SADIO Electronic Journal of Informatics and Operations Research

http://www.dc.uba.ar/sadio/ejs

vol. 7, no. 1, pp. 1-11 (2007)

A Review of Common Problems in Linguistic Resources and a New Way to Represent Ontological Relations

Antonio Vaquero[1] Fernando Sáenz[1] Francisco Alvarez[2]

[1] Universidad Complutense de Madrid
Facultad de Informática
Departamento de Ingeniería del Software e Inteligencia Artificial
C/Profesor José García Santesmases, s/n
E – 28040, Madrid, España
e-mail: {vaquero,fernan}@sip.ucm.es

[2] Universidad Autonoma de Sinaloa
Ángel Flores y Riva Palacios, s/n,
C.P 80000
Culiacán, Sinaloa, México
e-mail: fjalvare@fdi.ucm.es

Abstract

Existing lexical resources have taxonomic structure related problems that negatively impact the results of domain-specific applications. This is the result of an approach that focuses on implementation and content issues rather than on questions of design, semantic cleanness and application usefulness. Although taxonomy structuring methodologies have been developed to correct some of these problems, they remain too general and far from the problem-solving and domain-specific approach that ontology-based linguistic resources need in order to be problem-solving. Based on a short analysis of some common problems in lexical resources and of the available taxonomy structuring methodologies, we propose the use of a set of ideas that may be useful in the design and development of an application-oriented ontology-based linguistic resource, and help developers in refining the meaning of concepts and relations to avoid the common problems present in linguistic resources. In addition, we present an E-R model where these ideas have been integrated to better illustrate our proposal and enhance our methodology for the construction of ontology-based linguistic resources.

Keywords: Linguistic Resources, Methodologies, Ontologies, Relations, Structure, Taxonomy

1 Introduction

Despite that useful linguistic resources (LR) have been developed over time, most of them exhibit flaws in their structure that severely limit their reuse and integration, and have proven to be a major obstacle to obtain better results. Structure becomes crucial when developing ontology-based LR as it affects their reuse and integration. However, most of them have been developed relying essentially on the intuition of their developers, with no clear understanding of the ontological relationships involved, and without knowledge control and verification tools to prevent developers from making inappropriate and inconsistent modeling choices.

This inevitably leads to the misuse and confusion of taxonomic relationships, and produces improperly structured LR where relationships cannot be adequately interpreted or LR that could mistakenly be used as ontologies. Where do these problems stem from? The answer is: from the methodology used to build the LR. Different construction methodologies lead to very different features in the resulting resources.

In order to fill this gap in the development of ontology-based resources, taxonomy validating methodologies have been developed. Nevertheless, and although several ontologism have been developed using them, we believe that their respective foundations do not make them suitable for the development of problem-solving application-oriented ontology-based linguistic resources.

Therefore, based on an analysis of common problems in existing LR, and of the available taxonomy structuring methodologies we have developed a set of ideas intended to help developers to refine the meaning of relations in order to achieve a greater semantic precision. Furthermore, these ideas have been incorporated as part of an E-R in order to illustrate the ideas and to enhance our previous model as well as our methodology presented in [1].

The rest of the paper is organized as follows. In section 2, some common problems of LR are summarized. In section 3, the need for well-designed problem-solving application-oriented LR is pointed out. In section 4, the different taxonomy structuring methodologies are described. In section 5, we argue that the available taxonomy structuring methodologies are not suitable for the construction of problem-solving LR. In section 6, we depict a set of ideas intended to help developers to refine the meaning of relations. In section 7, a conceptual model that integrates the aforementioned ideas is introduced and described. Finally, in section 8 some conclusions and future work are outlined.

2 Linguistic Resources for Natural Language Processing

Lexical resources are the backbone of any NLP system. Natural language understanding and generation, translation, summarization or finding information in a document all require the knowledge that a lexical database (LDB) can provide. Through time, a LR has denoted machine-readable dictionaries (MRD), psycholinguistic-based word nets, ontology-based lexicons, etc. However, most of these resources have major flaws that severely limit their reuse and integration, and have proven to be a major obstacle to obtain better results. In this section we will make a brief review of some these flaws.

2.1 Machine-Readable Dictionaries

The use of machine-readable dictionaries (MRD) as a source of ready-made, comprehensive lexical knowledge was promoted by the need for robust lexical and semantic information to assist in realistic NLP applications. The goal was to obtain highly structured lexical knowledge bases from available MRD (e.g. LDOCE) [2]. The efforts towards this goal were based on two implicit postulates [3]: a) MRD contain information useful for NLP and b) this information is relatively easy to extract from MRD.

