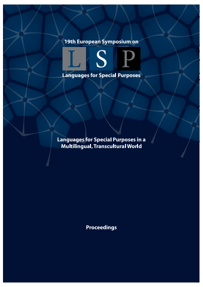

Languages for Special Purposes in a Multilingual, Transcultural World

Proceedings of the 19th European Symposium on Languages for Special Purposes, 8-10 July 2013, Vienna, Austria

http://lsp2013.univie.ac.at/proceedings

Terminology meets the multilingual Semantic Web: A semiotic comparison of ontologies and terminologies

Dagmar Gromann

Cite as: Gromann, D. (2014). Terminology meets the multilingual Semantic Web: A semiotic comparison of ontologies and terminologies. In G. Budin & V. Lušicky (eds.), *Languages for Special Purposes in a Multilingual, Transcultural World, Proceedings of the 19th European Symposium on Languages for Special Purposes, 8-10 July 2013, Vienna, Austria*. Vienna: University of Vienna, 418-428.

Publication date: July 2014

ISBN: **978-3-200-03674-1**

Terminology meets the multilingual Semantic Web: A semiotic comparison of ontologies and terminologies

Dagmar Gromann
Institute for English Business Communication, Vienna University of Economics and Business
Austria

Correspondence to: dagmar.gromann@wu.ac.at

Abstract. To pave the way towards a multilingual Semantic Web its resources such as ontologies need to be expressed in natural language in addition to logic. Ontologies are primarily designed for machine-readability of knowledge and often lack natural language content required for human users. Terminology provides an established model to accommodate specialized language and has the potential to bridge this gap towards multilingual Semantic Web resources. In addition, the technological infrastructure and tool support for ontologies have the potential of boosting terminology formalisms. To lay the ground for their integration a thorough comparison and multidisciplinary understanding of ontologies and terminologies is needed. This paper elaborates on their commonalities and divergences from a semiotic perspective by analyzing their main syntactic, semantic, and pragmatic aspects. Subsequently, results of this analysis are juxtaposed and implications for the ontology-terminology integration are presented.

Keywords. Terminology, ontology, semiotics, comparison, terms, ontology labels, multilinguality.

1. Introduction

The Semantic Web (SW) relies on ontologies and provides a fundamental body of knowledge that enables informed decision-making. As the main design goal of an ontology is to enable reasoning (Krötzsch et al. 2012), it lacks refined means of representing natural languages (Gracia 2012). This ability to represent knowledge irrespective of a natural language (NL) provides one of the greatest potentials for bridging communication barriers due to linguistic differences. An integration of terminology and ontology helps explore this potential. To provide the foundation for their integration, this paper undertakes a profound analysis of the similarities and differences of terminology and ontology from a semiotic perspective.

To provide clarity about the nature of both resources and to profit from their advantages requires more than a mere syntactic comparison and conversion. Semiotics is the study of signs and turned out to be the most adequate perspective for the comparison at hand. Semiotics provides guidelines for defining the representation and use of signs, but more importantly for translating from one sign pattern intended for one purpose to a sign pattern intended for another purpose (Sowa 2000: 55). The three main branches of semiotics – syntax, semantics, and pragmatics – provide the major categories from which comparison criteria are derived.

To provide a clear reference point and draw from community consensus, the major definitions of each resource as well as their syntactic, semantic, and pragmatic aspects are taken from standards (International Standards Organization) and specifications (World Wide Web Consortium). Although both standards provide visualizations, they differ in their use of UML and in granularity. This is why a data model for each resource has been developed in UML. The goal of this paper is to provide a very clear understanding of the nature of each resource as well as their differences and to produce requirements for a possible integration of both.

2. Comparison method

A deep understanding of two different data models and their functions presupposes a well-defined comparison method and analysis of both. The sheer number of varying approaches to defining both terminology and ontology render a clear point of reference for each resource inevitable, which here are standards (ISO) and recommendations (W3C). They reflect community consensus and best practices, being developed and curated by leading representatives of the respective field, and provide the input to the semiotic comparison criteria and modeling method.

2.1. Comparison criteria

Semiotics is the study of signs, i.e., it studies the structure and meaning of language including its non-linguistic sign system. Terminologies and ontologies take signs as a starting point and use them for organizing and representing knowledge. Given that both thus are based on semiotics, their semiotic comparison comes natural. With its three main branches – syntax, semantics, and pragmatics – semiotics provides guidelines for defining sign representation, organization, and use. Syntax regards the form or representation of signs and relates sign representations to one another. Semantics investigates the meaning of a concept and its determination (intension), its relation to physical or abstract objects (reference), and which objects can be instantiated by the concept (extension). In addition, semantics frequently investigates the truth value of statements and propositions. Pragmatics is mostly interested in what the representation effects in the interpreter – how meaning depends on context and use. From this semiotic perspective, the following main criteria form the basis for the present comparison method.

