Ontological Foundations for Structural Conceptual Models

Giancarlo Guizzardi

Enschede, The Netherlands, 2005

CTIT PhD Thesis Series, No. 05-74
Telematica Instituut Fundamental Research Series, No. 015 (TI/FRS/015)

ONTOLOGICAL FOUNDATIONS FOR STRUCTURAL CONCEPTUAL MODELS

PROEFSCHRIFT

ter verkrijging van
de graad van doctor aan de Universiteit Twente,
op gezag van de rector magnificus,
prof.dr. W.H.M. Zijm,
volgens besluit van het College voor Promoties
in het openbaar te verdedigen
op donderdag 27 oktober 2005 om 16.45 uur

door
Giancarlo Guizzardi
geboren op 22 april 1975
te Vitória, Brazilië

Chapter 1

Introduction

This thesis contributes to the definition of general ontological foundations for the area of conceptual modeling. This chapter presents the background of the thesis (section 1.1) and motivates the relevance of the work reported here (section 1.2). It also defines the main objectives of our research (section 1.3) and its scope (section 1.4). The chapter concludes by presenting the approach we follow to accomplish these objectives together with an overview of the thesis structure (section 1.5).

1.1 Background

Telematics is an area concerned with the support of the interactions between people or automated processes or both, by applying information and communication technology (ICT). In general terms, information and communication technology has a radical impact on its users, their work, and their working environments. In its various manifestations, ICT processes data, gathers information, stores collected materials, accumulates knowledge, and expedites communication. In fact, it plays a role in many, if not most, of the everyday operations of today's business world (Chen, 2000).

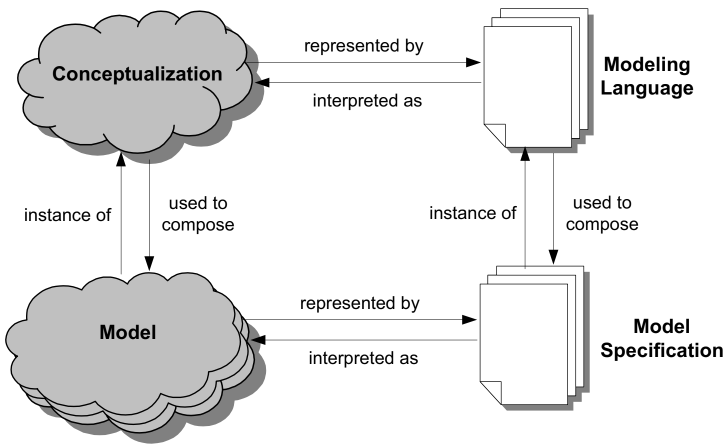
Telematic Systems are developed to support the enaction of *telematic services*. Users of telematic services are placed in a social context and, in order to satisfy the needs of these users, telematic services have to be strongly related to the design of the activities in the social context that these services support (Vissers et al., 2000).

An important constituent of the context in which a telematics service is embedded is the so-called *subject domain* (or *universe of discourse*) of this service. For instance, a *medical treatment reservation service* refers to concepts in a universe of discourse comprising entities such as patients, treatments, medical insurance, physicians, medical units, among others. The correct operation of this service, thus, depends on the correct *representation* of this subject domain. In particular, the representations of situations in reality used by a given system should stand for actual state of affairs of its subject domain. For example, if two people are said to be married in a system, or if a student is said to have graduated by a given university, these should reflect the actual state of affairs holding in reality.

Abstractions of a given portion of reality are constructed in terms of concepts, i.e., abstract representations of certain aspects of entities that exist in that domain. We name here a *conceptualization* the set of concepts used to articulate abstractions of state of affairs in a given domain. The abstraction of a given portion of reality articulated according to a domain conceptualization is termed here a *model.*

Conceptualizations and models are abstract entities that only exist in the mind of the user or a community of users of a language. In order to be documented, communicated and analyzed, these entities must be captured in terms of some concrete artifact. The representation of a conceptual model is named here a *model specification*. Moreover, in order to represent a specification, a *specification* (or *modeling*) *language* is necessary. The relation between conceptualizations, models, specifications and modeling languages is depicted in figure 1.1 below.

