Personal semantics of meta-concepts in conceptual modeling languages

by

Dirk van der Linden

Printed and published by Ridderprint BV, Ridderkerk, the Netherlands
Cover and lay-out by Dirk van der Linden
ISBN 978-94-6299-001-2

EE-Team dissertation series: 2015-01

The research reported in this thesis was conducted at Public Research Centre Henri Tudor in Luxembourg, under the auspices of the Institute for Computing and Information Sciences of Radboud University Nijmegen and the Enterprise Engineering Team. The research was partially supported by the Fonds National de la Recherche Luxembourg.

Personal semantics of meta-concepts in conceptual modeling languages

Proefschrift

ter verkrijging van de graad van doctor
aan de Radboud Universiteit Nijmegen
op gezag van de rector magnificus prof. dr. Th.L.M. Engelen,
volgens besluit van het college van decanen
in het openbaar te verdedigen op vrijdag 13 februari 2015
om 16.30 uur precies

door

Dirk Johannes Theodorus van der Linden

geboren op 31 oktober 1985
te Nijmegen

Promotor:
prof. dr. H.A. Proper

Copromotor:
dr. S.J.B.A. Hoppenbrouwers
dr. S. de Kinderen — University of Luxembourg, Luxembourg

Manuscriptcommissie:
prof. dr. Th.P. van der Weide
prof. dr. B. Henderson-Sellers — University of Technology, Sydney, Australia
dr. I. Hadar — University of Haifa, Haifa, Israel

Paranimfen:
G. Plataniotis, MSc.
L.C.J.M. Peters, MSc.

“Но ведь человек пишет потому, что мучается, сомневаеця. Ему все время надо доказывать себе и окружающим, что он чего-нибудь , да стоит. А если я буду знать наверняка, что я - гений? Зачем мне писать тогда? Какого рожна?”

“A man writes because he is tormented, because he doubts. He needs to constantly prove to himself and the others that he’s worth something. And if I know for sure that I’m a genius? Why write then? What the hell for?”

– Андре́й А. Тарко́вский (*Сталкер*, 1979)

Contents

Preface

The genesis of this research lies in a quickly aborted desire to integrate a bunch of modeling languages in order to create a holistic method for enterprise modeling (not entirely unlike UEML). It was quickly realized that such additional new efforts, while interesting from an academic point of view, have little practical value, or indeed, lastingness (as can be seen from the myriad of attempts lingering out there). From those initial forays into the analysis of conceptual modeling language semantics an idea was born that, perhaps, we should focus on techniques that give us more understanding into the users of these modeling languages languages, not the language itself.

The integration of multiple languages (say, one for goal modeling and one for value exchanges) is of course done for a good reason: because those two aspects need to meet, and their meeting will give rise to something more than just the sum of their parts. As such, the integration of languages is not a useless effort in itself. But what is really desired here, in a practical sense, is for the people from these different worlds to be able to relate to each other (at least for those models where the goal is to capture knowledge and reason about it together).

Along these lines we decided to forego the typical approach of defining yet another modeling language, or building yet another framework for integration. Instead, we opted to perform empirical work to focus on the knowledge questions that underlie many of these integration efforts: how do we know whether modelers understand each other, and how do we make such (converging or dissenting) understandings explicit? As a result, this thesis came from a couple of desires: the desire to add some insights to the field of conceptual modeling and its languages, and the desire to do empirical research instead of stipulating the latest and greatest way to do work, research, business, or just about anything.

We started eliciting the conceptual understanding modelers had of their modeling concepts, and began to refer to these individual understandings as 'personal semantics'. This was, as could have been expected, quickly dismissed. These quick dismissals of the mere notion of 'personal semantics' as being unnecessary because "*people simply have the right or wrong understanding of modeling concepts"* and it is in any sense trivial "*because we can just define the right semantics*" were the final stimuli we needed to focus and get down to work. With the additional studies we performed, the empirical data we have available, and the discussion of what consequences these have for modeling we hope to have convinced at least some modelers (researchers and practitioners alike) that insights into the way people think about modeling concepts are worthwhile.

The research reported on in this dissertation would not have been possible if not for the involvement of a diverse number of people. My co-supervisor, Stijn Hoppenbrouwers, always stimulated me to think of modeling languages from a linguistic point of view and treat them as actual languages, which led to the strong focus on *as-used* semantics. Whenever I was stuck in a conceptual dead-end or lost in my data an hour of critically thinking aloud at Stijn's whiteboard always led to new clues and angles to pursue.

Later on in the project, someone who I had been working with on and off for some time became more involved as my second co-supervisor. Sybren de Kinderen spent a significant amount of time with me editing this dissertation and gave invaluable input on many of the studies and articles themselves. The fact that this dissertation is at least slightly more readable than a collection of random research articles can be attributed to his diligent and at times hounding input on the coherence of everything you are about to read.

The different studies, papers, and ideas benefited greatly from the many discussions with my colleagues, friends and critics (often some combination of the three). I am indebted to all of them for their inspiring comments and discussions. Whether they were the inspiring academics I met at conferences, the friends that suffered reading through my writings, or my colleagues who were always there to discuss and critique my ideas and explanations, like Georgios Plataniotis and Marc van Zee while they were suffering and sweating with me in the pool and gym. My parents Hans and José gave me ample exercise in trying to explain what my research was actually about, which helped a lot to actually understand myself just what I was trying to do. The empirical work itself owes a lot to the insights Alina Lartseva gave in our many discussions on the empirical methods that would actually be feasible to use among practitioners in the field of conceptual modeling. The studies we performed would not have been possible without the many practitioners and students who participated in our studies. For their time investment, especially those students consistently participating in the longitudinal study performed at Radboud University the work itself, and I, are indebted to them.

Finally, I would like to thank my supervisor, Erik (sorry, Hend*erik*) Proper for all the usual stuff: allowing me the freedom to set out my own research, spending a lot time on discussions and critiques of my work, and knowing when to give me the freedom to find my own way. But more importantly: for stimulating me to work on other things than just my mind, doing a lot of swimming and running. I would like to make it be known that at the time of writing I could easily outswim him in the pool (thanks to my good friend Hervé Fokan). I consider it one of, if not *the*, crowning achievements of my time of doctoral research.

Luxembourg, Dirk van der Linden
December 15, 2014

List of Publications

Dirk van der Linden, Stijn J.B.A. Hoppenbrouwers, Alina Lartseva and Henderik A. Proper: Towards an Investigation of the Conceptual Landscape of Enterprise Architecture. In: T. Halpin et al. (Eds.): *BPMDS 2011 and EMMSAD 2011, LNBIP 81*, pp. 526–535. Springer, Heidelberg

Re-published in: Proc. of the 6th SIKS Conference on Enterprise Information Systems (EIS2011), *V. Dignum et al. (Eds.): CEUR-WS Vol 800, 2011, pp. 17–18*

Stijn J.B.A. Hoppenbrouwers, Wim van Stokkum, Maria-Eugenia Iacob, Ilona Wilmont, Dirk van der Linden, Chintan Amrit, and Maarten Joosen: Stakeholder Communication in Service Development. In: M.M. Lankhorst et al. (Eds.): *Agile Service Development – Combining Adaptive Methods and Flexible Solutions*, Springer, Heidelberg, 2012

Dirk van der Linden, Khaled Gaaloul and Wolfgang Molnar: Initial Results from a Study on Personal Semantics of Conceptual Modeling Languages. In: G. Bouma et al. (Eds.): *NLDB 2012, LNCS 7337*, pp. 360–365. Springer, Heidelberg, 2012.

Re-published in: *van Eck et al. (Eds.): Proc. of the 7th SIKS Conference on Enterprise Information Systems (EIS)*, Nieuwegein, 2012.

Dirk van der Linden, Alina Lartseva, Stijn J.B.A. Hoppenbrouwers and Wolfgang Molnar: Beyond Terminologies: Using Psychometrics to Validate Shared Ontologies. In: *Applied Ontology*, vol. 7, no. 4, 2012, IOS Press.

Dirk van der Linden and Stijn J.B.A. Hoppenbrouwers: Challenges of Identifying Communities with Shared Semantics in Enterprise Modeling. In: Sandkuhl et al. (Eds):

The Practice of Enterprise Modeling, LNBIP 134, pp 160–171, Springer, Heidelberg, 2012. (**Best Paper Award**)

Dirk van der Linden and Henderik A. Proper: On the cognitive understanding of language types in conceptual modeling languages. In: Jung and Reichert (Eds.): *Enterprise Modelling and Information Systems Architecture (EMISA 2013)*, LNI 222, pp. 149–162, GI, 2013

Dirk van der Linden: Categorization of modeling concepts: graded or discrete? In: Y.T. Demey and H. Panetto (Eds.): *OTM 2013 Workshops, LNCS 8186*, pp. 743–746. Springer, Heidelberg, 2013

Dirk van der Linden and Henderik A. Proper: Do domain-specific modeling languages accommodate enough conceptual distinctions? In: *Short Paper Proceedings of the 6th IFIP WG 8.1 Working Conference on The Practice of Enterprise Modeling (PoEM)*, CEUR-WS Vol 1023, pp 126–135, 2013

Dirk van der Linden, Stijn J.B.A. Hoppenbrouwers and Henderik A. Proper: On the identification of modeler communities. In: *International Journal of Information Systems Modeling and Design (IJISMD)*, vol. 5, no. 2, pp. 22-40, IGI Global, 2014

Dirk van der Linden and Henderik A. Proper: On the accommodation of conceptual distinctions in conceptual modeling languages. In: H.G. Fill et al. (Eds): *Modellierung 2014, LNI 225*, pp. 17–32, GI, 2014.

Dirk van der Linden, Henderik A. Proper and Stijn J.B.A. Hoppenbrouwers: Conceptual understanding of conceptual modeling concepts: a longitudinal study among students learning to model. In: L. Iliadis, M. Papazoglou, and K. Pohl (Eds.): *CAiSE 2014 Workshops, LNBIP 178*, pp. 213–218. Springer, Heidelberg, 2014

Dirk van der Linden and Henderik A. Proper: Category Structure of Language Types Common to Conceptual Modeling Languages. In: I. Bider et al. (Eds.): *BPMDS 2014 and EMMSAD 2014, LNBIP 175*, pp. 317–331. Springer, Heidelberg, 2014

Dirk van der Linden and Marc van Zee: On the Semantic Feature Structure of Modeling Concepts: An Empirical Study. In: *IEEE 16th Conference on Business Informatics (CBI)*, vol. 2, pp. 158–165. IEEE, 2014

List of Figures

List of Tables

Part I

Foundations

CHAPTER 1

Introduction

Abstract. In the introduction of this dissertation we set out the main subject by way of introducing Enterprise Modeling and some of its inherent challenges. We explain how those translate to the quest of integrating different Enterprise Models. We argue for the need to understand how individual people involved in the modeling process view the language constructs of the modeling languages they use, and relate this to existing work on ontologies. We then discuss more general aspects of conceptual models and concepts themselves necessary to understand our research objectives, and the research questions we will treat in the rest of this dissertation.

The content of this chapter is based on work published as:

Dirk van der Linden, Stijn J.B.A. Hoppenbrouwers, Alina Lartseva, Wolfgang Molnar: Beyond Terminologies: Using Psychometrics to Validate Shared Ontologies. In: *Applied Ontology, vol.7, no.4*, pp. 471-487, 2012

Stijn J.B.A. Hoppenbrouwers, Wim van Stokkum, Maria-Eugenia Iacob, Ilona Wilmont, Dirk van der Linden, Chintan Amrit, and Maarten Joosen: Stakeholder Communication in Service Development. In: M.M. Lankhorst et al. (Eds.): *Agile Service Development – Combining Adaptive Methods and Flexible Solutions*, Springer, Heidelberg, 2012

1.1 Different understandings of modeling concepts

Enterprise Modeling (EM) is the quest to model an enterprise in all of its complexity (e.g., describing its activities, people and their behavior, regulations, processes, resources, goals), and to understand how all those different aspects interrelate and affect each other (cf. Fox and Gruninger 1998, Dalal et al. 2004). Doing so supports the communication between (people working in) these different aspects, and the potential for linking between them in order to optimize the enterprise's overall behavior (Frank 2002). It is done both to describe and understand (i.e., the current state of an enterprise) and to stipulate and steer (i.e., towards the desired state of an enterprise).

The modeling of an enterprise thus typically comprises the modeling of many aspects (e.g., processes, resources, rules), which themselves are typically represented in a specialized modeling language or method (e.g., BPMN (Object Management Group 2010a), e3Value (Gordijn et al. 2006), RBAC (Ferrariolo et al. 1995)). Most of these languages share similar language constructs, or *meta-concepts* (e.g., PROCESSES[1], RESOURCES, RESTRICTIONS). However, from language to language (and modeler to modeler) the way in which meta-concepts are typically used (i.e., their intended semantics) can differ. For example, one modeler might typically intend RESTRICTIONS to be deontic in nature (i.e., restrictions that ought to be the case, but can be violated), while a different modeler might typically consider them as alethic conditions (i.e., rules that are strict logical necessities and cannot be violated). The modelers could also differ in whether they typically interpret RESULTS as being material or immaterial 'things'. Even for scenarios as simple as the delivery of a pizza these differences become apparent, as a pizza delivery can include alethic restrictions in order to observe temporal dependencies ("A pizza *cannot* be delivered before it is made."), deontic restrictions ("A pizza *should* be delivered within 30 minutes of its order."), and the result of the delivery can be a material thing (a certain amount of notes and coins of the local currency) or an immaterial one (a confirmation of payment on a debit card machine).

If one is to integrate or link models describing these individual aspects (i.e., the integrative modeling step in enterprise modeling (Lankhorst 2004, Kühn et al. 2003, Vernadat 2002b, Opdahl and Berio 2006, Delen et al. 2005)) and ensure the consistency and completeness of the involved semantics, it is necessary to be aware of the exact way in which such a meta-concept was used by the modeler. If this is not explicitly taken into account, problems could arise from, e.g., treating superficially similar concepts as being the same or eroding the nuanced view from specific models when they are combined and made (internally) consistent.

This challenge follows from the inherent collaborative nature of enterprise modeling (Ssebuggwawo et al. 2009, Rospocher et al. 2008, Frederiks and van der Weide 2006, Hoppenbrouwers et al. 2006; 2005), as it involves different people specialized in different aspects of the enterprise. These aspects have to be elaborated on to deal with the complexity of (re)designing modern day enterprises, with the added difficulty that it is up to the "people within the business to express their views in terms of a modeling language" (Barjis et al. 2009). Collaborative modeling in general deals with challenges like these (Rouwette et al. 2008, Hoppenbrouwers et al. 2009, Rittgen 2009) that arise because of the different people involved, such as optimizing the actual modeling process (Bidarra et al. 2001), ensuring its effectiveness (Dean et al. 2000) and dealing with problems that arise when integrating models made by different people with different viewpoints (Renger et al. 2008).

The particular EM challenge we are concerned with in this dissertation is the potential of *mismatched understandings between different modelers and stakeholders* (Kaidalova et al. 2012). Note that *mismatched understanding* does not only refer to misunderstandings

[1] To distinguish concepts from words used for them we print concepts in SMALL CAPS.

that the involved parties might be aware of. It explicitly also refers to the (more damaging) misunderstanding that the parties involved might *not* be aware of. People might disagree on what words to use, what they should mean, or use the same words without realizing they talk about different things. When these apparent or hidden disagreements extend to the words used by a modeling language (i.e., the meta-concepts), the produced models themselves might no longer reflect correctly or fully the conceptualization of the individuals involved. As models should be there to support the building of knowledge and its exchange (Stahl 2000), any threat to the validity and integrity of the models is a threat to the knowledge exchange itself. An often used strategy to deal with this is a priori agreeing on or working towards a set of standardized terminology and semantics. However, it is neither safe nor effective to simply assume that such expressed agreements, or even the models themselves, express correctly and completely the way a modeler conceptualizes them given that they are necessarily incomplete overviews of someone's conceptualization (Guarino et al. 1994).

1.2 Enterprise models

In this section we discuss in more detail the different understanding people can have of the language constructs of the models created to deal with their particular aspects. To deal with this challenge of mismatched understandings later on in the EM process we discuss how attaining an insight into the way in which people understand (their) conceptual models could be used to support the integration of (conceptual) models capturing various different aspects of an enterprise, originating from different stakeholders. Creating an overall enterprise model (or architecture) requires a deep understanding of not only of what the words used in models mean to their creators, but also how the various modelers interpret the meta-conceptual constructs they use (i.e., types in the modeling language). We argue this cannot be achieved by assuming *a priori* consensus on the used semantics (as is common practice), but that we need to incorporate methods that let us discover the existing conceptual understanding individual people, and perhaps communities (of particular kinds of modelers, practice, etc.). Thus, we propose that conceptual integration in enterprise modeling, and perhaps also in other fields, would strongly benefit from insights that are more concerned with figuring out how people actually understand their models than with stipulating how they ought to.

1.2.1 Integration of functionally distinct aspects

Holistic models, as discussed before, are needed to deal with the interplay between the various aspects of enterprises that cannot be determined solely from the partial, specialized (aspectual) models. For instance, strategic decisions, the analysis of emergent effects that are irreducible to constituent models and optimizing the interplay between the various aspects, all require an integrated model. Syntactically speaking, the integration of such models (i.e., model transformation) is less of a big issue. However, when it comes to semantics, a correct integration of the model that respects the understanding people have of both the domain concepts and the language constructs used to express those domain concepts (i.e., the meta-concepts) is far less trivial.

Most proposed solutions attempting to deal with this issue of integration in EM for focus on the (semantics of) the modeling languages. They generally either propose modeling languages as such (albeit with a wide scope meant for holistic models, e.g., ArchiMate (Jonkers et al. 2003); UEML (Vernadat 2002a, Opdahl and Berio 2006)) or they are methodologies that facilitate the integration of models (Kühn et al. 2003). However useful they can be, most of these solutions suffer limitations because of their focus on the modeling languages (a stipulated set of integrated definitions) instead of the *modelers* and how they actually interpret and use conceptual constructs. This stipulative focus causes them to be dependent on the explicitly expressed semantics of the language (assuming it is perfectly communicated to users and retained by them). In addition, every time the semantics of any modeling language involved changes, the solution (language, mapping) has to be revised as well, resulting in a never-ending update loop. Furthermore, the focus on the modeling languages becomes even more problematic if those languages have no agreed-upon standard semantics (e.g., i* (Ayala et al. 2005)) or suffer from underspecified or inconsistently defined semantics, e.g., BPMN (Dijkman et al. 2008a)) and UML (Breu et al. 1997). A focus on the modeler would counteract these issues by avoiding the problem of not knowing what (standard) semantics have been selected and to what extent someone adheres to them. Thus for the integration of enterprise models, in addition to the semantics of modeling languages it is important to have insight into the conceptualizations and interpretations of the various individual *modelers*.

1.2.2 Integration requires a focus on people

The need for insight indicated above should provide some of guidance for the process in which the enterprise modeler integrates models. Pre-defined specifications of either the modeling languages as such, or of the aspect-specific concepts used are not sufficient to understand the fine points of modelers' interpretations. Specification documents, whether they are ‘simple’ lists of the terminology used, or more detailed, formal structures (like the often invoked type of ontology suggested by Gruber (1993)) still only talk about pre-existing, agreed-upon meaning. But agreement is not the same as commitment, and agreement may be quite superficial (e.g., because of political reasons, for saving time, because people feel things could still be changed later on). While specification documents can be updated, they will rarely reflect or keep track of personal insights gained by a specific modeler in a specific context. Thus, one can never be entirely sure that someone claiming to understand a certain term by a certain definition actually does so. Rather personal insights are what actually shapes the semantics someone attributes to his models, as people tend to think and work in the semantics of their natural language and environment, not in what is dictated by a given formal specification (Sowa 2010).

Furthermore, the primary focus of communication for a given model in most modeling discussions is centered around the clarification of instance-level concepts (i.e., the concrete ‘things’ domain experts talk about). However, when we integrate models, we look at the semantics of the type-level concepts: the modeling language’s constructs. For example, if we have two models that both speak about users, we do not integrate the ‘John Doe’ and ‘Jane Doe’ instances they contain, but the underlying construct that represents those instances, e.g., ‘Person’, ‘Actor’, ‘User’. For this reason our focus on explicating meaning should be on those constructs (i.e., the *meta-concepts*) and how they are interpreted. Thus, the kind of insight into different modelers required by an integrating enterprise modeler should focus on the personal semantics of the type-level *meta-concepts* they use.

The need to understand the way in which language constructs are used intuitively leads us to some form of linguistic analysis. The simplest, and perhaps most obvious one would be some kind of terminological analysis. However, this involves the very problem we are trying to deal with: we cannot be sure any terminologies (i.e., specifications, agreements, standards) were properly adhered to, so we would have to ask each modeler to explicate their own terminology. Even assuming such a terminology would be conscientiously given, there is still no guarantee it would explicate every part, leaving the issue unsolved.

A better option might be some form of corpus analysis. We can define the corpus as the collection of (textually represented) models and related documents and process them to try and find useful information. Unfortunately, this kind of analysis will not necessarily render the meaning of the words as intended by the modeler: at best we can focus on collocation, figuring out which words are often used together with certain other words. While the idea that words that occur together in a similar context mean similar things (Harris 1954) is not bad *per se*, focussing on the models themselves implies that we really only find information about the co-occurrence of the words that were used, but not necessarily about how the modeler personally understood and intended them to be used.

Such a Firthian "you shall know a word by the company it keeps" (Firth 1957) view only works in our case if we actually relate these words and their company to the modelers and ask them what they actually mean for them. That, however, defeats the point of doing the corpus analysis, since if the lexicon we are interested in (i.e., the notation constructs) is small enough we can directly ask for the meaning of all of them right away. Furthermore, the corpora we can construct on the basis of models and related documents may very likely be too small for any useful analysis, given that we already have a fragmented playing field with each aspect modeling their small part of the total enterprise.

The strongest issue here is, as we said before, the focus on models instead of their modelers. Trying to deduce any detailed semantics from a model is a doubtful enterprise because, as Guarino (1998) puts it:

> "A set of intended models is [...] only a weak characterization of a conceptualization: it just excludes some absurd interpretations, without really describing the "meaning" of the vocabulary."

We cannot reason about things that are not in the model, as obvious as they might seem. If a certain modeler always uses an 'actor' construct to model human beings, e.g., they instantiate it with such terms as 'John Doe' or 'Business User', we know they find it permissible to see human beings as actors. But the fact that they never once instantiated it with terms like 'Sun Blade V100' or 'Dell Server' is simply not enough to conclude they find it *impermissible* to ever do so. Therefore we have to focus on finding the meaning at the source, i.e., in the modeler. For that, we will need a different kind of investigation than a purely linguistic one that bases itself on only partially expressed stipulative 'meaning'. What we need to know is what a construct actually *is* for people, i.e., how they conceptually understand it, and for this investigations into the conceptual understanding (e.g., psychometrical analysis) are needed to provide us with answers where sole terminological analysis cannot.

Focusing on how people understand modeling constructs is a more stable measure of what they mean than textbook definitions, that is, if we measure this kind of conceptual understanding instead of postulating it. We presume cognitive science's suggestion of a tight relationship between thought and the world, but a loose one between language and thought, cf. Wolff and Holmes (2011). What this means is that the language, in relation to the concepts it describes, changes far more rapidly and unpredictably than the actual concepts in relation to the world that it helps categorize and make sense of for a person. Focusing on measuring the conceptual understanding is therefore a safer bet for an understanding of what people mean than attempting to derive it from the language they use.

There is a strong analog to the (applied) ontology community here, as the measuring or explicating of such understandings is essentially the creation of personal ontologies. However, are ontologies not supposed to be a "formal explicit specification of *shared* conceptualization" (Gruber 1993) (emphasis added), instead of the *personal* conceptualizations we advocate? It is quite true that the eventual goal of most ontologies is to reflect and support the shared conceptualization. However, this can rarely be done effectively without first being aware of the personal conceptualizations of the individuals involved. The (applied) ontology community seems to agree that individuals have their own ontologies:

> "Every person (...) has an [often tacit] ontology (...) these ontologies pervade and underpin our deliberations, inform our decisions and guide our actions ..." (Uschold 2011)

Furthermore, it is also understood that these personal ontologies are an important asset in validating and strengthening the quality of (amongst others) shared ontologies:

> "it has become ever more important to make explicit these implicit ontologies thereby easing interoperability and improving operational effectiveness" (Uschold 2011)

It thus seems evident that explicating the conceptual understanding of individual people is a useful venture (cf. other research which explicitly uses personal ontologies to map or construct shared ontologies such as Dieng and Hug (1998) or Lacher and Groh (2001)). Furthermore, following on the argument from Sowa (2000) that "independently developed, but convergent theories that stand the test of time are a more reliable basis for standards than the consensus of a committee", we would argue that having an understanding of people's independently developed conceptualizations, or personal ontologies, may be a more reliable basis for understanding their semantics than 'agreed ontologies'.

1.3 Conceptual models

Because the different models used in Enterprise Modeling (used to capture and describe information of their relative aspects) are essentially conceptual models we will reflect on the structure of conceptual models, what is contained in them, and what in particular the referents of these models are. Since we want to investigate the conceptual understanding people have of the language constructs contained in them, it would do well to know just what we want to measure.

Conceptual modeling deals with the creation of models that describe a particular domain from the point of view of a number of involved stakeholders. At least: since recent times this has been the predominant use, while earlier it was predominantly used for information modeling in the scope of database design, before moving onwards to different aspects like processes, ontologies and other specializations (Wieringa 2008, Henderson-Sellers 2011). While conceptual modeling can also include e.g., models of solutions (software, tools, implementations) and their impact (Wieringa 2011), our focus is on *domain models*, those used to model and capture the information some people (stakeholders, observers, modelers, etc.) have of a particular domain under the spotlight.

Creating such models can be done both to facilitate communication between these people, to capture (current) knowledge of that domain, to reason with those models, and so on (Mylopoulos et al. 1990). In order for such models to be useful it is important they are of a high quality, which entails first and foremost that they are conceptually valid representations of what the stakeholders and modelers intended to represent (Robinson 2006). As modeling in this respect is such a human activity, it is important to be aware of all the personal factors that play into encoding such conceptualizations. For example, while a particular financial domain might have such concepts like 'insurance policy', 'payout regulation', and 'financial means', we need to use modeling concepts – the types from a modeling language – to capture them. For example, by modeling a 'payout regulation' as a specific type of restriction. Because people also have their own conceptualizations of these language types, where most discussions on clarification of terminology center around those from the domain, we need to be extra careful that a model truly represents what is meant by the modelers. Two modelers might for instance differ on their typical view on restrictions, the one seeming them as alethic conditions (i.e., statements of what logically has to be), and the other as deontic conditions (i.e., statements of what optimally should be). If left unspoken, and unresolved during the modeling sessions, this can lead to models that do not fully represent the original modeler's intent anymore.

1.3.1 What is modeled in a conceptual model?

Since we want to understand how people understand their models, especially the language constructs used by them, we need to focus on just what is contained in those models. What is the actual content of a conceptual model (i.e., the referents of the terms used in the model)? While intuitively one would assume that the referents here are *concepts*, hence, *conceptual* modeling, there are many people who seem to think such models refer either to mathematical objects, physical reality, or other like-wise interpretations. With the advent of the use of conceptual modeling shifting towards modeling domains and people's conceptualizations of the things therein, it seems to be valid to assume that concepts are the referent given the high focus on eliciting the understanding people have of some domain. This particular interpretation has been established as early as in the 1980s in fundamental reports that laid the basis for many conceptual modeling efforts (cf. Jardine 1984).

While we now can safely assume that it is concepts that are modeled in conceptual models, there is still a distinction we need to make that has been implicitly referred to so far. Essentially, models will contain at least two levels of concepts:

1. the domain concepts, and
2. the meta-concepts (i.e., the language constructs).

The domain-level concepts are the actual things from the 'Universe-of-Discourse' that are modeled (e.g., financial policies, types of money, the people involved in insurance payout claims). The language-level, or *meta-concepts*, however, are the actual concepts from the language that are used to express these concepts.

1.3.2 The distinction between concepts and meta-concepts

Let us explain the difference between domain concepts and meta-concepts a bit more clearly with some examples of how models are created.

To ensure the validity of domain models modelers often already attempt to reach consensus on what meta-concepts to use to represent elements from the universe of discourse, and what to call those representations. However, such discussions often take the form of agreeing on the concrete language construct to be used for such meta-concepts (e.g, "Let us model this pizza as a RESOURCE.", "Let us model the delivery of the pizza as a PROCESS."), while not necessarily going into detail on just what that construct ('resource' or 'process') actually *is* in this context. Instead, the deeper semantics often stay implicit. For example, the business process for pizza deliveries might involve restrictions that are logical necessities and cannot be broken (e.g., a pizza can not be delivered before it is made, thus correctly observing temporal dependencies) but also restrictions that can be broken, but perhaps *ought* not to because of moral or financial constraints (e.g., a pizza should be delivered within 30 minutes of its order). It is thus important that whatever meta-concept we choose to represent these restrictions accommodate the conceptual distinction between alethic (i.e., logical *necessity*) and deontic (i.e., how the world *ought* to be) modality. But, if we only agree that we model these kinds of restrictions as "rules", we leave this distinction implicit and depend on the (selection of the) modeling language to dictate what exactly the semantic status of both of these rules can be (let alone that users at times deliberately or unconsciously ignore a language's semantics and invent their own, not always explicit, semantics (Henderson-Sellers 2005)). As some languages accommodate less conceptual distinctions than others, that means relevant information might be lost.

Let us consider actors. If a modeler typically conceptualizes actors as human beings who take actions on their own accord, and wishes to model them so, it is necessary for whatever meta-concept is used from the modeling language to support the relevant parts of that conceptualization. In this case, making it explicit that any instantiation of an actor needs to be a single human being, and more specifically, a (in the given context) autonomous human being, for instance when modeling social decision making processes in an enterprise. If the produced model, however, is not explicit about these distinctions, there can be a host of conceptually related, but invalid, instantiations (e.g., an employee who needs permission for every action and can thus not be considered autonomous, a department with multiple persons). Such problems can be prevented by either using a modeling language that accommodates the necessary conceptual distinction, or by explicitly modeling such distinctions manually (e.g., by using unary constraints on the language construct for each necessary semantic feature). However, being aware of the relevant and necessary conceptual distinctions *before* the model creation process is often still lacking.

Thus, in order to ensure that a model can clearly and completely communicate its intended meaning, regardless of who is interpreting it at what time, a certain degree of conceptual alignment is needed. Specifically, an alignment between the conceptualization the original modeler has of their meta-concepts (i.e., the types used by a modeling language to represent the domain concepts) and the 'official', used, semantics of those modeling concepts (i.e., the semantics as found in the specification of the used language) makes it easier to ensure that the modeling language (or dialect thereof) which will be used accommodates the needed conceptual distinctions. To do so it is thus necessary to understand how the modelers (and stakeholders) involved in the modeling process understand the meta-concepts. Such understandings can be of value to the modeling process by e.g., providing model facilitators with input for the facilitation process (Rosemann et al. 2011), or by providing explicits focus points for annotations in the model, and could help enhance the quality of created models and improve consensus amongst stakeholders (Stirna et al. 2007). Or, as Wieringa (2008) put it: "The meaning of a notation must be appropriate for the intended use of the notation".

In our context of conceptual modeling the difference between a concept and a meta-concept is thus the level of abstraction and domain-specificity at which they are used, concepts being those things specific to a domain (e.g., combinatorial rules set for pizza toppings to ensure flavors complement each other), whereas meta-concepts are the more abstract, and general concepts used to reason about such concepts (i.e., the general notion of RESTRICTIONS). The conceptual understanding a person has of the meta-concept RESTRICTION here thus likely affects the way they will reason about the concept of PIZZA TOPPING COMBINATORIAL RULES, for instance by assuming that rules are typically of a deontic nature and *can* be broken, even if they should not be.

1.4 Concepts

Before we continue discussing why two modelers can differ so significantly on their understanding of the same concept we should more clearly define what we understand under the term 'concept', as even the field of conceptual modeling itself is often vague and leaves implicit just what concepts are. There is a wealth of discussion on concepts, their nature, whether they are useful in the first place, and why everybody (else) is wrong about them (cf. Fodor 1998, Laurence and Margolis 1999, Rosch 1999). Given that we do need to deal with them, whatever they are, we will take as a basis (fully understanding the irony) a definition given arguing for the abolishment of concepts by Machery (2009):

> "A concept of x is a body of knowledge about x that is stored in long-term memory and that is used by default in the processes underlying most, if not all, higher cognitive competencies when these processes result in judgments about x."

The concepts we have of things are thus strongly individual, being the main component of the way we make category judgments (i.e., categorize and judge everything around us from physical objects to the conceptual and the emotional). This strongly emphasizes the human side, as the categorization is primarily influenced by the personal experiences a person has in the world. The body of knowledge that we gather for each concept is derived from our interactions in the real world, as Geeraerts (2010) put it that:

> "...semantic memory for concepts is based on a subject's memories of past experiences with instances of those concepts."

Such judgment processes are thus influenced by the personal experiences we have in the world, and in their own turn again influence the judgment processes as they become participatory in reality (Gabora et al. 2008). This re-iterates the need to study how people understand the concepts in their models, as the different personal experiences that each modeler and stakeholder had forms and shapes the way they will interpret it, subtly or less so changing the meaning of its content for them.

1.4.1 On the structure and representation of concepts

Fundamental to any means of measuring the conceptual understanding people have of some particular concepts should be the question of the structure of those concepts. As we are dealing with categorization by human beings, we do not adhere to the discrete structure used so much in computing science, but instead to a graded structure, exemplified by the prototype or exemplar theory of concept structure (cf. Rosch 1973, Rosch et al. 1976, Ashby and Maddox 1993, Storms et al. 2000).