1. Syntax
   a. mechanisms for grouping objects into units and associating them
   b. permissible content of these units
   c. operations permissible on these units
2. Semantics
   a. object the units of knowledge/thought refer to
   b. mechanisms for expressing and interpreting meaning
   c. purpose of modeling intended meaning
3. Pragmatics
   a. usage of resource
   b. context of concept use/interpretation
   c. application scenarios

2.2. Reference definitions

The main organization fostering the advancement of Semantic Web technologies, i.e., also ontologies, is the World Wide Web Consortium (W3C). Its Web Ontology Language (OWL) enjoys great popularity and experienced a thorough revision in form of a recently published second version OWL 2, which will be used to analyze ontologies herein. The “Structural Specification and Functional-Style Syntax” (W3C 2012) represents the main source for creating a data model and conducting the syntactic analysis. As regards semantics, both versions of OWL are based on description logic – a fragment of first order logic. “Direct Semantics” (W3C 2012) and related publications specify its semantic aspects. Its actual usage will be taken from W3C documentation for users.

Concept-centered terminologies are standardized by the International Standardization Organization (ISO) in terms of theory and application (ISO 1087:2000), principles and methods (ISO 704:2009), representation and exchange (TBX - ISO 30042:2008). At the core of these various standards is a detailed analysis of a terminological metamodel in the Terminological Markup Framework (TMF - ISO 16642:2003), particularly targeted towards the re-use and exchange of terminological data. Combined with ISO data categories (ISO 12620:2009), the

TMF makes up the core of the Term Base eXchange (TBX) format – a language family for terminology exchange.

2.3. Modeling method

Visualizing syntactic elements of both resources by means of a graphical modeling language strongly supports and facilitates the human readability of their comparison. The Unified Modeling Language (UML) has become the de-facto standard for a graphical representation of computer artifacts. Both organizations, W3C and ISO, utilize UML to visualize their data models, but differently in terms of cardinality, associations, concept and constraint modeling, and designations. W3C diagrams mostly use cardinality restrictions at the end of directed associations. The TMF meta-model indicates numbers at both ends of the associations, TBX diagrams use a highly complicated method of associating this information to concepts rather than relations, and other terminology standards do no use cardinality at all. The use of associations is equally inconsistent across ISO documents from specific diagram types, e.g. tree diagrams for generic relations, to undirected associations. OWL always uses directed associations or generalizations to connect elements, which are modeled as classes. In ISO sometimes concepts are modeled as classes, then as free text without boundaries, or headings with notes.

Due to these differences in notation and differences in granularity, comparable UML class diagrams for each resource are provided herein. The three types of relations to be used are (1) generalization (the solid line with a hollow arrowhead), (2) aggregation/composition (diamond shape which is not filled/filled), and (3) undirected association (solid line). Generalization is the association of a specific to a less specific concept, where the former inherits the characteristics of the latter. Aggregation means that one element aggregates a number of other elements, which have an existence on their own. In contrast, composition aggregates elements that do not exist independently from the main element. Cardinality restrictions are only introduced where beneficial to the comparison and by no means seek to achieve a complete picture on multiplicity. One very nice modeling feature of UML is the ability to indicate abstract elements by italicizing their designation. This means that there is no syntactic equivalent in the targeted representation language for such an element.

3. Terminology

From its advent, the intent of terminology management has been a highly practical one and always related to a specific domain. Terminological resources are to fulfill a specific function and utilized towards a special purpose, e.g. consistent and homogenous asset management across the entire company. Thus, it has always taken a functionalist perspective, drawing from functionalism in linguistics (Budin 2010: 23). Although there are also semasiological or linguistic approaches to terminology, the one defined by ISO and of interest here is concept-oriented and onomasiological. It is defined as a set of domain-specific concepts designated by terms in a specialized language.

3.1. Terminology structure

In terminologies, concepts are the main mechanism for grouping objects into units of knowledge. Objects are abstracted into concepts, its properties into characteristics forming a definition, and one or several designations are assigned to the concept. The four main elements of a terminology according to ISO have been color-coded in Figure 1. Red elements represent relations that connect white elements, i.e., different types of concepts. Blue elements depict administrative and descriptive details added to concepts (associated information) as well as to the terminological resource (global and complementary information) and designations and definitions. Characteristics represented in green are integrated in the definition of a concept. However, due to their pivotal role in the concept and concept system formation process, characteristics have been assigned a separate color.