Figure 1-1 Relations between conceptualization, Model, Modeling Language and Specification



A language can be seen as determining all possible specifications (i.e., all *grammatically valid* specifications) that can be constructed using that language. Likewise, a conceptualization can be seen as determining all possible models (standing for state of affairs) admissible in that domain (Guarino, 1998). Therefore, for example, in a conceptualization describing genealogical relations, there cannot be a model in which a person is his own biological parent, because such a state of affairs cannot obtain in reality.

In this thesis, we are interested in the so-called class of *conceptual modeling languages*, as opposed to, for instance, languages aimed primarily at systems design and implementation. In a seminal paper, John Mylopoulos (Mylopoulos, 1992) defines the discipline of conceptual modeling as

"the activity of formally describing some aspects of the ***physical*** *and* ***social*** *world around us for purposes of* ***understanding*** *and* ***communication****...Conceptual modelling supports structuring and inferential facilities that are* ***psychologically grounded****. After all, the descriptions that arise from conceptual modelling activities are intended to be used* ***by humans, not machines...*** *The adequacy of a conceptual modelling notation rests on its contribution to the construction of models of reality that promote a common understanding of that reality among their human users."*

The specification of a conceptual model is, hence, a description of a given subject domain independent of specific design or technological choices that should influence particular telematics systems based on that model. Conceptual specifications are used to support *understanding* (*learning*), *problem-solving*, and *communication*, among stakeholders about a given subject domain. Once a sufficient level of understanding and agreement about a domain is accomplished, then the conceptual specification is used as a blueprint for the subsequent phases of a system's development process.

The quality of a telematics system and services, therefore, depend to a large extent on the quality of the conceptual specifications on which their development is based. The latter, in turn, is strongly dependent of the quality of the conceptual modeling language used in its description. For instance, if a modeling language is imprecise and coarse in the description of a given domain, then there can be specifications of the language which, although grammatically valid, do not represent admissible state of affairs. This situation is depicted in figure 1.2.

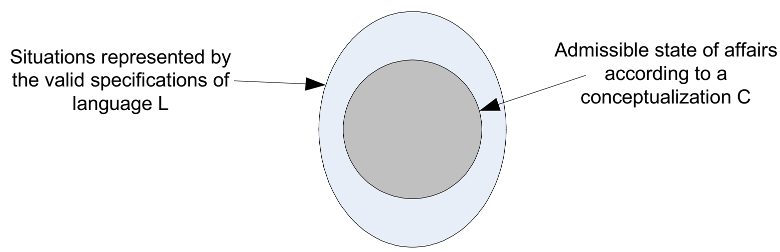


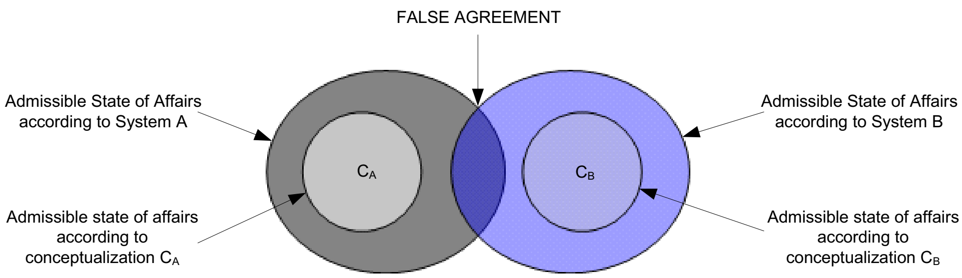
Figure 1-2
Consequences of a Modeling Language as an imprecise representation of a domain conceptualization

The difference between the two sets illustrated in figure 1.2 gives us a measure of the truthfulness to reality, or the so-called *domain appropriateness* of a given conceptual modeling language (Krogstie, 2000). In summary, we can state that the more we know about a given domain and the more precise we are on representing it, the bigger the chance that we have of constructing computational systems and services that are consistent with the reality of that domain.

Having a precise representation of a given conceptualization becomes even more critical when we want to integrate different independently developed models (or systems based on those models). Suppose the situation in which we want to have the interaction between two independently developed systems, which commit to two different conceptualizations. In order for these systems to function properly together, we must guarantee that they ascribe compatible meanings to real-world entities of their shared subject domain. In particular, we want to reinforce that they have compatible sets of admissible situations, whose union (in the ideal case) equals the admissible state of affairs delimited by the conceptualization of their shared subject domain. The ability of systems to interoperate (i.e., operate together), while having compatible real-world semantics is known as *semantic interoperatibility* (Vermeer, 1997).

