-checking of textual and logical definitions in ontologies.

VIII. Acknowledgment

Work on this paper was supported by the Swiss National Science Foundation (SNSF).

References

[1] S. Seppälä, "Contraintes sur la sélection des informations dans les définitions terminographiques: vers des modèles relationnels génériques pertinents," Ph.D. dissertation, Département de traitement informatique multilingue (TIM), Faculté de traduction et d'interprétation, Université de Genève, 2012. [Online]. Available: http://archive-ouverte.unige.ch/unige:21874

[2] ——, "An ontological framework for modeling the contents of definitions," *Terminology*, forthcoming.

[3] D. Marconi, *Lexical Competence*. The MIT Press, 1997.

[4] ISO 704:2009(E), *Terminology work – Principles and methods (ISO 704:2009)*, 3rd ed. Geneva: ISO, 2009.

[5] S. Seppälä, "Semi-automatic checking of terminographic definitions," in *International Workshop on Terminology design: quality criteria and evaluation methods (TermEval) — LREC 2006*, Genoa, Italy, 2006, pp. 22–27.

[5] For a list of (terminology) manuals that contain definition writing principles and recommendations, as well as other writings on definitions see https://sites.google.com/site/definitionsportal/literature.

[6] S. Pavel and D. Nolet, *Handbook of Terminology*. Canada: Public Works and Government Services - Translation Bureau, 2001.

[7] (2012) The pavel terminology tutorial. [Online]. Available: http://www.bt-tb.tpsgc-pwgsc.gc.ca/btb.php?lang=eng&cont=308

[8] R. Vézina, X. Darras, J. Bédard, and M. Lapointe-Giguère, *La rédaction de définitions terminologiques*, ser. Version abrégée et adaptée par Jean Bédard et Xavier Darras. Montréal: Office québecois de la langue française, 2009.