According to these theories, concepts are not simply judged in absolutes, but in a graded way. A tomato is *kind of* a (good example of a) fruit, just like a particular instantiation of some model is *kind of* a (good example of) that model. These are all judged in a graded way, with some things simply being better than others. The main difference between prototype and exemplar theory (less important to our efforts) has to do with the actual cognitive process for the judgments: while for prototypes a particular thing in the world is compared to an abstract, idealized set of features and information (the prototype), while according the exemplar theory it is compared to exemplary (previously experienced) instances of similar things. It is important to note that, regardless of whether one accepts prototype or exemplar theory, the judgments made are of a graded nature. Thus, they are not like the discrete in-or-out judgments so often used in computing science.

While most initial research into categorization and the theories derived thereof focused on physical objects (e.g., tables, chairs, utensils), later research has extended this to abstract concepts (e.g., the logical concept of if-then, love) and events (e.g., the act of cutting something, the act of handing something to someone) and showed they are categorized distinctly as well (cf. Adelson 1985, Majid et al. 2004, Malt et al. 2008). Thus, it should be clear that our concepts work on all these potential 'things', and thus also the abstract notions that we often deal with in conceptual models (e.g., processes, values). With this in mind we can study the personal semantics of common concepts in Enterprise Modeling notations without problem, physical objects (computers, human beings), abstractions (computational objects, organizational departments) and events (processes, transfers, exchanges) alike.

One final aspect that we would do well to keep in mind if we eventually attempt to generalize any findings on the conceptual understanding people have of concepts is something that prototype theory research has eschewed so far according to Geeraerts (2010), dealing with the distribution of knowledge that people have of concepts over linguistic communities. As Geeraerts (2010) stated:

> "...linguistic communities are not necessarily homogenous – semantic knowledge may be unevenly distributed over the members of the speech community.

> Prototype-theoretical studies generally tend to ignore the question of whether and to what extent the prototypical model of category structure might be plausibly interpreted as involving social variation over individuals rather than just psychological variations over contexts of use."

Thus, if we measure the conceptual understanding modelers and stakeholders have of the concepts in their models we should also keep in mind that it is quite likely these conceptualizations are different from person to person, and the *amount* of information in these conceptualizations (e.g., how detailed they are) can vary from person to person, even in a related group of modelers.

1.4.2 Characterizing concepts through dimensions

As a last point, we need to treat the idea of 'measuring' concepts, given our desire to figure out (likely measuring) how people understand the concepts in their models. We cannot fully measure the experience a person has with a concept, as the information related to such concepts available to a person is too wide and diverse to fully capture. The problem of having to communicate about these concepts with some signs, often words is that, as Malt et al. (2011) said:

> "Words can help identify the conceptual space of a domain, but they do not directly reveal bounded units of knowledge that can be labeled concepts"

Instead, the right approach is to characterize these understandings. A discriminatory *dimension* can help us narrow down the conceptual space of something we are attempting to characterize. Combinations of such dimensions narrow it down more and more until we get to some useful characterization that is specific enough to say something useful about how people interpret or experience the concept.

This approach is quite similar to arguing via connotations: say we wish to understand someone's concept of some fruit. We might provide them with words and ask to what degree they find them to have a connotation with the fruit, like 'yellow', 'is long and curved', 'tastes sweet', 'does not taste good', 'is expensive', 'goes well with chili sin carne'. All those different dimensions add something to our understanding of how that person sees this fruit: the dimensions of 'yellow' and 'is long and curved' hinting at the structure and composition (i.e., it likely being a banana), while dimensions like 'tastes sweet', and perhaps disagreement with 'does not taste good' hinting at their subjective attitude towards it (i.e., liking bananas or not). Oddly specific connotations like 'goes well with chili sin carne', combined with other knowledge gained from the earlier dimensions (i.e., knowing we are talking about bananas and the person likes it) could then even be used to tell us something about the context of the person answering, and identify the utterer of the phrase.

Back in the context of conceptual modeling and our research efforts, it thus becomes clear that if we can find a number of dimensions for the concepts we wish to investigate then we can characterize people's understanding of them equally as well.

1.5 Research objectives

We have discussed the difficulty that enterprise modeling faces while attempting to integrate the different models describing different domains, while also ensuring that the integrated model remains a valid representation of all the intended semantics. Moving on, we can now more clearly describe what we want to contribute to this issue. It is not our intention to stipulate some particular way of working, some way of integrating, or to create a tool or language that will solve this "if only all involved people would use it". Too often research stipulates such solutions, without necessarily validating them in practice (Wieringa and Heerkens 2006), let alone achieving any adoption by practice (cf. Kaindl et al. 2002, Davies et al. 2006, Regev 2014). Our focus will be on understanding the differences between modelers, particularly the different ways in which they understand conceptual modeling meta-concepts. We will investigate these understandings to focus on knowledge questions, which will help us achieve insights and generate data which can be used to more clearly understand, and guide modeling processes, while not stipulating and locking down every aspect of how those processes should be done.

1.5.1 Problem statement

Put briefly, the problem definition we work with is:

> The creation of an *a posteriori* integrated conceptual model which describes the different aspects of an enterprise is made more difficult by the different interpretations people working within each aspect have, but not necessarily explicate, of the meta-concepts (i.e., language constructs) they use to create conceptual models.

1.5.2 Research questions

To address the given problem and do so by achieving our objective of gaining a more fundamental insight into the conceptual understanding modelers have of conceptual modeling meta-concepts, we will concretely treat the following research questions. In Section 2.1.1 a more detailed exposition is given of the approaches to be used for the individual questions.

1. What are the meta-concepts (i.e., language constructs) shared between modeling languages used in enterprise modeling?
2. What is the conceptual understanding that people have of these meta-concepts?
   (a) To what extent do the categories for these meta-concepts more resemble a graded or discrete structure (e.g., artifactual or natural)?
   (b) To what extent do conceptual understandings differ on key points between different people (e.g., different sets of features, typicality)?
3. Does a person's conceptual understanding of a meta-concept change over time?
   (a) If so, does training in particular languages or techniques affect these base conceptual understandings?
4. To what extent do modeling languages allow for people to express their particular conceptual understanding of meta-concepts?
   (a) Do modeling languages make a distinction between explicitly or implicitly allowing them to do so?
   (b) Can particular (kinds of) conceptual understandings of certain meta-concepts be used to infer which modeling language or notation is best suited for a specific modeler?
5. Can particular (kinds of) conceptual understandings of certain meta-concepts be attributed to a community (of discourse, practice, etc.)?
   (a) Can this be used to predict a person's typical conceptual understanding of a meta-concept by finding their community membership?

1.5.3 Thesis contributions

By focusing on the research questions set out above, this dissertation contributes to the fundamental understanding of how people conceptualize the meta-concepts they use during conceptual modeling. Specifically, it gives insights into the different personal semantics people have in terms of how they make category judgments about them, attribute features to them, and what they see them as. It furthermore shows how this information can be used for practical benefits. Summarized, this thesis will make the following contributions:

- It derives a number of common meta-concepts shared between many modeling languages used in enterprise modeling.
- It demonstrates that the permissible understandings of these meta-concepts are quite flexible (i.e., their category membership is judged in a graded way), thus making the question of 'is x a valid y' a more difficult question to answer during model integration, especially when compared with situations where (personal) semantics are simply stipulated.
- It demonstrates feature sets which conceptual modelers (and those learning to model) attribute to these meta-concepts. These (combinations of) features can be used to identify how a meta-concept is typically seen, and what aspects are particularly salient.
- It demonstrates the different kinds of conceptual understandings that people have of the meta-concepts and how these understandings can affect the range of conceptually valid (model) instantiations. Furthermore, it demonstrates the way in which these understandings change over time, and how little correlation there seems to be with educational stimuli such as the learning of a new (modeling) language or technique.
- It shows that some specific conceptual understandings people have of these meta-concepts are not always accommodated by modeling languages, and can thus not be modeled explicitly, even though they may be vital to understanding how these people conceptualize the domain.
- It shows the common assumption held in practice that those who share certain superficial professional similarities think alike about modeling language constructs (i.e., the meta-concepts) cannot be easily backed up by empirical investigation of the personal semantics of such people.

1.6 What has been done already?

It would do well to also have some understanding of recent and ongoing research efforts that are practically related to our own investigations, and what has been done already by others. While our focus is strongly centered around answering knowledge questions and finding insights into the way people understand conceptual models, other work focusing on support (in the form of tools, formalizations, data structures, etc.) for capturing such information exists to some degree. For example, there are proposals for ontology formalizations that include the encoding of information about the typicality of a certain entity, which is information that can be elicited from investigations like the ones we perform. Furthermore, there is a strong link between our work and research efforts on (conceptual) model(ing) quality, which we will discuss, and clearly explain to what aspects of model quality our work might contribute.

Capturing personal semantics

Our research as reported here is centered around the need to understand how people think about certain things (i.e., meta-concepts). In the field of ontology research, in particular in *applied* ontology, this ties into the need for validity checking of created ontologies. Such validity checking is done to ensure that the semantic content of integrated models is consistent across individual models. In the field of ontology this is, for example, done by comparing whether personal ontologies (e.g., based on data elicited in individual interviews) are consistent with a shared ontology people have agreed on (cf. Almeida 2009). This is regardless of what expert opinion or official standards (e.g., ISO documents or OMG specifications) stipulate what the 'correct' semantic content should be. Doing so establishes a greater interoperability amongst all these models, which has been argued to be one of the most important benefits of using ontology (Neuhaus et al. 2011). For such efforts it is important that one can investigate and elicit detailed information about how people think about and understand some specific entities in order to have an empirical basis in which to ground for example these personal ontologies.

There have also been concrete efforts to create data structures and formalizations which can capture information relevant to such personal ontologies. For example, Liao et al. (2006) created a recommendation system for library purposes, which was based on corpus analysis of reference material. It uses this information to build personal ontologies by extracting relevant details from borrowing records. While based on corpus analysis and not specifically focused on capturing how people understand something, it does show a move towards capturing more personal information. Other efforts started to include more details relevant to this, such as for instance the inclusion of a distance metric which captures the semantic distance (or an approximiation thereof) between concepts in such personal ontologies (Yang and Callan 2008). Other efforts allow people to create ontologies in a more personal fashion by starting out from reference material and choosing data that fit their interpretation of some domain (Katifori et al. 2008), the result of which is used to support people in storing their personal preferences (Katifori et al. 2005). Personalization of existing ontologies has received more attention subsequently, with work by (e.g., Aimé et al. 2008; 2009) detailing their efforts of including a prototypicality gradient for both the conceptual (i.e., the elements in the ontology) and lexical (i.e., the words used for the elements) understanding people have of their domain. In order to capture this information, they created a formal structure supporting such information, in this case an implementation supporting different degrees of typicality. Other implementations of ontologies supporting the explicit modeling of prototypicality are for example those created by Yeung and Leung (2006; 2010), which encode likeliness and typicality, and explicitly deal with the different graded ways in which elements are valid examples of a particular concept. Their implementation has subsequently been critiqued and extended by Cai and Leung (2008), this time explicitly incorporating relationships to ensure one can reason about and infer knowledge between concepts in the modeled ontology.

The main difference between our focus and this existing research on personal ontologies is that they tend to focus on either *a priori* defining (formal) structures for such personal ontologies or using data mining (i.e., corpus analysis) techniques to acquire content. Our focus however, is on the empirical study of the content one would capture in such ontologies: the personal semantics people have of these captured things. The benefit of having such existing formalizations out there already is that they can subsequently be used to capture the potential different conceptual understandings and aspects related to them (e.g., different categorization judgments, typicalities) that we will investigate. The use of such formalizations does presume that it is known what the structure of particular categories or concepts is, and what the individual differences for their members are. It is necessary to have some quantification of the degree of difference between e.g., feature typicality, or how similar individual members are to each other before encoding them with these structures. If this information is not known, or worse, if it is not known whether this information is particularly relevant (i.e., in the case of strongly discrete categories), the use of such formalizations in a way that corresponds to the real way that people think is much more difficult. Furthermore, we should be clear that these existing efforts are often created with a general modeling focus, and not specifically targeted at supporting enterprise modeling, or other forms of (a posteriori) model integration.

Model quality

While our work is not directly aimed at analyzing or improving the quality of models or modeling languages, there are some strong links between ongoing research efforts and the contributions we make. Model quality is a large topic, with many directions that all receive attention, often doing so along the traditional lines of syntax, semantics, and pragmatics (cf. Lindland et al. 1994). Such proposals are created in order to have a more systematic way of reasoning about model quality instead of just listing sets of desirable properties for models. Extended work incorporating more semiotics-inspired lines such as physical, empirical and social can be found in efforts like the FRISCO reports and (e.g., Krogstie et al. 1995). Nonetheless, there seems to be little agreement on which quality framework (and thus way of looking at quality) is the best, as there is a large number of different frameworks available, all using their own specific terminology, with their own focus, resulting in a lack of adoption by practice and no truly standard ways of looking at model quality (Moody 2005).

Regardless of a specific framework for determining model quality, our work ties in with many of them as they often include pragmatic quality. Our focus on understanding how people view the language constructs of the modeling languages they use is closely related to this. In Lindland et al. (1994)'s original framework, model quality is evaluated by comparing different sets of statements, which are the **M**odel, **L**anguage, **D**omain, and audience **I**nterpretation. The audience interpretation refers to the statements that people involved in the modeling process (e.g., modelers and stakeholders) view the model as being composed of. Pragmatic quality is then given as the degree of correspondence between model and audience interpretation, ensuring that a model is qualitatively good when the audience interprets it as it is actually composed (e.g., the intended semantics are the same as the captured semantics). Another aspect of quality found in multiple frameworks, which is strongly related to our work is *perceived* semantic quality. Similar to our focus on *personal* semantics, research such as (Poels et al. 2005) focuses on ensuring perceived semantic quality of models by focusing on the way a model user perceives the degree of correspondence between a model and the domain its represents. Krogstie et al. (1995) defines the perceived semantic validity to have two aspects, namely the perceived validity of a model and its perceived completeness.

In order to determine these aspects, it is important to understand how the people involved in the modeling process view both the domain and the language they use to model it. In order to ensure that people's conceptualization of a domain is the same as the model that was eventually produced, insights like those we aim to generate about the way that people understand a modeling language's construct can be of value. The way in which they understand these language constructs constricts the amount of valid interpretations (or instantiations) for a given model, which has a direct effect on whether the created model can correctly represent the domain, and whether it can do so for all the elements important to the involved people (i.e., ensure also the perceived completeness).

Recent work by Fettke et al. (2012) more explicitly explores the understandability of models, which has received less attention as a goal on its own. Instead of stipulating a framework or set of criteria that define what 'quality' is for a model, Fettke et al.

(2012) look at it as something that is defined in discussions between the people involved in the modeling process, focusing specifically on what understandability means to those people. They propose to analyze such discussions for particular modeling sessions in order to figure out which quality aspects are relevant for them. Doing so leads both to a deeper understanding of what aspects are important for people involved in a specific modeling effort, as well to the general insight that people (and projects) all have different understandings of what comprehensibility means practically, as well as what the quality of a model is. These research efforts thus also come to the same conclusion as our work: that it is important to gain insights into how people think and conceptualize their domains.

The quality of the modeling languages we use to create models has also received attention, e.g., in work by Moody (2010) on a general theory of analyzing model language notations. This approach focuses strongly on the cognitive aspect, analyzing whether the visual symbols used clearly convey their meaning, are not ambiguous, whether visual metaphors are used, and so on. With this framework, multiple modeling languages have been analyzed and critiqued (e.g., Moody and van Hillegersberg 2009, Moody et al. 2010). Our work touches upon this as we also look at to what degree modeling languages accommodate the explicit capturing of conceptual distinctions (i.e., different conceptual understandings). If there is a semantically different understanding of a similar meta-concept, say the language construct used to model acting elements in a particular language, it would do well that they are represented by different visual elements. For example, distinguishing between human and non-human actors by using stick puppet character metaphors for the former. Our investigation into the plethora of different conceptual understanding people have of modeling meta-concepts gives an overview of what such distinctions exist, and thus provides valuable input for improving modeling languages in order to ensure their visual notations is discriminative enough.

Other less directly related efforts are for example focusing on people's individual differences using cognitive mapping techniques to ensure quality of conceptual modeling (Siau and Tan 2005), which is in line with our desire to generate datasets which can be used to more easily figure out how to make modelers communicate about their conceptual understandings. Much work exists that uses ontologies to ground existing modeling languages in reference ontologies (cf. Fettke and Loos 2003, Anaya et al. 2007), although work such as that of Clarke et al. (2013) also explores more the logical angle instead of ontological. This focuses not only on ensuring that concepts in models map to a properly ontologically grounded entity, but that the grammar of the modeling languages can properly represent all those concepts and their relations.

Understanding how people model

While we are focused on understanding the conceptualizations people have of the language constructs they use to model, understanding just how they actually model is an important aspect as well. Understanding the different ways people model and thus capture their knowledge of some domain in a model helps in understanding what is captured in those models, and how much (of the domain, of the modeler's total knowledge) is captured. There have been many studies into just what modeling is, most agreeing on it being some form of abstraction of knowledge and the explicit representation of that knowledge in some form (cf. Hoppenbrouwers et al. 2005). A particularly interesting work by Frederiks and van der Weide (2006) included a detailed analysis of the competencies required of people involved in the modeling process, and for what those competencies are necessary.

According to them, success of a modeling effort has to do with how well domain experts provide descriptions of the domain, and validate (paraphrased) descriptions thereof, and how well analysts afterwards map discussions onto concepts of models and evaluate validations of these models. One of the more poignant points related to our work is the neccesity that "*... analysts can handle implicit knowledge*". While the elicitation of knowledge is based on collecting significant information from the application domain, verbalizing it in a common language shared and understood by those involved, and then validating it, not all information is necessarily extracted. Thus, it is important to stay aware, and deal with potential consequences of such implicit knowledge. Our work and its contributions can be useful here, as the methods we use to elicit personal understandings can be used to make such implicit knowledge explicit. In other cases, particular conceptual understandings that have been found to be problematic for modeling sessions (e.g., people often having different, disagreeing understandings of what constitutes a concept like RESTRICTION) can guide a modeler or facilitator to steer discussions and ensure that all the potentially problematic areas have been thoroughly discussed and explored. In doing so, less knowledge is hopefully left implicit.

Other recent studies by e.g., Wilmont et al. (2010; 2013) were aimed at gaining more insight into the different cognitive styles people have while modeling. They focused specifically on understanding the different ways in which people abstract information to more general forms (Wilmont et al. 2012), and the different innate skills people have for such abstraction based modeling. This work is especially useful given the specific nature of abstraction in computer science, which is focused on *hiding* information instead of ignoring it (Colburn and Shute 2007). In turn, then, adapting modeling methods to more closely fit to how people realistically model has been attempted in some cases, such as for example with exemplary process modeling, where models are created based on best examples of processes instead of their abstract representations (Hofer 2011).

1.7 Thesis structure

Chapter 2 discusses the research approach we take in this work, and gives an overview of the different empirical methods that will be employed. In Chapter 3 we introduce the main meta-concepts and dimensions that will be used for studying the personal semantics people have. It gives an overview of the derivation of these meta-concepts and dimensions from an analysis of existing modeling languages and discusses the choices made therein.

The following couple of chapters then use these meta-concepts and dimensions to study in a more fundamental way the personal semantics people have of them. Chapter 4 uses the meta-concepts and the modeling terminology used for them as input for a category structure study, in which we determine whether membership judgments made for these meta-concepts are of a graded or discrete nature. A more detailed overview of the exact terminology that received many graded judgments is shown, and the implications on conceptual validity of models containing meta-concepts typically receiving graded judgments is discussed. Chapter 5 continues with the fundamental work by showing the results from a feature listing study and discussing the features most typical for the meta-concepts. It goes into the different kinds of features we elicited, and how (combinations of) these features provide ways of identifying what the meta-concept is for someone. Chapter 6 finally showcases two empirical studies into the conceptual understanding of both enterprise modeling practitioners, and computing and information systems students. The typical way in which they understand the meta-concepts is discussed in detail, practitioners are contrasted to students, and effects of the particular understanding that people have on the conceptual validity of a model is discussed.

This is followed by a more practical part of this thesis. Chapter 7 uses the meta-concepts and dimensions to analyze what combinations of meta-concept and dimensions are (not) accommodated in existing conceptual modeling languages used for modeling (different aspects of) enterprises. The implications of the lack of such accommodation, or languages implicitly encoding information in a specific way are discussed, and further implications on the way we choose languages, and actually model are reflected upon. Chapter 8 touches on the need for generalization by showing a study attempting to verify whether it is feasible to assume that people grouped together for some modeling effort based on superficial professional properties (e.g., similar backgrounds, shared knowledge of specific languages) will display similar personal semantics. We discuss the difficulty of properly identifying groups of people as having such personal semantics.

Finally, in Chapter 9 we conclude this thesis by reflecting on all the individual research questions and summarizing their answers. Implications for research and practice are discussed, and options for further research are explored.

CHAPTER 2

Research Methods

Abstract. In this chapter we outline the research approach we have used in the course of this thesis. We state our focus on investigating knowledge questions and performing research that aids in the creation of explanatory theories. For each of the research questions we detail how we will answer them, and elaborate on to what degree (and in which order) this should be, and actually was, done. Finally, we detail the empirical methods that will be used for the empirical studies eliciting the different kinds of data (categorization judgments, feature structure, and conceptual understandings).

The content of this chapter is based on work published as:

Dirk van der Linden, Stijn J.B.A. Hoppenbrouwers, Alina Lartseva, Wolfgang Molnar: Beyond Terminologies: Using Psychometrics to Validate Shared Ontologies. In: *Applied Ontology, vol.7, no.4*, pp. 471-487, 2012

Dirk van der Linden and Henderik A. Proper: Category Structure of Language Types Common to Conceptual Modeling Languages. In: I. Bider et al. (Eds.): *BPMDS 2014 and EMMSAD 2014, LNBIP 175*, pp. 317–331. Springer, Heidelberg, 2014

Dirk van der Linden and Marc van Zee: On the Semantic Feature Structure of Modeling Concepts: An Empirical Study. In: *IEEE 16th Conference on Business Informatics (CBI)*, vol. 2, pp. 158–165. IEEE, 2014

2.1 Research approach

As our focus is predominantly on answering knowledge questions (thus working towards explanatory theories in terms of Gregor (2006)'s classification of IS theories), we wish to measure and present data on how people differ in their personal semantics as objectively as possible. This is particularly necessary as the datasets generated by such studies can, and should be used by other researchers to both compare our findings and run further analyses for other research interests. Thus, the results should be as free as possible from interpretation from our side. Furthermore, we would like to attempt a generalization of findings, and have enough data to be able to compare people to each other, so that we can give pointers to trends or particularly typical ways people interpret meta-concepts.

For this reason we will eschew more interpretive approaches that work with the people themselves and understand why they do the things they do. Instead we will adopt a positivist approach, using traditionally dominant research methods like survey design and laboratory experiments (whether innate or adopted from other fields). This is quite common in the field of Information Systems, as it is becoming more and more interested in obtaining scientific knowledge in real world settings (Chen and Hirschheim 2004). We are thus, in terms of Chen and Hirschheim (2004)'s study and classification of IS research focused on deductive, hypothetical reasoning and generalization, while using empirical (often quantitative) methods, both of a longitudinal and cross-sectional nature depending on the individual study's requirements.

2.1.1 How to answer the research questions

Focus-wise there are two main parts to our work. Both of these parts have a number of related research questions (as set out in Section 1.5.2), which will be answered through various studies, utilizing different methods. They are:

1. the fundamental part, in which we determine the basic meta-concepts to investigate, and perform the empirical studies aimed at characterizing their conceptual structure; and

2. the practical part, in which we use the prior findings and their explanations in order to strengthen the theoretical backing of more practical aspects (i.e., modeling language selection, education of modelers).

For the fundamental part, the following questions will be investigated.

What are the meta-concepts shared between modeling languages used in enterprise modeling? This research question will be answered by the analysis of a number of modeling languages and methods used for different aspects of enterprise models, resulting in a list of categorized constructs and dimensions on which they differ between different languages and methods.

To what extent do the categories for these meta-concepts more resemble a graded or discrete structure? After having found out just what meta-concepts there are, research question 2a will be answered through a term categorization study performed among practitioners and students. The method used here is detailed in Section 2.2.1.

What is the conceptual understanding that people have of these meta-concepts? For the research questions to do with the (measurement of) the conceptual understanding people have of the meta-concepts, there are multiple things to investigate. These will be answered by a combination of longitudinal and cross-sectional studies employing semantic differentials, which are created based on the result of the analysis of modeling languages and methods for the first research question. The semantic differential we use for this is elaborated in more detail in Section 2.2.3.

To what extent do conceptual understandings differ on key points between different people?, *Does a person's conceptual understanding of a meta-concept change over time?* and *If so, does training in particular languages or techniques affect these base conceptual understandings?* These questions will be answered with a cross-section and longitudinal study, analysis of the data and comparison of datasets. Furthermore, a study into the feature structure of the meta-concepts will be performed, which adds to the ability to reason about the different ways people perceive the meta-concepts. The background of this feature elicitation is elaborated in Section 2.2.2.

The practical part is concerned with the final two research questions.

Do modeling languages allow for people to express their particular conceptual understanding of meta-concepts? and related subquestions (see Section 1.5.2). In order to answer this question we will use the data resulting from the modeling language analysis performed for research question 1, and the data from the semantic differential studies in order to perform a targeted analysis of what meta-concept and dimension combinations are supported, both implicitly and explicitly in the specification of a modeling language's semantics.

Can conceptual understandings be attributed to a community? and related subquestions (see Section 1.5.2). This final research question will be answered by further data analysis performed on the results from the semantic differential studies used for the earlier research questions. It will use clustering and component analysis in order to investigate similarities between the semantic differential results, cluster similar results, and then compare the found clusters to real-world properties of the people that were clustered.

The research this dissertation is based on aims to investigate knowledge questions in order to understand phenomena. Thus, the potential act of implementation (e.g., development of methods, tool support) is out of our scope. We will, however, use the data in order to generate insights into the way of working, and thinking, and in doing so reflect on the knowledge and synthesize it into useful insights. Implementation will remain out of the scope as it requires a significant effort on the validation of tool development that does not fit with the timeframe set for our approach.

2.1.2 Order

Certain research questions should logically be investigated before other angles are pursued. For example, investigating the feature structure of these meta-concepts, and attempting to determine their ranked structure hinges on the assumption that the category structure of such a meta-concept is of a graded nature. Thus, investigating whether the meta-concepts are of a graded or discrete nature should be done first. The order we have practically performed the different studies in is shown in Fig. 2.1. As can be seen, we were not always able to have a strict hierarchical order of studies, as depending on the availability of participants, data, and time constraints, many studies were run concurrently.

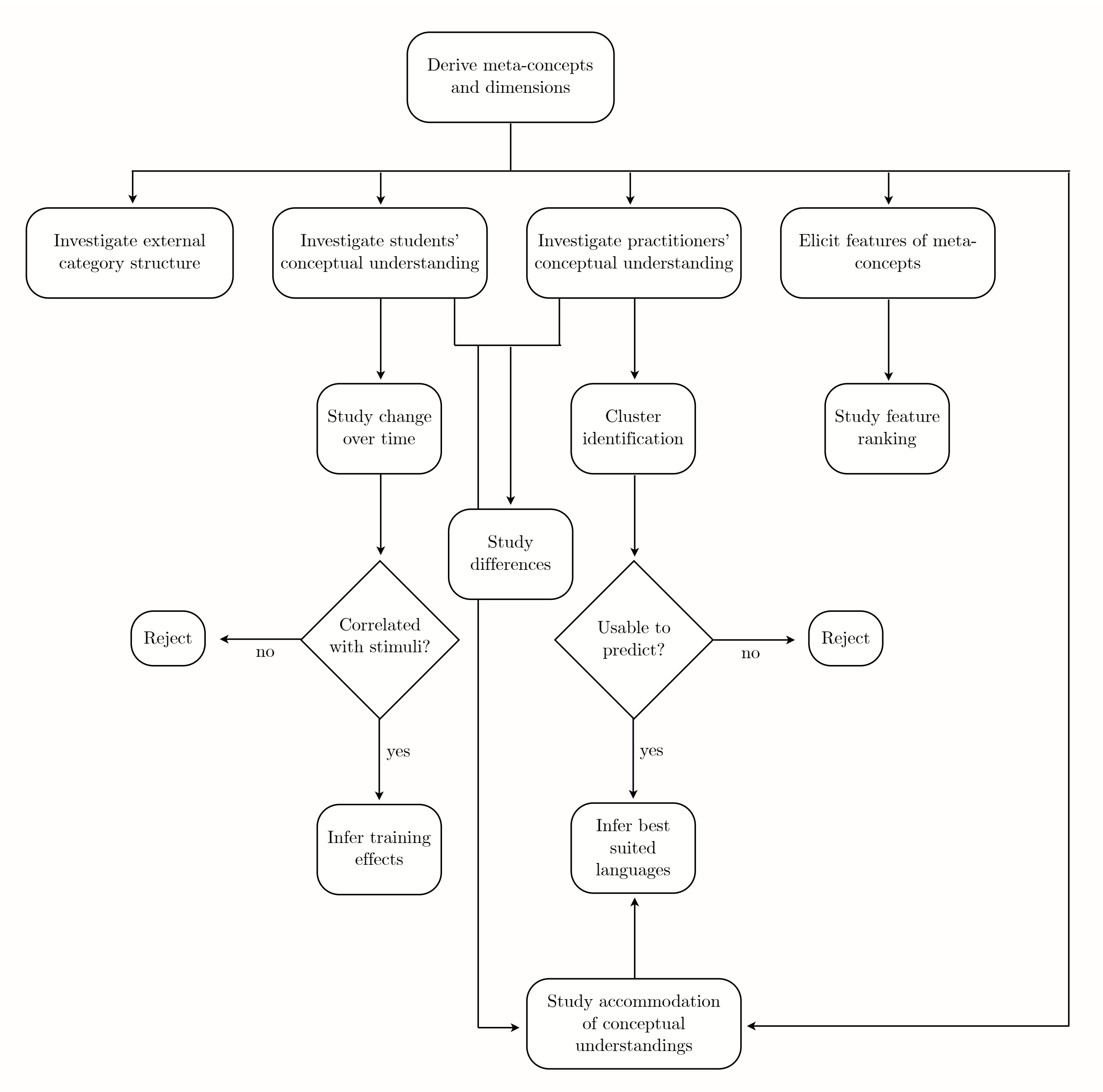


Figure 2.1: The structure of the research reported on in this dissertation.

2.2 Empirical methods

In order for the way we gather our data to have a basis in well grounded research, and to avoid the need to invent methods on the fly, we will use existing empirical methods from other fields (e.g., psychology, linguistics) and their related theories for our empirical work.

2.2.1 Category structure

Investigating the different ways in which people categorize everything from physical objects to events to emotions has been around for a long time. Such experiments reveal information about the category structure, specifically the occurrence or absence of graded category judgments.

The basic way of measuring category structure in terms of discrete or graded category judgments goes back to Barr and Caplan (1987). They used a numerical scale (1–7, with 1 meaning non-member, and 7 meaning full member) in order to investigate to what degree certain terms were considered to be members of a given category, and performed this for a large amount of basic categories (e.g., fruits, tools, vehicles) and terms. In their data analysis they assumed that responses that were neither 1 or 7 implied a graded category judgment (i.e., assuming x is graded if $1 < x < 7$). One can then also infer relative gradedness, or positioning of members by looking at the exact scores, comparing how far away from either negative or positive pole they are. From the many categories (Barr and Caplan 1987) investigated they found differences in the amount of graded, or partial member category judgments, with artifactual categories (i.e., categories reflecting non-natural, man-made 'things' such as tools or vehicles) displaying a higher proportion than natural categories.

Many studies followed using this approach, (e.g., Diesendruck and Gelman 1999, Kalish 1995). Most recently, Estes (2003) proposed an adaptation of Barr and Caplan (1987)'s method, making the measurement of graded category judgments more explicit. Their reasoning was to do away with the numerical scoring, and instead simply ask participants whether a certain term was to be considered as a full, partial or non-member (with the meaning of each of these options explained to participants beforehand). They also introduced other methods for testing gradedness which involve eliciting category judgments for two items presented at the same time, which were found to effectively show graded effects for specific categories (e.g., by showing two slightly dissimilar items like 'tomato' and 'apple' and enquiring at the same time what their membership status of the category FRUIT is). All these experiments had in common that artifactual categories elicited more graded category judgments than discrete categories. Uncertainty (or certainty for that matter) about membership status has also been investigated. For example Estes (2004) showed that the uncertainty hypothesis of people simply not knowing what some things were leading to more graded judgments was not necessarily correct. In their findings, one cannot infer that uncertainy necessarily leads to graded judgments. Thus, the individual level of expertise with the different terms originating from modeling languages should not affect the outcome of the category structure itself.

Other studies have also criticized the focus on only the natural and artifactual category dichotomy, reasoning that, for example, categories for emotional states are categorized distinctly, and can be differentiated from the others by their relative amount of graded and discrete category judgments (Altarriba and Bauer 2004). As a result, it can be concluded that investigating the category structure of categories that have not yet been classified before leads to potentially useful insight into the way these categories are experienced.

Procedure

We adopt the most recent procedure as used by Estes (cf. Estes 2003; 2004) in their studies. The steps to take for a typical investigation are explained below.

Select target categories. The first step is to decide which categories to investigate. For our focus this will be the meta-concepts investigated in the dissertation. One should keep in mind the width of the categories themselves, and how specific the investigation should be. For example, one could, in order of specificity, investigate FURNITURE, or INDOOR FURNITURE or WOODEN INDOOR FURNITURE.

Select target items. Second, it is necessary to select the terms for which the category judgments will be elicited. Depending on the nature of the experiment, one could attempt to find a balance of potential full, partial, and non-members (i.e., if the goal is to compare category structure between multiple categories). If the goal is to investigate the structure of a particular category, for which the category judgment status is not yet known, it is effective to select a large sample of terms related to the category. For modeling language constructs like the meta-concepts we investigate one can thus select all the terms used by modeling languages for different concepts (e.g., when investigating the category ACTOR, selecting terms like 'market segment', 'actor', 'business role', 'department').