However, researchers found out that in practice these two postulates were erroneous. MRD typically come to researchers in unusable formats. Hence, to make the MRD usable for research, considerable effort is often required [3]. Moreover, the information contained in dictionary definitions is flawed, because of the lack of consistency in constructing them, as well as space, readability, syntactic and phrasing restrictions. Thus, we have that even in the most straightforward case (detection of hypernyms for concrete objects, kitchen utensils) 50-70% of the information is garbled in some way in five major English dictionaries [2].

The flaws in definitions also cause certain information to be missing, unspecified or left to be implied by other parts of the definitions. Consequently, once a hierarchy has been extracted from a MRD, it is noticed that at the higher levels some concepts simply lack a term to designate them exactly. This lack of clear-cut terms for higher level concepts generates (at least) [2]: circular definitions yielding hierarchies containing loops, which are not usable in knowledge bases, and ruptures in knowledge representation (e.g. a utensil is a container) that lead to wrong inferences. In addition, MRDS do not have enough disambiguating power, because of their lack of contextual and world knowledge in terms of semantic relationships between words.

Finally, the MRDs have such an ambiguity overload, that in order to create a linguistic knowledge base for use in NLP from a MRD, it is necessary to apply sense disambiguation to the words in the dictionary, that is, it is necessary to have full NLP capabilities (including KBs) as a prior step [2].

2.2 WordNet

WN [4] is a resource that is a computational reflection of the lexical memory of native speaker. Its results are impressive with respect to the quantity of information it possesses. However, it has received a lot of criticism from various fronts.

Its biggest problem is the lack of structure. The reason is that it has not been built for knowledge representation purposes nor apparently according to basic taxonomy building principles and with consistency checking tools [5]. This may lead to ruptures in knowledge representation and thus impair the capability of reasoning from it [6]. Furthermore, WordNet seems to have often been made not simply on semantic grounds but also on lexical grounds, thus leading to a multiplicity of artificial categories or categories that should be connected but are not [6].

WN is commonly used as an ontology. The reason is that the obvious parallel between the hypernym relation in a lexicon, and the subsumption relation in an ontology, suggests that a lexicon is very similar to an ontology. Nonetheless, a lexicon, especially one that is not specific to a technical domain, is not a very good ontology [7]. Thus, WN is at best an ersatz ontology, where the relations between synsets do not always reflect an ontological relation because of the misuse and confusion of the subsumption relation [8,9].

It also has an asymmetrical granularity, and the fact that it is a general resource produces that the same synset includes specialized and non specialized variants [10]. For example, in EuroWordNet, the synset that includes the variant “hematoma” (a medical term in Spanish) includes also the variants “morado” and “cardenal” (colloquial terms for hematoma in Spanish) among some other variants belonging to a different specialization levels.

Furthermore, its utility remains under question, because of its lack of sufficient relations to properly tackle ambiguity (notice the resemblance with MRDs here) [11], and because outside the Computational Linguistics research community, no IT applications based on WN are available [12].

2.3 Mikrokosmos

In order to avoid the use of natural language (NL) words (the units of the dictionary of a NL) to represent NL meanings, a computational model of lexical knowledge was needed where these meanings were represented by ontological concepts. The premise is that any system for understanding or generating NL needs both a lexicon of the words in the language(s) in question, and some knowledge of world represented by a taxonomic hierarchy of concepts [13].

Mikrokosmos is a system developed with such ideas in mind. Nonetheless, it has been criticized, mainly because of its misuse of the subsumption relation [8, 9]. Hence MK has the same consistency problems with regard to the is-a relation as WN has. Moreover, just as WN, MK tries to be a general ontology; thus, it suffers from the same asymmetric granularity problem. It is not clear which is its actual coverage for domains different from those of interest for machine translation system [10] it was designed for. Its concept organization for other domains (e.g. medicine) is not completely clear and convincing.

In addition, Mikrokosmos represents semantic relations as nodes of the ontology ones (figure 1 illustrates this point by showing part of the Mikrokosmos ontology). This entails that such representation approach where relations are embedded as nodes of the ontology is prone to suffer from the is-a overloading problems described in [8, 9]. Moreover, it is unknown how these relations were added, and if any control mechanisms were used for their integration in the ontology.

