UNIVERSIDADE FEDERAL DO ESTADO DO RIO DE JANEIRO
CENTRO DE CIÊNCIAS EXATAS E TECNOLOGIA

Relatórios Técnicos
do Departamento de Informática Aplicada
da UNIRIO
n° 0019/2009

A Survey on Conceptual Modeling From a Linguistic Point of View

Lucia Castro
Fernanda Baião
Giancarlo Guizzardi

Departamento de Informática Aplicada

UNIVERSIDADE FEDERAL DO ESTADO DO RIO DE JANEIRO
Av. Pasteur, 458, Urca - CEP 22290-240
RIO DE JANEIRO – BRASIL

Relatórios Técnicos do DIA/UNIRIO, No. 0019/2009
Editor: Prof. Sean W. M. SIqueira September,2009

A Survey on Conceptual Modeling From a Linguistic Point of View

Lucia Castro[1] Fernanda Baião[2] Giancarlo Guizzardi[2]

[1,2]NP2Tec – Universidade Federal do Estado do Rio de Janeiro (UNIRIO)
[2]Depto de Informática – Universidade Federal do Espírito Santo (UFES)

{ lucia.castro, fernanda.baiao}@uniriotec.br, gguizzardi@inf.ufes.br

Abstract. Conceptual modeling remains a relevant research topic, even though more than thirty years have passed since Peter Chen enunciated his Entity-Relationship Model. Methods and methodologies for the creation of conceptual models have been the subject of studies and projects which goal is to produce clearer, complete and easier-to-read models. Several methods, modeling languages and tools, have been proposed over the years, some of which aim at creating and/or reading such models automatically, which can imply a simplification that might oppose the idea of semantic accuracy and completeness. The common denominator among all proposals is that, for a conceptual model to be effective and useful, a designer must learn the language used in the Universe of Discourse to be modeled, along with its underlying concepts, and then represent such concepts in a modeling language. Also, no matter the source of information, the knowledge about the scenario to be modeled is always passed to the designer in a natural language. For the resulting model to be both detailed and unambiguous, the modeling language must convey the semantics of such environment, in a way that anyone who is literate in this language can, from reading the model, get the same understanding as from the description in a natural language. In other words, the modeling language must be as rich and generative as the natural language in which the Universe of Discourse concepts are described. Several projects that focus on conceptual modeling have turned to linguistics as a support for the modeling process itself, relating natural language constructs to those of the adopted modeling language; what they all have in common is that their work is done from the perspective of the (meta)model itself. This report presents some of the aforementioned studies and proposes that the linguistic approach, invaluable as it is, should actually be applied to the modeling process from the natural language perspective.

Keywords: Conceptual modeling, ontology, foundational ontology, linguistic approach.

Table of Contents

1 Introduction

A model is a representation of something, or "… *an abstraction of reality according to a certain conceptualization."* [Guizzardi 2005]. In the field of information systems, the interest in modeling requirements and writing precise specifications for software arouse in the early 1970s; Parnas [1972] describes a specification scheme, and states that one of the issues related to software engineering at that time was the *"…lack of techniques for precisely specifying program segments…"* Peter Chen [1976] enunciated the Entity-Relationship Model that, in his own words, *"… adopts the more natural view that the real world consists of entities and relationships."* Since then, methods and methodologies for the creation of conceptual models have been the subject of studies and projects which goal is to produce clearer and easier-to-read models; several modeling languages and tools have been proposed over the years.

When it comes to data modeling, the common denominator among all proposals is that, for a conceptual model to be effective and useful, the designer must learn the language, i.e. the "jargon", which is the part of a natural language used in the scenario to be modeled, along with its underlying concepts, and then represent such concepts in a modeling language. For the produced model to be complete and unambiguous, the modeling language must convey the semantics of such environment, in a way that anyone who is literate in the modeling language can, from reading the model, get the same understanding as from, for instance, reading the same description in a natural language jargon. In other words, the representation language must be as rich and generative as the natural language in which the Universe of Discourse concepts are described; in other words, the modeling language must have consttructs to which natural language constructs can be unambiguously mapped.