While designations, characteristics, definitions, and information are represented in NL, relations and concepts are structuring elements in the concept system. Elements in italics are abstract because they are syntactically realized by their sub-elements. For instance, relations cannot be associative without specifying the type of proximity, such as time or space. The number of associative relations illustrated in the above diagram is by no means complete as there are many other thematic connections between ems. Terminological elements included in Figure 1 are the ones most interesting for a comparison.

Concepts mediate between real or abstract objects they refer to and designations and definitions, which make up the concept. The delimiting characteristics of an intensional definition unequivocally position the concept within the concept system. Not every concept might be included in the terminological resource, as the concepts and their interrelations constitute a coherent concept system. Moreover, the intensional definition determines the extension, the objects that meet all requirements of the intension. The pertaining designation(s) might either be linguistic expressions or non-verbal, such as images, formulas, code, diagrams.

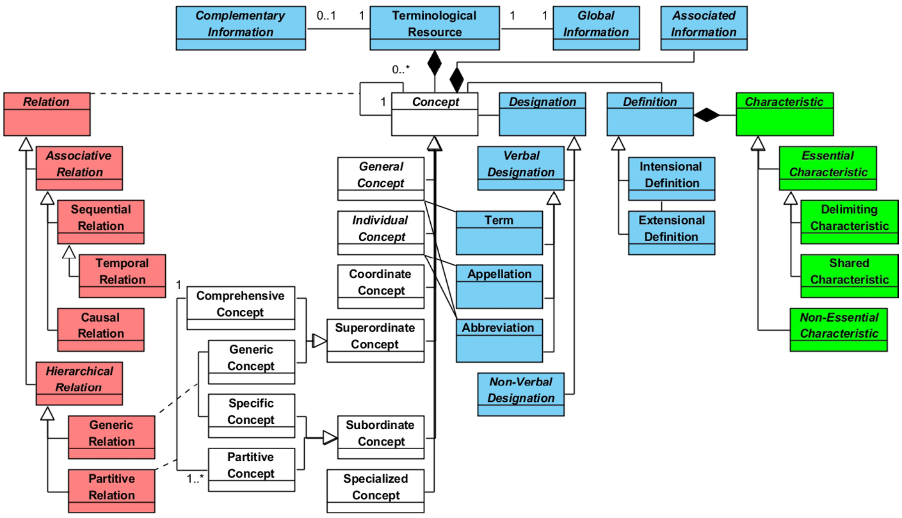

Figure 1: UML Overview Diagram of Syntactic Elements of Terminology

Apart from the concept relations illustrated in Figure 1, other operations can be performed on concepts and related designations. Definitions of such operations are not formal in the sense of machine-readable. They are contained in typed meta-data category such as reference, associative information, or description. Thus, operations are targeted towards human users only. It is possible to state a term is an antonym of another term, which is also possible on a conceptual level. However, synonymy is restricted to the designation level as it is presumed that synonym terms pertain to the same concept. For instance, equivalents in different languages pertain to the same concept and are separated into individual language sections. Two terms within the same language section are understood as synonyms. It is also possible to indicate partial equivalence by means of meta-data.

3.2. Terminology semantics

Conceptualization designates the abstraction of objects into units of knowledge called concepts. Concepts mediate between definition, designation, and objects. Terminology work entails an informed conceptualization of a specific domain, which means objects are grouped and described on the basis of their properties. In the terminological resource, objects are referred to by means of concepts, which mediate between objects and their definitions and designations. This reference is established by a set of essential and delimiting characteristics that uniquely define the concept intensionally and position it within the concept system.

Concepts are not isolated units but semantically located within a concept system, for which the context provides a framework. The basis for establishing a hierarchical ordering are subdivision criteria, such as material, usage, or specific traits of an object. In case there is more than one such criterion, the system is considered multidimensional. The semantics of the hierarchy are validated by means of the principle of inheritance, i.e., the subordinate concept inherits all characteristics of the superordinate concept. Associative relations are always thematic connections, such as by proximity in space and time. In fact, terminological relations in general connect concepts based on their senses (linguistic) rather than meaning (conceptual). Concepts and relations together establish a system that is supposed to be coherent, i.e., provide a coherent view on the specific subject field. Of course, this view depends on the specific context of the terminology.