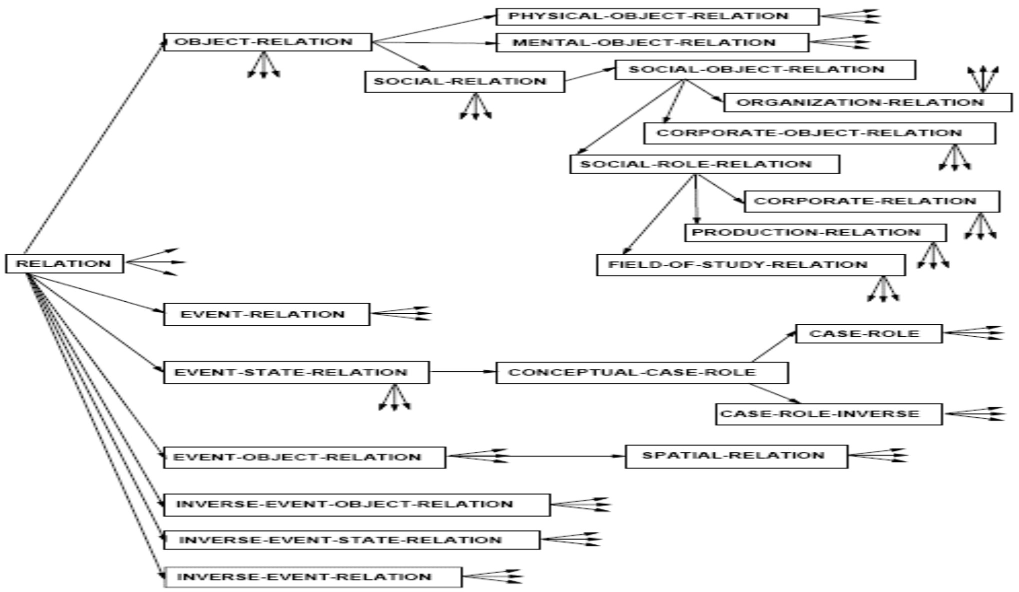

Fig. 1. Embeded Relations in the Mikrokosmos Ontology

2.4 SENSUS

It is accepted that NLP in general and specially machine translation (MT) requires a large amount of knowledge. However, large-scale knowledge bases (KB) are difficult to build because of the knowledge acquisition bottleneck. Thus, a plausible approach to overcome this problem is to merge existing resources. SENSUS [14] is a resource created with such a goal. It merges several resources (i.e. LDOCE, the Harper Collins Spanish-English bilingual dictionary, WN, the Penman Upper Model and Ontos), in order to create a KB big enough to be used in MT.

Nonetheless, the results of using machine readable dictionaries (MRD) like LDOCE, to produce wide-coverage lexical knowledge bases have already been pointed in section 2.1: mainly error, inconsistencies and circularities. Now, the use of WN to provide hierarchical structure is problematic, because of WorNet’s well-known structural problems mentioned in section 2.2, will be carried over to the new resource. The same applies to ONTOS, the ancestor of Mikrokosmos. Thus, it is unlikely that such an approach, that is, to focus only in coverage, would yield a better resource and increase its accuracy [15].

2.5 UMLS

Faced with the proliferation of differing vocabularies and conceptualizations in the biomedical domain, the Unified Medical Language System (UMLS) aims at providing a bridge across the existing schemes, by compiling and cross-linking all underlying concepts [16], and thus, help health professionals and researchers in the retrieval of biomedical information from different sources [10].

It has a three level structure composed by: a semantic network that acts as an upper ontology for classifying (categorize) the lower level concepts; a Metathesaurus where the terms from different vocabularies and classification systems that are “equivalent” in meaning are clustered to form concepts; and a lexicon that records syntactic, morphological and orthographic information for all the terms.

The quantity of information UMLS possesses is impressive. Nevertheless, the clustering of “equivalent terms” although producing a useful, unified structure (a directed acyclic graph), has circular hierarchical relationships as its side effect [17], with detrimental consequences in terms of graph traversal and knowledge representation. The reason is that UMLS preserves the structure of each source vocabulary. Therefore, hierarchical relations in the Metathesaurus reflect the organization principles of the source vocabularies [17]. Hence, it is not possible to assume that all the hierarchical relationships are taxonomic. Thus, what we have is another case of subsumption misuse and confusion because of the imprecise nature of relationships.

3 Problem-Solving Application-Oriented Linguistic Resources

We have come a long way from the days of MRD. However, still today, the focus is on coverage and time-saving issues, rather than on semantic cleanness and application usefulness. Proof of this are the current different merging and integration efforts aimed at producing wide-coverage general LR [15, 18, 19], and the ones aimed at (semi)automatically constructing them with machine learning methods [20, 21]. However, no amount of broad coverage will help raise the quality of output, if the coverage is prone to error [22]. We should have learned by now that there are no short cuts, and that most experiments aimed at saving time (e.g., automatically merging LR that cover the same domains, or applying resources to NLP that are not built for it, like machine-readable dictionaries and psycholinguistic-oriented word nets) are of limited practical value [23]. Furthermore, in the current trend of LR development issues such as how to design LR are apparently less urgent, and this is haphazard. More attention must be paid on how LR are designed and developed, rather than what LR are produced.

The experience gained from past and present efforts clearly points out that a different direction must be taken. As [3] pointed out back in the days of MRD: “rather than aiming to produce near universal LR, developers must produce application-specific LR, on a case by case basis”. In addition, we claim that these LR must be carefully conceived and designed in a systematic way, according to the principles of a software engineering methodology.