Several projects that focus on conceptual modeling have turned to linguistics as a support for the modeling process itself, relating natural language constructs to those of the adopted modeling language. Since [Chen, 1983] several studies have turned to parts of speech and syntactic structures of natural languages to "fill" the constructs of modeling languages, creating directives and/or heuristics for the modeling process. What all these projects have in common is that they see and perform modeling activities from the perspective of the model and the modeling language and use linguistics only as a support for the rationale behind their representation decisions.

This report discusses projects and methods that propose a linguistic approach to the conceptual modeling process and analyzes their linguistic bases. The main focus here is the process of describing known concepts in a modeling language, from a linguistic perspective. It is organized as follows: section 2 consists of a brief description of language constructs; section 3 presents some characteristics of natural languages that are relevant to the modeling process and defines semiotics and linguistics; section 4 addresses conceptual modeling itself]; section 5 presents linguistic approaches to the conceptual modeling process; and section 6 concludes the ideas.

2 Constructs and languages

Bunge [2003] defines construct as "*A concept, proposition, or set of propositions, such as a classification, a theory, or a moral or legal code.*" All languages, either natural or modeling ones, have inherent concepts and meta-properties that are tacitly understood by the community in which they are used. The semantic unit of a natural language is a word and its constructs can, consequently, be considered the parts of speech, that is, the classes into which the words are divided and classified. Each part of speech, meaning, nouns, verbs, adverbs, etc., are concepts in themselves. Similarly, modeling languages have their own constructs that also imply modeling language concepts - the entity-relationship model, for instance, has entities, relationships and attributes as constructs.

3 Natural Language, Semiotics and Linguistics

Pinker [1999] states that a [natural] language comprises "*… two tricks, words and rules.*" The words are "sounds", symbols or signs that convey ideas; the set of symbols of a language is called its lexicon. The rules are the grammar of the language, i.e., the way its or symbols can be organized so that meaningful messages can be passed among the members of a certain community. The science that studies signs or symbols is called semiotics, and it was firstly enunciated in 1915 by Saussure [2006].

Linguistics focuses on studying the linguistic signs, i.e., the ones that are part of natural languages. Linguistics include semantics (study of relations between signs and their meanings), syntax (study of the formal relations among signs, or the rules, in Pinker's words) and pragmatics (study of the relations between the signs and their users). Saussure [2006] proposed a dualistic view for the linguistic sign: the signifier (the uttered form of a word or a phrase) and the significant (the mental concept). In 1936, Ogden and Richards, in *The Meaning of Meaning*, [Ullmann, 1964] state that besides the significant and the signifier the analysis of the linguistic sign should also include the referent, or the thing in the world to which the sign or symbol refers; they present the semiotic triangle that, besides considering the significant and the signifier, introduces the concept of referent. According to this triangle, a symbol of the lexicon, represents a concept and refers to a thing in reality; the concept is an abstraction of the thing, i.e., of the referent. The semiotic triangle can be seen in figure 1.

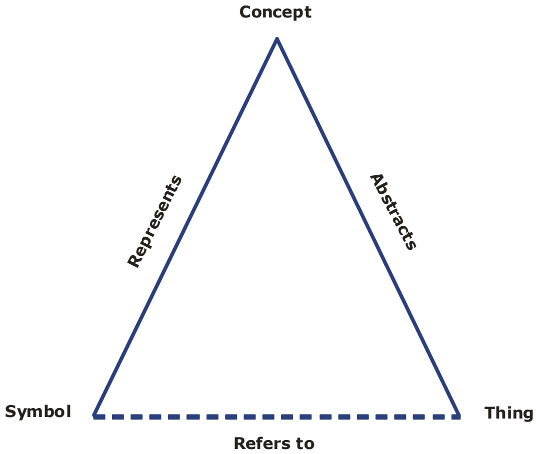

Figure 1. The Semiotic Triangle

In, perhaps, the most famous scene of all time, from the balcony Juliet says: *"What's in a name? that which we call a rose by any other name would smell as sweet;"* Pinker [1999], again, answers Juliet's question stating that "*What's in a name is that everyone in a language community tacitly agrees to use a particular sound to convey a particular idea.*" In other words, the symbol "rose" refers to a thing in reality and represents an idea within a certain community; the links from the symbol to the idea, or concept, and from the symbol to the thing are arbitrary and tacitly agreed upon, understood and shared by the members of such community.