Now, suppose we have the situation depicted in figure 1.3. C_A and C_B represent the conceptualizations of the subject domains of systems A and B, respectively. As illustrated in figure 1.3, these conceptualizations are not compatible. However, because these systems are based on poor representations of these conceptualizations, their sets of possible situations considered overlap. As a result, systems A and B agree exactly on situations that are neither admitted by C_A nor by C_B. To put it simply, although these systems seem to have a shared view of reality, the portions of reality that each of them aims at representing are not compatible together. This problem, termed *The False Agreement Problem* was first highlighted in (Guarino, 1998).

Figure 1-3 False Semantic Agreement between two Communicating Entities



Another important quality criterion for conceptual specifications is *pragmatic efficiency*. Since these specifications are meant to be used by humans, their conceptual clarity and ability to support communication, understanding and reasoning about the domain plays a fundamental role. This quality criterion of conceptual specifications is also termed *comprehensibility appropriateness* (Krogstie, 2000).

Thus, on one hand, a modeling language should be sufficiently expressive to suitably characterize the conceptualization of its subject domain, on the other hand, the semantics of the produced specifications should be clear, i.e., it should be easy for a specification designer to recognize what language constructs mean in terms of domain concepts. Moreover, the specification produced using the language should facilitate the user in understanding and reasoning about the represented state of affairs.

In this thesis we defend that the suitability of a conceptual modeling language to represent a set of real-world phenomena in a given domain (i.e., its domain and comprehensibility appropriateness) can be systematically evaluated by comparing the level of homomorphism between a concrete representation of the world view underlying the language (captured in a *specification of the language metamodel*), with an explicit and formal representation of a conceptualization of that domain, which is termed here a *reference ontology*.

In philosophy, ontology is the most fundamental branch of metaphysics. It is a mature discipline, which has been systematically developed in western philosophy at least since Aristotle. The business of ontology "…is to study the most general features of reality" (Peirce, 1935), as opposed to the several specific scientific disciplines (e.g., physics, chemistry, biology), which deal only with entities that fall within their respective domain. However, there are many ontological principles that are utilized in scientific research, for instance, in the selection of concepts and hypothesis, in the axiomatic reconstruction of scientific theories, in the design of techniques, and in the evaluation of scientific results (Bunge, 1977, p.19). Thus, to quote the physicist and philosopher of science Mario Bunge: "every science presupposes some metaphysics".

In the beginning of the 20th century, the German philosopher Edmund Husserl coined the term Formal Ontology as an analogy to Formal Logic. Whilst Formal Logic deals with formal[1] logical structures (e.g., truth, validity, consistency) independently of their veracity, Formal Ontology deals with formal ontological structures (e.g., theory of parts, theory of wholes, types and instantiation, identity, dependence, unity), i.e., with formal aspects of objects irrespective of their particular nature. The unfolding of Formal Ontology as a philosophical discipline aims at developing a system of general categories and their ties, which can be used in the development of scientific theories and domain-specific common sense theories of reality.

---

[1]The adjective *Formal* here refers to its more ancient meaning, namely, referring only to *Form*, in the sense of independent of *Content*. The use of *formal* as synonym for *precise* or *mathematical* originates from the fact that mathematical theories are typically *Formal* in the first sense.

More recently, ontology has been applied in a multitude of areas in computer science. In many cases, however, the term is employed with a more liberal meaning and, instead of referring to a general (i.e., domain-independent) system of categories, it is also used to refer to specific theories about material domains (e.g., law, medicine, archeology, molecular biology, etc.), named *domain ontologies*. Thus, if in the philosophical sense, ontology is the *study of existence and modes of existence in a general sense*, in computer science, a domain ontology is the study of what exists in a given domain or universe of discourse.

The activity of constructing domain ontologies is known in the literature as *Ontological Engineering*. An ontological engineering process typically comprises activities such as: *Purpose Identification and Requirements Specification*, *Ontology Modeling*, *Ontology Codification*, *Reuse and Integration*, *Evaluation* and *Documentation* (see, for instance, Falbo & Guizzardi & Duarte, 2002; Gómez-Pérez & Fernández-López & Corcho, 2004). Here, we consider a domain ontology as a special type of conceptual specification and, hence, ontology modeling as a special type of conceptual modeling.