Present benchmark. Before presenting the actual items to be investigated, it can be useful to include a benchmark for some categories of which the general distribution of partial and non-partial members are known. From existing datasets a selection of 5 full members, 5 partial members and 5 non members can be taken. For these items the participants are then asked to answer whether each item is a full, partial or non-member of the benchmark category. Beforehand it is explained that a full member means that the item definitely belongs to the category, partial member that it does, but only to a certain degree, and that non-member means that it does not belong to the target category. These benchmark results are useful later to compare and contrast a participant's results with known data in order to see if their category judgments are as expected – because if they are radically different, interpreting the actual investigated items should be done with (more) care.

Present items. The actual items that are to be investigated can now be presented together with their target categories. A list of items can be presented for a category (e.g., 20 items are shown for the category FURNITURE, followed by 20 items for the category FRUITS), and participants are asked for each item to respond whether they are full, partial or non-members, in the same way as during the benchmark. The order of both the terms and the categories that are investigated can be randomized here in order to prevent effects from bias or fatigue, and to make it easier to check for potential non-serious responses by finding unlikely long strings of repeating similar answers.

Process results. Depending on the way the target items have been selected, it can be difficult to make conclusive judgments about the category structure. Comparing the relative amount of full, partial and non-member responses to benchmark categories for which the average distribution of these responses is known is a good way, but cannot be used as effectively if terms were selected as seemingly random (for example, in our case of taking modeling language terms from specifications). This is because the distribution of *potential* graded and discrete judgments can be inherently different because, for example, the input data contained almost only terms that would be judged as discrete. Thus, it also becomes necessary to look at the terms and their scores in *context*. With modeling terms here it is for example a good approach to look at how common the terms are and how often they are used in different modeling languages. If the terms that are deemed partial members in an experiment are used by many languages, and thus occur many times, it is a strong hint of a significant gradedness effect taking place.

Example

Say that we wish to investigate a category relevant to conceptual modeling: SERVICE. We first select two benchmark categories for which we know one is typically discrete (a natural category, FRUIT), and one typically graded (an artifactual one, TOOLS).

First the benchmark was executed. Its results are as expected, with the investigated terms for fruit readily being judged in a discrete fashion: bananas and pineapples definitely being fruit, and tomato being one of the few contended partial members. The results for the tool category were as expected, with items like hammer, screwdriver and the like consistently being judged as full members, and the category itself displaying a much higher amount of partial member judgments (let us assume some not entirely unfounded possibilities: both 'gun' and 'pen' being considered partial members here). After crunching the numbers, we find percentages for the amount of full, partial and non-members for both benchmark categories, say, 9% for the fruit category, and 28% for the tool category.

Finally, for the actual target category we simply have a list of terms that people have used over the years in presentations, talks, and papers (ranging from the obvious like 'service' and 'service *something*', to the less obvious like 'fragment'). After processing the results from this category we find some interesting terms that are considered partial members, some non-members, and so on. We calculate a certain proportion, say 19% of partial member responses for the SERVICE category. This in itself is not enough to effectively conclude whether this category is of a graded or discrete nature – although we can conclude it tends towards the investigated artifactual benchmark category, and thus hints towards also being of a graded nature. The real contribution here is that certain terms, say 'fragment' and others were considered partial members by nearly all of the participants. Given that this term is used often in the context of services, we conclude that there is definitely some graded effect happening in this category, and more importantly, that some of the core terminology is of a graded, multi-interpretable nature.

Limitations of the method and its results

With this procedure, the items' membership status of a given target category is investigated out of their usual context (of use). This can have some effect on the actual judgments, as terms can be used in different ways, having different meanings. Nonetheless, the base understanding that people have of the terms will likely be the leading factor for their judgments. More important to keep in mind are the potential differences originating from a participant's cultural and social background, as these can lead to wholly different kinds of category judgments. For example, while in the Netherlands the term 'bicycle' can readily be expected to be a full member of the VEHICLES TO GO TO WORK

WITH category, one would not necessarily expect the same results while performing the same study in the United States. Such differences can at times be found in existing datasets, as they are necessarily reflective of both their era, and culture. The differences between groups can be used to a study's advantage though, as one can use this method to investigate the different ways in which *groups* find particular items to be discrete or graded members of a target category, which can be used to determine which terminology might lead to communication breakdowns if these groups are to meet and discuss. As noted in the description of the procedure, depending on the selection of the items it can be more difficult to conclusively determine the exact structure of a given category. This is especially when larger sets of items are taken from sources without processing them in balanced sets of potential full, partial and non-members. However, the more important finding of which terms themselves are considered graded can still be concluded in a valid way – which for conceptual modeling already leads to useful insights as it can be used to focus on particular terms which likely have multiple interpretations.

2.2.2 Feature listing

Feature listing is a way of attempting to characterize the conceptualization people have of some particular thing. It has been used in psychological research a long time (e.g., Rosch et al. 1976, Rosch and Mervis 1975), and the task of feature listing is understood to be dependent on the conceptual information people have about the item (Smith 1978). At times it is also referred to as property generation (e.g., Wu and Barsalou 2009). However, this can be confusing in some situations as the elicited features (or properties) are, from an ontological point of view, not always what one intuitively understands as being a property (e.g., some functional aspects, like how a particular item is used).

Feature listing has been found to characterize the conceptualization someone has of some investigated thing because it reveals information from two cognitive processes: (1) word association, and (2) situated simulation (Santos et al. 2011). The word association thus operates on information related to the linguistic system and information in the brain, while the situated simulation system operates on memory of events in which those things took place. Chaffin (1997) has also found that when people are familiar with the items that they are listing features for, features will be more likely produced based on events (i.e., their actual experiences with the thing – the situated simulation). For unfamiliar things the produced features seem to be more based on someone's definitional understanding of the thing (i.e., their linguistic knowledge – the word association).

Given that what we wish to achieve is the elicitation of data that helps us understand how people think about certain concepts in terms of their features, such elicitation studies are likely to yield useful data. Several studies have been undertaken on such premises alone already – that insights into the feature structure of concepts in itself can be useful for both research and practice (e.g., McRae et al. 2005). These elicited sets of featurs are useful because they also serve as a delimiter of what things are to people. Combinations of different features someone lists effectively reduce the potential instantiation space for a concept, so that it can be more easily understood what particulars (i.e., instantiations) they likely think of.

Nonetheless, it should be understood that such studies are temporal snapshots of someone's understanding of some thing, and that they are not necessarily stable. Over time, as people learn and find new insights the features they attribute, and their typicality to some thing might change. Gaillard et al. (2011) showed as well that the stability of such features furthermore depends on the nature of the concept – for some kinds of concepts the same conserved features are likely to be elicited time and time again, while for other concepts potentially each feature listing task might generate new data.

Procedure

We follow the basic procedure that other feature elicitation studies use. The three main steps and their considerations are explained below.

Select items. The selection of the things to be investigated beforehand is quite simple, as it comes down to choosing a particular concept, and then deciding on a logical lexical reference for it (e.g., 'actor' for the concept ACTOR). In order to ensure that participants are primed to the correct context (in our case, conceptual modeling), a short description which does not overly bias participants can be useful to have alongside them, in this case for example a description like 'an actor is some kind of thing that does something'.

Present items. For all of the items to be investigated, participants should be given the simple question of writing down any and all features or properties they associate with that item. Beforehand, an example can be given of different kinds of features to attempt to prevent a bias e.g., only structural or functional features. For example, when we use the example CHAIR, we can show participants potential features of a functional nature ('used to sit on', 'can put stuff on it'), structural ('has four legs', 'made of wood') or other natures like taxonomic classification ('is a kind of furniture'). In order to focus the features that are elicited, either a time restriction can be put in place or we can ask only for the n most typical features, so that participants are forced to give a specific amount and make a decision between those which are particularly salient for them and those that are less so. During the presentation of the items, care should be taken not to have items followed by semantically close items (e.g., eliciting features for 'hard goal' directly after 'soft goal'), as this can introduce unintended overlap between the elicited features.

Process results. The resulting elicited features generally have to be processed in at least two stages: translation and normalization. Where applicable, the results should be translated to a single language used for presentation of the results (commonly English). The elicitation of features in multiple languages can be useful in certain situations where participants have different first languages, as allowing them to generate features in their own language avoids potential issues like limited vocabularies, or unintended wrong selection of terms. Finally, for the normalization it becomes necessary to analyze the form of the words (e.g., collapse singular and plural versions of the same term) and potential homonyms in order to reduce the amount of redundancy. Afterwards, they can be classified according to further details, for example their modality (e.g., whether of an alethic or deontic nature), or their associated dimensions (e.g., whether of a sensory, taxonomic, or structural nature).

Example

As an example, assume that we are asking people to elicit their top 5 of the most typical features for the concept GOAL. The pre-processed result can include many terms, like 'has a time', 'is defined by someone', 'has name', 'has a name', 'has a label', 'has a type', 'has a relationship with a subgoal', 'becomes an evaluation criteria for future accomplishments', 'may be quantifiable', 'can be refined into a subgoal', 'is based on an objective'.

Certain terms clearly mean the same and can be collapsed, for example 'has name' and 'has *a* name' meaning the same. For some other features it is less clear cut whether they mean the same and should be collapsed into the same feature, like 'has a label'. Analyzing the kind of features in this list one sees some deontic aspects, a temporal feature, but primarily alethic and structural features. Thus, the first conclusion we could draw from this is that goals are apparently identified primarily in terms of their structural features. The actual features themselves, especially the repeated ones elicited from multiple participants then helps us to delimit what potential instantiations (or, particulars) the concept collapse to. For example, if the 'is based on an objective' feature is shared between almost all participants, we can reason that they see goals as traditional hard goals, and that as a result, soft goals where the satisfaction criteria is not known are not a typical goal for them. Further investigations can then be performed with the elicited and processed features in order to investigate their typicality in more detail, for example by executing a survey in which participants are now asked to rate on a numerical scale how typical the elicited features are for the target item. With such data clusters of features with similar typicalities could be found and used to reconstruct the most typical instantiations of what goals are to the participants.

Limitations of the method and its results

The primary thing to keep in mind while applying this method (and analyzing its results) is the potential for (introducing) unwanted contextual bias. It is necessary to prime the participants at least in some degree to the actual concept to be investigated – one does not want a participant to think of their favorite Hollywood actor when they are expected to elicit features for an actor in the context of conceptual modeling – but this priming should avoid introducing too much context. Short descriptions of the items that are investigated can be used to put participants on the right track conceptually, while avoiding introducing too much specificity or bias.

Finally, the results should not be taken as being a full measurement of someone's conceptualization of the investigated item, but as a (carefully constructed) characterization thereof. Likely not all aspects of the conceptualization are captured just by the feature listing, just as not all aspects of their conceptualization will be captured by other methods like the semantic differential explained below. Instead, the results from different methods together should be taken and looked at in a holistic way in order to paint the most accurate picture of the conceptualization that participants have of the investigated items.

2.2.3 Semantic differential

The Semantic Differential (Osgood et al. 1957) was conceived as a method to measure meaning through connotation with the dimensions of evaluative, potency and activity. It can be extended to any describable dimension. The test itself works by semantic association and is based on indirect questioning, i.e., using adjectives that are synonyms to the dimension under investigation. For instance, if we want to know whether a certain RESOURCE mentioned in a model is an artificially created one (e.g., money) or something naturally occurring (e.g., daylight) we have to figure out which adjectives carry the connotative meaning we want to investigate. Here we can use "natural" with its connotative meaning of something that is caused by nature, not by humankind. If we then ask someone whether a RESOURCE is "natural" or not, the connotative meaning carries over and we can deduce whether they find a RESOURCE to be something caused by nature or humankind. When this is done with multiple sets of adjectives ("natural – unnatural", "cultivated – uncultivated", etc.) for the same dimension we can create a significant claim about the understanding a person has of that category for that dimension.

Which dimensions to investigate can be decided upon in several ways. One can find inspiration from practical, real-world scenarios (e.g., observing conceptual modelers and seeing if disagreements about certain topics arise more than others), semantic differences in the methods and languages that people use (e.g., analyzing and comparing specifications of modeling languages to find differences) or directly eliciting the dimensions from participants themselves. This last scenario is quite useful if no problematic dimensions are known yet and one thus needs to figure out what exactly to investigate with the differential. In such exploratory cases the Repertory Grid technique is a useful complementary approach (Tan and Hunter 2002) to elicit such dimensions without biasing the participants towards a certain outcome by asking participants what constructs (comparable to dimensions in our work) they would use to describe a given concept (cf. its use as a tool for exploratory research in information systems research (Tan and Hunter 2002)). These constructs could then be used as input for the investigation by semantic differential. However, the downside to this approach is that if the participants do not provide enough shared constructs, there would be very few dimensions to investigate, making a systematic and targeted study a difficult endeavor.

The semantic differential is not a mere word game with no basis in reality, as can be seen in research that reveals these same structures in the subconscious semantic connections in the brain, cf. Luria and Vinogradova (1959; 1971). Because of its flexibility, which allows it to be used for any category and dimension, it is well documented and widespread in research. As such there are many well documented quality criteria that can be taken into consideration while constructing the test (Verhagen and Meents 2007, Garland 1990). Furthermore, it has been shown to be reliable and to stand up to test-retest validity considerations (Peter 1979, Di Vesta and Dick 1966).

Procedure

There are a number of steps to take in order to construct a differential that can be used to investigate personal semantics, as follows:

Determining the target subjects. The people targeted by the differential, i.e., those whose personal ontology we wish to work on, should ideally share a roughly similar cultural background and most importantly, the same mother tongue (first language, or L1). While differences from person to person should be expected, significant discrepancies need not be a problem, as the test works on an individual level, and communities or groups can be defined after the test, based on the results.

Determining the categories. The categories to be investigated can be general or specific. It is equally possible to investigate a category like 'RESOURCE' as it is to investigate 'RESOURCES THAT ARE BAD FOR THE ENVIRONMENT'. In general, broader categories are a better starting point if little is already known about the individual categorization, as these more detailed categories can be found through the results of the initial test.

Determining the semantic primes. In order to make sure each subject fills out the test according to their typical understanding of the category it is necessary to prime them to that understanding (i.e., set the context). This is done by offering a list of words that are often used for the entire range of the category. When investigating modeling languages one can use the words or names of the language constructs from specific notations, such as 'player', 'hardware', 'organizational unit' for the category ACTOR. During the test, the subject should be asked to select no more than 2 or 3 words from this list to describe what they would typically use to refer to that category.

Determining the dimensions. The dimensions are the actual distinctions we want to investigate. For instance, if we wish to know whether someone considers resources to be typically physical things (e.g., hardware, dollar bills) or non-physical things (e.g., software, meaning) we determine the dimension to be physical. A pragmatic approach to determining interesting dimensions is to find conflicting use of the same words, be it from practical origins (i.e., anecdotes of problems that arose in a project) or theoretical (i.e., semantic analysis of modeling language specifications).

Determining the initial set of adjectives. The adjectives are the indirect measurements of a dimension through connotative meaning. These correspond to the conceptualizations of the target subjects, and can be initially gathered from automated thesaurus searches and then refined by testing and asking for feedback from the subjects.

Piloting the Differential. A pilot implementation of the differential with more than the necessary amount of adjectives should be done with a (significant) sample of the test population. This helps establish the general polarities of the adjectives, and can be used to choose those that result in the strongest reactions. Testing should be done that verifies these results are as non-random and non-neutral in order to establish the final set of adjectives. Finally, via interviews or polling it should be tested that the adjectives actually reflect the desired conceptualizations.

Finalizing the Differential. A set of five or more adjectives per category should be selected based on their polarities in order to ensure a significant representation of the dimension (Verhagen and Meents 2007). In the test itself the order of categories investigated, and the order of adjectives should be randomized in order to prevent contextual contamination. A Likert scale is used to assess how much a subject agrees with an adjective. Labels could for instance be 'very much like', 'somewhat like', 'neutral', 'not much like', 'not at all like' (in the used language of the test) to enhance the understanding of what the scores mean. The actual differential can then be constructed, where for each category first the priming task is given, followed by the adjectives on for instance a 5 point scale. See Figure 2.2 for a graphical example of a possible implementation.

Process results The results for each category can be numerically represented and separately processed. The labels 'very much like' to 'not at all like' are given a score of 2 to -2, both extremes representing either complete agreement or disagreement on how applicable the dimension is. For example, a score of 2 for the dimension 'human' in the category ACTOR means the person strongly agrees that an actor is a human thing, while a score of -2 would be complete agreement that it is not a human thing. Neutral scores (e.g., $-2 < x < 2$)) can then be interpreted either as the subject having no opinion on the specific category (i.e., no familiarity), or that the combination of category/dimension is an unusual (i.e., irrelevant) one, where for both cases it may or may not be human. For instance some groups of people think the property of being human has little to do with rules. The median of the individual adjective scores for a dimension is calculated to represent the score for that category-dimension combination. The standard deviation can then be calculated in order to figure out cluster coherence.

Example

The results of a pilot study performed early on in the course of this dissertation are discussed here as an example of the method. Figure 2.3 most clearly shows a strong tendency of this group of students to associate many categories with being necessary and intentional but not vague. With the exception of GOAL, most categories received neutral (or irrelevant) results to the natural dimension. Most interesting is the strong individual deviation visible for the material dimension, where it is clear that the test subjects diverged in their opinion on certain categories. In general the large deviations from the median show clearly that the subjects we investigated cannot be considered a homogeneous group, even though they share most common attributes such as age group, gender and academic stage and focus. This demonstrates the danger of *a priori* definition of groups one may expect to share similar meanings. It seems safer to decide what groups exist on the basis of actual data.

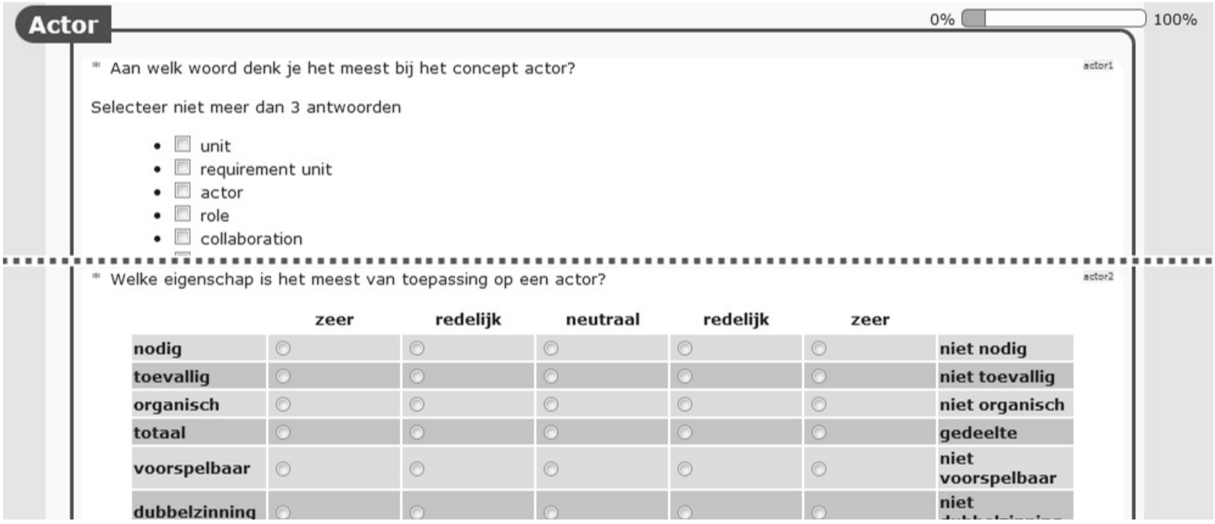


Figure 2.2: Example of the semantic differential (in Dutch) as we implemented it for our investigation. For each category we asked which words the subject associated most with that category. After answering they were presented with a list of bipolar adjectives (e.g., necessary – not necessary, whole – part) for which they had to answer to what degree the adjective applies to the category.

Figure 2.4 shows four individual categorization patterns for EVENT. At first glance there is less immediate agreement between these subjects, but some clusters for some dimensions can be seen. Most interesting is the strong divergence that can be seen between two subjects for their understanding of the dimensions composed, necessary and intentional. While the former subject sees an EVENT as being a composed, necessary and intentional thing, the latter sees it as being non-composed, non-necessary and non-intentional. The former interpretation could be that of someone who sees EVENTS as things we cause and for instance includes their consequence in their scope as well. The latter subject sees them perhaps more as things that just happen, regardless of what we do. From these incompatible understandings it should be clear that models produced by these subjects should thus be carefully compared in the likely case that they include any EVENT-type constructs.

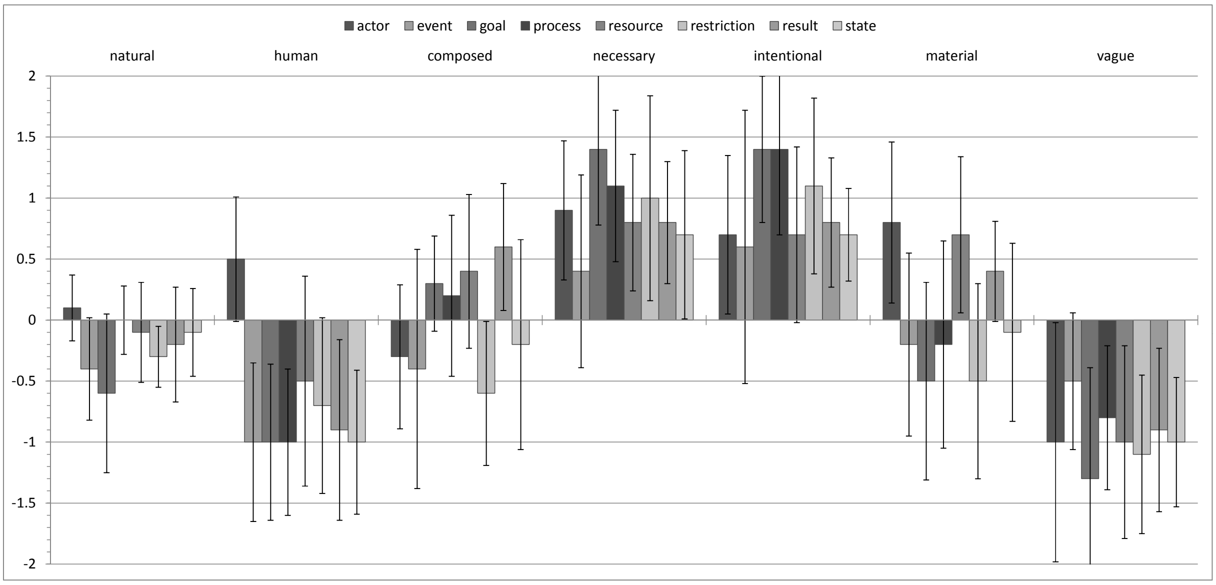


Figure 2.3: Visualization of the semantic differential results from an early pilot study. Shown is the median and standard deviation for each category-dimension combination grouped per dimension. To illustrate, the first cluster of bars shows in what respect all the categories (ACTOR,...,STATE) are considered *natural* things, followed by the next cluster that shows in what respect the categories are *human* things and so on. The higher a score (and thus the bar), the more a given category-dimension combination is considered applicable and vice versa the lower a score, the more a given category-dimension combination is considered not applicable.

Using Semantic Differentials

From the example investigation it should be clear that the main benefit of this method is that it can clearly show the different ways people interpret certain categories. As such, it can provide data that can be used to help verify that shared ontologies do not clash with the individual conceptualizations of people involved by comparing the (personal) conceptual structure that can be inferred from the data with the structure of the shared ontology. Our approach is thus meant to support and complement, rather than replace, existing practices of ontology engineering.

We envision that the main application of this method will be to generate datasets that characterize people and showcase different 'kinds' (groups) of modelers and their typical personal ontologies. These datasets are far removed from traditional formal ontology structures. They are not as easily exchanged or used directly as computational artifacts. Yet our focus is not on creating data as input for expert systems or other computational artifacts that would automatically reason with them, but instead to give humans working with a variety of models and modelers some amount of guidance for understanding others.

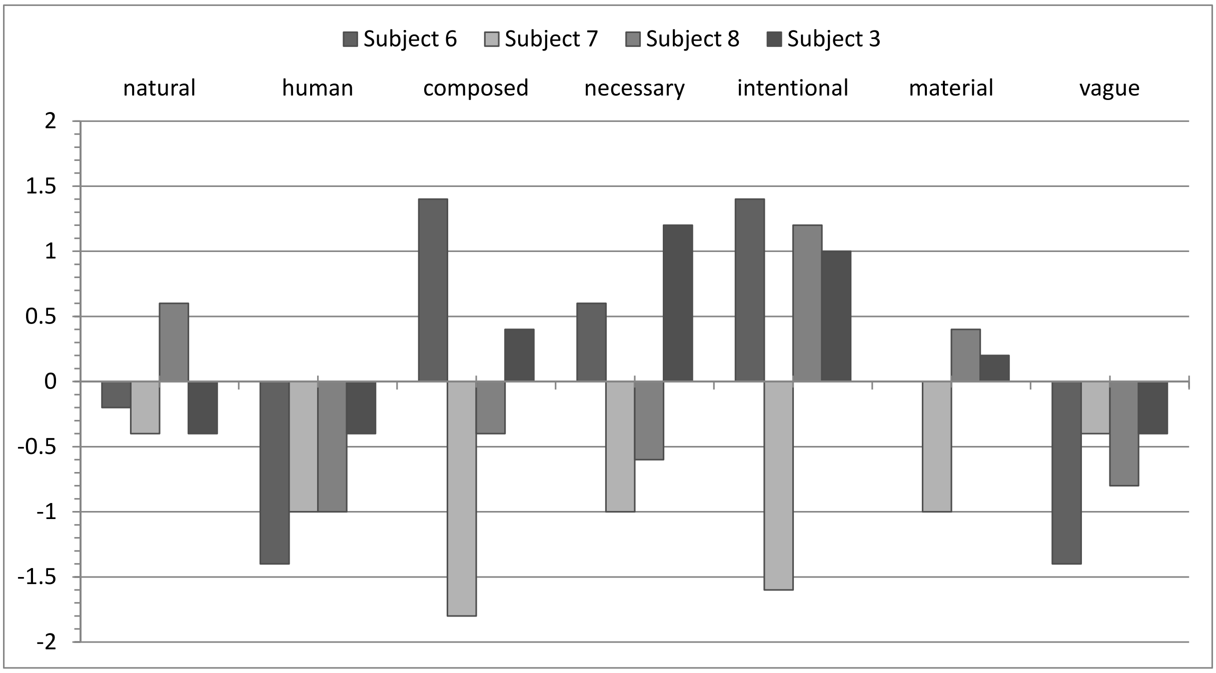


Figure 2.4: Comparison of how four subjects understand EVENTS. The bars represent the respective scores of each subject, grouped per dimension. To illustrate, subjects 6 and 3 score mostly positive on the dimensions *composed, necessary* and *intentional*, while subject 7 is the inverse of this and scores strongly negative for these dimensions. These differences clearly show the different interpretations subjects have of a category.

The method described here has potential for application in practice. When an enterprise modeler creates the integrated model of an enterprise they will use various different aspect models, specifications and other information to deal with basic syntax and semantic issues of notations they are not familiar with – and on top of that, the discovered categorization patterns for all the people involved in creating the aspect models. With this extra information they can connect the categorization patterns to the aspect models and thereby, with a certain probability, infer the semantic boundaries of the constructs used in each model. For example, where they have a model that contains many goal constructs, and the relevant categorization pattern shows a strong agreement on goals being 'not necessary', they might opt to model them as a 'soft goal' instead of the traditional mandatory hard goal. By doing this it will become clearer which constructs from the different aspect models can be linked together, i.e., which share roughly the same categorization pattern, and which do not, and thus have to be modeled or re-modeled with different constructs.

Another potential application is more technology-oriented and involves the use of a tool to guide the aspect modelers themselves as they create their models. It would still be based on the differential, but would guide modelers through what construct to choose from a given notation. This could be done by first asking broad questions concerning what they want to model, and then refining the possible constructs by going through all of the possible relevant distinctions (i.e., dimensions). The benefit of this approach could be that a modeling notation would be used more consistently and in line with its specification. This would heavily depend on the actual capability of the notation to represent all the distinct possibilities a modeler can arrive at. Finally, another possible application would be to validate the semantics of the notations themselves by directly having modelers evaluate them, with the same method as described above. This could make it possible to figure out whether certain notations would be used appropriately by the modelers, and if not, align the notation's specification with the modeler's interpretations while integrating. However, such applications are outside of the scope of this research.

Limitations of the method

The primary difficulty in applying this method is the need for a shared first language and a roughly similar cultural background among subjects, so that the adjectives used are assigned a sufficiently similar meaning and connotation. This does not mean one cannot investigate people with different mother tongues (a common situation, both in academia and industry) but it does necessitate more work on finding the most useful adjectives and constructing an effective test for separate language groups. The results for separate groups can, within reason, be used in combination. For example, results from separate investigations into tendencies of people from the Netherlands, Germany and Flemish Belgium or other countries with similar cultures and language groups, show enough similarity in conceptual background to suggest that the results can be mixed without becoming unreliable. What should not be done, however, is mixing results from groups that have strongly divergent languages or cultures, in the sense that one cannot realistically expect sufficiently similar conceptual backgrounds. For instance, comparing tests from Dutch speakers with Chinese speakers could potentially introduce too much of a conceptual gap to be reliable. This will have to be established experimentally.

Another limitation of the method lies in the way it is parametrized in view of the need for prior identification of (problematic) dimensions. Typically, we only investigate those dimensions which we already know, or suspect to be problematic, either because we have observed semantic mismatches or strongly expect them. As such, our method (without an exploratory phase based on, e.g., Kelly's Repertory Grid) has little predictive power in the sense of pointing towards hitherto unproblematic areas that may yet become problematic. However, if one's first concern is the resolution of already observed interoperability issues rather than preventing such issues from ever arising, this becomes less of an issue.

More importantly, the results of this method are not complete conceptualizations. They do not create a formal structure of the knowledge we capture, nor define any formal relations, yet the potential for resolving interoperability issues is there. We focus on finding 'some important differences' instead of finding 'everything that is similar'. To explain this by analogy, compare our issue to that of DNA-based paternity testing. It is not necessary to create a complete representation of the genomes of the individuals involved, but merely to find a significant number of points at which to find out whether patterns differ or not. We do not need full conceptualizations because for our goal of 'figuring out how to integrate the semantics of a model', the limited number of ontological distinctions we discover are in any case a "significant help to the characterization of their intended meaning" (Guarino et al. 1994).

Limitations of produced results

The data this method produces also has some limitations that have to be taken into account. Most importantly, because we measure personal understandings instead of postulating them, it is at best temporarily valid to attribute such an understanding to a particular person. As someone's understanding of 'the world and the things in it' changes over time, so does their personal ontology – and thus possibly the pattern of distinctions they exhibit. Because of this we should not simply rely on a single measurement of a modeler, but measure at regular intervals (for instance yearly, or at the start of a substantial new project).

What does not necessarily change over time is the pattern of distinctions we attribute to a given community of people. Suppose we find a cluster of people with certain properties, sharing a close age-range, interest and job function, we can attribute a specific pattern of distinctions to it, and define it as a community. If some people's understanding changes over time, and their personal ontology no longer fits this understanding, they are arguably no longer part of the community. However, this does not change the validity of that old pattern being correlated to that community. Instead, people move between communities as their understanding changes over time, perhaps eventually stabilizing, perhaps moving back and forth between communities. Given enough data, the real value may come from being able to identify and clearly specify the boundaries of a certain community in order to predict what kind of shared conceptual distinctions people might have.

2.3 Summary

In this chapter we have outlined the research approach we have followed during this thesis. We described the way in which we aim to answer the research questions detailed earlier, and in what order these questions would ideally be investigated. In order to ground the empirical methods used for our studies (into category structure, feature structure and conceptual understandings) these methods have been discussed and demonstrated through prior pilot studies and (fictional) examples of results.

Part II

Fundamental Findings

CHAPTER 3

What to Investigate?

Abstract. In this chapter we introduce the meta-concepts and discriminatory dimensions we will use as a starting point for the more detailed empirical and applied work on personal semantics. The selection of these elements is vital, because in order to study how people understand particular modeling meta-concepts, we first need to know which meta-concepts there are, and which of them are relevant in the context of conceptual modeling. The origin of these selected meta-concepts and dimensions lies in an exploratory study that was aimed at discovering the conceptual landscape of Enterprise Modeling and Architecture. It did so by virtue of an analysis of the notations used by modeling languages and methods, which resulted in a categorization of the used terminology.

The content of this chapter is based on work published as:

Dirk van der Linden, Stijn J.B.A. Hoppenbrouwers, Alina Lartseva and Henderik A. Proper: Towards an Investigation of the Conceptual Landscape of Enterprise Architecture. In: *T. Halpin et al. (Eds.): BPMDS 2011 and EMMSAD 2011, LNBIP 81*, pp. 526–535. Springer, Heidelberg

Re-published as an extended abstract in: Proceedings of the 6th SIKS Conference on Enterprise Information Systems (EIS2011), *V. Dignum et al. (Eds.): CEUR-WS Vol 800, 2011, pp 17–18*

3.1 Background – Why focus on languages?

It is important that we derive the meta-concepts and dimensions *a priori*, in contrast to eliciting them on an individual basis from participants during later studies. This is because we want to be able to compare and relate outcomes from different studies with each other. In order to do that, we need to have a fixed set of inputs (i.e., meta-concepts and dimensions) that all those studies can use. Even if every study were to include a specific concept elicitation phase before moving on to the study-specific details, there is no guarantee that the same, or indeed, similar, concepts would be investigated. For example, when we elicit conceptual understandings of students and practitioners and want to compare whether they think similarly or display significant differences, it is necessary that we actually investigate the same concepts (i.e., both studies investigating someone's understanding of ACTOR, regardless of what that understanding *is*). If we were to let participants decide themselves what meta-concepts are most relevant to them during a study, we would not be able to draw such comparisons. This would make it difficult to infer more practical consequences for conceptual modeling affairs, such as the effect from learning modeling languages (by comparing and contrasting the development of conceptual understandings from students to experienced practitioners). Thus, it is necessary to *a priori* elicit or derive the meta-concepts and dimensions we want to use.