4 Conceptual Modeling

Conceptual modeling is, obviously, the process of creating a conceptual data model. This process *"is by far the most critical phase of database design and further development of database technology is not likely to change this situation"*, [Batini et al. 1992] Its importance comes from the fact that the analyst has to learn the language, or "jargon", (its constructs, lexicon, semantics, syntax, etc) used by a community and create a model tha will be the basis for all subsequent information models and structures of data storage. The process of conceptual modeling involves two main activities: the first is the acquisition of concepts (and their corresponding symbols) used in the Universe of Discourse being modeled; and the second is the creation of the actual model in a modeling language.

The knowledge about the Universe of Discourse is obtained in a variety of ways, like from interviews with users, studying reports and functional documents pertaining to the environment, from observation of the group routine, etc. However, no matter the source of information, the knowledge about the scenario to be modeled is always passed to the analyst in a natural language. To develop a conceptual model, the designer must identify conceptual elements, understand how they relate with each other and then represent both conceptual elements and their inter-relationships in a modeling language [Gangopadhyay, 2001]. It is correct to say that the conceptual modeling process is a translation activity, i.e., identifying concepts that are represented by symbols that belong to a language, and then represent those same concepts with symbols that belong to a different language. Mounin [1975] defines the translation process as a contact between two languages, and the translator, in this case the analyst, as a bilingual person, i.e., someone who can communicate in both languages, someone who knows the constructs and symbols of both languages.

When developing a conceptual model (translation process), an analyst must find in the lexicon (set of symbols) of the modeling language, symbols that can express unambiguously the same ideas expressed by the natural language symbol used in the Universe of Discourse, as shown in figure 2. However, constructs cannot be forgotten, that is to say that the analyst must map natural language constructs to modeling language ones, considering the correspondence between their meta-properties.

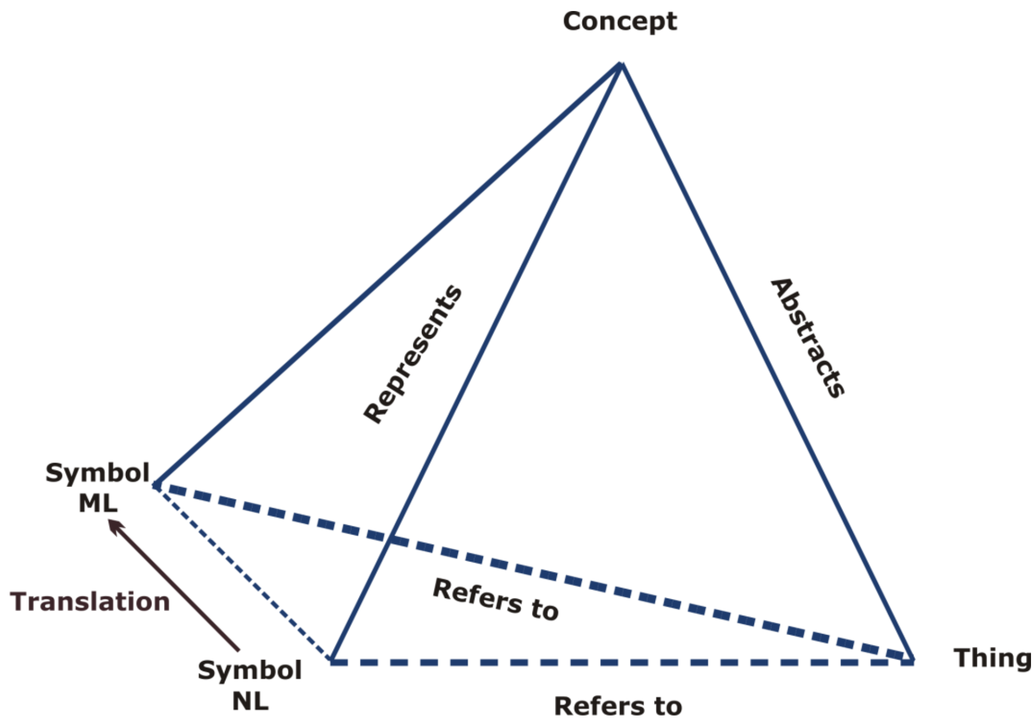

Figure 2. The Symbol Translation Process

5 Proposals and Researches

The challenges of the modeling process led analysts and scholars to turn to linguistics and to the fundamentals and structures of natural languages for support. Several methods and methodologies have been proposed that present rules, directives and heuristics as bases for the not simple task of creating conceptual data models.