Therefore, in figure 1.1, if by a conceptualization we mean a conceptualization of a material domain, then by modeling language we mean a *domain-specific modeling language*. In contrast, if in figure 1.1 by a conceptualization we mean a formal (i.e., domain- independent) conceptualization, then by a modeling language we mean a *general conceptual modeling language* (or *ontology representation language*).

The design of domain-specific modeling languages is a current and important research topic (Kelly & Tolvanen, 2000; Tolvanen, Gray & Rossi, 2004; Bottoni & Minas, 2003) in conceptual modelling and, as we show on chapter 2, some results of this thesis also contribute to the area of domain-specific language evaluation and design. Nonetheless, the focus of this work is not on domain-specific languages and domain ontologies but, conversely, on general conceptual modeling languages and their underlying formal conceptualizations, if only because (as we demonstrate in chapters 2 and 3), the design of the former presupposes the existence of a suitable general conceptual modeling language. Thus, henceforth we simply use the term *conceptual modeling language* when referring to a general conceptual modeling language.

Conceptual (Ontology) Modeling is a fundamental discipline in computer science, playing an essential role in areas such as database and information systems design, software and domain engineering, design of knowledge-based systems, requirements engineering, information integration, semantic interoperability, natural language processing, enterprise modeling, among many others. In particular, domain ontologies have a central position in the so-called *Semantic Web* vision (Berners-Lee, Hendler, Lassila, 2001). In this context, web resources (information nodes and computational services) have their semantics informed by association with one or more domain ontologies. For example, the systems A and B in the pattern of figure 1.3 could correspond to two independently developed *semantically annotated web* services (McIlraith & Son & Zeng, 2001). Since web services can be considered as special kinds of telematics services (Ferreira Pires et al., 2004), the results developed throughout this thesis amount to a contribution to the general area of telematic services. However, more generally, the systems A and B in the pattern of figure 1.3 could also correspond to two interacting software agents, social organizations, or human stakeholders. Therefore, the results presented here contribute more broadly to all the areas aforementioned in computer science in which conceptual modeling play an essential role.

In summary, we defend in this thesis that the truthfulness to reality of a given system, as well as the semantic interoperability of concurrently developed systems, strongly depend on the availability of conceptual modeling languages that are able of making explicit and precise representations of the conceptualizations of their underlying subject domains. Therefore, two central research questions are: How can we define a suitable *formal conceptualization* (and consequently a *formal ontology*) that a conceptual modeling language should commit to? How can we (re)design a *conceptual modeling language* that conforms to this formal conceptualization (ontology)? These questions are answered throughout this thesis.

1.2 Motivation

Nowadays, many languages exist that are used for the purpose of creating representations of real-world conceptualizations. These languages are sometimes named *domain modeling languages*[2] (e.g., LINGO), *ontology representation languages* (e.g., OWL), *semantic data modeling languages* (e.g., ER), among other terms. We shall refer to these languages as conceptual modeling languages henceforth.

Although these languages are employed in practice for conceptual modeling, they are not designed with the specific purpose of being truthful to reality. For instance, LINGO (Falbo & Menezes & Rocha, 1998; Falbo & Guizzardi & Duarte, 2002) was designed with the specific objective of achieving a positive trade-off between expression power of the language and the ability to facilitate bridging the gap between the conceptual and

[2] The term *domain modeling language* is used in this sense to refer to domain-independent languages which can be used to create specifications of different material domain conceptualizations, not to refer to domain-specific modeling languages as previously discussed.

implementation levels. This preoccupation also seems to be present in Peter Chen's original proposal for ER diagrams (Chen, 1976). OWL (Horrocks & Patel-Schneider & van Harmelen, 2003) has been designed with the main purpose of achieving computational efficiency in an automatic reasoning process. Some other languages, such as Z (Spivey, 1988) and CC Technique (Dijkman & Ferreira Pires & Joosten, 2001), take advantage of the simplicity of the well-defined mathematical framework of set theory. Finally, some of the languages used nowadays for conceptual modeling were created for different purposes, the most notorious example being the UML (OMG, 2003c), which initially focused on software design.