Having established a concept system, each concept is assigned with designations in a special NL. The term is said to derive its meaning from the concept, while further administrative and descriptive metadata (associated information in Figure 1) contribute to this meaning. This meaning - the set of characteristics making up the definition and its designation - corresponds to the initially introduced intension, while the extension is the set of objects being conceptualized. The truth value refers to the correspondence between object and concept. In case the meaning of the term, the concept, adequately describes the nature of the object we can talk about truth (Budin 2003: 73). Thus, the definition is the central vehicle for conveying and expressing conceptual meaning targeted to human users and currently not machine-readable per se. Nevertheless, the principle of term transparency demands the term to allow for a basic inference of the concept's meaning without further explanations.

Although the boundaries of a term and a concept system might not be clear-cut, each terminology targets a specific context. For instance, an accounting terminology for neophytes differs substantially from the corresponding expert resource. The perspective on the domain and its level of granularity depend on the purpose of and the terminologist(s) modeling the resource. This context also determines the degree of specialization a concept and its designation have. Marzá (2009: 104-05) categorizes this degree along the following five aspects: cognitive, semantic, semantic-pragmatic, semantic, syntactic, and formal. Cognitive refers to the fact that concepts have a clear position within a concept system. Semantic hereby refers to the fact that a term is always more specific in terms of its traits than an entry in a lexicon of the language. Text genre, subject field, way of handling the subject, target audience, and the communicative situation represent pragmatic parameters to determine an expression to be a term (semantic-pragmatic). The syntactic aspect refers to the nominal nature of terms. Formal states that updating terminologies requires a formation process similar to the initial concept formation.

3.3. Terminology usage

A terminology as a collection of concepts and terms is by definition user-oriented and can be applied to improve consistency and coherence in any setting of specialized communication. Thus, the targeted users of the terminology have to be identified prior to initiating terminology work. Terminology science as a field is highly multi- and interdisciplinary in that it crosses linguistic, conceptual, and communicative fields to provide a holistic approach to terminological entries. The three main functions that are generally associated with terminology are epistemic (support of knowledge acquisition), informational (managing information), and discourse-directive (optimizing specialized communication) (Budin 1996: 18). Many specialists, such as engineers, technicians, scientists, are attracted to the field of terminology in need of improving their professional communication. On the other hand, terminologies can also offer a systematic entry point to specialized fields.

One of the major functions of a term is its usage in a specific context. Its discourse-directed function allows it to serve as a source of reference for consistent terminology use throughout and across individual organizations or enterprises. Furthermore, concepts group multilingual terms, their variants, and synonyms, which is why terminologies are an invaluable working tool for translators and other communication specialists. A range of linguistic and terminological details can be added to this set of multilingual specialized terms. The ultimate goal of a terminological data collection is to be used in terminological applications and exchanged as terminological data for human users.

4. Ontology

Ontologies are the main building block of the Semantic Web, but are equally used in biomedical applications, software engineering, and information architecture. Ontologies consist of a set of concepts, relations, and meta-data, but are also grounded in logics, which is why they are equipped with formal semantics (Krötzsch et al. 2012). The most cited definition for ontologies is that of "formal, explicit specification of a shared conceptualization" (Gruber 1995: 907). It is shared in the sense that it should reflect community consensus and formal in that it is machine-processable. Explicit refers to logical propositions. W3C defines ontologies similarly as "formalized vocabularies of terms, often covering a specific domain and shared by a community of users". The main difference is the focus on "terms", logical specifications of entities, and the domain-specific reference. Ontologies vary along the line of expressivity from lightweight to formal and in terms of field of study from upper-level to application and domain ontology.

4.1. Ontology structure

The main elements for grouping objects in ontologies are concepts called ontology classes and derived by the abstraction process called conceptualization. The three main syntactic categories of the Web Ontology Language 2 (OWL 2) depicted in different colors on Figure 2 are entities, expressions, and axioms. This model describes the general structure of OWL and is independent of any syntax used to realize ontologies.