5.1 Chen's eleven rules

Chen [1983] proposes eleven rules to translate natural language information requirements from natural language (English) to an entity-relationship diagram, considering parts-of-speech and their functions in sentences. Chen's proposal seems to be the first attempt to apply linguistic concepts to conceptual modeling. However, as the author himself points out, these rules are directives for correlating certain patterns in the use of English to constructs in the ER meta-model. The eleven rules are:

- A common noun in English corresponds to an entity type in an ER diagram.
- A transitive verb in English corresponds to a relationship type in an ER diagram.
- An adjective in English corresponds to an attribute of an entity in an ER diagram.
- An adverb in English corresponds to an attribute of a relationship in an ER diagram.
- If the sentence has the form: "There are … X in Y," we can convert it into the equivalent form "Y has … X".
- If the English sentence has the form "The X of Y is Z" and if Z is a proper noun, we may treat X as a relationship between Y and Z. In this case, both Y and Z represent entities.
- If the English sentence has the form "The X of Y is Z" and if Z is not a proper noun, we may treat X as an attribute of Y and Z. In this case, Y represents an entity (or a group of entities) and Z represents a value.
- The objects of algebraic or numeric operations can be considered as attributes.
- A gerund in English corresponds to a relationship-converted entity type in ER diagrams.
- A clause in English is a high-level entity type abstracted fro a group of interconnected low-level entity and relationship types in ER diagrams.
- A sentence in English corresponds to one or more entity types connected by a relationship type, in which each entity type can be decomposed (recursively) into low-level entity types interconnected by relationship types.

Although one can relate to these rules and understand their meaning, they reflect Chen's experience and knowledge of the ER meta-model. Also, these rules are rather simplistic. Rule number 1, for instance, states that a common noun corresponds to an entity type but "car", "model", "color", "gas" and "size" are all common nouns; should they all be modeled as entity types?

5.2 Method based on the Conceptual Dependency Theory

According to Gangopadhyay [2001], there are many ways to elicit user requirements and it is the responsibility of the database designer to elicit the missing pieces of information that may be implicit in informal descriptions. Also, natural language descriptions tend to contain redundancies, which can be synonyms (same concepts with different names) or homonyms (different concepts with the same name). He proposes a methodology for the design of conceptual models from a functional model, or a functional description, of an information system expressed in natural language sentences.

In this method, first step consists of the construction of conceptual dependency diagrams (CDs) corresponding to each natural language sentence describing the functional model. (The Conceptual Dependency Theory, formulated by Roger Schank in 1975, states that different words and structures of a natural language that correspond to one same meaning should be represented equally.) An ER diagram is then created based on the conceptual dependency diagrams. Although this proposal focuses on semantics and integration, it implies an extra step – the creation of the conceptual dependency diagrams – what can be time consuming and, depending on the amount of information, a very laborious task.

5.3 LIDA Project

Overmyer *et al.* [2001] present LIDA, the Linguistic Assistant for Domain Analysis. They state that, as suggested by Chen, there are conventions in using parts of speech theory in the identification of objects and methods; there is a natural and practical association of nouns with classes, relationships with verbs and attributes with adjectives and prepositional phrases. However, in text documents describing the business environment, there are many more nouns, verbs, adjectives and adverbs than the ones necessary in the identification of objects and relationships.

The LIDA Analyzing environment works on imported documents, tagging the parts-of-speech of words. The analyst starts by the noun list, checking the ones that should be considered classes; then he/she searches the adjective list, searching for attribute candidates; then he/she proceeds through the verb list to identify possible methods and roles. After that the analyst proceeds to the LIDA modeler, that offers the functionalities to build a model (class diagram) from the model elements identified in the Analyzing environment.

The project aims at a semi-automatization of the modeling process by providing a text analyzing tool that tags parts-of-speech that allow analysts produce class diagrams. However, it focuses on system analysis and not on data modeling; also, it is based on Chen's eleven rules and, consequently, inherits their simplistic view of linguistic structures.