As we show in this thesis, the worldviews underlying these languages (their ontological metamodels), cannot be considered as adequate conceptualizations of reality. As a consequence, they fall short in offering their users suitable sets of modeling concepts for constructing precise and explicitly characterized representations of their subject domains of interest.

We defend here that the focus of a conceptual modeling language should be on *representation adequacy* (i.e., truthfulness to reality and pragmatic efficiency). Conceptual modeling is *primarily* about *"describing some aspects of the physical and social world around us for purposes of understanding and communication"*, not systems design. Moreover, conceptual modeling languages should be highly-expressive, even at the cost of sacrificing computational efficiency and tractability. After all, although conceptual modeling can greatly benefit from efficient tool support in activities such as model manipulation and visualization, storage, syntactic verification and reasoning, among others, *"the descriptions that arise from conceptual modelling activities are intended to be used [primarily] by humans, not machines."*

Currently, there is no commonly agreed language for describing real-world phenomena in computer science. For this reason, in order to overcome the deficiencies of existing modeling languages for this purpose, a number of recent research efforts have investigated the use of Formal Ontological theories to evaluate and redesign these languages, as well as equip them with adequate real-world semantics. Examples include (Shanks & Tansley & Weber, 2003; Evermann & Wand, 2001b; Bodart et al., 2001; Opdahl & Henderson-Sellers, 2001; Green & Rosemann, 2000).

The approach proposed here differs from the ones mentioned above in two main characteristics: First, each of the approaches presented focus on specific sets of concepts. For example, the ontological analysis presented in (Opdahl & Henderson-Sellers, 2001) is targeted at *part-whole relations*, the one of (Bodart et al, 2001) is targeted at *properties*, and the one of (Evermann & Wand, 2001b) analyses *classes*, *class hierarchies* and *properties* (among other non-structural concepts, such as interaction). Our approach is broader in scope and, hence, can be considered in this sense an extension of these efforts. Consequently, it provides a comprehensive set of ontological theories, which covers all fundamental conceptual modeling concepts, and tackles a number of conceptual modeling problems that have not yet been satisfactorily addressed by any of the existing approaches in the conceptual modeling literature.

Second, the type of ontological investigation carried out here is different from the investigation in these other approaches. One characteristic common to all the efforts aforementioned is that they employ the same ontological theory, namely an ontology named BWW (Bunge-Wand-Weber) based on the original methaphysics proposed in (Bunge, 1977, 1979). Mario Bunge is a physicist and a philosopher of science and his theory is meant to serve as a foundation for specific scientific disciplines. As a consequence, it subscribes to an approach of ontological investigation that is committed to capture the intrinsic nature of the world in a way that is independent of conceptualizing agents and, consequently, an approach in which cognition and human language play a minor or non-existent role.

As we demonstrate in the development of this thesis, an ontology that can be used for providing foundations for conceptual modeling should be a philosophically well-founded one, but also one that aims at capturing the ontological distinctions underlying human cognition and common sense. Nonetheless, this ontology should not be regarded as less scientific, in the sense that the very existence of its constituting categories can be empirically uncovered by research in cognitive sciences (Keil, 1979; Xu & Carey, 1996; Mcnamara, 1986) in a manner that is analogous to the way philosophers of science have attempted to elicit the ontological commitments of the natural sciences.

In summary, the position defended here subscribes to Mylopoulos' dictum (Mylopoulos, 1992) that *"[t]he adequacy of a conceptual modelling notation rests on its contribution to the construction of models of reality that promote a common understanding of that reality among their human users."*

1.3 Objectives

In this thesis, we aim at contributing to the theory of conceptual modeling and ontology representation. Our main objective here is to provide ontological foundations for the most fundamental concepts in conceptual modeling. These foundations comprise a number of ontological theories, which are built on established work on philosophical ontology, cognitive psychology, philosophy of language and linguistics. Together these theories amount to a system of categories and formal relations known as a *foundational ontology* (Masolo et al., 2003a).

Besides philosophical and cognitive adequacy, we intend our foundational ontology to be precise. Therefore, we make use of some modal logics concepts to formally characterize the entities that constitute our ontology. In case the ontological distinctions proposed cannot be properly characterized by standard formal approaches, we have proposed some extensions to these standard formal approaches to accomplish the characterization required.