Why derive meta-concepts and discriminatory dimensions by looking at modeling notations, and not by interviewing or studying people as they model? Arguably a modeling language is more of a reflection of the conceptualizations the language *designers* had or have of a given domain than the language *users* (although the two are not necessarily mutually exclusive). The users of a modeling language do not necessarily divide the world up in the same way (i.e., use the same kind of meta-concepts when conceptualizing a domain). However, there are some compelling arguments to derive meta-concepts through the analysis of existing modeling notations. Because of the sheer amount of languages and specifications that are in use, there is a wealth of data to use for the our purpose, which would make for a quicker start than a separate study dependent on a large number of people active in industry. We opt for the approach using an analysis of notations instead of working with people for a number of reasons.

First, by investigating the notations used by modeling languages and methods, we can also get an overview of the terminology used for the different concepts. These terms are useful for further studies, if not downright necessary for experimental work into categorization, structuring of concepts and feature-set elicitation. For example, in order to investigate a concept while ensuring participants are using it in the desired context, it is necessary to include a semantic priming task (see Sec. 2.2.3), for which either the terminology or the visual notation used during modeling would be ideal.

Second, having an overview of the terminology used in different languages, and access to the specification of those languages would also make it possible to derive discriminatory dimensions on which people differ in their conceptual understanding. We can do so by taking the semantics of the individual terms into account and for example, look for polysemy in order to find cases where superficially similar words are used in conceptually distinct ways, which can lead to the derivation of the necessary dimensions.

Finally, while a language user is often not the designer of that language, their choice for the language can be regarded as a reflection of their conceptual needs. It is likely that people interested in modeling a particular domain or aspect of an enterprise (e.g., the business processes) will opt for a language which allows them to model the things that are conceptually relevant to them, in the amount of detail they need. In this way, users of a modeling language constitute a specific *discourse community* (Hoppenbrouwers 2003). Their shared focus (i.e., their conceptual needs and points of interest), whether it existed prior to, or arose by influence of the language or method, are the differentiating factors from users of other modeling languages (and thus discourse communities). As Perelman and Olbrechts-Tyteca (1969) describe:

> "All language is the language of community, be this a community bound by biological ties, or by the practice of a common discipline or technique. The terms used, their meaning, their definition, can only be understood in the context of the habits, ways of thought, methods, external circumstances, and tradition known to the users of those terms."

This means that the language these people use has to be regarded at least as a reflection of their specific conceptual needs as well. By investigating these modeling languages that groups of people have chosen to use, we are effectively investigating the understandings and conceptualizations they have of their domains. While it remains less ideal from a research point of view to look at these languages as a reflection of people's conceptual needs compared to directly interviewing and surveying them, it nonetheless seems acceptable to derive the meta-concepts and dimensions we need for our further studies from the modeling notations they use.

By taking this approach it is true that we investigate the languages as they are on paper (i.e., in terms of their specification), which is not necessarily the same way in which they are used by people. It is a common occurrence that a modeling language is adopted, and its notation is used only partially, or with its semantics modified to suit the users (Henderson-Sellers 2005, Sowa 2010). Thus, if one wanted to study what the elements from the language (i.e., the meta-concepts) *meant* to the people using them, looking solely at the language specification would not be a good starting point. However, our requirements in this chapter are of a purely ontological nature (i.e., to investigate what meta-concepts there are). At this point we are not yet preoccupied with the more metaphysical concerns (i.e., to investigate what those meta-concepts actually *mean* for people). These latter concerns will be addressed in more detailed studies described in the following chapters. Because of this, whether the users of a modeling language use its meta-concepts such as specified or adopt their own interpretation has no impact on our selection, as right now we are concerned with just that – the selection of meta-concepts that will be studied in more detail in other investigations.

3.2 Selection of source languages

As we have opted to use modeling notations as input for the derivation of meta-concepts, we need to make a well-balanced selection of different modeling languages or methods that cover multiple (conceptually different) aspects. In order to prevent a conceptually narrow set of input we selected a wide variety of modeling languages and methods covering many different aspects, originating both from academia and industry. Where possible we based ourselves on official specifications (e.g., refereed and published works, official specification documents approved by standards management groups such as the Object Management Group (OMG)). As Enterprise Modeling covers many different aspects of modern day enterprises such as goals, processes, value exchanges, architecture, and IT, we included notations used for these different aspects, even if some of them are less widely used in terms of absolute users. We choose to particularly look at domain-specific languages as they focus on a specific application domain and often offer specialized terminology to deal with them (France and Rumpe 2005, Mernik et al. 2005). Given the many different aspects of an enterprise (e.g., processes, goals, value-exchanges, architecture) that modern enterprises contain and have to deal with, a plethora of domain-specific languages are often used to model them.

If no such standardized or widely agreed upon specification exists (e.g., as in the case of the goal modeling language i*, where there are multiple accepted and used 'standards' such as the RWTH Aachen wiki (Grau et al. 2007)) we opted for a widely used one. An overview of the languages and methods selected as input material is given in Table 3.1.

Table 3.1: The discourse communities we used as the initial input for our analysis.

Source / Ref.	Aspect	Reason for inclusion
ADeL (Patig 2011)	Software deployment	The Application Deployment Language aims at describing (and validating) deployment of IT, which is an aspect that is missed in many other languages.
ArchiMate 1.0 (The Open Group 2009)	Architecture	A standard for Enterprise Architecture (EA) modeling, it is one of the more expansive sources of EA-specific terminology.
ARIS / EPC (Scheer and Nüttgens 2000)	Architecture	The Architecture of Integrated Information Systems is a framework whose community's terminology is at the intersection of process modeling and EM.
Balanced Scorecard (Object Management Group 2010b)	Performance	An analysis framework intended for more than just (technical) modelers, it is a source of commonly used terminology by an important language community within EM; the not-necessarily technically inclined.

BPMN (Object Management Group 2010a)	Processes	The Business Process Modeling Notation is a standard for process modeling and useful as a source as its large community in industry and practice (Recker 2010) is heavily involved in, and promotes, abstract process thinking.
e3Value (Gordijn et al. 2006)	Value exchanges	The e3Value notation has a very specific economic viewpoint, offering terminology specialized for value exchanges.
Game theory (von Neumann 1944)	Behavior	von Neumann's original concept and its terminology is often used in lieu of more 'standard' modeling languages when dealing with economic aspects.
GRL, i*, KAOS (Liu and Yu 2001, Grau et al. 2007, Dardenne et al. 1993)	Specifications	GRL, i* and KAOS are widely used languages/methods used in goal oriented thinking.
ITML (Frank et al. 2009)	Implementation and deployment	The IT Modeling Language offers terminology specific to those implementing and deploying software, which many other languages lack.
RBAC (Ferrariolo et al. 1995)	Security	The terminology defined by Role-Based Access Control is a near defacto standard way of describing security issues, even outside of technical descriptions thereof.
VPEC-T (Green and Bate 2007)	Architecture (analysis)	The "Lost in Translation" approach focusing on Value, Policies, Events, Content and Trust offers an analysis framework that can be used by more than just those with technical expertise. For this reason it is worthwhile to take the terminology and its respective language community of non-technical users into account.

As we aimed to keep the analysis manageable we did not include many existing languages and methods, both general-purpose, and domain-specific. When it comes to widely used general-purpose languages for example, we did not include perhaps obvious choices such as ER or UML, nor methods like TOGAF, even if some of them are among the most used modeling techniques in practice (Davies et al. 2006). This was done because we initially wanted to ensure a wide diversity of input material that would reflect the different parts of an enterprise (i.e., the different Aspects listed in Table 3.1). Because of this we focused on first finding domain-specific languages so that we could ensure that elements from conceptually distinct worlds (e.g., specific terminology used only in the context of goals, or value exchanges, or business processes) would be included in our analysis. After considering (a number of) these languages we extended our scope and focused on including less domain-specific sources. We initially opted here to include modeling languages or methods originating not only from technical work in academia, but more business-oriented sources (e.g., Balanced Scorecards) to ensure a balance between conceptual elements originating both from the more technical as well as business worlds.

At this point, we found that we had an extensive amount of conceptually diverse input, with a seemingly significantly large number of terms. When we attempted to include more notations, whether domain-specific or general-purpose, we found that little was added in the sense of new conceptually distinct elements, or terminology being used in different ways. We then decided that the addition of more languages became unnecessary due to the saturation of semantically distinct terms. From the selected input we gathered all the terminology used by their notations and classified them into lists of nouns and verbs (i.e., entities and relationships). Figure 3.1 gives an example of the result of this process.

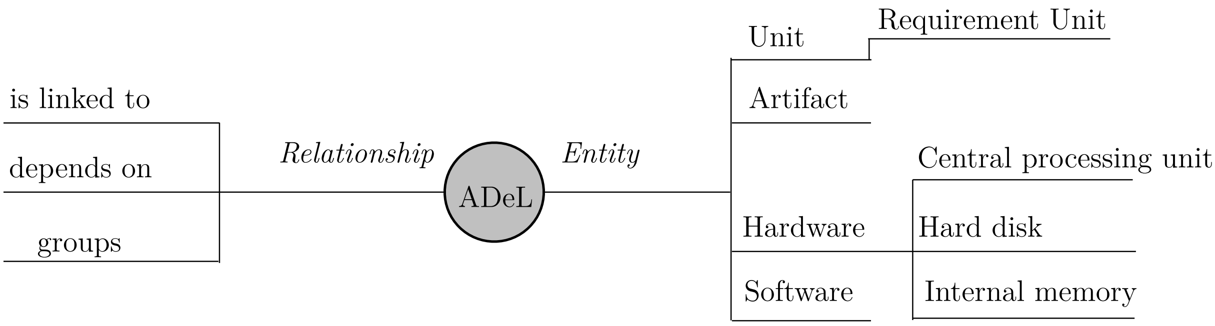


Figure 3.1: Example of the classified terminology for a modeling notation, here in particular for ADeL. On the left terminology used for relationships are denoted, while terminology for entities is given on the right, showing also the hierarchy of the elements from the specification.

3.3 Derivation of meta-concepts and dimensions

With the selected notations from Table 3.1 and their initial terminological classification we can now iteratively derive the main meta-concepts we will investigate, as well as the discriminatory dimensions we can find in those meta-concepts.

3.3.1 Derived meta-concepts

The meta-concepts (i.e., categories of the language constructs) were derived in an iterative fashion based on the terminology derived from the modeling notations. The requirements we had at this point were to have a practically manageable set in terms of number of concepts. Furthermore, they should remain recognizable to future study participants by being easily related to different domains of origin (e.g., goal modeling, value exchanges, process modeling). The derivation was performed with the involvement of three experienced conceptual modelers and researchers reaching agreement on the classification of each term in order to ensure less of a bias for the final categorization judgments. All people involved in this derivation had different professional backgrounds and foci in conceptual modeling, so that for the discussion on categorization judgments multiple interpretations were represented.

At first a simple 1 : 1 term to concept mapping was established, meaning that for each term found in the language we denoted the existence of one meta-concept. This led, as expected, to a large amount of meta-concepts with significant redundancy for almost all aspects. For example, meta-concepts as GOAL, SOFT GOAL, BUSINESS GOAL, BELIEF, were defined, as well as ACTOR, ACTORS, ARTIFICIAL ACTOR, and so on. This would lead to an overload of specialized (almost atomic) meta-concepts, making it necessary to start integrating the conceptually similar terms into their overarching concepts. We did so by collapsing these redundancies into the most common denominator and repeating this step for multiple iterations, so that we would end up with a limited set of high-level concepts from which it would be actually possible to study typicality and centrality effects.

For example, an initial mapping for acting elements included meta-concepts as BUSINESS ACTOR, PERSON, HARDWARE, SOFTWARE, which would iterate into smaller sets of meta-concepts such as HUMANS, ARTIFICIAL ACTORS, ABSTRACT ACTORS, again be reduced to HUMANS ACTOR, NON-HUMAN ACTOR and eventually end as the over-arching meta-concept "ACTOR". Through this iterative process, each term was classified into a final set of high-level meta-concepts.

We used the semantics of each term as defined or described in its specifications as the guideline for the classification here, not what might be intuitively expected of the term itself. For example, *end event* being a RESULT according to BPMN, not an EVENT. While it would be quite feasible to continue to collapse the set of selected meta-concepts until one arrived at nothing more than, say ENTITY and RELATIONSHIP in the vein of Chen (1976), or going even further by ending up with a single particular in the sense of blobjectivism as proposed by Horgan and Potrč (2000), there would be little practical benefit in doing so. Given the need for the meta-concepts to be recognizable, and in doing so offer enough different points of investigation, we finished with the classification found in Table 3.2. It is broad enough to fit the explicit differences found in the source material, while still remaining minimal in the amount of meta-concepts.

Table 3.2: The meta-concepts and the words used for particular instantiations of them resulting from our analysis of the languages and methods. Numbers denote the originating community as following: 1=ArchiMate, 2=ADeL, 3=BPMN, 4=e3Value, 5=GRL/i*/KAOS, 6=ITML, 7=ARIS, 8=Balanced Scorecard, 9=Game theory, 10=RBAC, 11=VPEC-T.

Meta-concept	Members
ACTOR	unit[1], requirement unit[2], actor[1,4,5], role[1,5,10], collaboration[1], player[9], infrastructure/application component[1], device[1], application software[7], organizational unit[7], position[7,5], perspective[8], market segment[4], hardware role[6], software role[6], hardware[6], software[6], organizational role[6], environment/software agent[5]
EVENT	event[1,5,7,11], behavior[1], function[1,7], interaction[1], activity[3], task[3], business rule/service task[3], transaction[3], start event[3], intermediate event[3], value activity[4], value interface[4], value offering[4], connection[4], move[9], contribution[5], correlation[5], dependency[5], means-ends[5], control[5], goal refinement[5], monitor[5], operation[10], operationalisation[5], performance[5], operation[5], initiatives[8]
GOAL	goal[5,6,7], hard-goal[5], soft-goal[5], business goal[6], achieve goal[5], assignment[5], avoid goal[5], cease goal[5], expectation[5], maintain goal[5], requirement[5], consumer need[4], belief[5], value[11], target[8]
PROCESS	organizational/infrastructure service[1], information/other service[7], service[1], IT service[6], process[3], sub/business/process flow[3], business process[6], dependency path[4], game[9], task[5]
RESOURCE	artifact[1,2], location[2,6], hardware[2,6] , cpu[2], hd[2], memory[2], software[2], value[1], data/business object[1], object[5], node[1], network[1,6], network device[6], representation[1], meaning[1], device[1], computer hardware[7], machine resource[7], environmental data[7], data input[3], input[5], value object[4], information[9], resource[5], content[11], value port[4]
RESTRICTION	interaction[1], contract[1], interface[1], message[7], catching[3], throwing[3], boundary[3,5], rule[9], decomposition[5], belief[5], priority[6], license[6], boundary condition[5], conflict[5], domain property[5], permission[10], value[11], policy[11], trust[11], (non)interrupting[3], strategy[9], measure[8], strategic objective[8]
RESULT	product[1], human/material/service output[7], data output[3], outcome[8], end event[3], payoff[9]

It is important to stress the need for the selection of these particular meta-concepts with their high level of abstraction. This is done because the things we wish to investigate through other studies (e.g., feature clusters, centrality and typicality of conceptual elements, as explained in more detail in Sec. 1.4.1) hinge on a prototypical view, for which it is necessary to investigate the central, more conserved regions of a concept. The selection of more specialized meta-concepts would make that impossible, as it would not be possible to as easily compare different clusters of typical conceptual understandings if they were already separated into their distinct own concepts (e.g., separating human and non-human actors). While more specific (and thus detailed) meta-concepts would of course lead to more detailed studies, it should also be considered that the selection of those meta-concepts a priori would introduce an unwarranted bias originating from our own ideas of how the world (and in this case, conceptual modeling) is made up.

Consider the example of the concept *kitchen equipment*. If we know already that what we really want to investigate is kitchen equipment used to prepare food, we would be tempted to just skip to the more detailed concepts, such as CUTTING BOARD, FRYING PAN, or WATER KETTLE. However, those (subjective) choices are strongly reflective of the way that we see the world, whereas other people might divide the world in different ways, and end up with different sets of specific concepts: for example simply MORTAR and PESTLE. By making such a priori choices for detailed concepts we unnecessarily bias ourselves to a particular view of the world and lose the ability to investigate how people other than ourselves view it. This works in the same vein for our modelings concepts: if we choose a very specific interpretation to investigate, for example a business process modeling concept of goal, then we would forego any potentially interesting results that we might have found for the goal concept in more studies of value exchanges. For these reasons we will use these higher level, abstract meta-concepts which can encompass all these interpretations in our studies.

It is not our goal to give stipulative definitions for the selected meta-concepts, but for clarity's sake they can roughly be described as being: things that act (in the broadest sense of the word) (ACTORS), things that happen (EVENTS), things that someone or something wants to achieve (GOALS), things that happen in some way (PROCESSES), things that are used for something (RESOURCES), things that limit something or someone's actions (RESTRICTIONS), things that are the end state of something (RESULTS).

Additional considerations

Note that the classification is limited to the terms describing entities (i.e., nouns), because initial classifications that included relationships (i.e., verbs) significantly increased the amount of members of the EVENT meta-concept compared to the other meta-concepts. We found it undesirable to (superficially) cloud further analyses for the debatable virtue of including the many existing (strongly similar) relationships within this meta-concept. Since it is necessary to use these words in further empirical research, having such an unbalanced amount of words for the EVENT meta-concept (with no direct clear benefit to having them), would make it more difficult to ensure participants in these studies do not become demotivated and give up because of question overload.

Initially, we had also selected the meta-concept of STATE, as there were some (albeit a rather small) amount of terminological references to it. However, after several early pilot studies and feedback sessions with potential participants for our longitudinal study into conceptual understanding (described in detail in Ch. 6.4) we decided not to include it. We found that the limited amount of terminological input was a detractor to not biasing people into a specific interpretation, which made it difficult to get useful readings. This lack of different interpretations might be because most people had at least some background in (discrete) mathematics, and were familiar with the basic notions of states and transitions. This did not lead to many different opinions on the metaphysical nature of a state itself, as most semantic focus was often given to a *thing* that is in some state. Whatever properties it has while being in such a state were then often attributed to that thing, not the state itself as an ontologically separate entity.

3.3.2 Derived dimensions

As it is commonly accepted in cognitive and psychological literature that we can best investigate concepts by characterizing them instead of attempting to exhaustively measure someone's conceptual understanding (Malt et al. 2011, Pinker 2007), we also need dimensions on which to characterize them by. This means that instead of attempting to exhaustively write down what exactly a concept is for someone, that we simply find some dimensions on which people (commonly) agree or disagree for that concept, and by that characterize typical understandings. For example, we could characterize the concept of SPAGHETTI ALLA BOLOGNESA on such dimensions as 'tastiness', 'ease-of-making', 'costliness', 'easy-to-eat'. Such sets of dimensions thus make it easier to see when people differ in particular aspects of their conceptual understanding.

We derived such an initial set of dimensions for our meta-concepts by analyzing the classification of the terminology and looking for terms which, according to their specification's definition, could function as antonyms. We analyzed all the terminology from Table 3.2 in this way and derived a number of dimensions that serve as a means to characterize people's understanding of the meta-concepts. The results of this analysis and the selected discriminatory dimensions are shown in Table 3.3.

Table 3.3: Dimensions we found which can be used to differentiate the different lexical members.

Dimension	Explanation
natural	Whether something was intentionally created (be it with a purpose or not) or occurred naturally. For example, an artifact versus a location or an actor versus some hardware.
human	The human condition appears both in the sense of actors, for example a (human) actor or some software. It also seems to appear in composed structures, for example an organizational unit or market segment being necessarily subject to the human condition, a collaboration not so.
composed	If something is composed of multiple parts (be it an aggregation or complex structure). For example, a single actor versus a organizational unit, or a cpu being a single resource, while a piece of hardware is seen as more than one.
necessary	The difference between logical (im)possibility and probability. For example, a belief being a restriction that can be broken, while a (logical, or natural) rule is logically impossible to be broken.
material	Something exists either as a physical, material object or is a (metaphysical) conception. For example, a materially existing resource can be a piece of hardware, while a piece of information can be just as much of a resource without existing as a material object.
intentional	Something is done, or provoked (be it with or without a reason) instead of spontaneously occurring. For example, an event can be a spontaneous occurrence while a transaction is always intentionally provoked.
vague	Something is either determined explicitly or is vague and left open for interpretation. For example, a hard goal or a soft goal.

Combinations of these dimensions can be used to characterize the meta-concepts, effectively arriving at the particular interpretations that some term from the investigated modeling languages use for them. Different combinations of dimensions thus discriminate between these different interpretations. For example, while ‘hardware’ and ‘software’ are both a singular, non-human, non-natural ACTOR, they differentiate on the ‘material’ dimension. When the same term is used for a different interpretation of the same concept, different combinations of these dimensions can be used to discriminate between them. An ‘actor’ is often a non-composed, non-vague thing, but whether it is human often varies. That particular dimension can thus be used to discriminate between the different usages of the term. It should be noted that doing this based on the specification given by modeling languages becomes more difficult when the specification of a term is ambiguous or contradictory. For example, according to the iStar Wiki (Grau et al. 2007) an ‘Agent’ is an ACTOR with a concrete physical manifestation (i.e., in line with our dimension *material*). However, in the same specification it is explained that the term ‘Agent’ was chosen “*. . . so that it can be used to refer to human as well as artificial (hardware/software agents)*”. The latter of these, software agents, would be difficult to interpret as actually existing in a material sense. As such, figuring out which dimensions to use to characterize the term thus becomes more difficult.

Clarification of the dimension derivations

To go into more detail about how to derive such dimensions, we now explore how we derived each of the dimensions from Table 3.3. The term ‘actor’ could, depending on which notation it originated from, be used both to refer to a human being performing some action, as well as a non-human entity performing an action. Based on this distinction it becomes possible to infer that there is a dimension, tentatively to be called *human* on which we can characterize someone’s conceptual understanding of the concept ACTOR. Often dimensions do not apply only to one meta-concept, as the dimension described here also can be found in RESOURCES, where e.g., BPMN has a specific subtype for resources that indicates the resource is a ‘Human Performer’.

The RESOURCE meta-concept also seems to include (words for) things that ‘just’ exist and can be used like that, like a ‘location’. Others are dependent on some kind of manufacturing process, that is, they have to be made, such as ‘hardware’. This indicates that there is a dimension on which we can distinguish between those things that are *natural*, and those which are created by artificial, or man-made means (e.g., processes that do not occur spontaneously in nature, like the building of a car, or the intentional creation of a new chemical element).

For some meta-concepts there are both things that are seen as either singular things or collections of things. For instance, depending on the abstraction level that a person thinks on, an ACTOR can be such a specific individual thing, or a collection of things. The different terms like ‘organizational unit’ or ‘market segment’ are often taken as collection of individual actors (albeit approached at one interface point), while a ‘player’ or some piece of ‘hardware’ is regarded as a singular thing. From this we can derive that there is a dimension on which people distinguish whether things are *composed* or not.

We found as well that a distinction is made between things that physically exist in the world versus things that do not. When it comes to ACTORS, there are physical objects like pieces of ‘hardware’, but also abstract entities used for the same, like ‘roles’, or ‘software’. In some cases words that one might assume to not be physical objects are actually defined as being so, as is the case with the way the i* language uses ‘agents’, as they are defined as being an entity *“with a concrete physical manifestation”*. For RESOURCES there are also distinctions to be found for this dimension, with clearly physical(ly needed) objects like ‘hardware’, ‘cpu’, or some kinds of ‘business objects’, while the non-physical occupies equally as much: ‘information’, ‘representation’, ‘meaning’, and so on. It becomes evident that there is thus a dimension on which we distinguish between things being *material* or not.

We find that for the meta-concept GOAL there is a clear dimension on which people distinguish them: whether they are well-defined (e.g., having specific, measurable achievement criteria), or whether they are more open-ended (e.g., having fuzzy or non-measurable achievement criteria). This is reflected in most goal-modeling languages having both ‘hard goals’ and ‘soft goals’. From this we take the dimension of whether things are *vague* or not.

For some meta-concepts there are also differences whether something was intended or simply occurred, like for example with rules being either because of environmental conditions (i.e., laws of nature and logical necessity), or whether they were intentionally created to perform some specific function. When it comes to EVENTS it is also sometimes the case that an ‘event’ spontaneously occurs, while some other things like ‘transactions’ are always triggered by an intended source. Because of this we also include a dimension of *intentionality.*

Finally, while not necessarily reflected as clearly in the different words used by the modeling languages, there seems to be a distinction between things that are logical necessities and those that are not. For example, the many different kinds of RESTRICTIONS put on entities to control their behavior can, depending on what that entity is, either be logical rules that simply cannot be broken (e.g., operational conditions in an algorithm) or a deontic rules that shouldn’t be broken, but can (e.g., a guideline to be followed during some business process). Some general-purpose languages like Object Role Modeling (ORM) 2.0 explicitly model this with the use of different colors for the deontic or alethic status of a restriction. We thus chose to include a dimension of *necessity* to reflect these differences.

Additional considerations

It is not our intention to claim the dimensions selected and described here are exhaustive for the different ways in which people see the meta-concepts. Instead we aim to have a wide-range of dimensions available on which to investigate people's conceptual understandings. While it is likely there are more dimensions on which people differ in their understanding of these meta-concepts, we consider that the selected set offers an effective starting point with enough semantic diversity to support us in our investigations. Furthermore, the dimensions are grounded in existing distinctions found in modeling languages in use in practice. In line with our argument for using modeling language notations to derive the meta-concepts, we find using them equally useful as a starting point for deriving these dimensions.

Other existing classifications, ontological and metaphysical alike, offer far more extensive and detailed classifications of what kinds of things exist, and what properties they have, which could be used to generate these dimensions. However, how would we make a selection out of such sources without either overly biasing us to a specific (kind of) dimension that may or may not be that relevant to conceptual modeling? Given that many of these existing classifications are of a significant size and complexity (e.g., SUMO (Niles and Pease 2001)), or focused on providing upper ontologies from which the distinctions often remain too abstract for our more domain-oriented requirements (e.g., UFO (Guizzardi and Wagner 2004)) it is difficult to make a balanced selection, even if they are more specific to our context like the UEML class taxonomy (Anaya et al. 2010). We would again have to go back to the modeling language notations themselves and verify whether the dimensions proposed by some ontology are relevant to it.

An example of one of the detailed distinctions that such ontologies provide (which might be too abstract for our use) is the distinction between endurant and perdurant things. In the case of a perdurant thing only some part of it is seen to exist when it is observed at some specific point of time. An endurant thing exists fully even when observed in a single point of time. A HUMAN BODY, for instance, could be regarded as such. The way that human body moves, its locomotion, however, is perdurant, as WALKING or JUMPING cannot be readily and fully observed in a single point of time. Thus, one could make a distinction between concepts on this dimension of endurant/perdurant. But, in the context of conceptual modeling, such distinctions quickly become esoteric and of a less direct interest, as they do not relate to more concrete domain concerns on which practitioners actually discuss (e.g., "*do we know the satisfaction criteria for this goal?*", "do we allow server uptime to be seen as a necessary resource?"). Thus, this particular dimension might be too academic in nature to be salient for studying dimensions experienced by actual people involved in the creation and use of models.

For these reasons we have opted for the derivation of the dimensions based directly on the notations and semantic differences found between their terminologies. Doing so led to a timely and practical derivation of semantically distinct dimensions that can be used right away for further studies and analyses.

3.4 Hints of conceptual disagreements

With the set of selected meta-concepts, their related terms and the discriminatory dimensions we now have a decent base that we can use as input for our empirical methods. From the simple overviews presented here we already see hints both at a common, shared categorization of concepts between different aspects of EM and at a difference in preferred, or central terminology. Thus, there seems to be a common conceptual ground, with perhaps typical understandings correlated with particular aspects.

Some examples of this are terms used for the meta-concept ACTOR. In e3Value an *actor* is necessarily a being with some very specific properties (i.e., economic independence), while in a language often used with it, i*, it can also be other things, like a computer. Thus, while users of both languages likely share the common conceptual ground of an actor being some active entity, someone using e3Value would quickly associate it with what they see as economically independent beings (i.e., humans), while someone using i* can be expected to have a more generic abstract association. Another example concerns the meta-concept RESOURCE. A language with a focus on deployment of software has many words for describing *hardware* that is needed, and as such a speaker of ADeL, ITML or other languages with comparable focus might readily associate some RESOURCE with material objects that are needed to do something. Someone with other concerns, say architecture as expressed in ArchiMate, or more abstract process design like BPMN, might more quickly associate RESOURCE with some *information*, *knowledge*, or *environmental data* (i.e., immaterial objects) needed for something to be produced.

Yet more examples can be found with the meta-concept RESTRICTION, as many terms used for it differ in how they deal with 'necessity' of compliance to some restriction. On the one hand, languages tasked with implementation or technical factors often deal with logical, computational laws, and therefore use words like *rule* in the alethic sense: restrictions that logically cannot be broken. On the other hand, those speakers dealing with human aspects often find themselves prescribing deontic rules or restrictions which can be violated, even though this is improper. Depending on the view of a speaker, their interpretation of the meta-concept RESTRICTION could be biased towards rules that are either alethic or deontic in nature. Some counterintuitive classifications can also be found in a number of meta-concepts. Some descriptions or types of things are sometimes seemingly seen as specific kinds of those things. This happens with terms for ACTORS, like ARIS' *position*, ArchiMate's *organizational role* or RBAC's *role*, which are used not in the sense of *a role*, but as *some [actor] entity with the role of*. Similarly, depending on the focus of a language, words that are commonly seen as properties (for instance, *location*), can also be regarded as concrete concepts, like *location* being a (limited yet necessary) RESOURCE for people dealing with the (physical) deployment of machines.

While it is true that some of these differentiations of meaning between languages might not be intended, factors like historical compromises in the design or implementation of a language (such as ArchiMate's awkward handling of information as a physical entity, cf. The Open Group (2009), ch. 6.4), still allow for such differences to occur. In time those differences can become (unintentionally) entrenched in the conceptual landscape of a language and its speakers, especially if they are not quickly corrected. As a result, they will drift away from a shared understanding with once similar communities, and risk becoming a distinct community.

Knowing all this, we can see how important it is to discover these differences in meaning to know exactly what is meant by a model. For example, an e3Value *actor* and an ArchiMate *organizational unit* both are some kind of ACTOR. However, in terms of Table 3.3's dimensions, *actor* in this sense is naturally occurring, human and not composed, while *organizational unit* is not naturally occurring, human, and composed. Thus, while both languages speak of superficially the same thing, the mapping shows they are not interchangeable. Different words mapping to the same concept can be found in the same way. Take for example the meta-concept RESULT. An *outcome* in ARIS is often seen as a non-naturally occurring, non-materially intended 'thing', which would be the same as an *end event* in BPMN. Thus, in this context the mapping shows the words are interchangeable.

3.5 Summary & outlook

In this chapter we have answered our first research question ("*What are the meta-concepts shared between modeling languages used in enterprise modeling?*") with the selection of meta-concepts and dimensions detailed in Tables 3.2 and 3.3. We can now move on to more detailed investigations using them. We will first investigate a specific aspect to do with the conceptual understanding that people have of these meta-concepts, namely, whether the structure of these concepts more resembles that of a graded or discrete concept (i.e., whether specific things or instantiations are regarded as being absolute members or whether partial membership is permissible).

CHAPTER 4

On the Category Structure of Modeling Concepts

Abstract. In this chapter we investigate the category structure of the meta-concepts we just derived in order to study whether they more closely approximate a discrete or graded category. It is important to be aware of this distinction because it can have a significant effect on the way models (and the modeling languages used to create them) should be formalized and reasoned about. Furthermore, when a model contains (terms from) graded categories, the (conceptual) validity of instantiations of that model likely becomes a graded matter as well (i.e., while two instantiations can both be conceptually valid and fit with someone's view on the domain, some are more valid than others). To this end we performed an empirical study involving a group of modelers with varying levels of experience in order to discover their actual categorization habits. Overall, we find that most of the investigated meta-concepts exhibit a graded structure, with experienced modelers displaying this even more strongly than the other groups.