5.4 Color-X

Burg *et al.* [1997] present COLOR-X, COnceptual Linguistically based Object-Oriented Representation Language for Information and Communication Systems, which is a modeling environment and the first phase of a much larger project that aims at the generation of OO programming code. The project relates to some previous work, such as, the Natural Language Information Analysis Method (NIAM); Conceptual Prototyping Language (CPL); Conceptual Graphs (CG) and Explanation Modeling Language (EML) [Burg *et al*. 1997].

The project is based on the WordNet, an online lexical English database where parts-of-speech are organized in synomym sets, which is a project of the Cognitive Science Laboratory of the Princeton University. Color-X relates concepts (constructs) described in WordNet to OO constructs - the authors state that the objects that play an important role in the UoD must de identified and that they are probably nouns. Besides objects, relationships also must be identified and that these are probably verbs. [Dehne *et al*. 2001]. Color-X uses linguistic concepts aiming to create system models that reflect both static and dynamic aspects of the referred system. Its main goal, however is the generation of better OO code. Also, it does not perform an actual natural language constructs analysis; the mapping between WordNet and Color-X constructs resamble Chen' eleven rules.

5.5 OntoLT

Buitelaar *et al.* [2004] describe OntoLT, which is a plug-in designed to be used with to be used with Protégé, a widely used ontology tool. Aiming at ontology learning, OntoLT allows for the definition of mapping rules so that concepts and attributes are automatically extracted from annotated texts. It starts from the term extraction of a corpus through statistical process that determines their relevance. The terms are then clustered into groups for the identification of taxonomy potential classes. Relations between those are also statistically measured from the "connectedness" between the identified clusters.

However, the linguistic annotation resource is not a part of OntoLT, but SCHUG (Shallow and CHunk-based Unification Grammar Tools), a rule-based system for German and English analysis. SCHUG provides annotation of part-of-speech, morphological inflection and decomposition; it can read the results of several natural language processing tools and transform them into XML documents [Declerck 2002]. OntoLT provides a statistical preprocessing step that helps ensure that only information that is relevant to the domain be extracted: besides the pre-configured mapping rules, users can generate new ones from the results of this statistical preprocessing,.

The project, although focused in natural language processing, aims at machine-oriented products and not technology independent conceptual models. Also, since it is based on an statistical analysis, semantics is given a considerable importance.

5.6 English Structures and EER

Hartmann and Link [2007] reviews Chen's eleven rules for the correspondence between English grammatical structures and EER constructs - they re-organize and extend those rules in twelve heuristics, as follows:

- Heuristic 1 summarizes Chen's rules 1 and 2, and states that transitive verbs relate common nouns.

- Heuristic 2 addresses specialization and states that when an "is a" phrase connects two common nouns, the first noun represents a subtype of the second.
- Heuristic 3 addresses roles and states that in prepositional verbs prepositions refer to an object in the sentence, in which case the preposition corresponds to the role name of the component corresponding to the object.
- Heuristic 4 generalizes Chen's rules 3 and 4 and states that an adjective corresponds to an attribute of a component, whereas an adverb corresponds to an attribute of a relationship type.
- Heuristic 5 corresponds to Chen's rule 5, adding that X can also correspond to a collection type that is component of the object type corresponding to Y.
- Heuristics 6, 7 and 8 follow Chen's rules 6, 7 and 8.
- Heuristic 9 generalizes Chen's rule 9 and states that a gerund (nominal form of a verb) corresponds to a relationship type that is the component of a further relationship type.
- Heuristic 10 states that a clause refers to a relationship type and its components.
- Heuristic 11 says that a complex sentence corresponds to a relationship type which has at least another relationship type as its component.
- Heuristic 12 addresses alternatives and asserts that in a sentence like "… Z is either X or Y…" Z is a cluster type which components are X and Y.
- Heuristic 13 states that in a sentence of the kind "… X is a collection of Y …", X corresponds to a collection type which component is Y.

Despite the extensions and a few details, this approach follows the directives presented by Chen [1983], being rather simplistic as well.