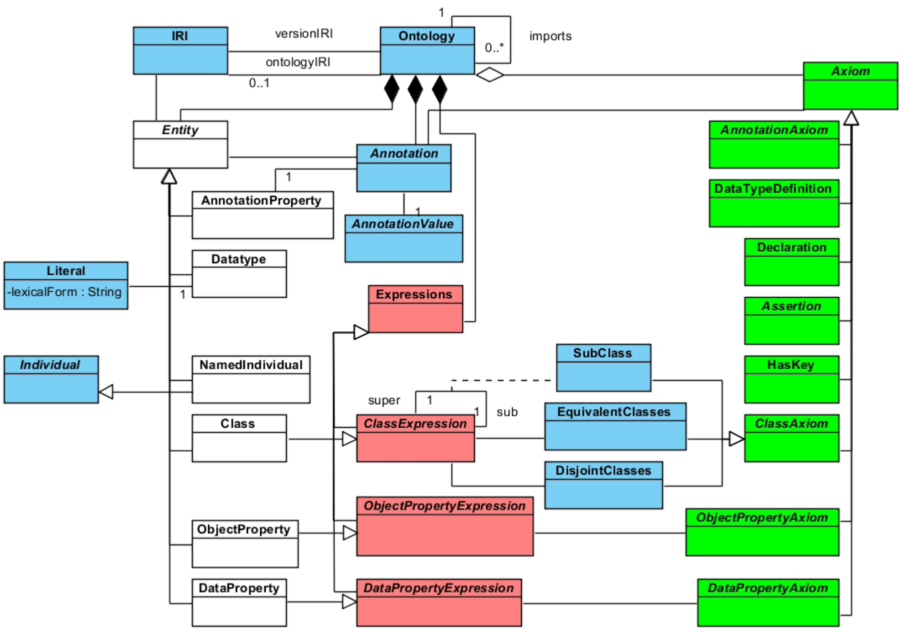

Figure 2: UML Overview Diagram of Syntactic Elements of OWL Ontologies

Entities are classes, named individuals, properties, and datatypes. Each entity is uniquely identified by an Internationalized Resource Identifier (IRI). IRIs are a generalization of the Uniform Resource Identifier (URI) allowing Unicode characters and uniquely identify names and resources on the Web. Classes represent a set of individuals. While named individuals are entities, OWL 2 considers individuals actual objects from the domain. Object properties connect two individuals, while data properties connect one individual to a data value. Datatypes represent sets of such data values, while literals can be considered individuals that denote data values. Literals are lexical forms that provide a string for individuals. Literals, data properties and object properties cannot be used directly to associate information to a class or property. This is why there are annotation properties, which can relate any entity to any value. Annotation properties assign NL content to the ontology, its entities, axioms, and annotations themselves as metadata. Some of them are built-in with a predefined significance, such as *rdfs:label* for NL expressions, *rdfs:comment* for NL definitions and comments, *owl:versionInfo* to specify the version of the ontology. As metadata they are beyond reasoning, i.e., cannot be used for or influence automated inferences.

Axioms are statements or logical propositions that declare what is true in the specific domain of discourse and are generally categorizes as assertional (ABox), terminological (TBox), or relational (Rox) axioms (Krötzsch et al. 2012). The ABox fills the conceptual world with instances of objects, such as assertions, which relate individuals to classes. The TBox describes relations between classes, such as subsumption, equivalence, or disjointness. The RBox does the same for properties and also allows further specifications such as transitivity, symmetry, etc.

Most elements are never referenced directly but by means of expressions, which can be considered either their parent element or a combination of elements. Object property expressions represent relationships between pairs of individuals, of which (inverse) object properties are a subtype. Data property expressions relate individuals and data values, but only have data properties as subtype. Classes and object properties connect to class expressions, called complex concepts in logic. They formally specify conditions on the properties of individuals they represent.

The permissible content of entities in ontologies is restricted by the use of expressions and axioms. For instance, the object property *hasAmountReceiveable* can be restricted to only connect to instances of the class *Creditor* in a specific ontology. As regards permissible operations on these units, OWL 2 provides an ample set of primitives that can connect class expressions. These operators are as follows, whereby the expression in brackets refers to the equivalent Boolean operator or a short explanation: intersection (and), union (or), complement (not), enumeration of individuals (providing a list of individuals), equivalence (same as), and a range of restrictions. These restrictions can either be applied to data properties or object properties, such as universal quantification (for all or every), existential quantification (for some or there exist some), and cardinality restrictions.

4.2. Ontology semantics

In the process of ontology engineering observable or conceivable objects are abstracted by means of conceptualization. Its intension refers to the signature of the ontology, i.e., the set of entities contained (Guarino et al. 2009), as well as the pertaining axioms, which further formally specify the instantiation of the ontology. Ontologies are able to represent knowledge including the objects it relates to by means of individuals, which are added to the ontology in a process called ontology population. Individuals constitute the extension of the ontology (Guarino et al. 2009).

Meaning of OWL 2 ontologies are based on direct semantics every implementation needs to comply with in order to be OWL 2 conformant. This formal semantics allows for the use of logical deduction to infer additional information from the facts represented in the ontology (Krötzsch et al. 2012). A logical consequence is a statement that can be considered true in terms of what is said in the ontology. Thus, direct semantics specify for which possible states of a world a particular set of OWL statements is true. Instead of making default statements about axioms, ontologies specify all possible situations where the depicted axioms hold, i.e., are true or satisfied, which is called the open world assumption in description logics (Krötzsch et al. 2012). Thereby, everything that is left out is simply omitted and not considered true or false.