The content of this chapter is based on work published as:

Dirk van der Linden: Categorization of modeling language concepts: graded or discrete? In: Y.T. Demey and H. Panetto (Eds.): *OTM 2013 Workshops, LNCS 8186*, pp. 743–746. Springer, Heidelberg, 2013

Dirk van der Linden and Henderik A. Proper: Category Structure of Language Types Common to Conceptual Modeling Languages. In: I. Bider et al. (Eds.): *BPMDS 2014 and EMMSAD 2014, LNBIP 175*, pp. 317–331. Springer, Heidelberg, 2014

4.1 Background

We categorize the world around us in different ways depending on the subject matter. Some things we categorize more discretely, like natural things (e.g., things that naturally occur like fruits, plants, or animals), some things we categorize in a more graded way, such as artifactual things (e.g., manmade things like tools and vehicles). These different categorization tendencies have been shown many times in research, starting around the time of Rosch (1973) and Rosch and Mervis (1975). Also, they have been investigated by many others explicitly elaborating on the category structure for a number of natural and artifactual categories (cf. Diesendruck and Gelman 1999, Barr and Caplan 1987, Estes 2003; 2004). While some work investigating this has had difficulties in finding significant differences in categorization tendencies between artifactual and natural categories (Kalish

1995), other larger studies have re-affirmed it (Diesendruck and Gelman 1999). There are also arguments that only the natural/artifactual distinction is not enough, and we should also distinguish between other categories like emotional states (Altarriba and Bauer 2004). Regardless of the debate on whether particular kinds of categories are usually categorized in a specific way, it is clear that *we do not categorize everything in the same way.*

The categorization we speak of here deals with membership judgments. That is, whether a certain thing is judged to be a member of a given category. For example, most people would have no problem saying that an apple is a member of the category FRUIT, and they will likely reject the notion of a newspaper being so. In these cases, it is is easy to make a discrete, or absolute judgment. However, when borderline cases are introduced interesting effects occur (Hampton 1998, Hampton et al. 2006). Given, for instance, cases that do not have clear or crisp boundaries, like tomatoes or rhubarb, people have more difficulty deciding with certainty whether they are FRUITS or not. In such cases people often tend to give *graded judgments* – things being members of a category only to a certain degree. This prevalence of (strongly) graded membership judgments is then often correlated with the structure of the category being graded. It should thus be clear that the things we reason about while modeling or developing systems are also subject to such categorization effects, even when they are abstract things like IF-THEN statements (Adelson 1985). Given that many of our modeling efforts (be they the creation of domain models, ontologies to formalize knowledge or support reasoning with, databases to implement schemata, etc.) require us to be as exact as possible about what we aim to model, being aware of differences in membership judgments is thus an important aspect of properly representing a given domain and the things in it.

The importance of being aware of these different judgments starts during the modeling phase, particularly in settings where there is collaborative modeling and integration efforts (e.g., enterprise modeling). The uncertainty of membership judgments (i.e., what is a valid instantiation for this type, is this instantiation as valid as others) creeps into models, and is often lost, unless explicitly elicited and written down. The effect this has on the validity of a model can occur on two levels, the level of the categories from the domain (i.e., the concepts from the universe of discourse) and the level of the categories from the language (i.e., the meta-concepts used by a modeling language). Domain categories – the concepts from the universe of discourse – often receive great attention in discussions between modelers and stakeholders as well as in discussion between modelers themselves. This ensures (to some degree) that modelers know what things the stakeholders want to see in a model, and that they understand those things in the same way (Hoppenbrouwers 2003). However, categories from the language receive such detailed attention far less often, e.g., by asking "*What exactly is this type 'actor' from the language we are using?*

Does it allow us to model the acting elements from the universe of discourse we know about?". Instead, we often end up using our own intuitive or naive interpretations as we use the semantics of our own natural language (Sowa 2010) – together with all the category structures and nuances that come with doing so. Because of this, the modeling language that ends up actually being used often differs from the (formal definition of the) modeling language that is used on paper (Henderson-Sellers 2005). For example, a modeling language might formally define an actor as a rather specific thing (e.g., requiring it to be a singular abstract entity, and whatever other features might apply), which makes it fairly easy to determine whether something is a valid instantiation of that type – a human being here definitely not being one. On the other hand, one of the modelers (or simply any reader) might not use or be aware of those semantics, and instead see the type as having a different range of conceptually valid instantiations. This is problematic because it means that important semantics of the model might be lost when it is interpreted by other people not involved in the original modeling process (e.g., during model integration), or stakeholders who were not aware of some of the not explicated particularities. This is exacerbated by the fact that we do not have an insight into the structure of these categories *as used* by people, because not only do we not know what is considered valid, we do not know whether some things are considered more valid than others.

Thus, in this chapter we aim to clarify whether the categories common to many modeling languages and methods (i.e., those meta-concepts used by a language to instantiate domain concepts by) are categorized in a discrete or graded fashion. The implications of this for model creation and usage (particularly for models used to capture and document a certain domain) are important to be aware of. If a category from a language is typically judged in a discrete fashion, the semantics of models are likely easier to communicate, formalize, and keep coherent. However, if such a category is typically judged in a graded fashion, communicating it to others becomes more involved, also requiring more explicit discussion. Furthermore, the formalizations and tools we use need to explicitly support such structure (e.g., by using ontologies with support for features as typicality and centrality).

4.2 Experimental parameters

What we wish to do is to examine whether a number of categories more closely resemble graded or discrete categories. We can do this by performing a category membership experiment for the target categories and a number of benchmark categories of which we know whether they are typically judged in a discrete or graded fashion, and to what extent their members are so judged. The important things to find out are the overall membership judgments for the categories, and specifically the amount of non- and partial membership judgments for the individual terms.

This is typically done by letting participants rate terms (i.e., the non, partial and full category members) either on a numerical scale (cf. Barr and Caplan 1987, Diesendruck and Gelman 1999, Kalish 1995), or by directly asking them whether the exemplars are full, partial or non-members (Estes 2003). Foregoing the numerical scales for this direct questioning membership status has some advantages, as there is less ambiguity both from the participants' side, as well as from the researchers' side when it comes to interpreting the exact meaning of the numerical data. The use of these methods in practice and more methodological details about the exact measuring scales and potential confidence issues are explained in more detail in Chapter 2.2.1.

4.2.1 Considerations

There are a number of considerations to take into account with this investigation. First is the issue of the potential participants and their (natural) language. Most importantly, when we ask whether a certain thing is a member of a category or not, we would optimally do that in the participant's native language. However, as the terms used by most modeling languages and methods (i.e., the terms we will use in our experiment) are in English, we need to either use them as-is, or translate them. Given that most modelers use the terms as given by languages (i.e., in English), albeit often appending their own semantics, we will perform the experiment with the terms without translating them to the participants' native language.

For the benchmark we will use datasets from existing research. However, an issue with these existing and still often used datasets is that they can be outdated (e.g., the commonly used Barr and Caplan dataset was published in 1987), and they can be sensitive to cultural differences. Category judgments can shift as certain objects fall out of common use and are replaced by entirely different things, as well as certain objects that can be seen differently in different cultures. For example, while in Barr and Caplan's dataset bicycles are found to not be strong members of the category VEHICLE, repeating the experiment with Dutch, Danish or German participants (who are far more likely to use a bicycle as a mode of transport, cf. Pucher and Buehler (2008)) will likely lead to significantly different results. As such, care will have to be taken when interpreting the results from the benchmark categories to place them into the correct frame of time and culture.

While there are other datasets available that were gathered from non-English native speakers (e.g., Ruts et al. 2004, who performed an exemplar generation study amongst Belgian students) that might be used to create a more even dataset, they often only include full members and lack the necessary borderline and non-members.

4.2.2 Method

Participants: In total fifty-six people participated in the present study. Twenty-one of them were advanced (3rd or 4th year) students at an undergraduate university of applied science with a focus on computing science and modeling. Thirty-five were professional modelers employed at a research institute with a focus on information technology, and used modeling languages and tools to varying degrees. According to this level of use they were further subdivided into a group of fifteen beginning modelers and twenty-one expert modelers. All participated voluntarily and received no compensation for their participation.

Materials: The materials used for the benchmark in the experiment were based on the list of exemplars reported on by Barr and Caplan (1987). We used 5 full, 5 partial and 5 non-members terms for both of the benchmarks. They were translated and presented in Dutch for the twenty students, but presented in English for the participants at the public research center, given that this was the only shared language between all participants and all participants were sufficiently fluent. In this text we consistently refer to them in English. For this benchmark we included the categories FRUIT and VEHICLES (see Table 4.2). For the modeling part of the experiment we investigated the meta-concepts as found in Table 3.2: ACTOR, EVENT, GOAL, PROCESS, RESOURCE, RESTRICTION and RESULT. More information on these meta-concepts, terms and their derivation can be found in Chapter 3.3.1. The terms used for the members of these categories are the terms as used by the modeling languages and methods, based on the official (or most-used) specification.

Procedure: The procedure was based on Estes (2003)'s experimental approach. Participants completed the task through on online survey. In this survey, participants were instructed to judge whether a list of given terms were either full, partial or non-members for the current category. Participants were informed beforehand that partial member scores meant that the exemplar belonged to the category, but to a lesser degree than others. This was first done for the two benchmark categories, and followed in the same way for each of the investigated categories from the modeling languages. Additionally, for the categories from the modeling languages participants were instructed that the terms were to be considered in the context of conceptual modeling. The orders of the terms in each category were randomized for each participant. Care was taken to validate that participants filled out the survey seriously by comparing results and checking for long strings of repeating answers that the randomization of question order should have prevented from occurring.

4.3 Data

4.3.1 Benchmark concepts

The proportion of graded membership judgments exclusively for the borderline terms (i.e., expected partial member, as determined by the original datasets) used in the benchmark are shown in detail in Table 4.1. What was to be expected is that the typically discrete category (FRUIT) would show lower proportions of graded judgments compared to the typically graded category (VEHICLES). The given scores indicate the proportion of partial member judgments (e.g., 19% of students, 13% of beginning modelers, and 30% of expert modelers considered an avocado as a partial member of the FRUIT category). Shown are respectively the scores for students, beginning modelers, expert modelers, and the scores as reported by Barr and Caplan (1987), and Estes (2003).

Table 4.1: Partial member proportions for the partial member terms of the benchmark.

Category	Term	Student ($n = 20$)	Beginner ($n = 15$)	Expert ($n = 21$)	(Barr and Caplan 1987)	(Estes 2003)
FRUIT	avocado	0.19	0.13	0.30	0.37	0.16
	coconut	0.24	–	0.05	0.38	0.37
	tomato	0.33	0.27	0.25	0.34	0.05
	cucumber	0.19	–	0.25	0.23	0.21
	rhubarb	0.14	0.20	0.15	0.45	0.26
VEHICLES	gondola	0.24	0.20	0.20	0.50	0.21
	tricycle	0.14	0.13	0.10	0.64	0.58
	wheelchair	0.29	0.27	0.50	0.70	0.63
	horse	0.48	0.27	0.55	0.54	0.50
	husky	0.38	0.27	0.55	0.27	0.21

Table 4.2: The categories and terms for the benchmark as adapted from Barr and Caplan (1987) and Estes (2003), followed by the used Dutch translations for the student group.

Category	Term
FRUIT (discrete)	apple, pear, plum, banana, pineapple, avocado, coconut, tomato, cucumber, rhubarb, carrot, onion, potato, rose, spinach
VEHICLES (graded)	bus, car, truck, van, taxi, gondola, tricycle, wheelchair, horse, roller skates, husky (dog), lawnmower, bus driver, carton, newspaper
FRUIT (discrete)	appel, peer, pruim, banaan, ananas, avocado, kokosnoot, tomaat, komkommer, rabarber, wortel, ui, aardappel, roos, spinazie
VEHICLES (graded)	bus, auto, vrachtwagen, busje, taxi, gondel, driewieler, rolstoel, paard, rolschaatsen, husky (hond), grasmaaier, buschauffeur, doos, krant

4.3.2 Modeling concepts

An overview of the average amount of full, partial and non-member judgments for all investigated categories is given in Table 4.3. The results are given for each investigated group (students, beginning modelers and expert modelers), and indicate the proportion of membership judgments. For example, students considered 47% of the presented terms for the ACTOR category to be full members, 18% to be partial members and 35% to be non-members. The primary points of interest here are the higher scoring partial and non-member results, as they indicate words actually used by modeling languages that are either only considered to be partially reflective of their category (e.g., a 'market segment' would be only considered somewhat an ACTOR), or are considered not to be exemplars of that category (e.g., a 'requirement unit' would not be considered an ACTOR).

Table 4.4 gives a detailed overview of specific modeling language terms considered partial members by at least 30% or more of one of the investigated groups. We decided on an initial cutoff value of roughly at least one-third of a group for partial membership based on the data of Barr and Caplan (1987), judging it to be an effective cutoff point for inferring a significant partial membership effect occurring. On average the expert modelers displayed a much higher amount of graded judgments than the students or beginning modelers. For the terms listed in this table, students on average considered 15% of the investigated terms to be partial members, while beginning modelers did so for 32% and expert modelers considered 83% to be partial members.

Table 4.3: Average amount of membership scores (full, partial and non-members) for each group of investigated categories. The average proportion of partial judgments for the modeling concepts are bold, as well as the proportion of partial responses for the natural and artifactual benchmark category listed below them for comparison.

	student ($n = 20$)			beginner ($n = 15$)			expert ($n = 21$)		
Category	full	partial	non	full	partial	non	full	partial	non
ACTOR	0.47	0.18	0.35	0.30	0.14	0.55	0.41	0.25	0.35
EVENT	0.46	0.14	0.41	0.39	0.16	0.45	0.29	0.19	0.51
GOAL	0.65	0.11	0.23	0.60	0.16	0.24	0.56	0.20	0.24
PROCESS	0.66	0.14	0.20	0.62	0.22	0.16	0.41	0.32	0.28
RESOURCE	0.59	0.19	0.22	0.62	0.19	0.20	0.54	0.22	0.24
RESTRICTION	0.50	0.21	0.29	0.55	0.18	0.27	0.39	0.24	0.37
RESULT	0.73	0.16	0.11	0.86	0.07	0.08	0.76	0.16	0.09
average	0.58	**0.16**	0.26	0.56	**0.16**	0.27	0.48	**0.23**	0.29
FRUIT	0.44	**0.10**	0.45	0.47	**0.05**	0.42	0.49	**0.09**	0.41
VEHICLE	0.48	**0.14**	0.37	0.49	**0.13**	0.37	0.51	**0.20**	0.29

Table 4.4: Terms considered partial members by $\geq$ 30% of the people from at least one group (students, beginning modelers, or expert modelers), considering the cutoff considerations discussed in 4.3.2. The terms listed here are *only* those considered partial members, not including the terms primarily considered full or non-members. Raw datasets containing the complete set of terms in terms of full, partial and non-members can be found in Appendices A.1 through A.7.

Category	Term	Student	Beginner	Expert
ACTOR	unit			✓
	requirement unit			✓
	infrastructural component	✓		✓
	organizational component			✓
	device			✓
	application software			✓
	organizational unit			✓
	hardware			✓
	software	✓		✓
EVENT	behavior			✓
	function			✓

	interaction			✓
	activity		✓	
	task		✓	✓
	service task			✓
	value activity	✓		✓
	contribution			✓
	operation			✓
GOAL	expectation	✓	✓	✓
	requirement			✓
	consumer needs			✓
	target			✓
PROCESS	organizational service			✓
	infrastructure service			✓
	information service			✓
	other service		✓	✓
	IT service		✓	✓
	service			✓
	sub flow		✓	✓
	process flow			✓
	dependency path		✓	
	game		✓	✓
	task			✓
RESOURCE	artifact		✓	✓
	hd			✓
	location		✓	
	data object		✓	✓
	business object			✓
	object	✓	✓	
	data input			✓
	input			✓
	value object	✓		✓
	network device		✓	
	representation		✓	
	value port		✓	
	device	✓		
RESTRICTION	belief		✓	
	priority		✓	
	value		✓	

	interface	✓		
	catching			✓
	throwing	✓		✓
	license			✓
	trust			✓
	interrupting			✓
	non-interrupting			✓
	strategy			✓
	strategic objective	✓	✓	✓
RESULT	end event		✓	✓
	payoff			✓

4.4 Discussion

We will first discuss the results in general, showing how they support the assumption that there are categories in modeling languages that are of a graded nature. Finally, we also discuss a number of limitations of our current study that should be kept in mind when interpreting the results. A discussion of the more practical consequences these findings have for conceptual modeling can be found in Chapter 7.4.2.

4.4.1 General discussion

It was expected that the partial member judgments for the natural and artifactual benchmark categories would show a difference, with the artifactual category displaying a higher proportion of graded judgments. Although compared to the results from Barr and Caplan (1987) and Estes (2003) the overall amount of graded judgments seems to be lower, the relative distribution still seems intact. This is the case for both the beginning and expert modelers (the proportion of some graded judgments for VEHICLES being at least twice as large compared to the ones for FRUITS). This is not the case for the student group, as the difference between the benchmark categories there was found to be much smaller. This could be explained by the lower amount of experience with (and exposure to) modeling and languages (and modeling languages) that students have. This is further reflected in Table 4.4 where there are far less words considered to be partial members by students as compared to experienced modelers.

On average the proportion of partial member judgments is 0.16 for students, 0.16 for beginning modelers, and 0.23 for expert modelers. When we compare these scores to the average proportion of partial member judgments for the discrete and graded benchmark categories in Table 4.3 (respectively 0.10 and 0.14 for the students, 0.05 and 0.13 for the beginning modelers and 0.09 and 0.20 for the expert modelers), we can see that for the two groups of modelers most scores shown for the categories from modeling languages more clearly reflect the graded benchmark category than the discrete one. Thus, as a careful first investigation we seem to have found support that most categories from modeling languages are of a graded nature. Perhaps more importantly the results from Table 4.4 corroborate this idea, as it can be seen that there is a large number of real modeling terms that are deemed to be partial members by those modelers with a lot of experience. Furthermore, given that the distribution of terms for these categories was not the same as the benchmark categories (i.e., the benchmark categories were made up of equal amounts of full, partial and non-members, while for the categories from the modeling languages we were unaware of this distribution, with them likely containing proportionally more full members) this makes it all the more acceptable to support the idea described in the introduction that *these categories can be seen as exhibiting a graded structure*.

Another interesting finding is the high amount of non-member judgments found in many of the categories. It is striking that the terms we have used in modeling languages and methods are sometimes considered absolute non-members of their related category. In particular, it can be seen that EVENTS are the largest category for non-members across all groups (respectively $0.41, 0.45$, and 0.51), while ACTORS and RESTRICTIONS also have a high amount of non-members in some groups. A possible explanation for this is that people are quicker to judge about things they are specialized in, for example a process modeler having more snap judgments about concepts to do with processes, and thus also being more willing to rule out terms. In practice this means that the terminology we use originating from some languages might not reflect our innate category judgments at all, raising the question whether this is a bad thing (e.g., because the terminology is far away from our naive understanding and semantics) or perhaps not that much of a problem (e.g., because the mismatch between a term and our understanding of it in a given context makes it easier to 'redefine' and use it in that context).

As already hinted at and most clearly visible in Table 4.4, there is a striking difference between the groups we investigated when it comes to the proportion of partial member judgments. The expert modelers have a far higher amount of partial member judgments compared to the students, and in a lesser degree to the beginning modelers. An exception to this are PROCESSES and RESOURCES, which are judged more comparably between beginning and advanced modelers. This might be explained by the fact that the department of the research institute which the majority of the participants were working in has a strong focus on service science and is thus focused on many efforts involving processes (e.g., process modeling). An explanation for the difference between these groups might be that students simply have had less exposure to modeling terminology and are thus more likely to give absolute judgments. On the other hand, there is also the possibility that the (expert) modelers are, through training and experience, cognitively better equipped to deal with situations with abstract and vague concepts (cf. Wilmont et al. 2012), which could manifest in a higher amount of graded judgments.

4.4.2 Limitations

While it is good to find that our results hint towards the modeling language categories (i.e., the meta-concepts) having a graded nature, care must be taken not to immediately extrapolate this finding and use it to judge the structure of the investigated modeling categories in general. Because the data originates from a a single study, with two of our groups of participants being people with *professional* experience in conceptual modeling, repeating it with additional groups of (experienced) people to validate whether they share the same graded structure would be a prudent thing to do. Nonetheless, we see no immediate threats to the internal validity of the study we have performed.

It should also be noted that the study presented here talks about the structure of the category in terms of it being graded or discrete, but does not aim to give a representation of the *internal* structure in terms of where each member belongs in relation to each other. It is very likely that the internal ordering of the categories (regardless of the graded or discrete structure) is specific to different groups of people. For example, it can be expected that process modelers will have a different central core for a number of categories than, for example, goal modelers.

Furthermore, as categorization judgments are something inherent to people, it would also be useful to perform this study on specific subgroups of modelers (e.g., process modelers, enterprise architects, goal modelers) to analyze whether the proportion of graded responses is different for specific categories or not (i.e., test whether categories that modelers are focused on receive less partial member judgments). One could for instance hypothesize that people who are specialized in a topic have less semantic flexibility in regards to the categories of that topic.

Related to the terms we used, it might also be interesting to see whether the introduction of model context (i.e., presenting the terms while being used in a model) instead of the isolated terms themselves would yield different results. Nonetheless, the results from our study investigating the terms in isolation also provides useful insight into the number of terms that would typically not be considered as good representatives of their functional category. Furthermore, this might provide an additional source of complexity for participants, as with the amount of terms we used in the study, a large amount of modeling languages would be used, some of which participants are likely not familiar with.

Finally, as referred to earlier, the distribution of the terms for the modeling categories was not optimal (i.e., not evenly divided between full, partial and non-member), which makes it more difficult to infer detailed general statements about the internal ordering of the terms. However, this is a reflection of the fact that we performed a study on terms as they are used in practice (by modeling languages), and not on an artificially prepared text corpus. Such work on the detailed structure of these categories like those described above can be undertaken in further research, where the individual category members are rated on typicality and centrality in order to attempt to build an actual representation of a shared category structure. Such findings could then be used to create a more evenly distributed set of modeling terms for further membership judgment experiments and other purposes like discovering internal ordering of terms. Some of these potential avenues for further research will be discussed in Chapter 9.2.1.

4.5 Summary & outlook

In this chapter we have presented a study into the categorization of categories used by most modeling languages. It showed that many of these modeling categories (i.e., the meta-concepts) are likely of a graded nature (that is, some things are considered to be better members than others), which can have an effect on the semantics of models and their derivatives. We have discussed the implications for validity of models, and argued that it is important to be aware of these different categorization judgments. The main contribution of this chapter to our research questions has thus been empirically showing that *the categories we use to model are likely of a graded nature*, which before was only assumed (or worse, ignored). More specifically, we have shown that the modeling terminology from actual modeling languages and methods are affected by this graded nature as well. With this we have found an answer to our research question, "*To what extent do the categories for these meta-concepts more resemble a graded or discrete structure?*". Now that we know the structure of the meta-concepts that we investigate in this dissertation is of a graded nature, we can pursue further investigations into other aspects which also exhibit such gradedness. A first focal point will be the features that people attribute to these meta-concepts (e.g., a resource is made of wood), and attempting to infer which features are shared between multiple people and thus are more or less typical for each meta-concept. This will be the focus of Chapter 5.

CHAPTER 5

On Features of Modeling Concepts

Abstract. Is an actor typically considered a human being? What about an autonomous entity? In this chapter we investigate the typical feature structure of Chapter 3's meta-concepts in order to create an empirically grounded description of the semantic feature structure that people implicitly use while reasoning about, and with such concepts. Apart from the insights into the meta-concept structure that this chapter presents, consequences for the quality of models and use of modeling languages are discussed. These results will be used in Chapter 7 to discuss in more detail how the process of modeling stands to benefit from more insights into how the individual modelers see the basic modeling concepts shared between them. This is done in the context of when such a modeling process involves multiple people with different backgrounds, modeling different aspects (i.e., *enterprise modeling*).

The content of this chapter is based on work published as:

Dirk van der Linden and Marc van Zee: On the Semantic Feature Structure of Modeling Concepts: An Empirical Study. In: *IEEE 16th Conference on Business Informatics (CBI)*, vol. 2, pp. 158–165. IEEE, 2014

5.1 Background

Things are identified by their features, whether they are physical objects, abstract elements or psychological entities like concepts (Malt 1990, Bloom 1998). For example, I might identify a certain object as a *table* because it has some properties I have come to expect of tables (e.g., it has four legs, is made of wood, is of the right height to sit at, and there happens to be a full set of dinnerware on top of it). This holds for conceptual entities as well, where, for example, we identify the creation of a good research paper as a *goal* because we see certain features like wanting to achieve it, the possibility of it being achievable, and we want to achieve it, that it can be achieved, and that there is some way to achieve it. The features we identify things by are often of a structural (an aspect of a thing's physical or logical composition: the table being made of wood) or functional (an aspect of what a thing is used for, or does: the table being used to eat at) nature. Earlier work (Malt and Johnson 1992) found that non-natural concepts (i.e., human-made or abstract things) were also mostly identified by their structural features, as examples with similar functionality but different structure were still excluded as being part of those concepts. This might mean that the modeling concepts we use are as well strongly identified by their structural features (e.g., a process being composed of steps, an actor being a human thing, a resource being physical material).

This kind of identification through features is fundamental to human nature (Harnad 2005), as we categorize everything around us. It is how we decide *just what things are*. When something does not have the properties we typically associate with it, or differs from them too much, we simply reject it as possibly being so. For example, a typical Japanese dining table would be far less of a typical dining table for most Europeans given the difference in structure (being significantly lower), and function (being interacted with in a fundamentally different way: sitting on the floor in front of it instead of on a chair). As such, two different people could have a fundamental disagreement about whether a specific thing is a table or not. The identification of things in this way is an important aspect of determining whether something corresponds to our expectations – whether it is a correct model of our subjective reality. Such judgments are not discrete, but most often graded: things are not simply considered to 'be' a thing or not, but they are so to a certain degree. When we see a particular thing with some set of features (for instance the Japanese table) we thus make a graded judgment deciding to what degree it is a table or not. Whether this is done in the brain by comparing to an abstract set of features (in case of prototype theory), or to specific exemplars with a set of features (in case of exemplar theory) is still debated (cf. Storms et al. 2000, Ross and Makin 1999). However, it is clear that a comparison to some set of observed features is made.

The meta-concepts we use so often in modeling tasks are no different from this. For example, someone could simply find it incorrect to view a 'department' as being an actor, because they are not singular human entities that can easily be attributed moral responsibility for their actions. When that person is forced to model a department as an actor in some modeling language, because the language dictates it to be done in such a fashion, the produced model is not a truly valid representation of that modeler (or stakeholder's) conceptualization anymore. If we had more of an insight into what features are most common and typical for these modeling concepts we would have a better chance of judging whether a produced model corresponds to the involved modelers' and stakeholders' conceptualizations. Perhaps more importantly, if we know how stakeholders and modelers alike identify things in terms of their features (i.e., *what the modeling concepts really are for them*), we could reason whether the modeling languages we use let us (or can be extended to) model the world as they see it.

Investigating what features are most linked with common modeling concepts is thus directly relevant to the practice of conceptual modeling. In particular, the quality of models and modeling languages can benefit from a deeper understanding into these features, as it would be easier to ensure models are a valid representation of the conceptualization someone has of a particular domain. While quality is a term with many different meanings and aspects (Fettke et al. 2012, Houy et al. 2012), determining whether a model is able to convey the intended meaning of the original modeler (Robinson 2006) is surely relevant for all purposes. This holds especially when models are used for building and exchanging knowledge (Stahl 2000) (i.e., communicating between different parties), as is often the case with the different aspects modeled in Enterprise Modeling, making it important to be aware of the validity of just what knowledge is being exchanged.

An empirically derived understanding of the features that modelers typically associate with modeling concepts would help us in this, especially so because we tend to model with the semantics of our own natural language (Sowa 2010), and all the bias that comes with doing so. The primary use would be in knowing how people (in the context of conceptual modeling) *actually* judge whether some element from the universe of discourse is a typical example of some concept, or whether it is barely so. For example, if resources are typically resources because they are material, we know that we need to be very careful when we want to model an abstract piece of 'information' as a resource for some business process, and perhaps discuss whether we select instead a material representation of that information (i.e., a set of documents). By analyzing what kind of features are used to make these judgments (e.g., structural or functional, or other yet unknown modalities), we could also gain more insight into the fundamental nature of these concepts themselves. Furthermore, from analyzing the most typical features (e.g., those often repeated or shared between different modelers) *and* atypical features, we could find specific areas that would benefit from extra attention during modeling sessions. If we know what there are a large number of varied views for a particular concept (e.g., many people holding different views on what constitutes a resource), focusing extra discussions during the modeling process on those concepts would help enhance the validity of produced models, and likely have a positive effect on the agreement between the people involved (Stirna et al. 2007). Finally, knowing what features are most typical would make it possible to analyze where modeling languages might go wrong (by forcing people into an atypical use of a concept), or can be made better (by not allowing for the explicit modeling of some particular conceptualization).

5.2 Experimental Parameters

In this section we give an overview of the exact procedure we followed for our study, what materials we used, how we gathered the participants, and how we processed the results. We chose to perform a simple elicitation study with clear semantic priming at every question, where it was made clear to participants that all questions and answers were to be given in the context of conceptual modeling.

5.2.1 Considerations

Feature listing studies (sometimes also referred to as property generation), as described in more detail in Chapter 2.2.2, are a fairly simple type of study, where our primary requirement is to reach enough participants. For this reason we opt to create an open survey to which to invite people, and in doing so attempt to achieve a wide spread of results from different professional backgrounds. Similar to other studies we have performed, the main consideration here is the language used to survey the participants. As we want to reach a wide variety of people, and also include multi-national companies in which employees might have a different first language, it would be more practical to have a single shared language for the survey. Given that the practitioners we attempt to reach *a)* in general speak and understand English to a high professional working capacity, and *b)* the modeling languages they typically use are predominantly in English, it seems safe to assume the language used for surveying participants can be English.

However, a more troubling consideration for this particular study is what language should be required of the participants for the features they are asked to list. While participants might clearly understand the concept and descriptions given to them in English, an association task such as this feature listing in which they are asked to list any features that come to mind might yield more (or simply, faster) results in their native language. With such a setup it becomes necessary to 'normalize' the resulting features to a single language before analyzing and comparing them. This can introduce a translation and interpretation bias originating with the researchers performing this step. The use of multiple people, both independently translating non-English features to English, and afterwards comparing and contrasting their interpretations seems to be a safe compromise here. For this reason we will allow for participants to list features in their own language if they wish to do so (which practically means features can potentially be in English, Dutch, German, or French), and afterwards have at least two researchers translate the listed features back into English and reach consensus on any contended translation.

5.2.2 Method

Materials The basic examples of a chair and its functional and structural features were inspired by common works in cognitive science and categorization studies (e.g., Barr and Caplan 1987). We use the meta-concepts from Table 3.2, and use the direct lexical references for them (i.e., actor, event, goal, process, resource, restriction, result), together with a simple description for each meta-concept that was non-specific enough to not prime participants on a particular interpretation which already exists in the context of conceptual modeling such as hard goals or soft goals (see the descriptions given in Section 3.3.1 given on page 60).

Participants We invited professionals from academia and practice with significant conceptual modeling experience to participate in our study. Where available, we selected people who had participated in our earlier studies, and had at least an intermediate level of experience with modeling and modeling languages. In total 45 people participated, all of whom were experienced modelers employed at various internationally operating businesses located throughout Europe and the United States. All participated voluntarily and received no compensation for their participation.

Procedure The survey was executed online with the use of LimeSurvey. Participants were instructed in detail what the experiment was about, and what kind of information was needed. First we elicited what modeling languages the participant had significant experience with, followed by a longer explanation of the feature questions. For the studied set of seven modeling concepts we then asked participants to write down any and all features they would typically attribute to this concept, in the context of modeling. In order to encourage as much intuitive responses, participants were allowed to respond such features in a number of major languages locally used (English, German, French, Dutch). The results of this step were first (where necessary) translated into English, and then iteratively analyzed in order to standardize the grammatical and lexical form of the feature (e.g., standardizing 'human', 'is human', 'a human thing' to 'is human'). After this, another iteration followed in which features with the same meaning were clustered together (e.g., 'is a person', 'is a human being', 'is a human' into 'is human'). We then counted the amount of times that a feature was expressed by participants in order to determine which features were most commonly given. Each of these features were also analyzed to determine their modality and whether they were of a structural or functional nature (e.g., a RESTRICTION 'limiting behavior' is functional, while 'exists as an abstract thought' is structural.).

5.3 Data

The diverse amount of views on the different concepts we investigated was ensured by the wide variety of participants and the languages they were specialized in. While most common were general purpose language such as UML, ER and their variations, ArchiMate was also widely used. The total list (see Table 5.1) encompasses a wide range including also a host of formalized languages, methods and frameworks, and improvisational approaches (e.g., sketching models with pen and paper or in Microsoft PowerPoint).

Table 5.1: Used modeling languages, as reported by participants.

language
Object-oriented programming languages, IDEF, UML, ER, ArchiMate, BPMN, NIAM, Essence, Petrinets, BPEL, Amber, Merode, Turtle, DEMO, Dataflow diagrams, The Decision model visual modeling for decision logic, EPC, IE, CBM, BMC, TOGAF, IDEF, IDEF0, IT City Planning, Flowcharts, Value stream mapping, Value chain diagram, SIPOC, BPEL, ARIS, Proprietary language, Microsoft Powerpoint, ORM, Business object diagrams, Natural language, Plateua planning, ISAC, Yourdon Ward en Mellor ERD, Evolutionary NIAM, CaseTalk, JBF, i*, Secure Tropos, CORAS, VCL, VHDL, WSDL, e3Value, RBAC, XACML, EMF, BMO, PERT.

As the complete, raw list of elicited features is rather large (see Table 5.2 for their respective sizes), we give an overview of cleaned up results, with redundant features removed from them. (The full list is found in the Appendices.) We will focus specifically on the reoccurring features (i.e., those likely more typical as they were shared between multiple participants), which are given in Table 5.3. An overview of the different modalities expressed in the results is given in Table 5.4 (see later discussion in Section 5.4.1). Full datasets can be found in Appendices B.1 through B.7.