Once constructed, we have used this foundational ontology as a *reference model* prescribing the concepts that should be countenanced by a well-founded conceptual modeling language, and providing real-world semantics for the language constructs representing these concepts.

In the reference ontology proposed, we have focused on providing foundations for the most fundamental and widespread constructs for conceptual modeling, namely, types and type taxonomies, roles, attributes, attribute values and attribute value spaces, relationships, and part-whole relations.

Besides the theoretical work, we have addressed existing conceptual modeling problems, and contributed to the creation of sound engineering tools that can be used in the conceptual modeling practice. These have been realized in the form of ontological design patterns, capturing standard solutions to recurrent conceptual modeling problems, and methodological directives. However, more importantly, we have instantiated the approach defended here, by proposing a concrete conceptual modeling language that incorporates the foundations captured in our reference ontology.

The ontology proposed serves as a reference for designing new conceptual modeling language, but also for analyzing the ontological adequacy of existing ones. However, to conduct these activities in a principled manner, we have established a systematic relation between a modeling language and the ontology representing the real-world conceptualization of a given domain. Once this relationship has been precisely understood, we have analyzed and redesigned a specific modeling language, namely, the Unified Modeling Language (UML) (OMG, 2003c). The objective has been to propose an ontologically well-founded version of UML that can be used as an appropriate conceptual modeling language. The choice for UML lies on two main points: (i) the current status of UML as *de facto* standard modeling language; (ii) the growing interest in its adoption as a language for conceptual modelling and ontology representation (OMG, 2003a; Kogut, 2002). Because of these reasons, the re-designed version of UML is in itself an important research contribution of this thesis.

Finally, in order to demonstrate the suitability of the conceptual modeling language proposed we have developed a case study in a domain where we can exercise both (i) the capabilities of the language in precisely characterising the domain elements; (ii) the use of the language in supporting the semantic integration of different domain models. In particular, in (ii), we have shown the importance of a suitable conceptual modelling language in making explicit the ontological commitments of the conceptualizations underlying the individual models and, consequently, in helping to prevent false agreement in their integration.

In summary, the objectives of this thesis have been:

1. To establish a systematic relation between a *modeling language* and a *reference ontology*, and to propose a methodological approach to analyze and (re)design modeling languages to reinforce *representation adequacy* exploiting this relation;

2. To construct philosophical and cognitive *foundational ontology* for conceptual modeling and to formally characterize the elements constituting this ontology;

3. To demonstrate the usefulness of the ontological categories and theories that were proposed to address existing *conceptual modeling problems*;

4. To demonstrate the adequacy of the approach proposed in (1) and of the foundational ontology constructed in (2) by analyzing and redesigning an existing conceptual modeling language for representation adequacy;

5. To demonstrate the adequacy of the ontologically well-founded conceptual modeling language produced in (4) in the activity of improving the domain representation of existing conceptual specifications, and supporting their semantic integration.

1.4 Scope

The focus of this thesis is on general (i.e., domain independent) conceptual modeling languages. For this reason, we focus here on the construction of formal ontological theories instead of (domain-specific) material ones.

Our objective is to provide foundations for *structural* (i.e., static) aspects of conceptual modeling languages, as opposed to dynamic ones. This class of languages includes languages known as *data modeling frameworks*, *ontology representation languages*, *knowledge representation languages*, *semantic data modeling languages*, among others. To put it in simple terms, we restrict ourselves here to *objects*, the *types* they instantiate, the *roles* they play in certain contexts, their *constituent parts*, their *intrinsic and relational properties*, and the *structures in which their features are valued*, among other things. In contrast, we do not elaborate on *processes* and *events*. To put it in philosophical terms, the foundational ontology developed here is an *ontology of Endurants* (*continuants*) not one of *Perdurants* (*occurents*) (van Leeuwen, 1991; Masolo et al., 2003a). This is far from denying to the latter the status of ontological entities. Actually, in (Guizzardi & Wagner, 2005a), we elaborate on the role of an *ontology of perdurants* as an extension of the work presented in this thesis. In summary, the restriction of the discussion promoted here to the ontological category of endurants is merely a matter of scope.