4 Taxonomy Structuring Approaches

Although the taxonomic structure problems represent a serious obstacle in the developing of ontologies and OBLR, we found out that there are only two proposals that actually deal with such problems: The OntoClean [24] and Archonte [25, 26, 27] methodologies. These methodologies are similar but at the same time, as it will be seen below, follow opposite approaches to attain the same goal.

4.1 The Metaphysical Approach of OntoClean

In order to understand what OntoClean is all about without getting into the details of the S5 modal logic it uses, it is necessary to understand what the fundamental idea of ancient Greek culture was, as well as the notions of essentialism and psychological essentialism.

4.1.1 The Marrow of Reality

According to [28], the key to understand ancient Greek culture, the central category that introduces a deeper meaning and from which all others depend, the idea that is self-evident, is that behind reality there is a set of models, structures or essential forms around which it is organized. Furthermore, these essential forms are perfectly comprehensible (e.g. they can be labelled) and rational (e.g. they can be expressed using modal logic). For instance, cows are born and die, but what is permanent is the form of the cow. Behind the changes of concrete objects there is the permanence of the eternal forms.

4.1.2 Essentialism

In Philosophy, it is most commonly understood as a belief in the real, true essence of things, the invariable and fixed properties which define what a given entity really is [29]. An essence characterizes a substance or a form, in the sense of the “forms” or “ideas” of Aristotle and Plato respectively. Thus, an essence (i.e. the set of properties) is permanent, unalterable, and eternal; and present in every possible world.

Therefore, for any specific kind of entity (e.g a tiger), it is theoretically possible to specify a finite list of properties (e.g. the rigidity, identity and unity metaproperties of OntoClean) all of which any entity must have to belong to a specific group or natural kind (e.g. see the table of properties and the taxonomy of kinds in [24]. Natural kinds are the "forms" or "ideas" mentioned above, that have always existed, and that are part of a realm different from our physical world: the metaphysical realm.

4.1.3 Psychological Essentialism

The basic tenet of psychological essentialism, as conceived by [30], is the idea that key human cognitive processes, those which determine how we approach experience, reflect a basic belief that unobservable essences are causally responsible for the surface features we observe. As such, the world is divided up into essences from which preset associated properties can be inferred (e.g. the metaproperties mentioned above).

Furthermore, these properties play a key role in our everyday reasoning and categorization tasks, that is, they guide and back our inferences about (natural) kind membership. For instance, a chair is member of the kind "chair" because it possesses a set of (meta) properties that intentionally defines it as member of that kind, and although it could be used as a doorstopper or a ladder (e.g a ladder "is-a" chair), it will always be a chair because that is why it came into being. Seen this way psychological essentialism can be understood as a reasoning heuristic and inference system that states that categorization is theory-driven, whereby concepts must conform to an overall world-view, rather than just a delineation of phenomena by lists of observable attributes (see [31])).

4.14 Providing a Theory of the World

OntoClean provides the world-view theory mentioned above by stating that all domains share a same set of ontological commitments and semantic prescriptions that makes them ontologically compatible regardless of the "state of affairs. It allows retaining models that are ontologically compatible according to a top-ontology (i.e a taxonomy of kinds) that represents the structure of reality as a whole, subsumes all domains, and specifies the constraints that notions must respect (through an axiomatization that presents itself as an absolute truth) so they can be integrated in an ontology.

4.2 The Linguistic Approach of Archonte

While OntoClean has a clearly cognitive and philosophical background that rest on the thesis that there is one and only one central world to which natural cognition relates, that is, a core of interconnected beliefs called "common sense" that exists entirely independent of human beings and their physical world, Archonte claims that it is natural language the one that allows expressing what the common sense of a domain is, and that common sense is what is understood and known for every actor of the domain once he/she has mastered both the discipline and its language.

Therefore, natural language conveys the "common sense" knowledge that any system reasonably "intelligent" needs in order to adopt a problem resolution behavior adapted to its context of use [24]. Archonte relies on the work of [32], that states that even for well-defined domains, the norms that fix the meaning of a word and of its reference (e.g. its concept) cannot be foreseen, and that the meaning of words is immanent to a given situation and context of usage. Hence, Archonte proposes a hermeneutic approach that provides concepts with a domain and task-dependent meaning by means of a "differential semantic".

4.2.1 The Differential Paradigm

This paradigm can be seen as a process of normalization that establishes a context of reference to provide an appropriate meaning: the point of view of the task or problem. It states that the meaning of a linguistic unit is given in terms of the similarities and differences it has with other neighboring units in the same context of usage, and that this differences and similarities can be found by analyzing a corpus.