5.7 KCPM

The KCPM – Klagenfurt Conceptual Predesign Model – is a conceptual modeling project from the University of Klagenfurt, Austria; there are several articles and documentation produced by this Klagenfurt group, of which this report will discuss a few. [Fliedl *et al*, 1996, 2000], [Kop and Mayr, 2002]

Aware that modeling methods force analysts to make early design decisions in terms of which construct should be associated to a certain concept of the Universe of Discourse, and that bad decisions need to be rectified, what can be expensive, this research group proposed a new phase in conceptual modeling, that is, the *conceptual predesign*. According to them, the analysis phase should be focused on capturing as completely as possible the relevant aspects of the environment being modeled. During this new phase, relevant information is collected from natural language sources, and then translated into a glossary-like predisgn schema, which is then translated to a conceptual model, according to a set of 18 transformation rules. The NIBA – natural language information requirements analysis – project that uses the NTS (Natürlichkeitstheoretische Syntax), developed in the linguistics department of the Klagenfurt University, provides for the linguistic analysis of collected information.

NTS is a German grammar model based on generative grammar theories in which the descriptions of grammatical phenomena are represented by trees expressing both constituency and dependency relations. The head of a construction is usually a lexical category. The linguistic analysis comprises four steps, that is, words are compiled in a lexicon, word- and morphosyntatically interpreted microstructures are assigned to words and verb groups; X'-mechanism of NTS analyses (sentence) syntax; and linguistic concepts are mapped to concepts of the conceptual predesign. Since the predesign activity is time consuming, a parallel project named NIBA (Ntürlichsprachliche Informationsbedarfsanalyse, or, natural language requirement analysis) aims at providing tools based on related theories, in order to automate this step. In fact, the three main objectives of KCPM project are to provide for a form of requirement documentation that is user centered; to automate as much as possible the process of creating the predesign scheme; and to automate as much as possible the mapping from the predesign scheme to the conceptual scheme.

KCPM is based on the German language, which grammar is much more rigid than the English one, for instance. Although the articles presented in this report do not introduce the details of the natural language analysis, it is known that German is a declined language, which means to say that word endings determine their syntactical function; the word order is not relevant. Such grammar structure, despite being hard on the speaker, and mostly on a foreign learner, is easier for a computer to work on. In English, as well as in Portuguese, the syntactical function of words is given by the position of such word in the sentence. Consequently, the linguistic analysis in KCPM may not apply to other languages.

The projects that advocate a linguistic approach to conceptual modeling presented in this report share the fact that they work from the modeling language perspective, and linguistic considerations are basically a rationale that justifies the choice of a construct over the other. In the eleven rules, for instance, the natural language constructs are described in terms of the constructs of entity-relationship model: "*a common noun corresponds to an entity type* " or "*an adjective in English corresponds to an attribute of an entity*". If one considers these two directives from the perspective of the natural language, it becomes clear that they should be rephrased. In English, as well as in Portuguese, the class of nouns is completely specialized in common nouns and proper nouns; since proper nouns are never present in a conceptual model, the first rule states that all nouns can correspond to an entity type. Also, the other rule in the example above states that an adjective corresponds to an attribute; in fact, an adjective characterizes a noun but in a conceptual model it corresponds to one of the values in the attribute domain – an attribute is a concept in itself, and in a natural language is represented by a noun, most certainly, an abstract noun.

Another interesting example would be heuristic 2 in [Hartmann and Link, 2007]: from its enunciation, the two sentences below would be structurally identical and present a specialization:

This car is a vehicle.

This doctor is a teacher.
Car is, in fact, a specialization of vehicle; but a doctor is not a specialization of teacher.

6 Conclusion

Conceptual models are invaluable tools for the understanding of the Universe of Discourse for which a new information system is to be created and/or updated. It also plays *"an important role of conceptual models is in facilitating interaction through discussion about the problem and its solution."*. [Veres and Mansson 2005]. Moreover, semantically accurate conceptual models are the basis for data integration. To create conceptual models that can actually be used as described above, the analysts need to learn the language of the Universe of Discourse being modeled; no matter the source, they will always come across descriptions done in a natural language. Consequently, it was only natural that analysts and scholars would eventually turn to linguistics for support.

Several studies, starting from Chen's eleven rules [1983], have presented heuristics and directives for the modeling process from the linguistic analysis of source narratives. While such studies advocate the need for semantically accurate and unambiguous models, they also discuss the need for automatic generation and understanding of such models; those can be opposing ideas. Some of such initiatives were presented in the section 5 of this article, and, in fact, most of them adopt the eleven rules as a basis. What all the works have in common, though, is that, first, they all are the result of their authors' experience in conceptual modeling; and second, they all treat the linguistic analysis from the perspective of the (meta)model. Moreover, all projects presented only go as far as the word classes, or parts-of-speech, not taking into account their semantic subclasses.