Table 5.2: Amount of features elicited per concept. These amounts are given excluding repeated features (i.e., exactly the same term was elicited from multiple participants), the number of which are given at the bottom of the table.

	actor	event	goal	process	resource	restriction	result
total	131	98	97	51	105	75	79
% structural	52%	70%	82%	73%	82%	52%	85%
% functional	43%	30%	18%	27%	18%	48%	15%
repetitions	46	35	13	4	28	16	8

Table 5.3: The most often reoccurring features we elicited in our study. Full datasets can be found in Appendices B.1 through B.7.

concept	feature
ACTOR	has a name, has a role, has skills, can perform an action / actions, has a responsibility / responsibilities, has capabilities, carries knowledge, has permissions, has rights, is human, is independent, carries out actions, is machine, is a person (with a role in a company), is an organization, initiates actions, is autonomous, is active, has responsibilities, is part of a business function
EVENT	is triggered by something, has consequences, has a trigger, can trigger an action, has a duration, triggers a process, occurs at a time, can be observed, has a name, happens, is a process end, triggers a task, has a description, has a source, is atomic, is a trigger, is repeatable, is an occurrence, is temporal, can trigger a process
GOAL	has a related stakeholder, has a name, can be achieved, becomes an evaluation criteria for future accomplishments, is clear, is measurable, has to be achieved, gives direction, is desired, is in the future
PROCESS	has input, has output, has related actors, has steps
RESOURCE	is used for something, is material, is immaterial, is countable, is scarce, has a name, has a value, has costs, has a source, can be human, is human, is a machine, is needed for an activity, has a type
RESTRICTION	has a type, limits something, has a name, has conditions, is natural, limits something, limits possible solutions, is quantitative, has a source, is measurable, can be temporal
RESULT	is measurable, has a name, has a quality, has a description, is realized, is observable, has value for someone

5.4 Discussion

5.4.1 General discussion

In this section we will first come back to the research questions we stated, and then discuss some of the more interesting findings in more detail.

Elicited and most typical features

The results given in appendices B.1 through B.7 are a complete, processed overview of all the elicited features that conceptual modelers use to identify common concepts. They are unweighted terms which, for at least one practitioner are in the top 5 most important features which they use to identify that specific concept by. The total amount of terms here is a collection of the results from all the individual participants and is thus not meant to be a requirement for what every actor has to be. These results can be found in Table 5.3, where only the features that were repeated by multiple participants are listed. For example, in appendix B.1 the first features are 'has a name', 'has permissions', 'has a role', and so on. This means that, in order to be a conceptually valid actor, an entity should have some kind of name that identifies it, it needs to have permissions on something, and it needs to hold a specific role. Almost all of the concepts have at least some repetition of features that allow for identification, whether it is the existence of some specific name or label, or them being capable of being identified or observed. Furthermore, most concepts can be dealt with on an instantiation level, as often they are seen in terms of specific instantiations.

The most often recurring features for the concept ACTOR center around, as could be expected, its ability to act. Specifically its logical and/or ethical capability to act when it should do so (e.g., having responsibilities), and its role of acting as an independent or autonomous entity. This seems to point the common view of an actor more towards human beings and (smart) agents, not simple reactive agents or automata. Furthermore, this is expressed again in the often reoccurring features that identify exactly *what* acts, including such features as stating whether it is a human being, a specific person, a machine or an organization.

When it comes to EVENTS, the most conserved features are that of being a trigger for something, and needing to be triggered itself. This shows that most modelers conceptualize of events as intentional things: things that are specifically made to happen. Environmental or contextual events, even if they might impact on something (e.g., some market conditions impacting a process' efficiency) are thus typically not considered to be events in this context, which is in line with the often used assumption of closed world semantics.

GOALS are most importantly identified by knowing whether they can be achieved, and to what degree they can be verified to have been achieved. This is in line with the often used distinction between hard and soft goals, as it is often important to know whether a goal's satisfaction criteria can actually be known to have been met.

PROCESSES are primarily identified as things with a clear beginning, middle, and end. They also typically have related actors that drive its events.

For RESOURCES there are a number of important features. They are identified by their nature (e.g., material or immaterial), the availability (e.g., scarcity) and the need of them having to be used for something. This implies strongly a view that resources exist primarily as a contextualized entity that only exists in the context of some other entity (e.g., a process) using them for a specific purpose.

RESTRICTIONS are, like most other features identified by their ability to be properly identified, described, and measured. However, this goes into more detail than some other concepts, as features like the specific conditions of what it limits, and for how long it does so are found to be important for something to be a restriction.

RESULTS are quite similar to restrictions in terms of their identifying features: the ability to be properly identified, described, temporal conditions like when a result is achieved. Furthermore, just like resources it seems that results are another contextually existent entity, that they exist in the context of being the result of some specific thing, often a process.

It is important to re-iterate that the feature sets as described here are only a small subset of the total amount of elicited features. All of the features we elicited, even those mentioned by only a single participant and not repeated by others, are used by practitioners in actual situations for the identification of concepts. Thus, each of these features should be taken seriously as a possible restriction of what specific modeling concepts can be in professional contexts. While the discussion we gave based on those repeated features gives a likely overview of how multiple people identify modeling concepts, there are far more possible features, and thus ways of seeing the world out there (as evidenced by the plurality of the elicited features), meaning that we should not simply close our eyes to them because they do not represent the majority opinion.

Functional and structural nature

As Table 5.2 shows, the largest amount of features we elicited were of a structural nature, excepting the concept of actor, which also had a significant amount of functional features (likely because actors are inherently associated with what they do: their functionality). Most of the features that can be seen in the other Tables corroborate this quite clearly, since most often they deal with some structural aspect a concept has (e.g., having a label, being some specific kind). This is in line with the findings from Malt and Johnson (1992) where it was also found that most artifactual categories were identified primarily by structural features. The reason given for this difference is that most concepts have multiple (often directly observable) structural features by which we identify them (even for abstract objects), but often only few main functions that we associate with them.

Table 5.4: The main modalities we found features of in the elicited features. We distinguish between alethic (what is or has to be), temporal (what will be), deontic (what should be), and mereological (what is part of what).

modality	feature prefix	example of feature
alethic	is ...	an actor is human, an event is triggered by something, a goal is based on an objective, a process is composite, a resource is used by an actor, a restriction is valid until some time, a result is material
	has ...	an actor has capabilities, an event has consequences, a goal has a related objective, a process has steps, a resource has a quality level, a restriction has to be made explicit, a result has value for someone
	does not ...	an event does not have a duration
temporal	always has ...	an actor always has an input, an actor always has an output
	becomes ...	a goal becomes an evaluation criteria for future accomplishments
deontic	has to ...	a resource has to be protected, a goal has to be fulfilled, an event has to be monitored, an actor has to work under constraints
	can be ...	an actor can be human, an event can be caught by something, a goal can be refined into a subgoal, a resource can be available
	may ...	an event may be notified to listeners, a goal may be quantifiable, a process may involve roles, a resource may be decomposed
mereological	is part of ...	an actor is part of a business functional

This results in a smaller amount of functional features elicited from each participant. This means also that these feature sets are likely less (modeling task) context-dependent than if they were identified primarily by functional features, since the structure exists independently of the actions taken or the environment in which the entity is found.

Modalities found in the results

Apart from the distinction between structural and functional features, we found some reoccurring modalities expressed in the dataset. Table 5.4 gives an overview of these, with some specific examples. A large group contains features of an *alethic* nature, which talk about things that are or are not the case (both functional and structural); for example, an actor being a human thing, an event being triggered by something. These types of features clearly state that some feature has to be the case, and thus needs to be true for some thing in order for it to be that concept.

Some *temporal* features were also present in the data set, dealing with features having specific time-bound aspects. These temporal features are sometimes a stronger version of a similar alethic feature, like for example with the feature 'an actor always has an input', which is a stronger version of the logical requirement that 'an actor has an input'. Others are interesting because they transform structural features of some concepts, like a goal becoming a specific thing (e.g., an evaluation criteria) which specializes (and perhaps narrows) the set of needed and sufficient features for it to be a goal.

The other largest class of features we found were those of a *deontic* nature. These deal with features that should be the case, like a 'resource has to be protected', 'an actor has to work under constraints'. The difference of these features compared to their alethic counterparts is that they are less strong in stating what is the case in the subjective reality of the modeler, and are thus likely weaker features for rejecting something as not being a proper instantiation of a concept. In this case, even though a resource has to be protected, some particular thing that is not protected, but has enough other features that correlate with resources, will likely still be identified as being a resource. Thus, the features of modal nature are less strong when it comes to rejecting the identification of something as a specific concept.

Finally, we found some *mereological* features, which express part-whole relationships. For example, the 'an actor is part of a business function' is stronger than a similar subset/type, as here the business function cannot exist without the actor. However, compared to the total amount of results these kind of features were few and far between and thus likely are not widespread amongst practitioners. It is interesting that the amount of mereological features is so low, as they could be inferred from some other features (for instance, if an actor is an absolutely necessary thing for some process, then it makes sense to see that actor as a part of the whole business process; the process collapsing and not existing without that actor). The simplest explanation for this would be the use of terminology from which it cannot be unambiguously inferred that a mereological feature was intended. If people are simply used to express these types of features also with more general terminology like, e.g., 'belongs to', or 'fits in', it is difficult to estimate, from their verbalizations alone, whether a statement about some mereological structure was intended. Reasons for this lack of features could be that either people do not conceptualize these more stringent part-whole relationships and thus do not verbally express them either. A stronger focus on thinking of part-whole relations in terms of sub and super typing originating from widespread database thinking could be an explanation for that. A lack of support from modeling languages or methods could also be intuitively argued for as a reason. However, this does not hold as a widespread language like UML already supports the definition of constrictions on the structure of things with 'containment' relationships, where the whole seizes to exist once a single element is removed. Thus, there seem to be existing ways to both reason and encode mereological structure in modeling languages.

5.4.2 Limitations

The primary concern we have for the internal validity of the study discussed in this chapter and the presented results is the multi-lingual component of the answers, and the related difficulty in ensuring that meaning was correctly preserved during translation to a common language for the analysis phase. However, given that most responses were given by people fluent in both English and their native tongue, and the translations were validated by multiple readers, we are confident that no errors were introduced here.

The external validity of this study is ensured partially by the varied background of the participants: we have responses from people from multiple kinds of sectors (government, for-profits, educational institutes), and from different countries (USA, the Netherlands, Germany, France, etc.). We aim to increase the external validity (that is, the generalizability of the sets of often repeated features) by incorporating another quantitative study that will further clarify the exact typicality of each of these features. However, because of the nature of the feature sets and the significant amount of terms that need to be investigated, this is a significant research effort which will take a longer running time in order to gain enough results, as will be discussed in Chapter 9.2.

5.5 Summary & outlook

In this chapter we have shown the results of a study into what features conceptual modelers typically identify common modeling concepts by. We discussed some interesting details of the data, including common modalities found in them, and the need for modeling languages to explicitly accommodate them. Related to one of our main research questions, "*To what extent do conceptual understandings differ on key points between different people (e.g., different sets of features, typicality)?*" these findings as discussed thus offer an insight into some of the key aspects – the typical set of features – of the conceptual understanding people have of the modeling meta-concepts.

These features can be used as a guideline for what aspects are important to focus on during modeling sessions, as they represent the subtly different ways in which conceptual modelers interpret common modeling concepts, and might thus, unbeknownst to them, disagree with each other. This will be discussed in more detail in Chapter 7.4. Now that we have performed these investigations into the external category structure and a first glance at the internal structure of these meta-concepts, we will go into more detailed studies of the internal structure with the in-depth semantic differential studies described in Chapter 6.

CHAPTER 6

On Conceptual Understandings

Abstract. The aim of this chapter is to have a more in-depth look at the internal structure of the meta-concepts. It incorporates two semantic differential studies into the conceptual understanding people have of the meta-concepts. First a small group of modeling practitioners, all professionally active in an enterprise modeling context is investigated to see how they view the meta-concepts on the basic set of dimensions. Second, the results of a longitudinal study among computing and information systems science students into the conceptual understanding of the same meta-concepts is presented. Of particular interest here was the potential evolution of the conceptual understanding that students had over the course of their studies. Finally, we compare the results of both practitioners and students. We initially expected that practitioners would have more specific conceptualizations related to the particular languages they use, the domain or aspect they operate in, whereas students, being far less experienced, would have more general conceptualizations of the same meta-concepts. These assumptions both proved to be false, of which we will discuss the potential cause and consequences of as a reflection on both studies.

The content of this chapter is based on work published as:

Dirk van der Linden, Khaled Gaaloul, Wolfgang Molnar: Initial Results from a Study on Personal Semantics of Conceptual Modeling Languages. In: G. Bouma et al. (Eds.): *NLDB 2012, LNCS 7337*, pp. 360–365. Springer, Heidelberg (2012)

Re-published with extended abstract in *Proceedings of the 7th SIKS Conference on Enterprise Information Systems (EIS2012)*, Nieuwegein, the Netherlands

Dirk van der Linden, Henderik A. Proper: On the cognitive understanding of types in modeling languages. In: Jung and Reichert (Eds.): *Enterprise Modelling and Information Systems Architecture (EMISA 2013), LNI 222*, pp. 149–162, GI, 2013

Dirk van der Linden, Henderik A. Proper and Stijn J.B.A. Hoppenbrouwers: Conceptual understanding of conceptual modeling concepts: a longitudinal study amongst students learning to model. In: L. Iliadis, M. Papazoglou, and K. Pohl (Eds.): *CAiSE 2014 Workshops, LNBIP 178*, pp. 213–218. Springer, 2014

6.1 Background

In the previous Chapters we have set out a detailed overview of the meta-concepts common to conceptual modeling languages and on what semantic distinctions they can be discriminated. We then showed that people judge these meta-concepts to be of a graded nature, where some particular things are better examples of such a concept than others, further detailing this by eliciting the features that people most typically associate with these meta-concepts. While this gives us some insight into the structure and understanding of these concepts, it does not yet give us a detailed hint at the attitude people have towards them – what these meta-concepts *are* to people.

It is important to be aware of what these meta-concepts are to people, because the models in which we use them are there to support the building and exchange of knowledge (Stahl 2000). While the quality of such models are dependent on many aspects and can be interpreted differently depending on what is expected from the model (cf. Fettke et al. 2012, Houy et al. 2012), it is contingent on their ability to communicate any intended meaning clearly and completely (Robinson 2006). Thus, we need to be able to be sure that the semantics in a model reflect the semantics intended by the modeler. This goes further than merely ensuring that the model is capable of expressing those semantics that the original modeler wanted to (with, e.g., test scenarios). We need to additionally ensure that the model does not allow one to express situations that, by the original modeler and stakeholders, are considered conceptually invalid (or simply illogical).

For this to be possible it is necessary to first have more detailed understanding on how people typically understand these meta-concepts, what particular instantiations make sense to them, and what particular instantiations do not. This can be done effectively by characterizing their *conceptual understanding* of the meta-concepts (in regards to the found semantic dimensions), as those directly affect the amount of conceptually valid (instantiations of) models. Such characterizations offer other benefits as well, as performing these studies among both practitioners and students allows for the comparison of typical conceptual understandings between those with more and less modeling experience, experience in particular languages and techniques, and so on. Furthermore, measuring the development (or potential lack thereof) of such conceptual understandings over time can give some insights into the stability of such understandings. This can be used to reason about how fragile the consideration of a particular model's instantiation as conceptually valid is, or if one needs to worry about models becoming conceptually invalid as soon as a modeler changes their mind or has some new insight. Thus, this chapter will focus on studies attempting to understand more clearly what these meta-concepts are to people, in terms of how they find the dimensions relate to the meta-concepts.

6.2 Experimental parameters

This section gives an overview of the experimental parameters for both of the studies discussed in this chapter. Because they share the empirical method used for them, as well as a significant amount of materials (i.e., both investigated the same meta-concepts and dimensions) we discuss the setup of both studies here at once. If some details are specific to one of the studies (e.g., just the practitioners or students), they will be labeled as such.

As explained earlier in Chapter 1, in order to investigate the conceptual understanding people have of a concept we need to understand that we cannot completely measure concepts, but instead can characterize them (cf. Malt et al. 2011, Pinker 2007) by looking at which semantic dimensions resonate strongly to them. This can be used in order to reason about the range and boundaries of the concept. For example, if we know that for someone the dimension 'human' strongly resonates with their view on the concept actor, we can deduce what are typical and atypical examples of that concept for them, and thus clearly characterize 'what' an actor is for them. If we have a number of semantic dimensions to measure for each concept, we could thus clearly characterize someone's conceptual understanding of a particular concept, and distinguish between understandings different people have.

6.2.1 Considerations

The size of the semantic differential we use for our study is quite large. To investigate the seven meta-concepts on seven distinct dimensions we require a total of $(7 \times (7 \times 5) =)$ 245 bipolar scale questions. While such questions can (and should) be answered quite quickly without too much conscious consideration, the sheer amount of questions does represent a significant time investment expected from a participant. Pilot studies showed that on average half an hour to 45 minutes were necessary to fully fill out the differential. This is a significant investment for practitioners active in industry. Because of this, it will be difficult to get a large amount of participants from professional settings. However, reducing the length of the survey is difficult, as it would require reducing the amount of terms below the threshold of statistical significance, or removing entire dimensions altogether. Only investigating a subset of meta-concepts and dimensions per participant would be a potential solution, but this would require a far greater amount of available (and willing) participants than we had access to at the time of investigation.

For the longitudinal study there were some additional considerations to keep in mind. Since the people involved were required to participate in a survey at the end of each semester, and do so for several years in a row, it was necessary 1) to ensure the drop-out rate would be low, and 2) that participants did not skip any surveys. Dealing with a decrease in the number of participants because of drop-out was expected to some degree, as we started the study at the beginning of the participants' undergraduate program, and some of them were bound to drop out, switch majors, or stop responding due to different reasons. In order to deal with this from the beginning we included incentives for students to respond in the form of gift coupons which were raffled among participants that fully concluded the survey during each semester. In order to deal with participants skipping one or multiple surveys of the study (which would lead to an incomplete dataset of their conceptual change, and thus become unusable for comparison with others) we ensured that we reminded participants up to 3 times for each survey if they had not yet completed it.

6.2.2 Method

Materials: The concepts we investigate are the meta-concepts as found in Table 3.2: ACTOR, EVENT, GOAL, PROCESS, RESOURCE, RESTRICTION and RESULT. More information on these meta-concepts, terms and their derivation can be found in Chapter 3.3.1. The terms used for the members of these categories are the terms as used by the modeling languages and methods, based on the official (or most-used) specification. The dimensions we investigate these concepts on are the dimensions presented in Table 3.3, namely: natural, human, composed, intentional, necessary, material and vague. A more in-depth description of these dimensions can be found in Section 3.3.2, page 63. Any combination of these features can then be taken as a characterization of a given concept, for instance an actor being a natural, human, non-composed material thing.

Participants – practitioners: We report on a study investigating practitioners ($n = 12$), which was performed amongst employees of two internationally operating companies that provide support to clients with (re)design of organizations. The participants all had several years of experience as enterprise modelers, creating and using conceptual models and using diverse modeling techniques.

Participants – students: We report on a longitudinal study amongst computing and information science students at Radboud University Nijmegen. We initially gathered students in the very first session at the beginning of their studies, at which 46 students volunteered to participate. Of these, 19 actually participated in the first phase. Over the course of the study, several students either stopped responding (without specific reason given), stopped because they changed their study program, or because they dropped out entirely. At the final measurement, 9 people participated. However, because one of them had not participated in an earlier phase we were forced to reduce the total set down to 8 complete measurements of the total timespan.

Procedure: We use a semantic differential (elaborated on in section 2.2.3) in order to investigate the conceptual understandings which participants have of the selected meta-concepts. For each combination of a meta-concept with a dimension (e.g., ACTOR–*human*) we used five adjectives (found through an earlier pilot study among the same student population) which relate to the specific semantics of the dimension to be investigated, so ensuring statistical significance (Verhagen and Meents 2007). An overview of these adjectives can be found in Appendix C.1. We then constructed a differential with a page for each meta-concept in which we included 1) a priming task to ensure participants responded in the context of conceptual modeling, and 2) a differential in which a bipolar scale of the adjectives was presented to each participant in a random order. They were asked to rate how well each adjective related to the particular meta-concept on a 5 point Likert scale.

Procedure – students: For the longitudinal study, we started the study at the beginning of the students' studies so that we would have a null measurement. From then on, at the end of each semester, students received an email inviting them to a digital implementation of the semantic differential, where they were also asked to detail what courses they had followed, and what new languages or techniques (if any) they were introduced to. In each phase we sent out 3 reminders to participants if they had not yet responded, and afterwards reduced the set of active participants down to those that participated.

Processing: The resulting data from the semantic differential was processed to calculate an average score for each concept-dimension combination based on the individual adjectives used to describe that dimensions. From this we constructed a vector for each concept, which contained scores ranging from 2.0 to -2.0, describing for each dimension how it relates to that concept. We considered scores ≥ 1.0 as positive judgments, and scores $\leq$ -1.0 as negative judgments. Other scores were considered as neutral. These judgments were then used to calculate a percentage wise breakdown of the amount of different polarities (i.e., negative or positive connotations) found for each concept.

6.3 Study among practitioners

This section will focus on how enterprise modeler practitioners understand the meta-concepts as used in most modeling languages in terms of their attitude to the various investigated dimensions. We are particularly interested in any conceptual understandings that display a specific interpretation of a particular meta-concept, where that would affect the range of conceptually valid instantiations (e.g., ACTORS should not be instantiated with human things). We are furthermore interested in studying whether these experienced modelers on average display a specific interpretations for some meta-concepts, or whether these remain more non-specific, and thus more open to different interpretations. Thus, we will treat the following specific questions in this section:

1. What conceptual understanding do practitioners have of the meta-concepts?
2. Are there particular understandings for some meta-concepts that stand out?
3. Are these conceptual understandings significantly specialized or more generic?

With these results we also hope to demonstrate the use of having insights into how people conceptualize modeling concepts in general, and our method being a valid systematic way of uncovering them. In Chapter 7 we will discuss in more detail what consequences our findings have for modeling languages, and the use and creation of models themselves, especially in an inherently collaborative effort like enterprise modeling.

6.3.1 Data

A visual overview of the scores from the semantic differential for all meta-concepts and dimensions is given in Fig. 6.1. These results from the study were processed into the amount of different polarities, and are shown in Table 6.1. The percentages are an aggregate reflecting the amount of negative, neutral and positive results. The amount of neutral responses can be taken as a measure of how open people's typical conceptualization of a concept is, in that it allows for more flexible (and possible amount of) instantiation. Negative and positive responses indicate that an instantiation of the concept would either need to, or not need to display a certain dimension. For example, when it comes to ACTORS, there are 9% negative responses, 53% neutral and 38% positive. This means that about 9% of the responses indicate a dimension that *has to be false* for a typical actor instantiation (e.g., a typical actor should not be a composed thing), 38% of the response indicate a dimension that *has to be true* for a typical actor instantiation (e.g., a

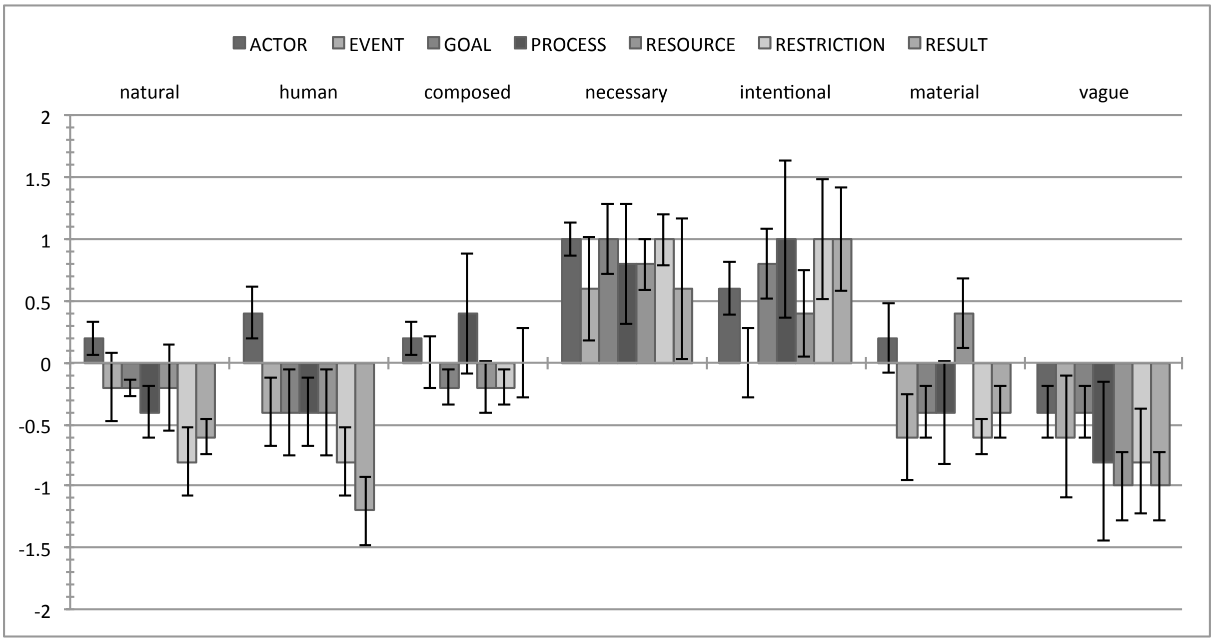


Figure 6.1: Visualization of the semantic differential results from our study amongst practitioners. Shown is the median and standard deviation for each meta-concept and dimension combination grouped per dimension. To illustrate, the first cluster of bars shows in what respect all the meta-concepts are considered *natural* things, followed by the next cluster that shows in what respect the categories are *human* things and so on. The variance in the results of the meta-concepts is respectively 0.38, 0.57, 0.68, 0.93, 0.73, 0.94, and 0.92.

typical actor should be a human thing), while the remaining 53% allow for dimensions to be either false or true (e.g., an actor can be either material or immaterial thing).

Table 6.2 shows a number of specifically strongly polarized meta-concept and dimension combinations. The percentages here denote how many of the participants judged the meta-concept to need (or need not) to display the given (semantic) dimension. For example, 91% of practitioners judged goals to need the 'necessary' dimension, which means that the majority conceptualize goals as things that need to be achieved (or perhaps gained, if goals are also conceptualized as material objects). It can also be seen that 64% of practitioners judged processes to be non-vague things, meaning that they are (or should) be well-described things.

Table 6.1: A detailed overview of the polarities of the responses from the practitioner study.

pol.	nat	hum	com	nec	int	mat	vag	avg.
				actor				
neg	0%	9%	9%	0%	0%	0%	45%	9%
neu	73%	55%	73%	27%	45%	55%	45%	53%
pos	27%	36%	18%	73%	55%	45%	9%	38%
				event				
neg	36%	36%	0%	0%	9%	55%	55%	27%
neu	64%	55%	73%	27%	64%	45%	45%	53%
pos	0%	9%	27%	73%	27%	0%	0%	19%
				goal				
neg	9%	45%	9%	0%	9%	36%	45%	22%
neu	82%	45%	82%	9%	36%	64%	45%	52%
pos	9%	9%	9%	91%	55%	0%	9%	26%
				process				
neg	36%	36%	9%	18%	9%	45%	64%	31%
neu	64%	55%	45%	18%	36%	45%	18%	40%
pos	0%	9%	45%	64%	55%	9%	18%	29%
				resource				
neg	36%	36%	36%	18%	9%	9%	64%	30%
neu	64%	55%	64%	18%	45%	55%	27%	47%
pos	0%	9%	0%	64%	45%	36%	9%	23%
				restriction				
neg	73%	73%	18%	9%	9%	55%	82%	45%
neu	18%	18%	82%	18%	36%	36%	18%	32%
pos	9%	9%	0%	73%	55%	9%	0%	22%
				result				
neg	64%	73%	18%	18%	18%	36%	91%	45%
neu	36%	18%	64%	27%	9%	45%	9%	30%
pos	0%	9%	18%	55%	73%	18%	0%	25%

6.3.2 Discussion

General discussion

As briefly discussed in the previous section, the results as seen in Tables 6.1 and 6.2, directly affect the total amount of conceptually valid instantiations for a given meta-

Table 6.2: Some strongly (> 55%) polarized meta-concept – dimension combinations.

meta-concept	*is ...*	score
goal	necessary	91%
process	necessary	64%
	not vague	64%
resource	necessary	64%
	not vague	64%
restriction	not natural	73%
	not human	73%
	necessary	73%
	not vague	82%
result	not natural	64%
	not human	73%
	intentional	73%
	not vague	91%

concept. The higher the amount of neutral scores, and vice-versa the lower the amount of polarized (and more to the point, strongly polarized) scores, the lesser dimensions need to be exhibited by an instantiation of a meta-concept. Most interesting for our purpose are the polarized scores, as they indicate a dimension that has to be exhibited by the meta-concept, and thus restricts the amount of conceptually valid interpretations for that meta-concept.

The results also show that around half of the judgments for the meta-concepts are of a polar (either positive or negative) nature, indicating that there are many dimensions that are conceived of as either strongly typical or atypical. On an individual level this may vary, as specific people might find all dimensions (a)typical for a given meta-concept, or find no dimensions particularly typical for a given meta-concept. Such differences can be correlated with someone's expertise in a specific area, causing them to more closely scrutinize the meta-concepts relevant to that specific area.

Discussion of meta-concepts

Goals are one of the few meta-concepts where there is but one absolutely needed dimension. Specifically, goals are considered by most (91% of practitioners) to be necessary. What this means is that goals are considered to be logical necessities in the sense of their being achieved, i.e., they are the logical consequence of following a certain procedure to achieve them, and should be achieved then (and likely only) then. This is in contrast to the other option of goals being things that 'merely' *should* be achieved. As a consequence, this seems to imply an implicit bias towards hard-goals, as soft-goals are often used to model goals for which the achievement status is (or cannot) be known a priori. The absence of any other needed dimensions is interesting as well. It could have been expected that, in line with the implicit bias towards hard-goals, the majority of practitioners would consider goals as being non-vague (i.e., well-described), yet only 45% does so.

A slight majority of practitioners deem RESOURCES to be necessary and not vague. This seems to imply that resources, are commonly seen as things that need to be well-defined, before they can be reliably (and systematically) used for some other activity. Such typical interpretations might clash with the way resources are handled in some languages, e.g., the information object in ArchiMate (The Open Group 2012), which allows you to represent a functional piece of information needed by some business process, without requiring you to be explicit about what information it contains.

RESTRICTIONS already have a larger typical dimension set. They are typically considered non-natural, non-human, necessary and non-vague things. The non-vague dimension seems obvious in that it demands us to be clear about what a restriction actually does (perhaps making rules more typical restrictions than something like, say, architecture principles). Non-natural clearly implies that they are not naturally occurring restrictions (e.g., natural laws like gravity) but more like restrictions that are 'created' to control or restrict other things. Related to this restrictions are also not considered to be human things, which in this context might mean that the general way of conceptualizing a restriction is seeing it as the abstract entity itself, and not, for instance, the human being enforcing it. More interesting from a modeling language point of view is the, perhaps obvious, view of restrictions being logically necessary things. This means that a typical restriction is something that has to be adhered to, and not something that can be broken. To accommodate that, a modeling language should either have its restriction meta-concepts be inherently alethic, or allow for a way to explicitly distinguish restrictions with different modality. While there are not many (especially domain-specific) languages that accommodate this, Object Role Modeling (ORM) (Halpin 2005) does have explicit support for denoting elements as being either alethic or deontic, and thus could be a useful choice of language when a domain involves those non-typical restrictions: ones that do not necessarily have to be adhered to.

RESULTS are typically non-natural, non-human, intentional and non-vague. This seems to fit with the common conceptualization of results, because they are seen as 'new' things, and thus shouldn't be already existing naturally occurring objects. This requires a certain level of ontological scrutiny though, because a tree is certainly a naturally occurring entity, as well as a tree branch that naturally fell off from a storm, but a log that results from cutting the tree into parts with a chainsaw is not. Furthermore, a branch that merely breaks off from a tree would not be considered a result in our sense because its breaking off was not intentional. Some results as modeled by languages might seem not to fit and evoke discussion on their ontological status, as, for instance, the human output concept in ARIS (Scheer and Nüttgens 2000). At first glance, it would seem not to fit the characterization as it is likely a human kind of thing. However, ontologically speaking, the actual thing here is the activity performed by a human, not the human itself. Say, the successful delivery of a pizza. As results are not typically considered material things, this seems thus to be a conceptually valid instantiation of a result.

Effects on conceptual validity

To detail how certain combinations of dimensions can change the amount of conceptually valid instantiations for a given meta-concept, some examples are shown in Table 6.3. In line with the explanation of context (in this case, the dimensions) changing the amount of conceptually valid instantiations, we show how some of the common meta-concepts are affected by the results we found.

On the one hand, when a meta-concept has only few absolutely needed dimensions, the amount of conceptually valid instantiations is significant. Conversely, the amount of conceptually invalid instantiations is quite small. A goal is a good example of this, as the only needed dimension is it being a necessary thing, which means we can instantiate it as an actual physical object to be achieved (i.e., the goal is the pizza that will be baked), the activity of having done so (i.e., the goal is the final state of the pizza baking process) or even more abstract, an abstract object that follows from its creation (i.e., the goal is the information gained when the pizza is baked). The few counter-examples are those where the goal is not necessarily achieved, which could be either something that is not achieved, or something that was achieved by accident (e.g., as a side-effect from another process). The main problem here is that the counter-examples (and examples themselves too, perhaps) are so broad that it is difficult to figure out whether an instantiation was actually a valid instantiation or not.

On the other hand, when a meta-concept has many needed dimensions, it is easier to find out whether something is a conceptually valid instantiation or not. When it comes to results, certain objects are clearly valid instantiations, like a freshly baked pizza or a specification of a modeling language. The latter can actually be a result in two ways, as both the actual specification of the language is a valid result, as well as the representation

Table 6.3: Examples of difference in conceptually valid instantiations for some concepts reflecting some of Table 6.2's results.

type	is a . . . feature-wise	example	counter-example
goal	necessary thing	any thing to achieve, whether physical (e.g., producing a representation of a model) or non-physical (e.g., producing the information needed for it) entity to be acquired or some state to be reached, and so on	a non-necessary goal, or thing that is achieved as a side-effect of achieving something else
restriction	not natural, not human, necessary to adhere to and well-defined thing	legal rules given in a state's laws, well-known and documented natural laws	etiquette, informal dinner rules, or natural laws like gravity
result	not natural, not human, intentionally achieved well-defined thing	a baked pizza, a sawed log, a specification of a modeling language, a representation of the specification of a modeling language	a newly hired person, the outcome of trial-and-error testing

of the language (i.e., the actual written documents describing it). Counter-examples are then also easy to come by, for instance, the actual person hired as the 'result' of a hiring process is not typically considered a result (by virtue of being human). Furthermore, any 'results' that were the outcome of trial-and-error approaches are also on shaky grounds, as they are not typical results by virtue of not having been necessary.