It refuses to follow the common extra-linguistic semantic approaches to characterize the meaning of words: the referential and psychological ones [27]. In the referential approach, each linguistic unit has an associated referent that can be an object (extensional reference) from a universe of individuals, or a concept (intensional reference) from an abstract universe. Thus, understanding a word means finding its correspondent object or concept. In the psychological approach linguistic units have an associated psychological representation or mental image. Consequently, understanding a word means to build a representation of it.

The creators of Archonte argue that since the above approaches associate each linguistic unit to a non-linguistic entity (object or representation), this poses the problem of defining what these objects and representations are. In other words, the objects of the world must be ontologically defined (in the philosophical and formal ontology sense followed by OntoClean) or the psychological structures that make possible linguistic comprehension must be established (e.g WordNet).

4.2.2 Differentially Structuring a Taxonomy

Since Archonte follows an inter-linguistic method, it is necessary to use terminological extraction tools to discover candidate concepts. Once this has been done, the ontologist has to structure the ontology using a set of principles [26] based on differential semantics. This will create a differential tree-shaped ontology where the differences and similarities are expressed in natural language. The principles are the following: a) similarity with parent principle (SWP); b) similarity with siblings (SWS); c) difference with siblings (DWS) and d) difference with parent (DWP).

Since these principles are attached to concepts, herein lies the similarity with OntoClean. In order to properly structure a taxonomy or ontology, concepts (not relations) must have a set of properties that determine the structure of a domain and if an is-a link can exist between any two concepts.

4.2.3 The Semantics of Relations

In addition, Archonte states that any binary relation is defined by the concepts it links, and by an intrinsic semantic content [27] that we will call intrinsic properties of the relation. These intrinsic properties are used to structure a tree of relations (by following the same differential principles above) like the one Mikrokosmos has. Thus, Archonte not only controls the structuring of the ontology but also the definition of relations as part of the ontology itself.

5 Problems with Taxonomy Structuring Methodologies

As appealing and sound as the OntoClean approach seems to be, it stumbles into a few problems. First, although it can argued that psychological essentialism does not entail metaphysical essentialism[30], one cannot help noticing that the description of psychological essentialism above resembles Kripke's essentialism and causal theories of reference (see [33]). Furthermore, in [33] the authors provide evidence that due to cultural differences, a global theory of reference like the one OntoClean provides is not possible. Recent work in cultural psychology has shown systematic cognitive differences between East Asians and Westerners, and some work indicates that this extends to intuitions about philosophical cases [33]. In other words, there is proof of cultural variation on a number of basic cognitive processes.

As researchers in history and anthropology have long maintained, one should be wary of simply assuming universality without evidence. We suspect that philosophers take their own intuitions regarding the referents of terms, and those of their philosophical colleagues, to be universal, and evidence suggests that it is wrong for philosophers to assume a priori the universality of their own semantic intuitions.

Second, practical domains are a reality of practices expressed in natural language. This reality is empiric and does not correspond to a set of “a priori” defined objects determining their structure. Hence, practical domains (e.g. medicine) are non-theoretical, and are structured around precise tasks. For instance, medicine is structured around diagnostic, monitoring and prescription tasks, and domain knowledge does not exist outside this set of tasks that puts this knowledge to work. In this context, knowledge is but a possibility to act as part of the fulfillment of a given task.

As for Archonte, it takes a step down the ladder by rejecting and criticizing the metaphysical approach of OntoClean [25, 27], and grounding the structuring of the taxonomy on an intra-linguistic semantic theory. Apparently this is due to the conception of AI that its creators have. They see AI as a technological project that aims at the computational resolution of problems or the computational automation of tasks that need knowledge that is normally expressed in natural language.

Such a standpoint seems to coincide with our work. However, as much as they criticize the philosophical bias of OntoClean, it seems that in order to be general enough for any domain, Archonte is partially based on

Husserl's, Kant's and Locke's philosophy (see [25, 27]). Hence, some of the foundational aspects of Archonte remain incomprehensible or fuzzy at best for someone that is not philosopher. Furthermore, Archonte controls semantic relations using the same set of principles used to structure the taxonomy. Is it because these relations are extracted from texts or is there any other reason?

6 A Problem-solving Approach to the Representation and Control of Relations

Ontoclean proposes a universality that we do not believe is possible, and Archonte proposes a standard method, much in the spirit of problem resolution methods that consider that applications and domains just provide a different vocabulary to perform tasks that are always the same (e.g. taxonomy structuring).

Our approach will be problem-solving and application-oriented. From the standpoint of LR, only when we know the actual specific use we intend to do of a specific resource can we build the very best linguistic resource for that use, and although it can be argued, as [34] do, that usefulness for applications is too expensive and not realistic, we claim that only at the local level can we enforce the kind of control and verification that [35] states that is needed, and that "not realistic" is waiting for an ever coming standard or universal resource.