A correct linguistic analysis, however, is not enough to guarantee that the model produced is complete, easy to read and unambiguous; the modeling language to be used must have the constructs necessary for the model to be semantically equal to the natural language description. Consequently, the choice of a language for the representation of concepts is determinant on the resulting model [Lopes *et al* 2009]; a well founded ontological language, like OntoUML, seems to be the best choice. [Guizzardi 2005]. A linguistic approach that works from the perspective of the natural language and that takes into account its constructs raises the possibilities of, not only creating complete and unambiguous conceptual models, but also evaluating the completeness of the modeling language itself.

References

Batini, C., Ceri, S., Navathe, S., 1992. Conceptual Database Design, *Benjamin/Cummings*.

Buitelaar, P., Olejnik, D., Sintek, M., 2004. A Protégé Plug-In for Ontology Extraction from Text Based on Linguistic Analysis. *1st European Semantic Web Symposium (ESWS04)*.

Buitelaar, P., Sintek, M., 2004. OntoLT Version 1.0: Middleware for Ontology Extraction from Text. *International Semantic Web Conference*.

Bunge, M., 2003. Philosophical Dictionary. *Prometheus Books*.

Burg, J., Riet, R., 1997. The Impact of Linguistics on Conceptual Models: Consistency and Understandability. *Data and Knowledge Engineering*, 21.

Chen, P., 1976. The Entity-Relationship Model: Toward a Unified View of Data, *ACM Trans. Database Systems* 1, No. 1

Chen, P., 1983. English Sentence Structure and Entity-Relationship Diagrams, *Elsevier Science Publishing*.

Declerck, T., 2002. A set of tools for integrating linguistic and non-linguistic information. Proceedings of the SAAKM Workshop at ECAI.

Dehne, F., Steuten, A., Riet, R., 2001. WordNet++: A lexicon for the Color-X-method, DKE

Fliedl, G., Kop, Ch., Mayerthaler, W., Mayr, H. C., Winkler, Ch., 1996. NTS-based derivation of KCPM cardinalities: From natural language to conceptual predesign, *Application of Natural Language to Information System*, IOS Press.

Fliedl, G., Kop, Ch., Mayerthaler, W., Mayr, H. C., Winkler, Ch., 2000. Linguistically based requirements engineering – The NIBA-project, *Data and Knowledge Engineering*.

Gangopadhyay, A., 2001. Conceptual Modeling from Natural Language Functional Specifications, *Artificial Intelligence in Engineering*, Volume 15, Issue 2, pp. 207-218.

Guizzardi, G., 2005. Ontological Foundations for Structural Conceptual Models, *Telematica Instituut Fundamental Research Series* 15, Universal Press.

Hartmann, S., Link, S., 2007. English Sentence Structures and EER Modeling, *4^{th} Asia-Pacific Conference on Conceptual Modeling*, Volume 67.

Jackendoff, R., 2002. Foundations of Language, *Oxford University Press*.

Kop, Ch., Mayr, H. C., 2002. A User Centered Approach to Requirements Modeling, *Modellierung.*

Lopes, M., Guizzardi, G., Baião, F., Falbo, R., 2009. Reverse Engineering a Domain Ontology to Uncover Fundamental Ontological Distinctions, *ICEIS*.

Mounin, G., 1975. Les Problèmes Théoriques de la Traduction, *Cultrix* (in Portuguese).

Overmyer, S., Benoit, L., Rambow, O., 2001. Conceptual Modeling through Linguistic Analysis Using LIDA, *23rd International Conference on Software Engineering*.

Parnas, D., 1972. A Technique for Software Module Specification With Examples, *Communications of the ACM*, Volume 15, Number 5.

Pinker, S., 1999. Words and Rules, *Perennial*.

Quirk, R., Greenbaum, S., 1973. A University Grammar of English, *Longman*.

Saussure, F., 2006. Course in General Linguistics, *McGraw-Hill*.

Ullmann, S., 1964. Semantics – An Introduction to the Science of Meaning, *Fundação Calouste Gulbekian* (in Portuguese).

Veres, C., Mansson, G., 2005. Cognition and Modeling: Foundations for Research and Practice, *Journal of Information Technology Theory and Application (JITTA)*, 7:1.