The objective of this thesis is also to evaluate the suitability of languages to represent phenomena in a given domain. In terms of quality criteria for modeling languages, our scope is on *expressiveness* and *clarity*. Thus, it is not the objective here to discuss specific language technologies related to the definition of metamodel specifications, concrete syntax or formal semantics. Moreover, we do not discuss aspects related to systems design and, in particular, we do not address the impact on design choices of the modeling concepts proposed here. Finally, the target of our work is on conceptual modeling concepts and languages conceived for representation adequacy, aimed at being employed by human users in activities such as communication, domain understanding (learning) and analysis. Therefore, the study of properties such as computational efficiency and tractability of these languages fall outside the scope of this work.

1.5 Approach and Structure

The structure of this thesis reflects the successive elaboration of the objectives identified in section 1.4. The approach followed here to accomplish these objectives is detailed in the sequel.

(O1). Objective 1: To establish a systematic relation between a *modeling language* and a *reference ontology*, and to propose a methodological approach to analyze and (re)design modeling languages to reinforce *representation adequacy* exploiting this relation

This objective is accomplished in chapters 2 and 3 of this thesis. We start chapter 2 by discussing the various aspects that comprise a system of representations, or simply, a language. After briefly discussing the issues of (abstract and concrete) syntax, (formal and real-world) semantics and pragmatics, we concentrate on the definition of an evaluation framework that can be used to precisely evaluate the suitability of a language to represent phenomena according to a given real-world conceptualization. In our approach, this property can be systematically evaluated by comparing the level of homomorphism between a concrete representation of the world view underlying the language (captured in the specification of a *metamodel of the language*), and an explicit and formal representation of a conceptualization of that domain, or a *reference ontology*. The framework proposed comprises a number of properties (*lucidity, soundness, laconicity, completeness*) that must be reinforced for an isomorphism to take place between these two entities.

Although the focus of our work is on general conceptual modeling languages, the framework and the principles presented can be applied to the design of conceptual modeling languages irrespective of the generalization level to which they belong. In particular, they can also be used for the design of domain-specific modeling languages. In chapter 2, the approach presented is illustrated with a small case study in the design of a domain-specific visual modeling language for the domain of genealogy. The evaluation and redesign of a general conceptual modeling language is the main case study of this thesis, which is presented in chapter 8.

In chapter 3, we elaborate on some of the concepts of this framework by presenting a formal characterization of a *conceptualization* and its *intended models* (the models standing for *admissible state of affairs*), the *ontological commitment* of a language, and of the role of an *ontology* to approximate the valid specifications of a language to the intended models of its underlying conceptualization.

The main objective of chapter 3 is, however, to discuss the topic of ontologies both from philosophical and computer science points of view. We first give a historical perspective on ontology from a philosophical perspective, and discuss the importance of ontological investigations for science, in general, and for conceptual modeling, in particular. The formal characterization aforementioned is also used in this chapter to harmonize the original uses of ontology in philosophy with the several senses the term is employed in computer science. By doing this, we offer a precise definition of the meaning of the term, which is assumed for the remaining of this work.

(O2). Objective 2: To construct a philosophical and cognitive *foundational ontology* for conceptual modeling and to formally characterize the elements constituting this ontology.

(O3). Objective 3: To demonstrate the usefulness of the ontological categories and theories that are proposed to address existing conceptual modeling problems.

The accomplishment of these objectives constitutes the core of this thesis. The construction of the foundation ontology proposed here is organized in four complementary chapters in the following manner:

(a). In Chapter 4, we provide a theory for defining ontological distinctions on the category of conceptual modeling object universals, as well as constraints on the construction of *taxonomic structures* using these distinctions. By using a number of formally defined meta-properties, we can generate a *typology of universals*, which in turn can be used to give real-world semantics for important conceptual modeling concepts such as *types*, *roles*, *phases* and *mixins*. Besides providing unambiguous definition for these concepts, the elements of our theory function as a methodological support for helping the user of the language to decide how to represent elements that denote universal properties in a given domain. The usefulness of this approach is demonstrated in this chapter by showing how the theory can be used to evaluate and improve the conceptual quality of class hierarchies and concept taxonomies. Finally, in order to provide a suitable formal characterization of the ontological distinctions and postulates present in this theory, we present some extensions to a traditional system of modal logics;