Therefore, instead of relations with an unclear meaning we propose the use of relations with well-defined semantics, up to the granularity needed by the ontology developer. This implies that instead of giving concepts a set of properties (like OntoClean and Archonte do) the semantic load will be put on relations themselves, through two sets of properties that we call intrinsic properties and algebraic properties of relations.

6.1 Algebraic Properties of Relations

The meaning of each relation between two concepts must be established, supported by a set of algebraic properties from which, formal definitions could be obtained (e.g., transitivity, asymmetry, reflexivity, etc.). This will allow reasoning applications to automatically derive information from the resource, or detect errors in the ontology [36]. Moreover, the definitions and algebraic properties will ensure that the corresponding and probably general-purpose relational expressions are used in a uniform way [36]. Tables 1 and 2 (taken from [36]) show a set of concrete relations with their definitions and algebraic properties.

Relations	**Definitions**	**Examples**
C is-a C_1	Every *C* at any time is at the same time a C_1	*myelin is-a lipoprotein*
C part-of C_1	Every *C* at any time is part of some C_1 at the same time	*Nucleoplasm part-of nucleus*

Table 1. Definitions and Examples of Relations

Relations	**Transitive**	**Symmetric**	**Reflexive**
is-a	+	-	+
part-of	+	-	-

Table 2. Algebraic Properties of some Concrete Relations

6.2 Intrinsic Properties of Relations

How do we assess, for a given domain and task, if a specific relation can exist between two concepts? The definitions and algebraic properties of relations, although useful are not enough. Thus, for each relation, there must be a set of properties that both a child and its parent concept must fulfill for a specific relation to exist between them. We call these properties, intrinsic properties of relations.

7 A Conceptual Model for an Ontology-based Linguistic Resource

In order to better illustrate our ideas, we present an E-R model, shown in figure 2, which incorporates the aforementioned ideas.

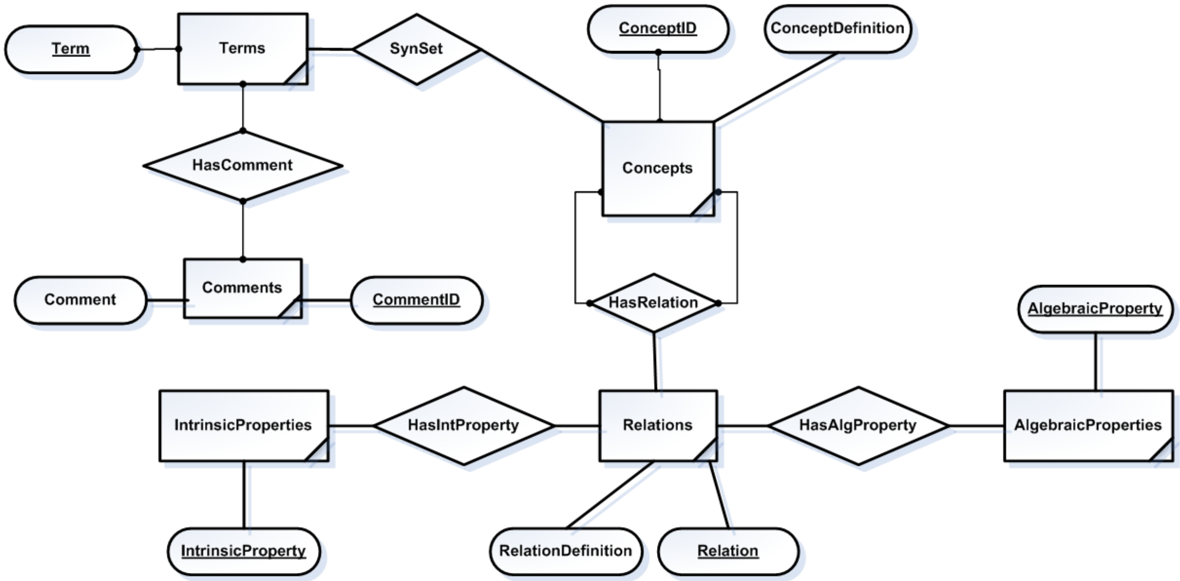

Fig. 2. Conceptual Model for an Ontology-based Linguistic Resource

The entity set Concepts denotes the meaning of words, and it has two attributes: ConceptID (artificial attribute intended only for entity identification), and ConceptDefinition, intended for the textual definition of the meaning (informal semantics).

The entity set Terms represents the terms that will be associated to each concept. We have to remember that an isolated ontology is by no means an OBLR and that creating the ontology first and then "enrich" it with lexical information from another LR has proved to be rather difficult [15].

The entity set Relations represents the set of relations that can exist in an ontology, and it has two attributes: Relation that captures the textual name of each relation (e.g., is-a, part-of, etc.), and RelationDefinition for the textual definition of relations (informal semantics).