It has to be noted that the semantics of OWL 2 does not require any NL in order to be specified. However, in order to be intelligible to human users, NL expressions are needed. Labels are the only means of representing terms in ontologies, but there is no standardization of or best practice for content of labels and the only built-in annotation of such a label is its language by means of an *xml:lang* tag.

4.3. Ontology usage

As the interrelations and interaction of axioms can be very subtle and difficult for human users to understand, logical consequences can be inferred automatically. The computation of inferences is called reasoning, which requires tools – reasoners. The higher the expressive power of the selected ontology language the more complex the reasoning task. The trade-off between expressive power and complexity of reasoning needs to be evaluated for each ontology. For this purpose there are three profiles of OWL 2 EL, QL, RL, which are all a subset of description logics and created for a specific purpose. The most common profile for large ontologies with low expressive power is OWL 2 EL specifically designed with terminologies in mind, i.e., complex structural descriptions, large number of classes, heavy use of classification, and application to large amounts of data.

The computation of logical consequences and inferences is not only the main usage scenario for ontologies, but at the same time provides the main context for its interpretation. Reasoning is one important design goal for ontologies (Krötzsch et al. 2012). The reasoning algorithm for automating the activity needs to consider the used subset or profile of description logics. Each reasoning task processes a certain axiom, which means a range of reasoning tasks are required to check the ontology:

- Satisfiability: checks whether the formal concept definition is meaningful
- Classification: computes the subsumption hierarchy for classes and properties
- Consistency: checks whether the knowledge is represented consistently
- Equivalence: infers equivalence of classes
- Querying: provides query language and inference-based answers, including instance retrieval

In case there is no NL content in an ontology, the output of a query or the result of a consistency check is difficult to understand by humans. This is why several approaches seek to verbalize logical consequences as well as content of ontology entailments by using controlled NL, such as the SWAT project (Nguyen et al. 2013). Other approaches investigate ontology visualization as a means of facilitating the human interpretation of ontology content (Fu et al. 2013).

Application scenarios for domain ontologies range from machine translation to the detection of similarities of e.g. products in a product classification ontology or web page contents when managing a corporate website based on ontologies. Most applications nowadays are to be found in the biomedical domain.

5. Comparison of terminology and ontology

While concept-oriented terminologies are defined as a set of domain-specific concepts designated by terms in a specialized language, ontologies are most frequently specified as formal and explicit knowledge representation based on description logic. To explicate the most basic difference in this definition it has to be stated that ontologies refer to and relate objects, while terminologies treat concepts and terms.

5.1. Terminology and ontology structure

A number of elements in the ontology have no correspondents in a terminology: all types of expressions and axioms, assertions. A flat representation of strings in labels without any internal structure is not equivalent to a domain-specific consistently used term in a terminology, but the only option for adding terms. Associative and partitive relations have no direct correspondents, but depending on the context might be expressed by expressions or restrictions. Class assertions would require an instance relation, which is currently not needed in terminological modeling.

One of the clearest equivalences can be established between the generic relation and the subclass axiom. This suggests that some terminology concepts might coincide with ontology classes or individuals despite of the semantic differences discussed in the next section.

Concepts in ontologies are sets of individuals, while terminologies abstract away from objects, but cannot be described as sets of objects. Instead terminologies are defined as sets of concepts designated by terms. As has been argued before, terms are not concepts. In terminology, concepts do not exist independent from their definition written in NL. Ontologies, however, do not even require NL at all, even if its omission is detrimental to user-based applications. Smith et al. (2005: 648) delineate definitions of the term *concept* as psychological (mental entities), linguistic (meaning of general words), epistemic (units of knowledge), and ontological (abstraction of kinds, attributes, or properties). Terminological concepts in this classification clearly are epistemic, while ontologies belong in the last category.

5.2. Terminology and ontology semantics

The process of conceptualization in terminologies entails the reference to real-world object by placing concepts within a clearly defined net of semantic relations and defining it in NL. Assertions and expressions in ontologies make statements about real-world objects and allow relating these objects (individuals) by means of properties. Thus, ontologies realize the extensional definitions, while terminologies fully focus on intensional definitions of concepts. As regards the semantic perspective on truth values, ontologies are based on the open-world assumption, i.e., everything that is not in the ontology is not assigned a truth value, whereas depicted axioms need to be true. This also implies that isolated parts of the ontology are valid as long as they are satisfiable. This usually holds not true for terminology, which depends on the coherence of the entire concept system in one resource. From a semantic perspective, commonalities of both resources are reduced to the linguistic equivalence of contents of an ontology label with a term in a terminology.