Limitations

Even though the results we have shown are interesting and could be used for a multitude of purposes, care must be taken not to immediately extrapolate the results of this relatively small study and infer general truths from them. Different groups with varying backgrounds and specializations might have different conceptualizations, and thus further studies might uncover additional and conflicting conceptualizations. As such, the contribution of this work has been in demonstrating the possibility to systematically uncover such conceptual distinctions and laying the ground for further investigations.

6.3.3 Conclusion

In this section we have shown an overview of how practitioners conceptualize the common meta-concepts. We discussed how these conceptualizations can affect the range of conceptually valid instantiations for such a meta-concept, which can have a negative effect on the validity and usefulness of created models if we do not take care to discover and communicate these conceptualizations. Furthermore, some languages might be more suited to deal with certain conceptualizations by virtue of explicitly expressing conceptual distinctions that other languages leave implicit. Care should thus be taken to also select an appropriate modeling language that fits most with the conceptualizations of the involved modelers and stakeholders. This will be discussed in more detail in Chapter 7.1.

6.4 Longitudinal study among students

In this section we will discuss our investigation into the conceptual understanding that students have of the common meta-concepts used for conceptual modeling. In contrast to the study among practitioners, our focus here was more on if and how those understandings change over time during a student's progress through their academic curriculum. To this end we followed a group of students starting computing and information science studies at Radboud University Nijmegen (see Section 6.2 for experimental details). We followed them from the beginning of their studies as they learned new theories, techniques, and languages for modeling. We focused on investigating whether their conceptual understandings changed as they became acquainted with new languages and techniques, and whether there were correlations between the introduction of such educational stimuli and changes in their conceptual understanding.

6.4.1 Background – on training modelers

For concepts we often use in a professional setting (like the meta-concepts we focus on in this dissertation), education and training come into play already early on in the career of a modeler. We attempt to shape the way people see certain things by exposing them to particular viewpoints, and interpretations of things, in order to shape their conceptual understanding and make it easier to communicate and work together. This is meant to ensure that people's conceptual understandings are close enough so that no mismatched understandings between different modelers and stakeholders remain, which otherwise could have dire consequences for the modeling process and any produced models (Kaidalova et al. 2012). An example of such a process is the use of 'Semantic Reassurance' (Hoppenbrouwers 2003) during modeling sessions, where modelers and stakeholders clarify their interpretations of a particular concept to each other and attempt to come to a common ground for that concept.

However, before any of these techniques and processes come into play, there is the education of the people involved in the modeling process. For new students in computing and information science (who represent a significant source of later professional conceptual modelers in the Netherlands) this is where they are first introduced to the core modeling concepts and start shaping their understanding of them. The curriculum teaches them both new ways of thinking (e.g., object-orientation in terms of programming, fact-orientation in terms of modeling), new ways of working (e.g., specific analysis methods), and the tools to use with them (e.g., specific languages like ORM, BPMN, UML). Afterwards, much training is done in-house to educate people in a specific language/methodology, and way of thinking/working, as was our experience working with practitioners and companies. In other words: apart from familiarizing them with specific tools, companies/practitioners aim to ensure that (new) people think and conceptualize in the way they want.

Much of the effort of this kind of training specifically for conceptual modeling seems to rely on a basic assumption: *that such training has a controllable effect on the conceptual understanding the trainee has of the modeling concepts.* However, there is little in the way of studies attempting to verify whether such effects are systematically achieved. For this reason, we are interested in understanding whether students develop specific conceptual understandings, or perhaps conceptual prejudices, when it comes to the meta-concepts. More concretely, we want to investigate whether the curriculum that a student follows might push them into a specific interpretation of these meta-concepts, and if so, what kind of effect they have on their conceptual understanding. Because most academic programs are focused on training well-rounded people who can orient themselves in new conceptual environments, we can assume that the point is not to steer people into specific, narrow conceptual understandings (i.e., strongly biasing people into accepting one kind of thing as correct), but instead to focus on opening their minds to many different, equally valid viewpoints from which they can analyze multiple situations (i.e., to steer them in a direction where their conceptual understanding allows for many possible correct things). On the one hand, academic training seems to promote being open-minded, (i.e., by treating many different viewpoints and approaches), on the other hand, it might also work against that (i.e., by training people in specific languages with specific conceptual biases). To focus on this, we will treat the following specific questions in this section:

1. Does the conceptual understanding that students have of modeling concepts become more refined or nuanced as they progress through their studies?
   (a) If there is such a change, is it of a discrete or continuous nature?
   (b) If there is such a change, is it one-directional or reversible?
2. Is there a correlation between the educational stimuli students receive and the possible change in their conceptualizations of modeling concepts?
   (a) Do conceptualizations take the form of the semantics of a specific language or approach?

In order to treat these questions, we will investigate the base conceptualizations that students have of the selected modeling concepts. In particular, this means that we study the core connotations they have of these concepts when primed in the context of conceptual modeling, without focusing on a specific kind of model or modeling issue. This should allow us to see whether the training they receive affects the actual cores of the conceptualizations, without dealing too much with the ever-changing content of the concepts due to contextual effects. For example, a specific focus on goal modeling which might give rise to radically different conceptualizations compared to a specific focus on process modeling.

6.4.2 Data

Conceptual understandings

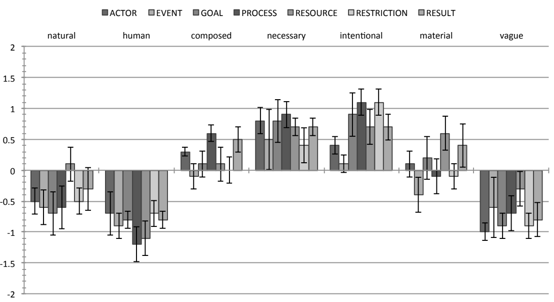
The polarities we calculated show the relative amount of positive, neutral and negative responses to each meta-concept-dimension combination and are shown in Table 6.4, with some potentially interesting ones detailed later in Table 6.6. We present a visualization of the concept-dimension scores in Fig. 6.2. It shows the averaged results for each concept-dimension combination for each phase, with error bars showing the range of individual results. A more detailed overview of the specific meta-concept and dimension polarities for each phase of the longitudinal study can be found in Appendix C.3.

Table 6.4: Average polarities over all concept-dimension responses for each phase of the longitudinal study. Polarity scores of individual concept-dimension combinations are excluded due to space constraints, but are available upon request.

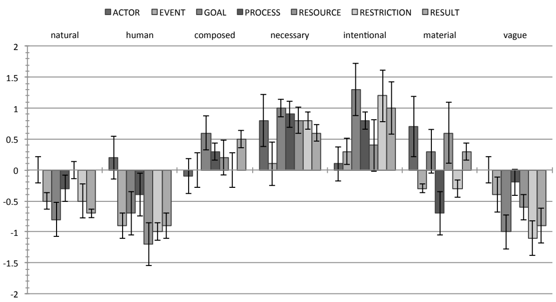
polarity	actor	event	goal	process	resource	restriction	result
				phase 1			
neg	26%	31%	31%	27%	21%	30%	20%
neu	53%	57%	46%	43%	47%	51%	54%
pos	21%	11%	23%	30%	31%	19%	26%
				phase 2			
neg	10%	30%	26%	20%	17%	26%	27%
neu	63%	57%	46%	63%	56%	51%	49%
pos	27%	13%	29%	17%	27%	23%	24%
				phase 3			
neg	7%	36%	31%	24%	24%	30%	26%
neu	57%	56%	41%	53%	46%	49%	49%
pos	36%	9%	27%	23%	30%	21%	26%
				phase 4			
neg	10%	37%	30%	30%	20%	29%	29%
neu	60%	49%	37%	37%	49%	44%	36%
pos	30%	14%	33%	33%	31%	27%	36%
				phase 5			
neg	17%	33%	31%	29%	24%	33%	24%
neu	59%	59%	39%	51%	44%	46%	47%
pos	24%	9%	30%	20%	31%	21%	29%

Educational stimuli

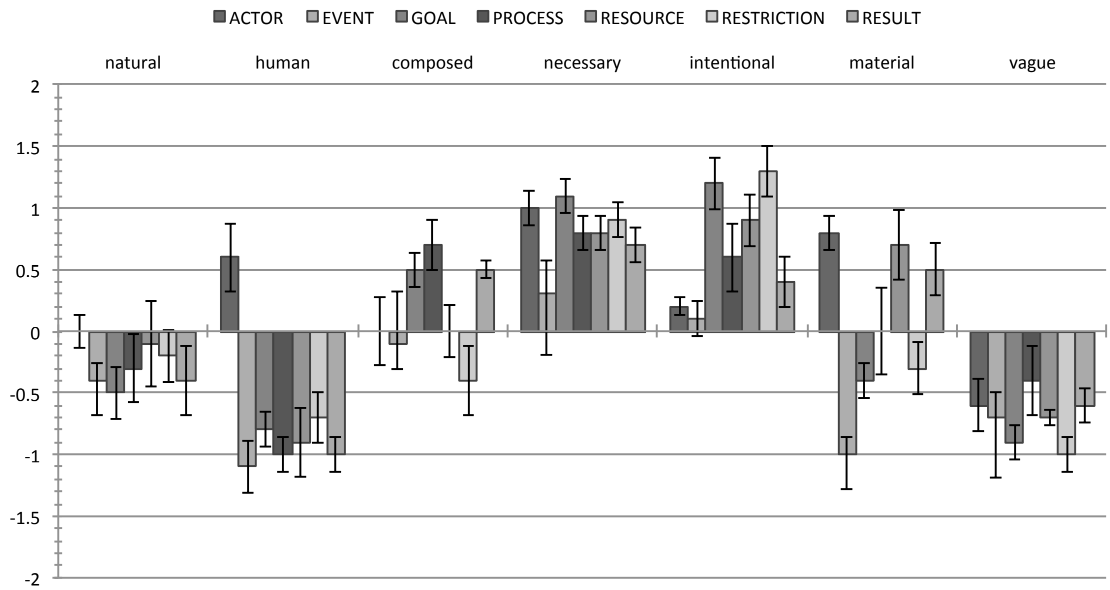
The educational stimuli are detailed in Table 6.5. They include the courses that students followed, and languages and techniques they were introduced to in the semester before that particular measurement was made.



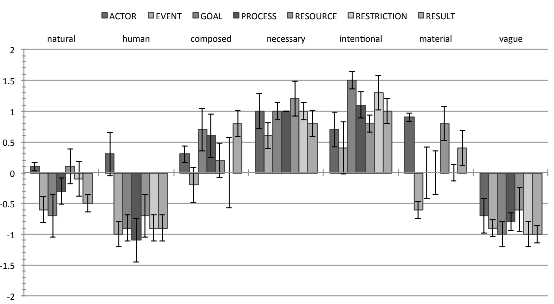
(a) Phase 1



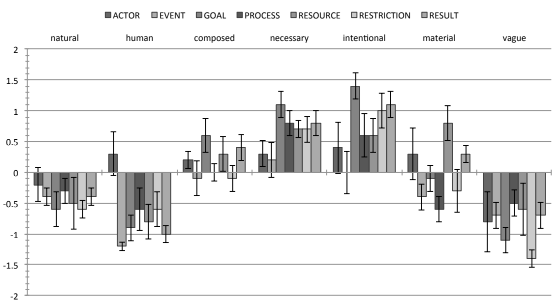
(b) Phase 2



(c) Phase 3



(d) Phase 4



(e) Phase 5

Figure 6.2: Visualization of the average concept-dimension scores and individual variations for each phase of the longitudinal study.

6.4.3 Discussion

General discussion

We will first answer the questions we posed before moving on to more detailed discussion of the individual results. Our primary question to do with whether teaching students

Table 6.5: Overview of educational stimuli students reported receiving before the measurement. Each phase was measured at the end of a semester. Phase 1 is a null measurement at the start of the participants' studies, hence no stimuli were reported.

at ...	languages	course topics
phase 2	C++, SQL, ORM, UML, Finite automatas, Formal Grammars, XML	Domain Modeling, Formal Thinking, Programming, Discrete Mathematics
phase 3	Java, ORM, UML, C++, Assembly, Ruby, Clean, Logic	Digital Information Architecture, Mathematics, Research & Development, Object Orientation, Processors, Applied Logic, Business Process Modeling, Languages & Automata, Artificial Intelligence, Requirements Engineering
phase 4	Turing Machines, Haskell, Clean, Ruby, Clean, Prolog, abstract ORM, ORC, Clean	Information Systems Analysis, Functional Programming, Mathematics, Security, Efficient Storage of Data, Reliable Storage of Data, Transformation and Implementation
phase 5	UML, Prolog, NXC, Clean, Prolog, Structural Operational Semantics, Natural Semantics	Innovation Management, Human-Machine Interaction, Semantics & Correctness, Requirements Engineering, Knowledge Representation, Research & Development, Functional Programming

languages and techniques for modeling affects their conceptual understandings was "*Do the conceptualizations students have of modeling concepts become more refined or nuanced as they progress through their studies?*". We can answer this by looking at both Fig. 6.2 and Table 6.4. In Fig. 6.2 we see that there is not a clearly obvious shift for any of the concepts or individual concept-dimension combinations to a particular understanding. For this to happen, the bars should either gradually or suddenly switch from ranging to one of the extremes to the other, or stay neutral in the middle. However, as we can see over time the general pattern of all the results stays similar, not having any of its constituents change too much. The semantic dimensions natural, human, and vague stay mostly negative for most concepts, while the dimensions composed, necessary, intentional stay positive. The dimension material is the one dimension in which we clearly see both positive and negative polarities for different concepts, although these particular concept-dimension combinations still do not seem to change much over time.

We can look in more detail at Table 6.4 to compare the actual distribution of the polarities for the concepts through time to verify this lack of systematic change. Here we also see that, while there are subtle variations from phase to phase in the relative amount of negative, neutral, and positive responses, there does not seem to be a significant gradual change increasing or decreasing over time for any of them. However, some specific concept-dimension combinations, do seem to have gradual shifts in one direction, which are documented in Table 6.6.

Table 6.6: Notable shifts of conceptual understanding in the results of the average (i.e., all participants) polarity scores.

polarity	p1	p2	p3	p4	p5	primary trend
		humanity of results				
neg	60%	60%	80%	70%	80%	
neu	40%	40%	20%	30%	20%	stronger negation
pos	0%	0%	0%	0%	0%	
		necessity of results				
neg	0%	10%	10%	0%	0%	
neu	50%	40%	30%	30%	30%	stronger acceptance
pos	50%	50%	60%	70%	70%	
		vagueness of actor				
neg	60%	20%	40%	50%	50%	
neu	40%	80%	60%	50%	50%	slight decrease in negation
pos	0%	0%	0%	0%	0%	
		naturality of actor				
neg	40%	0%	0%	10%	20%	
neu	50%	70%	80%	90%	70%	increase in neutrality
pos	10%	30%	20%	0%	10%	

As a result, our first subquestion, "*If there is such a change, is it of a discrete or continuous nature?*" becomes irrelevant. However, the second subquestion "*If there is such a change, is it one-directional or reversible?*" is still interesting to look at, as the data do show that there are sometimes shifts for specific concept-dimension combinations where the polarity changes, and reverses again over the course of our study. While this might also be attributed to individual or contextual factors, it can hint at the flexibility of the students in their conceptual understandings while focusing on a specific way of thinking and working (e.g., because in one semester they work in a different paradigm than the others).

Our second main question, "*Is there a correlation between the educational stimuli students receive and the possible change in their conceptualizations of modeling concepts?*" and the related subquestion "*Do conceptualizations take the form of the semantics of a specific language or approach?*" can be answered by also taking into account the data given in Table 6.5. There do not seem to be specific systematic shifts that can be correlated with educational stimuli, nor do they seem to be systematically widening or refining to fit a specific way of thinking that could be attributed to them (e.g., the strong fact-oriented thinking approach of ORM (Halpin 2005)). Given that students used several languages and techniques almost from the beginning of their studies until the final measurement, one could have expected to see some kind of development towards fitting those ways of thinking. However, given the lack of specific shifts into particular conceptual understandings discussed for question 1, this seems unlikely as well.

As touched upon earlier, there were some specific concept-dimension combinations that did show a development towards a specific conceptual understanding. Notable ones are shown in Table 6.6. These are specifically a stronger rejection of the possibility of results (of something) being human things, a much stronger view that results are necessary to achieve, a very slight decrease in rejecting a vague thing being an actor, and a strong increase in neutral views on whether actors can be natural things. These patterns all show an example of a different polarity gaining or losing ground, which all translate into the willingness of a specific person accepting or rejecting a particular thing as being a good example of that modeling concept. When we see that someone has a much stronger negative view on a particular thing (i.e., here the humanity of results), during modeling sessions those might come to the foreground when people clash on their conceptualization of the universe of discourse. Finding such specific strong polarized concept-dimension combinations might thus be a useful aid in steering such discussions to avoid communication breakdowns.

Limitations

Due to the amount of people that dropped out during the study, and the generally low initial response rate we cannot guarantee a strong external validity due to the lack of statistical generalizability. This could have potentially been prevented by including multiple, parallel groups of students (originating from different universities). However, this would lead to a strong heterogeneity of the results because different academic institutes and programs focus on different aspects. As such, whether those results could be first combined in order to create a larger coherent set of data is debatable as well. Nonetheless, the results here are still a thorough examination of specific individuals, and can be used to reason about the effects found in them, and to what degree measurement of their conceptual understanding is a feasible, and useful endeavor.

6.4.4 Conclusion

In this section we have discussed the results from our longitudinal study which investigated how the conceptual understanding a group of students have of the meta-concepts changed over time. Surprisingly, we found little systematic change in their conceptual understanding, especially when attempting to correlate such changes with specific educational stimuli (e.g., courses taken, languages learnt). These findings can have some consequences for the way in which we train people to be modelers, and the (ease of) adoption of specific conceptual modeling languages. However, we will first use Section 6.5 to go into the differences (or lack thereof) between the conceptual understanding that practitioners and students have of the meta-concepts.

6.5 Comparing practitioners and students

Given that we performed these studies in two groups with quite different levels of experience in conceptual modeling, it would be interesting to pay some explicit attention to the differences and similarities between their conceptual understandings of the same meta-concepts. As the study with practitioners was performed with a number of people who all had multiple years of experience in applying conceptual modeling, were intimately familiar with some modeling languages and techniques, and were usually specialized in some particular domain (e.g., finance, government, healthcare), one could expect some differences when comparing them to undergraduate students just beginning in computing and information science. Because many students in this program go on to have a career in industry in these kinds of companies, seeing whether they develop into the already experienced practitioners would be interesting. From our basic assumptions on the difference between practitioners and students regarding their experience and knowledge about techniques and modeling languages, we would have expected a difference in their conceptual understandings in one of two ways:

1. Practitioners displaying (much) more polarized conceptual understandings
2. Practitioners displaying (much) more neutral conceptual understandings

While this might seem contradictory, there is a fitting explanation for both possibilities. On the one hand, if practitioners were to display significantly more polarized understandings (i.e., much more strong judgments stating that a particular dimension has to be reflected for a given meta-concept: an ACTOR *has* to be human), this could be in line with their larger amount of experience and familiarity with modeling languages and techniques. If such a person has, for example, worked in the finance domain for a decade, using BPMN to model taxation processes, his typical conceptual understandings might have strongly shifted in favor of those valid and most used for that domain. Thus, one could expect such strongly polarized responses from practitioners as they all share this larger amount of experience and expertise compared to the investigated students.

On the other hand, there is also an argument to be made for the case where practitioners would display significantly more neutral conceptual understandings. That is, they would be far less likely to be absolute in stating that a particular dimension has to be reflected for a given meta-concept (e.g., an ACTOR can be a human, but also a machine, or an abstract thing). With the large amount of experience practitioners have, they are likely to have been exposed to many different viewpoints and interpretations of the domain they work in. It can be expected of such a modeler that they are capable of identifying and understanding the way someone thinks. If they have done so for many years, dealing with different people all with different ideas, one could expect them to become more open-minded, and less likely to rule out particular conceptualizations as being 'incorrect'.

Either way, one would expect some noticeable difference between the conceptual understandings that practitioners and students have. As can be seen in Table 6.7 this is hardly the case. While the proportion of responses for the student study varies between phases, the average trend that can be seen in the column 'average' seems to remain around ∼30% for both polar scores, and between 40–50% for neutral scores. On an individual level there are, expectedly, large differences to be found. Similarly, there are differences for particular meta-concepts, as it can be seen that practitioners are less neutral about RESTRICTIONS and RESULTS. The main difference here is that practitioners have more negative scores for dimensions that do not apply to them (e..g, a RESTRICTION cannot be a naturally occurring thing). However, it is the difference between the groups and averages that would have to be significant in order to be able to say anything about how different practitioners and students truly are. Given the lack of significant difference in the average proportion of scores, it thus does not seem to be the case that there is a systematic difference between how polarized or neutral practitioners and students are in their conceptual understandings of the meta-concepts. In Table 6.8 we also compare the development of students' conceptual understandings for a number of specific meta-concepts and dimensions (based on highly polarized combinations found in the practitioner study). As can be seen, and was

Table 6.7: Comparison of the average amount of polar (*neg*ative, *pos*itive) and neutral (neu) responses between practitioners ($n = 12$) and the final set of student responses ($n = 8$). Both studies yielded comparable proportions of polar responses ($25 - 30\%$ positive and negative), with only a slightly larger amount of neutral responses ($\sim 5\%$).

	actor	event	goal	process	resource	restriction	result	average
				Practitioners				
neg	9%	27%	22%	31%	30%	45%	45%	30%
neu	53%	53%	52%	40%	47%	32%	30%	44%
pos	38%	19%	26%	29%	23%	22%	25%	26%
				Students (phase 1)				
neg	26%	31%	31%	27%	21%	30%	20%	27%
neu	53%	57%	46%	43%	47%	51%	54%	50%
pos	21%	11%	23%	30%	31%	19%	26%	23%
				Students (phase 2)				
neg	10%	30%	26%	20%	17%	26%	27%	22%
neu	63%	57%	46%	63%	56%	51%	49%	55%
pos	27%	13%	29%	17%	27%	23%	24%	23%
				Students (phase 3)				
neg	7%	36%	31%	24%	24%	30%	26%	25%
neu	57%	56%	41%	53%	46%	49%	49%	50%
pos	36%	9%	27%	23%	30%	21%	26%	25%
				Students (phase 4)				
neg	10%	37%	30%	30%	20%	29%	29%	26%
neu	60%	49%	37%	37%	49%	44%	36%	45%
pos	30%	14%	33%	33%	31%	27%	36%	29%
				Students (phase 5)				
neg	17%	33%	31%	29%	24%	33%	24%	27%
neu	59%	59%	39%	51%	44%	46%	47%	49%
pos	24%	9%	30%	20%	31%	21%	29%	24%

discussed earlier, the average tendency of the students does not (systematically) change much over time. Furthermore, the scores are quite similar to each other, with practitioners and students sharing the same polarity (i.e., either being positive or negative) for the listed meta-concept and dimension combinations.

Table 6.8: Comparison between some specific strongly (> 55%) polarized meta-concept and dimension combinations scores for practitioners and students. For students average scores from each phase are shown in chronological order.

meta-concept	*is ...*	score	score (student)
goal	necessary	91%	70%,60%,60%,70%,70
process	necessary	64%	70%,50%,50%,70%,60
resource	necessary	64%	60%,60%,70%,70%,70
	not vague	64%	30%,50%,70%,50%,50
restriction	not human	73%	50%,70%,50%,50%,50
	necessary	73%	40%,60%,60%,80%,50
	not vague	82%	70%,50%,70%,80%,70
result	not human	73%	60%,60%,80%,70%,80

What could be the reason behind this lack of strong difference between practitioners and students? We have to acknowledge that with the small dataset presented here (in particular, 12 practitioners compared to 8 students), it would not be fair to generalize and claim this specific comparison holds in general. However, even with that in mind, the sheer lack of overall difference between practitioner and student is difficult to understand.

It could be the case that polar responses are overrepresented in the results from the students because, even though they were not sure about some meta-concept and dimension combinations they felt pressured to give non-neutral responses (e.g., not wanting to click the equivalent of "*don't know*" to every question). While the data of the surveys were treated confidentially, so anonymity in regards of other students or instructors being aware of responses was guaranteed. Controlling for such an effect beyond this would be difficult. Yet, equally the same argumentation could be used for overrepresentation of neutral answers, as it could be reasoned that students did not want to 'be wrong' about their answers, even though it was explained that the personal attitude and feeling participants had was the focus of the study. Exit interviews held with the students after each measurement (and processing of the data) could have been used to corroborate the findings and figure out whether this was the case.

Another option would be to execute a more contextualized investigation where practitioners are asked for their conceptual understanding of meta-concepts not in the general context of modeling, but specifically in the context of the domain they focus on, and the particular kinds of models they work on. However, that would lead to more difficult comparisons with student data, as with students it is feasible to investigate the base conceptual understanding of the meta-concepts, but as they lack experience in applying it in a specific domain, contextualizing the investigation in the same way would be difficult.

This lack of significant difference between how practitioners and students conceptualize the meta-concepts returns as well in a different study discussed in Chapter 8, where the clustering of people by their real-world properties (e.g., modeling languages used, educational background, professional focus and operation domain) and measured semantics was compared, and for both groups little correlation between measured semantics and real-world properties was found.

Going from student to practitioner

This lack of difference between the conceptual understandings of practitioners and students is of particular interest now for further discussions on how to best educate and train people to become effective conceptual modelers. As there does not seem to be much of a difference (and perhaps, implied, not a necessary difference) in how students and practitioners think about meta-concepts, perhaps the main difference is in their familiarity and skill with particular languages and techniques. The findings of the longitudinal study, that people do not seem to be affected in their conceptual understanding by learning new languages and techniques resounds with other research (e.g., Recker and Dreiling 2007). Such recent work into the training and education of conceptual modelers has discussed similar observations, and argued that teaching and training should thus not focus on teaching specific tools and techniques, but instead on different understandings and how people view the world.

6.6 Summary & outlook

In this chapter we have presented the results of two semantic differential studies which investigated the conceptual understanding that people had of the common meta-concepts. The goal of these studies was to gain a more detailed understanding of how different people understand these meta-concepts, in terms of a number of semantic dimensions. While the results of both studies showed that there is a wide variety of understandings differentiated to effectively the personal level, there are some conclusions that can be drawn. In regards to our research questions as given in Section 1.5.2, we can reflect on some specific questions:

What is the conceptual understanding that people have of these meta-concepts? The conceptual understanding people display of these meta-concepts seem to be highly personal and potentially vary wildly from one person to another. Some attitudes of particular dimensions to specific meta-concepts do stand out (e.g., shared strong attitudes towards almost all meta-concepts being non-vague, highlighting the desire to be exhaustive and correct in our descriptions of things), but overall the lesson to be drawn here seems to be that each modeler is a person in their own right, with their own particular conceptual idiosyncrasies, and during the modeling process should thus be approached and understood on a personal level.

Does a person's conceptual understanding of a meta-concept change over time?" While our longitudinal study has shown that the conceptual understanding that someone has of these meta-concepts changes over time, there seems to be little rhyme or reason to *how* it changes. The subquestion "*If so, does training in particular languages or techniques affect these base conceptual understandings?*" is thus particularly important here, as in the longitudinal study there did not seem to be any correlations between the way conceptual understandings changed and what educational stimuli (the learning of a new modeling language, adoption of a new technique, and so on) were presented.

Part III

Practical Findings

CHAPTER 7

Accommodating Conceptual Understandings

Abstract. In this chapter we will elaborate on some of the implications that our findings have for (the practice of) conceptual modeling. While the findings themselves as they have been presented in the previous chapters are more of a fundamental nature, they relate to a number of more practical aspects. When people wish to model, they need to *select an optimal language*, which we will treat from the aspect of conceptual accommodation: analyzing a number of modeling languages to see what kind of meta-concept and dimension combinations they (implicitly or explicitly) accommodate. Finally, we reflect about *when we model*, the way in which we do so, and the artifacts we use to capture our knowledge and support our communication processes (e.g., the languages, ontologies, and other formalizations) need to support the different structural aspects such as graded category judgments, typicality differences between various model instantiations, and so on.

The content of this chapter is based on work published as:

Dirk van der Linden and Henderik A. Proper: Do conceptual modeling languages accommodate enough explicit conceptual distinctions? In: J. Grabis et al. (Eds.): *Short Paper Proceedings of the 6th IFIP WG 8.1 Working Conference on the Practice of Enterprise Modeling (PoEM) CEUR-WS Vol 1023*, 2013, pp. 126–135

Dirk van der Linden and Henderik A. Proper: On the accommodation of conceptual distinctions in conceptual modeling languages. In: H.G. Fill et al. (Eds): *Modellierung 2014, LNI 225*, pp. 17–32, GI, 2014.

Dirk van der Linden and Henderik A. Proper: Category Structure of Language Types Common to Conceptual Modeling Languages. In: I. Bider et al. (Eds.): *BPMDS 2014 and EMMSAD 2014, LNBIP 175*, pp. 317–331. Springer, Heidelberg, 2014

7.1 Background

It is important that the different conceptual understandings people have of a modeling language's meta-concepts can be modeled distinctly by that language. It would not do well for the overall clarity and semantic quality of a produced model if we conflate semantically different conceptual understandings (e.g., human beings, abstract entities and material objects) under the same banner (e.g., 'actor') and pretend that they are one and the same thing. Yet, in modeling languages often these different conceptual understandings are not made explicit. Frequently, the designers of a modeling language define a type (e.g., actor) and allow it to be instantiated with a wide diversity of entities (humans, hardware, abstract and mathematical entities) which have no common conceptual basis. Sometimes modeling languages do accommodate (some of) these conceptual distinctions, but then do so only implicitly. That is, in their specification or meta-model they assume a particular interpretation. As such, all instantiations of a model are then implicitly assumed to abide by that interpretation (e.g., all actors in the given model are assumed to be human things, all goals are assumed to be hard goals). An example we have used before of a modeling language doing so is the i* specification as found in the Aachen RWTH wiki (Grau et al. 2007), which defines agents (the acting entities) as having "*a concrete physical manifestation*". This implicitly makes it semantically incorrect to use abstractions (e.g., agents as they are commonly understood) and furthermore, perhaps ontologically incorrect to use composite agents – market segments – as the composition itself is not physically manifested.

It is more useful if a modeling language accommodates such conceptual distinctions *explicitly*, to the extent needed in relation to its expected and planned use. That is, instead of relying on the underlying semantics to define every concept they allow (or perhaps require), to use a notation that explicitly encodes information about our interpretation – and do so by providing distinct notational elements for all the important different conceptual distinctions. This can mean for instance, having exclusive (visual) elements to represent such distinct concepts by (e.g., the amount of 'stick puppets' in in ArchiMate actor type denoting whether it is a single actor or a collection of them). This is important from a cognitive point of view as it improves the quality of the notation by ensuring there is no notational homonymy. Many researchers have proposed methods and frameworks to analyze the degree to which languages are complete in this sense (Gemino and Wand 2004, Burton-Jones et al. 2009), often ontological in nature (e.g., UFO (Guizzardi and Wagner 2010), Bunge-Wand-Weber (Wand and Weber 1990) and their applications (Fettke and Loos 2003, Guizzardi et al. 2003)), although some have been criticized as being poorly suited when applied to the information systems domain given their dependency on a materialist ontology (cf. Wyssusek and Klaus 2005).

A major effort on this topic was undertaken by Moody (2009) in his work on a general 'physics of notation'. Several modeling languages have been analyzed to estimate their cognitive quality in terms of this framework (e.g., i* (Moody et al. 2010), BPMN (Genon et al. 2011b), UCM (Genon et al. 2011a), and UML (Moody and van Hillegersberg 2009)). However, most of these analyses are aimed at the semantics of the (visual) syntax, and forego a more detailed analyses of the semantics of the individual elements of meaning themselves. By this we mean that they analyzed the semantic quality of the formalization of grammar or the syntax (i.e., which elements interoperate in what way), but gave less attention to the question what the elements arranged by this syntax (i.e., the meta-concepts) actually mean to the users of the language. From a quality perspective, important related issues touched upon by such work are the *semiotic clarity* (one-to-one correspondence between semantic constructs and graphical symbols) and *perceptual discriminability* (symbols should be clearly distinguishable) (Moody 2009). This issue comes into play more clearly with domain-specific modeling languages than it does with general-purpose languages like UML, ER or ORM (even though these languages were originally designed for specific purposes like software and database engineering) because they have more native specialized semantic elements (i.e., types) to represent the important aspects from their domain by. It is thus important that these domain-specific languages have the ability to explicitly express important semantic distinctions that might arise in needed specific situations.

The goal of this chapter is not to provide detailed individual analyses of all the languages involved, but to explore whether there is a trend in modeling languages to support enough distinctions or not, and on basis of that argue what kinds of research and engineering efforts are needed to deal with optimizing the conceptual completeness of modeling languages. Hence the purpose here is to gain a deeper (empirical) understanding of the issues and challenges involved, rather than 'jumping' to the creation/suggestion of mechanisms to possibly deal with them. Therefore, we specifically look at the cognitive quality of a number of modeling languages and methods in terms of the semiotic clarity of their semantic constructs. These constructs can be both visual (for visual notations) and textual (for textual notations), but both require a proper correspondence between semantic constructs and symbols used for them.