The entity set AlgebraicProperties represents the algebraic properties of relations (formal semantics), and it has one attribute: AlgebraicProperty that denotes each algebraic property. The entity set IntrinsicProperties conveys the set of intrinsic properties that each relation could have and has one attribute: IntrinsicProperty which represents each intrinsic property.

The ternary relationship set HasRelation is used to represent that two concepts in an ontology can be linked by a given relation. The relationship set HasAlgProperty is used to convey that relations could have attached a set of algebraic properties; the same applies for the relationship set HasIntProperty, but for intrinsic properties. In addition, we have decided to group terms in synonym sets (represented by the SynSet relationship) as this is a useful feature, mainly because synonymy is one of the fundamental linguistic phenomena that influence the structure of the lexicon [37].

8 Conclusions

We have made a short analysis of some significant approaches for the construction of linguistic resources. The analysis shows that they all have severe flaws due to the absence of a rigorous, systematic and well-documented methodology for their development. Moreover, the resources analyzed have been developed as universal ones, and used in a wide range of tasks and problems that clearly needed an application-oriented and task-dependent resource. Thus, the results obtained from their usage have not been as fruitful as expected.

This shows that, for a given application, the appropriate resource must be developed in order to have acceptable results. Hence, we point out to a new direction, and underline the need for a problem-solving application-oriented approach based on ontological semantics, where the ontological choices at the relationship level can be controlled and verified. Nevertheless, the approach we propose needs the help of state-of-the art terminology extraction tools, in order to identify candidate concepts and relations, as well as a more or less deep analysis of the domain and problem in order to identify (if possible) the algebraic and intrinsic properties of relations.

References

1. Sáenz, F. and Vaquero, A. 2005. Applying Relational Database Development Methodologies to the Design of Lexical Databases. Database Systems 2005, IADIS Virtual Multi Conference on Computer Science and Information Systems (MCCSIS), ISBN 972-8939-00-0, 2005.
2. Ide, N., and Veronis, J. Extracting Knowledge Bases from Machine-Readable Dictionaries: Have we wasted our time? In Proceedings of the First International Conference on Building and Sharing of Very Large-Scale Knowledge Bases. Tokyo. 1993.
3. Evans, R., & Kilgarriff, A. MRDs, Standards and How to do Lexical Engineering. Proc. of 2nd Language Engineering Convention (pp. 125–32). London. 1995.
4. Miller, G., Beckwith, R., Fellbaum, C., Gross, D., Miller, K. & Tengi, R. Five Papers on WordNetTM . CSL Report 43, July. 1990.
5. Martin, P. Correction and Extension of WordNet 1.7. ICCS 2003, 11th International Conference on Conceptual Structures,(Springer Verlag, LNAI 2746), pp. 160-173), Dresden, Germany, 2003.
6. Burgun, A. and Bodenreider, O. Aspects of the Taxonomic Relation in the Biomedical Domain. In: C Welty C. Second international conference on formal ontologies in information systems, 2001.
7. Hirst G. Ontology and the Lexicon. In Staab and Studer (eds), Handbook on Ontologies, Springer: 209-230. 2004.
8. Guarino, N. Some Ontological Principles for Designing Upper Level Lexical Resources. A. Rubio et al. (eds.), Proceedings of the First International Conference on Language Resources and Evaluation, Granada, 28-30 May, Paris: ELRA, 527-534. 1998.
9. Welty, C. and Guarino, N. Supporting ontological analysis of taxonomic relationships", Data and Knowledge Engineering vol. 39, no. 1, pp. 51-74, 2001.
10. Feliu, J.; Vivaldi, J.; Cabré, M.T. Ontologies: a review. Working Paper, 34. Barcelona: Institut Universitari de Lingüística Aplicada. DL: 23.735-2002 (WP), 2002.
11. Nirenburg, S., McShane, M. and Beale, S. The Rationale for Building Resources Expressly for NLP. Proceedings of LREC 2004, 2004.
12. Navigli, R. Automatically Extending, Pruning and Trimming General Purpose Ontologies. In Proceedings of the International Conference on Systems, Man and Cybernetics. 2002.
13. Hirst, G. Near-synonymy and the Structure of Lexical Knowledge. In Working Notes for the AAAI Spring Symposium on Representation and Acquisition of Lexical Knowledge: Polysemy, Ambiguity, and Generativity. Technical Report SS-95, Standford University, CA. 1995.
14. Knight, K and Luk, S.K. Building a Large-Scale Knowledge Base for Machine Translation. In proceedings of of the 12th national conference on artificial intelligence (AAAI'94), Seattle, USA, July, 1994.
15. Oltramari, A.; Prevot, L.; Borgo, S. 2005. Theoretical and Practical Aspects of Interfacing Ontologies and Lexical Resources. In Proceedings of SWAP2005.
16. Bouaud, J., Bachimont, B., Charlet, J. and Zweigenbaum, P. Acquisition and Structuring of an Ontology within Conceptual Graphs. In Proceedings of the ICCS'94 Workshop on Knowledge Acquisition using Conceptual Graph Theory, University of Maryland, College Park, MD, 1994.
17. Bodenreider O. Circular Hierarchical Relationships in the UMLS: Etiology, Diagnosis, Treatment, Complications and Prevention. In Proceedings of the AMIA Symposium. 2001.
18. Hongyan Jing, Yael Dahan Netzer, Michael Elhadad, Kathleen McKeown. Integrating a Large-Scale, Reusable Lexicon with a Natural Language Generator. Proceedings of the 1st International Conference on Natural Language Generation, pp.209-206, 2000.
19. Philpot, A., Hovy, E. and Pantel, P. The Omega Ontology. In IJCNLP Workshop on Ontologies and Lexical Resources, pp. 59-66, 2005.
20. Makagonov, P., Ruiz Figueroa, A., Sboychakov, K. and Gelbukh, A. Learning a Domain Ontology from Hierarchically Structured Texts. In Proc. of Workshop “Learning and Extending Lexical Ontologies by using Machine Learning Methods” at 22nd International Conference on Machine Learning, 2005.
21. Makagonov, P., Ruiz Figueroa, A., Sboychakov, K. and Gelbukh, A. Studying Evolution of a Branch of Knowledge by Constructing and Analyzing Its Ontology. In Christian Kop, Günther Fliedl, Heinrich C. Mayr, Elisabeth Métais