(b). In Chapter 5, we concentrate on the topic of *part-whole relations*. First, we extend the insufficient axiomatization offered for these relations in present conceptual modeling languages, by considering a number of theories of parts from formal ontology in philosophy (Mereologies)(Simons, 1987). Thus, by building on the literature of meronymic[3] relations on linguistics and cognitive sciences, we extend the formal notion of parthood to a typology composed of four different conceptual part-whole relations. The elements in this typology are also characterized by additional formal meta-properties (e.g., essentiality, exclusiveness, separability, transitivity);

(c). In Chapter 6, we present the core of the foundational ontology proposed here, by addressing the categories of *attributes, attribute values and attribute value spaces, relationships* and *weak entities*. This fragment of our ontology is presented in a parsimonious theory, which is used to provide unambiguous real-world semantics for these concepts. In particular, this chapter offers a simple, precise and ontologically well-founded semantics for the problematic concept of relations, but also one that can accommodate more subtle linguistic distinctions. As it is demonstrated, this foundation for relations has a direct impact in

---

[3]**Meronym**: a word that names a part of a larger whole; *brim* and *crown* are meronyms of *hat*. The contrary idea is that of **Holonym**, i.e., a word that names a whole of which a given word is part. In this example, *hat* is a holonym for *brim* and *crown* (WordNet, 2005).

improving the representation of these entities in conceptual specifications. Additionally, it provides a principled basis for an ontological interpretation and for the specification of structured datatypes;

(d). In Chapter 7, we employ some of the results of Chapter 6 to fully describe the modeling concept of roles. The chapter also serves as an exemplification of the usefulness of the categories proposed in our foundational ontology. Firstly, by employing the categories and postulates of the theory of universals constructed in Chapter 4, we propose an ontological design pattern capturing a solution to a recurrent and much discussed problem in role modeling. Secondly, with some definitions offered in Chapter 6, we have investigated the link between some of the formal meta-properties defined for part-whole relations and those meta-properties by which roles are characterized. Thirdly, by borrowing some results from Chapters 4 and 6, we have managed to harmonize some different conceptions of roles used in the literature. Finally, by building on an existing theory of transitivity of linguistic functional parthood relations, and on some material from Chapter 6, we have proposed a number of visual patterns that can be used as methodological support for the identification of the scope of transitivity for the most common type of part-whole relations in conceptual modeling.

(O4). Objective 4: To demonstrate the adequacy of the approach proposed to fulfil (O1) and of the foundational ontology constructed to fulfil (O2) by analyzing and redesigning an existing conceptual modeling language for representation adequacy.

(O5). Objective 5: To demonstrate the adequacy of the ontologically well-founded conceptual modeling language produced to accomplish (O4) in the activity of improving the domain representation of existing conceptual specifications, and supporting their semantic integration.

In Chapter 8, we present the two major case studies of this thesis. As a first case study to exemplify the adequacy of the framework and foundation ontology proposed, we use the latter as a reference for analyzing the ontological appropriateness of the Unified Modeling Language (UML) for the purpose of conceptual modeling. Moreover, by employing the systematic evaluation method comprising the framework, we have identified a number of deficiencies and recommended modifications to the UML metamodel specification accordingly. As a result of this process, we have managed to produce a conceptually cleaner, semantically unambiguous and ontologically well-founded version of the language.

As an attempt to shield as much as possible the user of a conceptual modeling language from the complexity of the underlying ontological theory, we take the approach of (whenever possible) representing the ontological principles underlying a language in terms of syntactical constraints of this language. As a consequence, we manage to produce a modeling language whose grammatically valid specifications approximate as much as possible the intended models of its underlyimg conceptualization.

Finally, in Chapter 8, we also carry out a second case study, which uses the version of UML proposed in that chapter to analyze and integrate several semantic web ontologies in the scope of a context-aware service platform. This case study aims at demonstrating the adequacy of this version of UML as a conceptual modeling language and as a tool to support minimizing the false agreement problem previously discussed. Accordingly, it has also demonstrated the suitability of the ontological foundations underpinning this language for these purposes.

An overview of structure of this thesis is presented in figure 1.4 below.

Figure 1-4 Overview of the thesis structure relating the objectives of this thesis with the chapters in which they are accomplished