The RDF(S) label property system specifies a many-to-many relation to ontology elements. Several labels can be associated to one ontology element, while one label can refer to several elements, i.e., might be ambiguous. This system presupposes that each label has at least one correspondence to at least one ontology element, which raises several issues on the paradigmatic and syntagmatic level. In terms of paradigmatic aspects, there is no specification of the type of semantic relation a set of labels associated with one concept share. On a syntagmatic level, components of NL expressions cannot be associated with a specific concept and thus, are not considered at all in the current label system.

Naming a domain ontology element requires the assignment of linguistic signs to elements. There is no standardized procedure or best practice for this process and it is an optional choice to assign labels. In addition, time pressure and multiple agents (several experts, engineers, and semi-automated term extraction procedures) often lead to a lack of consistency in the naming of ontology elements. Terminology standards and best practices offer established term formation processes and terminological principles that foster consistency of terms in terminological data collections.

5.3. Terminology and ontology usage

The emphasis of both resources in terms of what they can describe and express about the described knowledge differs. Terminologies describe, define, relate, and model the terminological differences and similarities within and across languages, something that has not yet been accomplished by ontologies. On the other hand, ontologies allow for informed decision-making due to the logical statements depicted about the state of the world represented. In a nutshell, terminologies are interested in how terms are used in specialized communication, while ontologies seek to draw inferences on the specific state of a certain domain of discourse.

While terminologies are classified as knowledge organization systems, ontologies are categorized as knowledge representation systems. This differentiation can be explored from a pragmatic perspective. Knowledge organization has its origin in the library science community and is concerned with the construction of semantic tools for information retrieval (Friedman and Thellefsen 2011). Related processes entail classification, indexing, and description of documents and knowledge, which are performed by information specialists and computer algorithms. In knowledge representation the two main types of representation relevant for the purpose at hand are the represented world and the representing world. While the former is the world *about* which we make statements and assertions, the latter is the one *in* which we make statements and assertions (Friedman and Thellefsen 2011). Thus, representation targets a system of reasoning that assigns truth values by exchanging logical arguments.

6. Requirements to ontology-terminology integration

The proposed semiotic comparison clearly shows that terminologies and ontologies cannot be treated equivalently as they differ from a syntactic, semantic, and pragmatic perspective. The most striking syntactic difference is the refined modeling of semantic relations and fine-grained NL contents of terminologies, which have no corresponding elements in the ontology. Thus, a requirement to the integration definitely is the generation of a flexible model to accommodate features specific to individual NLs and differences between languages. The syntactic differences already indicate that the terminological data model requires a formal semantics to benefit from ontological advantages. Adding formal semantics would bring several elements, such as synonymy relations, to the conceptual level of terminologies and might be beneficial to terminological data modeling. However, simply replacing the concept system of the terminology by a formal ontology is not feasible as the pragmatic and semantic comparison shows. A profound integration framework considering all three levels is needed.

As the number of differences is substantial, a manual integration of terminology and ontology is out of question. Such an endeavor clearly calls for automating the process. In addition, granularity is one aspect that requires further considerations, e.g. is it even necessary to dispose of an existential quantifier in terminology or can the purely NL elements of terminologies be formalized. The last aspect is particularly interesting, as characteristics and subdivision criteria might profit substantially from a formal representation.

7. Related work

With the target of accelerating the ontology design process, controlled vocabularies, thesauri, and terminologies have been re-engineered into ontologies syntactically (Villazón-Terrazas and Gómez-Pérez 2012) based on patterns. Kless and Milton (2010) provide a semiotic comparison similar to the one proposed in this paper, but compare thesauri and ontologies. The results thereof were then turned into a methodology for re-engineering thesauri into ontologies (Kless et al. 2012). Integrating terminologies with ontologies focuses mostly on linguistic aspects. Resulting ontology-terminology models either provide meta-data for the ontology (e.g. Aussenac-Gilles et al. 2008) or generate a completely new socio-cognitive linguistic categorization framework for terminological descriptions (Temmerman and Kerremans 2003). Interoperability of concept-oriented terminology and ontology engineering standards and formats is yet to be achieved. The three main strands of semiotics provide an excellent foundation for an interoperability framework of knowledge representation and terminology standards.

7.1. Conclusion and discussion

This paper describes, analyzes and compares terminologies and ontologies from a semiotic perspective. These resources were found to be substantially different in all three aspects: syntax, semantics, and pragmatics. From a syntactic perspective only the generic relation and associated metadata of the terminology are fully matched by an axiom and annotations in the ontology. Semantically speaking there is basically no congruence - not even optionally used ontology labels can be equated with terms. From a pragmatic angle the main difference lies in the goal of fostering specialized communication in terminologies and enabling automated reasoning for ontologies. Furthermore, the former is a knowledge organization and the later a knowledge representation system – a differentiation that is worth being explored from a pragmatic side.