(eds.). Natural Language Processing and Information Systems. 11th International Conference on Applications of Natural Language to Information Systems, 2006.

22. McShane, M.; Nirenburg, S. and Beale, S. An implemented, integrative approach to ontology-based NLP and interlingua . Working Paper #06-05, Institute for Language and Information Technologies, University of Maryland Baltimore County, 2005.
23. Nirenburg, S., McShane, M., Zabludowski, M., Beale, S. and Pfeifer, C. Ontological Semantic text processing in the biomedical domain. Working Paper #03-05, Institute for Language and Information Technologies, University of Maryland Baltimore County, 2005.
24. Welty, C. and Guarino, N. Supporting ontological analysis of taxonomic relationships. Data and Knowledge Engineering vol. 39(1), pp. 51-74, 2001.
25. Bachimont, B. Hermenutique Materielle et Artefacutre. These presente pour obtenir le grade de Docteur de l'Ecole Polytechnique. 1996
26. Bachimont, B., Isaac, A., Troncy, R. Semantic Commitment for Designing Ontologies: A Proposal. Knowledge Engineering and Knowledge Management. Ontologies and the Semantic Web: 13th International Conference, EKAW 2002, Siguenza, Spain, October 1-4, 2002.
27. Bachimont, B. Arts et Sciences du Numerique: Ingenierie des Connaissances et Critique de la Raison Computationnelle. Memoir presente pour obtenir une habilitation a diriger des recherches. Universite de Technologie de Compiegne. 2004.
28. Schwanitz, D. La Cultura: todo lo que hay que saber. Editorial Taurus. 2002.
29. Barrett, H.C. On the functional origins of essentialism. Mind and Society, 3, Vol. 2, pp. 1-30, 2001.
30. Medin, D. L., & Ortony, A. Psychological essentialism. In S. Vosniadou and A. Ortony (Eds.), Similarity and analogical reasoning (pp. 179-195). New York: Cambridge UniversityPress, 1989.
31. Xu, F. & Rhemtulla, M. In defense of psychological essentialism. In B.G. Bara, L. Barsalou, and M. Bucciarelli (eds.), Proceedings of the 27th Annual Conference of the Cognitive Science Society (pp.2377-2380), 2005.
32. Rastier, F., Cavazza, M., Abeillé, A. Sémantique pour l'analyse. De la linguistique à l'informatique, Masson, Paris, 1994.
33. Machery, E. et al. Semantics, Cross-Cultural Style. Cognition 92, Pp. B1-B12, 2004.
34. Calzolari N., Zampolli A. Syntactic/semantic lexicons for the European languages: Towards a standardised infrastructure. In proceedings of VEXTAL: Venezia per il Trattamento Automatico delle Lingue, Pp. 39-45, 1999.
35. Hicks, R.C. Knowledge-based Management System-Tools for Creating Verified Intelligent Systems. Knowledge Based Systems 16, Pp. 165-171, 2003.
36. Smith B, Ceusters W, Klagges B, Köhler J, Kumar A, Lomax J, Mungall CJ, Neuhaus F, Rector A, Rosse C. Relations in Biomedical Ontologies. Genome Biology, 6(5), 2005.
37. Edmonds, P. and Graeme, H. Near-synonymy and lexical choice. Computational Linguistics 28(2):105-144, 2002.