For terminology to truly meet the multilingual Semantic Web, an interoperability framework considering all semiotic aspects and requirements introduced above is required. Terminology fosters the accommodation of specialized natural languages, while ontologies command of ample tool support. Furthermore, proven linking mechanisms of the Semantic Web are beneficial to terminology harmonization. As a future research activity the ontology-terminology integration would be a mutually beneficial encounter.

8. References

Aussenac-Gilles, Nicole, & Szulman, Sylvie, and Depres, Sylvie (2008). The Terminae Method and Platform for Ontology Engineering from Texts. In Paul Buitelaar & Philipp Cimiano (eds.), *Ontology Learning and Population: Bridging the Gap between Text and Knowledge*. Amsterdam, Netherlans: IOS Press, Vol. 167, 199-223.

Budin, Gerhard (1996). Wissensorganisation und Terminologie - Die Komplexität und Dynamik wissenschaftlicher Informations- und Kommunikationsprozesse. In Hartwig Kalverkämper (ed), *Forum für Fachsprachen-Forschung*. Vol. 28, Tübingen: Gunter Narr Verlag Tübingen.

Friedman, Alon, & Thellefsen, Martin (2011). Concept Theory and Semiotics in Knowledge Organization. *Journal of Documentation*, 67 (4), 644-74.

Fu, Bo, Noy, Natalya F, & Storey, Margaret-Anne (2013). Indented Tree or Graph? A Usability Study of Ontology Visualization Techniques in the Context of Class Mapping Evaluation. *International Semantic Web Conference 2013*. Sydney, Australia: Springer

Gruber, Tom (1995). Toward Principles for the Design of Ontologies Used For Knowledge Sharing. *International Journal of Human-Computer Studies*, 43 (5-6), 907-28.

Guarino, Nicola, Oberle, Daniel, & Staab, Steffen (2009). What is an Ontology?. In Steffen Staab & Rudi Studer (eds.), *Handbook on Ontologies*. Second Edition, International Handbooks on Information Systems: Springer, 1-17.

Kless, Daniel & Milton, Simon (2010). Comparison of thesauri and ontologies from a semiotic perspective. *Advances in Ontologies*, Vol. 35, 35-44.

Kless, Daniel, Jansen, Ludger, Lindenthal, Jutta & Wiebensohn, Jens (2012). A Method for Re-Engineering a Thesaurus into an Ontology. In Maureen Donnelly & Giancarlo Guizzardi (eds.), *Proceedings of the 7th International Conference on Formal Ontology in Information Systems*, Vol. 239, Graz, Austria: IOS Press, 133 – 146.

Krötzsch, Markus, Smiancík, Frantisek, and Horrocks, Ian (2012), 'A Description Logic Primer', CoRR, abs/1201.4089.

Marzá, Nuria Edo (2009). *The Specialised Lexicographical Approach: A Step Further in Dictionary-making*. Vol.102, Bern, Wien: Lang.

Nguyen, Tu Anh T, Power, Richard, Piwek, Paul, & Williams, Sandra (2013). Predicting the Understandability of OWL Inferences. In Philipp Cimiano, Oscar Corcho, Valentina Presutti, Laura Hollink, & Sebastian Rudolph (eds.), *The Semantic Web: Semantics and Big Data.* Volume 7882, Springer, 109-23.

Smith, Barry, Ceusters, Werner, & Temmerman, Rita (2005). Wüsteria. *Studies in Health Technology and Informatics,* 116, 647-52.

Sowa, John (2000). Ontology, Metadata, and Semiotics. *Conceptual Structures: Logical, Linguistic, and Computational Issues, Proceedings*, 1867, 55-81.

Temmerman, Rita & Kerremans, Koen (2003). Termontography: Ontology Building and the Sociocognitive Approach to Terminology Description. In Eva Hajicova, A. Kotesovcova, & J. Mirovsky (eds.), *Proceedings of* CIL17, Prague, Czech Republic: Matfyzpress.

Villazón-Terrazas, Boris & Gómez-Pérez, Asunción (2012). Reusing and Re-engineering Non-ontological Resources for Building Ontologies. In Mari Carmen Suárez-Figueroa, Asunción Gómez-Pérez, Enrico Motta, & Aldo Gangemi (eds.), *Ontology Engineering in a Networked World*, Berlin, Heidelberg: Springer, 107-45.