7.2 Analysis

The different aspects that are focused on in enterprise modeling, typically have a number of (not necessarily overlapping) specific conceptual distinctions, which are important to be aware of. For example, a motivational model describing the things to be achieved by an enterprise and the reasons for wanting to achieve them is likely to require more detail (and thus fine-grained conceptual distinctions) for what goals are than, say, a model describing the related process structure. In previous Chapters we have elicited and described such distinctions in more detail, such as for instance whether goals absolutely have to be achieved, whether the 'victory' conditions for achieving it are known, whether the goal itself is a physical thing to be attained or not, and so on. On the other hand, a model describing the process undertaken to achieve a certain goal (e.g., bake a pizza) might require conceptual distinctions like whether the actors involved are human entities or not, whether it is one or more actors responsible for ensuring the goal's satisfaction, and so on. Thus, not all conceptual distinctions that are relevant to one aspect (and the modeling language used for them) will be as relevant (and necessary to model explicitly) for other aspects.

Table 7.1: A cross-section of aspects of modern enterprises, and some modeling languages used, or usable to represent them. The selected languages here are based on the initial selections from Table 3.1. While there the focus was on gathering terminology, here the focus is on conceptual differences, and thus we extended the list with more languages having different foci, relevant for the analysis of conceptual accommodation.

Aspect of an Enterprise	Related languages
Architecture	ArchiMate (The Open Group 2012) (1.0, 2.0), ISO/DIS 19440, ARIS
(Business) Processes	BPMN (Object Management Group 2010a), (colored) Petri nets, IDEF3, EPC (van der Aalst 1999)
Design decision-making	EA Anamnesis (Plataniotis et al. 2012), NID (Gal and Pfeffer 2003), OMG DMN (Object Management Group 2014a)
Deployment of IT artifacts	ADeL (Patig 2011)
Goals & Motivations	i*, GRL, KAOS (Dardenne et al. 1993), TROPOS (Giunchiglia et al. 2003), ARMOR (Quartel et al. 2009), ArchiMate 2.0'S (The Open Group 2012) motivational extension, OMG BMM (Object Management Group 2010b)
Management of IT artifacts	ITML (Frank et al. 2009)
Strategy & Capability Maps	TBIM (Francesconi et al. 2013), OMG BMM (Object Management Group 2010b), Capability Maps (Scott 2009)
Value exchanges	e3Value (Gordijn and Akkermans 2003), REA-DSL (Sonnenberg et al. 2011), VDML (Object Management Group 2014b)

This increased amount of focus on specific aspects has thus, amongst other factors, led to a plethora of modeling languages, methods and frameworks (some of which we have analyzed for their terminological use in Chapter 3.2). Many of these aspects have a large amount of dedicated modeling languages available, which differ only slightly in their actual notation or specification. This is evidenced, for example, by the large amount of overlap between the notations used in goal modeling such as i*, GRL, KAOS, TROPOS, which at times use the same terminology to refer to conceptually different things. The languages we look at to investigate what conceptual understandings are accommodated are given, together with the aspects of enterprises they are, or can be used for, in Table 7.1. We chose these specific languages based on the work done in Table 3.1, the languages integrated into the Unified Enterprise Modeling Language (UEML), and a number of additional languages that were not relevant for the analysis in Chapter 3, but here might offer some points of interest when it comes to accommodation of conceptually distinct understandings.

In order to systematically talk about whether these modeling languages accommodate different conceptual distinctions we will again use the meta-concepts and dimensions from Chapter 3 as a guiding framework. We thus took the meta-concepts from Table 3.2 (ACTORS, EVENTS, GOALS, PROCESSES, RESOURCES, RESTRICTIONS, RESULTS) and for each of them looked at whether the selected languages allow for the dimensions from Table 3.3 (*natural, human, composed, intentional, necessary, vague*) to be explicitly modeled. We started with a full list including each possible conceptual distinction for each concept, resulting in many different possible points of analysis. We then went through all the concepts and removed the distinctions that we deemed less relevant or interesting in the context of how domain-specific languages were typically used, and what earlier findings from the semantic differential studies (reported on in Chapter 6) had shown not to be particularly salient (e.g., whether a process is human, whether a result is intentional, and so on). For the resulting list it is then shown why it can be useful to be aware of this dimension for that particular meta-concept, and what modeling language supports doing so. Doing so led to the creation of the overview shown in Table 7.2.

Table 7.2: This table gives an overview of a number of relevant conceptual distinctions for common modeling. For each of the concepts, we list relevant conceptual distinctions, show what they are useful for, and what languages support modeling them explicitly, might support it, or (where relevant) make a specific implicit interpretation.

Dimension	Useful to ...	Supported by ...
	ACTOR	
human	Distinguish between actors that can be more fickle than pure rational agents.	BPMN through the explicit use of a 'Human Performer' resource type, VDML does contain a 'Person' subtype of Actor which is specified to be human, but does not distinguish in the visual notation between types of Actors.

Dimension	Useful to ...	Supported by ...
composed	Distinguish whether an actual entity acts or whether a group of them does, which impacts responsibility judgments for actions	ArchiMate, TROPOS via 'composite Actor', somewhat as well with differentiation between 'role' and 'position', e3Value somewhat through differentiation between actor and market segments, VDML distinguishes between an 'actor' being a singular participant, and modeling 'collaboration' or 'participant' as potentially multiples.
material	Know whether an actor physically interacts with the world (and can thus be affected by it directly – think hardware vs. software)	i* assumes that an agent is an actor "*with a concrete physical manifestation*" (iStar Wiki)
intentional	Know whether an actor is considered an explicit part of a system, i.e., is expected to act or not on certain things, in contrast to actors from outside the systems scope which may act but were not regarded or thought of to do so	Implicit in most languages, mentioned as such in TBIM, depending on interpretation could be argued to be explicit in OMG BMM with differentiation between internal and external influencer.
specific	Knowing whether an actor is a specific thing (i.e., an instantiation) or a general thing (i.e., a role)	Supported by some (e.g., ArchiMate), through type/instantiation dichotomy, explicit in TBIM by the claim that an agent "*represents a concrete organization or person*" ArchiMate, implicit in e.g., e3Value and RBAC by automatic use of roles (types).

Dimension	Useful to ...	Supported by ...
	EVENT	
intentional	Distinguish between events that should, or will happen given a set of circumstances, and events that happen (seemingly) unprovoked.	Arguably explicitly supported by BPMN through the use of 'None' type triggers for Start Events.
	GOAL	
composed	Distinguish between complexity level of goals, i.e., whether they are an over-arching strategy or directly needed goals.	TBIM explicitly models composite goals as 'business plan' types, implicit in some other languages focused on strategy/tactics (e.g., OMG BMM).
material	Distinguish between objects and their representations, i.e., is the goal to achieve an increment in the integer on a bank account, or to hold an n amount of currency.	
necessary	Distinguish between goals that have to be attained and those that should.	
specific	Distinguish between goals for which the victory conditions are known and not, i.e., hard vs. soft goals.	Most goal modeling languages/methods/frameworks (e.g., i*, GRL, KAOS, ARMOR) support this explicitly. Surprisingly[1] ArchiMate's motivational extension does not.

[1]Given that it was derived from ARMOR, which does explicitly support soft/hard goal distinctions

Dimension	Useful to …	Supported by …
	PROCESS	
composed	Distinguish between black (closed, singular) and white (open, composed) boxes.	Arguable either way for BPMN with the use of pools, which can function as black boxes, however, those do not allow for linking sequence flow to it, and are thus self-contained.
intentional	Know whether they are part of an intended strategy or something that has to be dealt with (i.e., negative environmental processes)	
specific	Know whether the structure is (supposed to be) clear (deterministic) or not (fuzzy).	
	RESOURCE	
natural	Know whether a resource requires a 'fabrication' process.	Somewhat related, TBIM explicitly models resource types as being either animate or not.
human	Know whether resources can act on their own and produce issues, e.g.., be unreliable, not always generate the same outcomes	

Dimension	Useful to ...	Supported by ...
material	Distinguish between objects and their representations, i.e., whether a given resource a collection of paper and ink blobs or the information contained within them.	Explicit in ITML through the use of hardware/software dichotomy.
	RESTRICTION	
natural	Distinguish between restrictions we cannot do anything about and those we can.	
intentional	Distinguish between restrictions we stipulate from those that arise holistically (whether good or bad).	Some languages implicit, e.g., EA Anamnesis, and BPMN through use of 'Potential Owner'.
necessary	Distinguish restrictions that can be broken from those that cannot.	(supported by some GPML, e.g., ORM 2.0).
specific	Distinguish restrictions for which we know when they are broken and not.	
	RESULT	
natural	Know whether a result requires some kind of 'fabrication' process	

Dimension	Useful to ...	Supported by ...
material	Distinguish between an object and its representation, i.e. whether the physical pizza or the status update in the IS saying a pizza was baked is the result of a given step in the pizza making process.	
specific	Know whether a result is (supposed to be) clear (deterministic) or not (fuzzy).	Arguably supported in BPMN through the use of 'None' type End Events.

7.3 Discussion

7.3.1 General discussion

Languages used for specific aspects do seem to explicitly accommodate some basic (and often widely accepted) necessary conceptual distinctions. For example, the de facto language used for process modeling, BPMN, has explicit support for differentiating between human and non-human actors, which can be important to know for critical steps in a process. Most modeling languages used for motivations and goals also accommodate the distinction between goals with well-specified victory conditions and those with vague or unknown conditions by means of separate hard and soft-goal elements. These explicit distinctions in the notation are likely correlated with the conceptual distinctions being widely accepted as important and having become part of the basic way of thinking. However, taken overall, there does not seem to be a consistent or systematic pattern behind what language explicitly accommodates (or lacks) which conceptual distinctions.

As such, there are a number of conceptual distinctions for which we found no explicit support by any modeling languages. For example, we found no support for explicitly modeling goals and results as being material things. It also did not seem possible to explicitly model goals as being a logical necessity in the investigated languages. The distinction whether results were things that naturally occurred or fabricated was also not supported. When it comes to processes we found no support to model them explicitly as being intentional, and distinguishing between specific (i.e., well-defined) processes and processes more fuzzy in their structure. Modeling resources as being humans was also not supported, while this is likely not an unthinkable interpretation – effective management of 'human resources' being important for large enterprises. Finally, we found no explicit support for modeling restrictions as naturally occurring and specific things. We will discuss some of these distinctions in more detail.

While the scope of this chapter is not to present an overview of such deficiencies based on our data, we can discuss some of them as an example. For some concepts the primary features that have been discussed are all explicitly supported, like for instance goals being differentiated between those for which we know when we achieve them and those for which we do not. However, not every language allows for such distinctions, as ArchiMate 2.0 for example still does not incorporate the hard/soft goal distinction, even though it was based on a language (ARMOR) which itself did. When it comes to events, the important features of intentionality are clearly supported in the most common languages dealing with events, like for instance BPMN's extensive dealing of triggers, and the possibility of modeling an unintentional event by using a 'None' type trigger. However, some of the features we found for resources are far less often supported by languages, like for example the identification of their nature in being material or immaterial. This does not seem to be explicitly supported in most modeling languages, usually being implicitly assumed to be a specific kind. A rare exception to this is ITML which at least enforced the distinction by using a hardware/software dichotomy where it is immediately clear whether a resource is a physical object or not. While not every feature listed once by an individual modeler should be part of each modeling language, it does seem to suggest that a combined effort of updating languages with the features important for their users would be a useful venture.

7.3.2 Some unaccommodated conceptual distinctions

Surprisingly, we found no explicit support for differentiating between goals with varying levels of necessity and obligation. While many common methodologies (e.g., the MoSCoW technique (Clegg and Barker 1994) of dividing requirements into must, should could, and would haves) call for such distinctions, many modeling languages conflate them all into a single kind of goal. Arguably in certain aspects it would make sense to make an implicit choice, as in e.g., process modeling it is necessary for certain steps in a flow to be reached before the flow continues, which can be seen as an analog to logically necessary goals. However, goal models in dedicated languages seem not to make this distinction, even though there is a strong focus on differentiating between hard and soft-goals, which seem correlated with different levels of necessity (e.g., one cannot as certainly rely on a soft-goal to be achieved compared to a hard-goal, especially for mission critical goals).

Another seemingly unaccommodated distinction is the necessity of restrictions, that is, whether some restriction (e.g., a rule, principle, guideline) is an alethic condition that cannot be broken or whether it is not and thus can be broken. While in the context of enterprise modeling there is a strong differentiation of terminology used for different kinds of normative restrictions that can be considered breakable, or at least not strictly enforceable (e.g., principles, guidelines, best practice), these often seem to be used outside of modeling languages in their own approaches – e.g., architecture principles (Proper and Greefhorst 2010). It seems problematic that many languages used for aspects of enterprises, and languages used to describe the actual enterprise architecture like ArchiMate do not have explicit notational support for these different kinds of restrictions. Many models that are analyzed a posteriori (e.g., when they are integrated in other models, and the original modelers are no longer involved or available) then become difficult to interpret, as the notation of different kinds of restrictions can be ambiguous and lead to situations where it is not clear whether a restriction can be relaxed or not. Surprisingly the only language that seems to support this conceptual distinction is ORM (in particular version 2), which supports the explicit modeling of restrictions as being either alethic or deontic conditions through its visual notation.

Another conceptual distinction that is typically not accommodated by most languages is whether something is material or not. In particular, the material status of resources is often defined in a conceptually ambiguous way. For example, in TROPOS, resources are stated to be "*physical or informational entities*", which makes it difficult to know whether a modeled resource is the actual 'object' (e.g., some information) or its representation (e.g., a collection of paper and ink blobs that represents that information). It is important to be aware of this distinction as this has consequences for the way in which the resource can be interacted with, and in what way it can be manipulated, and possibly consumed. For example, if we have a process in which a human actor performs a certain task for which they need clear instructions, we can see those instructions as being a vital resource. Modeling them as the physical representation – a paper printout of the instructions – means that this specific resource is only available to one actor. On the other hand, if we model it as the actual informational object, it is available to more than just one actor at a time. Furthermore, when a resource is material, it also has the possibility of being consumed. For example, when we model the process of baking a pizza, some of the resources involved (i.e., the ingredients) are consumed. It is important to be aware of this, as that means it is necessary to keep track of stock levels, and perhaps optimization thereof via e.g., system dynamics models.

Finally, a conceptual distinction that is not explicitly accommodated by many languages is whether an actor is a human being or not. BPMN was the only language we analyzed which explicitly supports it by having a notional element 'Human Performer' which is only used for human actors (albeit called a resource in BPMN jargon). Apart from this some languages could be argued to offer partial, or implicit support, like for example ArchiMate having certain assumptions that are only valid for specific layers, such as acting entities not being machines or resources in the business layer. It is important to be aware of this, especially when responsibility comes into play, which is for instance done in KAOS models, by making some agent responsible for some goals. However, the *actual* responsibility for any given thing cannot, from a legal and social perspective be placed on a non-human entity. At the end of the day (or chain of responsibility), there is always a person held responsible (and accountable) for some given action. In the case of software engineering, for example a programmer is held responsible for the downtime caused by bugs, in the case of a building collapsing after a summer breeze the architect is held responsible for not properly analyzing the environmental and soil conditions, and so on. When responsibility is modeled, it thus seems prudent to know whether an actor is the actual responsible party or whether it defers its responsibilities to a different, human entity. Another important aspect of human beings is that they are not necessarily rational and reliable. Thus, when a given task or process depends on a specific human actor, it is quite possible that the process is not performed as well as needed, or at all. As such, knowing that a process involves human actors, a certain level of fault tolerance and redundancy would be needed. Conflating human actors with non-human actors makes it far more difficult to know where this is necessary, and could thus lead to models (and predictions made with them, e.g., efficiency or execution time of a process) not holding true to the real world situation.

7.3.3 Additional implications and their potential solutions

The overall lack of explicitly accommodated conceptual distinctions (of which there might be more than just those we discussed) in many modeling languages are especially relevant for enterprise modeling. It makes it much more difficult to ensure that integrated models are valid (or complete) representations of the semantics intended by the original modelers, as sometimes these modelers simply lack the notational elements to express important semantics. While it is possible to 'simply' denote this information by annotating the models with extra text, it would be a more ideal solution if modeling languages supported these distinctions. Furthermore, while some languages do offer explicit notational support, their specification or meta-model does not necessarily enforce correct use of these elements (e.g., ArchiMate does not enforce correct distinction between composite and singular actors). There are many languages we analyzed which have an implicit interpretation of some of the conceptual distinctions, sometimes specific (e.g., i*'s handling of agents as having a concrete physical manifestation) sometimes vague (e.g., TROPOS' handling of resources), which further complicates matters, as this interpretation of the language might not be the interpretation a modeler wishes to take for a given context. The fact that some languages have semantics which are considered to remain vague (e.g., GRL (Heymans et al. 2006), i* (López et al. 2011)) only adds to this. Most languages seem to have a well-defined and formalized semantics of the syntax, while lacking much, if any, formalization of the semantics of the elements of meaning themselves (e.g., EPC, Kindler 2004).

Thus, it seems necessary to stimulate a move towards more explicit focus on (formalization of) the semantics of the elements of meaning of modeling languages. The lack of coverage for some of the distinctions shown in Table 7.2 makes it clear that more work is needed on extending the specification of relevant languages with the ability to explicitly distinguish between these different conceptual understandings. This could, and perhaps should, be done in accordance with the actual practitioners in the field, by investigating what conceptual distinctions are important for them, and what they need to be able to explicitly model, instead of solely relying on analyzing languages with pre-existing reference material like Bunge-Wand-Weber or UFO, which are more useful for analyzing the overall ontological and semantic quality of a language. However, to ensure that a modeling language reflects the conceptual needs and desires of the people in the field we cannot rely solely on such reference materials because by their nature they cannot take each and every (potentially significant) personal difference into account. It is important to not do this just once, but keep up to date with the changing conceptual distinctions that the practitioners and stakeholders have in order for our modeling languages to stay relevant and capable of representing the real world. Given the existence of a large number of different dialects of modeling languages sometimes only differing slightly (e.g., i*, GRL, TROPOS for goal modeling), it seems that supporting many different conceptual distinctions in a single notation would be welcomed by many. An argument made for the research needed for this is given in Chapter 9.2.3.

It is necessary to nuance this call for an increased amount of semantically distinct constructs in languages somewhat, as it would not do for any and all modeling language to be maximally expressive, as doing so would increase cognitive load, and thus the complexity of both using and interpreting the language and its produced models too much. Moody (2009) also found such relations between multiple quality aspects, where the increase of one aspect would have a negative effect on others. Instead, languages used for specific aspects should be optimized in such a way that the conceptual distinctions important to them (e.g., the hard/soft goal dichotomy in goal modeling) are explicitly accommodated, but other things can be left implicit. For languages used for multiple aspects or wider domains, one can take inspiration from, for example, the ArchiMate language which has a guiding principle, elaborated on in Lankhorst et al. (2010), of keeping the language as simple as possible, by ensuring that 80% of a domain can be modeled with 20% of the language's available constructs. Furthermore, even though we found some conceptual distinctions that were barely supported in existing modeling languages (e.g., distinguishing between alethic and deontic restrictions), this does not say anything about whether such distinctions are actually desired by the users of the language, and would thus end up being used in practice. The moral here is to strike a good balance between the complexity of the language and the amount of semantics that can be explicitly expressed by it.

Apart from being more adaptive to different conceptualizations people have in the modeling process, we can also ensure that our modeling languages are inherently more suited to explicitly deal with them. A possible strategy to deal with this could be to 'upgrade' the concept of *view* as used in e.g., the field of Enterprise Architecture (cf. The Open Group 2012) or systems and software engineering (cf. IEEE 2011). Traditionally, a view provides a model of a domain from a specific (set of related) concern(s). This could be extended with an articulation of all the expressed (and preferably shared) understandings of the modeling concepts used in the view. Even more, one should consider the joint creation (by a group of stakeholders or modelers) of a view as the joint creation of a model of the domain and the meta-model of the modeling concepts used in that view. This is essentially a form of domain/purpose specific modeling language. When modeling a single domain in terms of a 'swarm of views', where each view is modeled by a specific group, from the perspective of a (set of related) concern(s), an integrated or joined model of that domain could then be constructed as a shared (and traceable) understanding among the different views. Such approaches to constructing models by integrating views can be found in (e.g., Dijkman et al. 2008b, Brandt and Hermann 2012). At first this might sound as a laborious task. However, as our research has indicated, when we do not respect the group-based and personal understanding of modeling concepts and or domain concepts, there is a risk of (implicit) misunderstandings. Such misunderstandings can have severe adverse consequences in an enterprise and information systems engineering context. Thus, a certain level of cost/benefit tradeoff for ensuring these misunderstandings are prevented should be kept in mind.

Furthermore, it would be useful to know how specific domain or purpose-specific modeling languages really need to be, and on the other hand, how general-purpose modeling languages can be while not conflicting with people's conceptualizations. This correlates with the (limits of) someone's semantic flexibility, which can be investigated by testing the limits of their conceptualizations. This likely affects their ability to easily use languages that do not accommodate their needed distinctions (e.g., a modeler who typically only uses the concept of human actors). This can be investigated, for example, through validation by instantiation testing to see to what degree people can accommodate non-matching uses of their conceptualizations as defined in a language's specification.

7.4 Reflection on creating and using models

From the moment when we model (whether in informal discussions or with the aid of a properly selected modeling language), we communicate and interact with many people. Given this inter-subjective nature, many potential communication errors and terminological mismatches are bound to occur. Anything that helps us find potential weak spots where such events are likely to occur before committing models to paper would be a useful venture. Doing so can prevent far more costly rectifications in later stages, when models cannot simply be amended (e.g., when a business process has actually been implemented, when a software program has been made, when specific domain experts or stakeholders have moved on and are no longer available).

Our findings into the conceptual understandings that people have of the studied meta-concepts can thus be used practically for *steering the communication in the modeling process.* What we can take from the studies discussed in this dissertation is that individual modelers all have different conceptual understandings of the concepts they use to model elements from a domain with. The actual core of these conceptual understandings, i.e., the most conserved features, come to be so by a lifetime of experience, and are unlikely to just change during a single session in which people communicate and share their understandings. This is corroborated by findings from other studies, like Recker and Dreiling (2007) who performed a study on how people deal with picking up a new process modeling language without being formally taught them. In this study it was shown that if someone was familiar with one particular language, the threshold to go to a different, similar one was very low (e.g., switching from one goal modeling notation to another). As many of such similar languages often have a strong overlap in their conceptual make-up, people who are strongly familiar with one can easily pick up the others because it matches their conceptual understanding, even if they have not yet mastered its specific syntax or visual notation.

While the categorization judgments we make are often of a graded nature (for example something being an actor to a certain degree), meaning that there is a certain amount of flexibility between people holding different conceptual understandings, it should be clear that it is worthwhile to investigate just what these typical understandings for people are as to preemptively avoid terminological or conceptual mismatches. As the most often found kind of feature in our datasets were those of an alethic or and structural nature, people seem to identify concepts very often by what they actually are for them. This implies strongly the benefits of using ontologies during modeling work, especially when those ontologies are actual reflections of the understanding involved people have of the domain (Guarino 1995), for which studies like these can give valuable input. Such ontologies are often useful to define the specific fine-grained semantics modelers hold of the concepts they use for a particular task, which is especially useful given that most often languages like ER and UML are used (Davies et al. 2006), in which we need to stereotype and define these detailed semantics in order to properly document them.

7.4.1 Implications for the way we model

The results we have presented in earlier Chapters (see e.g., 6.3 and 6.4) have implications for the creation and use of models. As it is clear that there are some common type-feature combinations amongst the people we investigated (while the total amount of conceptual diversity is far greater than that), we should reflect on the way we model in order to deal with the effects of these different conceptual understandings.

While the data presented and discussed reports on average and shared conceptualizations, the conceptualizations that individual people have can vary wildly. It is important that we take a moment in the modeling process before we start creating models to deal with these differences. A way to do so is by looking at the types that are needed to model the domain under investigation, and then discuss whether certain features (and combinations thereof) apply to it. This can be done by taking known lists of features that have proven to be a source for disagreements and conceptual misunderstandings, such as the features reported on in this work. For example, when we model the process of baking pizzas, we might focus on the resources for a moment. We can then take some time before creating actual models to discuss what we find conceptually valid resources and what we do not (i.e., what we would rather not be possible to express in the model). From such discussions we gain a clearer understanding of how people see such a process and want to model it, which helps us to focus (e.g., by abstracting from such potential resources as time, only modeling material things as resources). By doing so the semantics of the created model are shared and well understood by all the people involved in the process, not just those who are intimately familiar with the specification of the used modeling language.

Such discussions can go into more detail feature wise as the focus on what to model becomes narrower. For example, when there is already a strong focus on processes, the types most relevant to a process (e.g., processes and restrictions) can be explored more closely by looking at more relevant features in greater detail. For instance, a process model might benefit from having a level of traceability on the restrictions in it (Plataniotis et al. 2012), which requires us to discuss whether restrictions are always made by specific people, or departments, and whether that implies that restrictions carry responsibility with them. These aspects require a more fine-grained characterization of the process type than if it is used in a general way. Equally so, when the focus is on modeling value exchanges, then the focus can be on, for instance, characterizing types like resource and actor more clearly by investigating more detailed features (e.g., value exchanges can be analyzed from an ethical point of view, and thus require actors to be moral agents).

Others have proposed different ways of modeling that are in line with what we have discussed about our fundamental findings, such as for example using a process modeling approach which is centered around the typical examples of processes (i.e., specific instantiations) as used, not their abstract descriptions. The notion of prototypes and graded conceptual validity we found for the structure of the meta-concepts supports this as a likely effective way to reason about models, given how close it is to how people actually conceptualize the concepts from those models. In their proposal for a different way of doing business process modeling, Hofer (2011) described a way of not looking at processes as algorithms, but by focusing on exemplary process instances, which also reduced the total amount of information that has to be captured to a more manageable amount.

7.4.2 Implications for what we use to model

Applicability of our findings

Based on the findings from our categorization study as discussed in Chapter 4 we see also a number of implications for model creation and use, most relevant to domain-specific languages. With the plethora of domain-specific languages (e.g., ArchiMate, BPMN, e3Value, i*, ITML, ADeL) in active use today all with their own focus (e.g., enterprise architecture, processes, value exchanges, goals, IS implementations, IS deployments) our findings could have consequences for many modeling efforts. The consequences we discuss should thus be taken to be most relevant for domain-specific modeling languages like these and any artifacts based on the models created with them.

When it comes to the modeling languages and models that are affected by our findings, we see a number of different possible effects depending on the purpose of the models:

1. *models used to communicate* between, and with different modelers and stakeholders (e.g., conceptual models)

2. *models used to formalize* information from a given domain, for whatever purpose (e.g., ontologies, models as documentation)

3. *models used to transform* by non-human systems (e.g., compiled source code)

This list is not intended as a taxonomy of models, nor as an exhaustive list of the different kinds of models that are affected by graded category structures. It is merely a starting point to reason about the different consequences we see our work having for different kinds of models. We furthermore do not mean to imply that these kinds are mutually exclusive (e.g., that models used to communicate are never used to formalize or transformed into executable models).

Models used to communicate

Models used to communicate involve conceptual models of many possible purposes (e.g., capturing a domain, models used to guide decision making). As we have shown that the categories used by modeling languages are likely of a graded nature, the models created by them necessarily also contain categories of a graded nature. The most important consequence here is that an instantiation of a model is not just simply valid or invalid, but will display degrees of validity as well. If the category goal is seen as a graded structure, with some things being better goals than others, it is thus possible to instantiate a model that contains some goal type with two different cases that are both valid, but not equally so. As the formal semantics of most modeling languages do not explicitly support such degrees of validity, it is important that we are clear about the limits of conceptual validity of our models. In other words, to ensure people read and use models in a similar way, we need to ensure that we provide clear examples of possible valid instantiations, and perhaps more importantly, clear examples of that which we consider invalid as well.

For example, while ‘hardware’ and ‘software’ are both considered partial members (by at least the experienced modelers in our study) of the category ACTOR, the exact degree to which they are both considered so is something that is likely different for different (groups of) people. If we are creating a model used for the implementation of an information system, which would likely incorporate such terms for the things that act to support and execute business activities, we need to be clear to what degree they can both be seen as ACTORS. For instance, the modelers or stakeholders might envision the hardware as the actual acting part, with the software providing the instructions for doing it so, and thus find a model where ‘hardware’ is said to act out a business function more valid than where 'software' does so. However, others might disagree and see 'software' as the actual thing that acts. As these interpretations can be different from group to group, it is thus important to involve explicit discussions about the degrees of validity for different things we use in our models during model creation.

Models used to formalize

Models used to formalize are for instance models that capture knowledge about a certain domain and attempt to formalize it in order to reduce the amount of ambiguity. A formalization involving graded categories needs to ensure that membership requirements are not discrete, and more importantly, take into account the relevant properties of a graded category (e.g., centrality and typicality of members). There is work in the field of ontology engineering that strives towards explicitly supporting these structures, (e.g., Aimé et al. 2008) and explicit modification of ontology formalizations to incorporate the noted features (Yeung and Leung 2006; 2010), and critiques and extensions of proposed work, (e.g., Cai and Leung 2008). If such formalizations are not used, and instead a classical approach based on discrete judgments is used, much semantic information about the domain and the judgments from the original modelers is lost. This can lead to misinterpretations by other readers and users of the model if there is no communication between them and the original modelers anymore. For instance, someone might consider a horse as a VEHICLE (albeit an atypical one) and thus consider it to be somewhat of a valid vehicle in their created ontology. However, when this is formalized discretely, any other member of the VEHICLE category (e.g., a car) would be considered on equal footing with the horse, which a large number of people would likely disagree with. As such, the formalization can no longer be considered a correct representation of the real world and loses a lot of its value. Given our findings it thus seems important to choose the right formalization for the job so as to have the ability to correctly reflect the intended semantics. This further strengthens the importance and relevance of the research into ontology formalizations by e.g., Yeung and Leung that we have mentioned.

Models used to execute

Models used to execute are for instance source code which is run by an interpreter, or compiled (i.e., transformed) and then executed. Other options are models interpreted by model provers, expert systems, or ontologies used for automated reasoning and so on. For example, a model used by an expert system to check for a number of possible cases (e.g., a medical advice system) might need graded structures and judgments in order to correctly reason with the real-world information. A number of formalizations for e.g., descriptive logics have been proposed to incorporate graded features like typicality and centrality (Britz et al. 2009, Giordano et al. 2008, Freund et al. 2004). These models are affected in a similar way to the ones used to formalize, meaning that their formalizations need to support any graded structures found in them. This is all the more important to ensure here, as executable models are often no longer read and interpreted by people, and thus any errors or oversights in them are less likely to be corrected.

A common solution given for the problems we have discussed here is to 'simply' define the terminology in detail beforehand. Such definitions are expected to avoid issues with ambiguity and possible different interpretations. However, the act of defining terminology for such purposes is far from trivial, requiring a level of detail that one cannot know beforehand. It is difficult to know which details are or become important in order to discuss and define them. Insights into the categorization habits of modelers themselves can be complementary to such processes, and aid them. They can do so by pointing out areas that require more attention (i.e., categories with high partial member judgments, terms that receive many partial member judgments) and thus guide modelers to points of attention that should be discussed in detail before a definition is decided on.

7.5 Summary & outlook

In this chapter we have discussed implications of some of our fundamental findings for the more practical aspects of conceptual modeling. Specifically, the importance of explicitly modeling conceptual distinctions and analyzed a number of modeling languages to investigate what kind of distinctions they support. We showed that, while some conceptual distinctions are explicitly supported by relevant modeling languages, there are still a large amount of potentially relevant distinctions that are not accommodated, or implicitly interpreted in a specific way by modeling languages. In doing so we have found an answer to our research questions centering around "*to what extent modeling languages allow for people to express their particular conceptual understandings of meta-concepts.* We proposed that research on active practitioners should be done regularly to keep up to date with conceptual distinctions deemed relevant and important by modelers and stakeholders alike, especially if such data would be used to infer which modeling language would be best suited to use for a particular task. Furthermore, we discussed implications for the way in which models are created and used in communicative settings, and their underlying support structures. One consequence of our research into personal semantics which has been left implicit so far is the potential for predicting what kind of typical personal semantics a person would exhibit. As a final step, we will report on a study about this in the next Chapter.

CHAPTER 8

Attributing Identity to Conceptual Understandings

Abstract. In this chapter we discuss the use and challenges of identifying communities with shared personal semantics. People tend to understand meta-concepts (i.e., a modeling language's constructs or types) in a certain way and can be grouped by this understanding. Having an insight into the typical communities and their composition (e.g., what kind of people constitute a semantic community) can make it easier to predict how a conceptual modeler with a certain background will generally understand the meta-concepts they use, which is useful for e.g., validating model semantics and improving the efficiency of the modeling process itself. We have observed that in practice decisions to group people based on concept use are often made, but are rarely backed up by empirical data demonstrating their supposed efficacy. We use data from two studies discussed earlier in Chapter 6 involving experienced (enterprise) modeling practitioners and computing science students to find such communities. We also discuss the challenge that arises in finding common real-world factors shared between their members to identify them by and conclude that there is no empirical support for commonly used (and often implicit) grouping properties such as similar background, focus and modeling language.

The content of this chapter is based on work published as:

Dirk van der Linden, Stijn J.B.A. Hoppenbrouwers: Challenges of Identifying Communities with Shared Semantics in Enterprise Modeling. In: *Proceedings of 5th IFIP WG8.1 Working Conference on the Practice of Enterprise Modeling (PoEM2012)*, Rostock, Germany, 2012

Dirk van der Linden, Stijn J.B.A. Hoppenbrouwers and Henderik A. Proper: On the identification of modeler communities. In: *International Journal of Information Systems Modeling and Design (IJISMD)*, vol. 5, no. 2, pp. 22-40, IGI Global, 2014

8.1 Background

As we have argued in previous Chapters, to deal more effectively with the issue of misunderstandings and terminological mismatches in the modeling process it is necessary to have an insight into the conceptual understanding of the people involved in the modeling process; in particular, how they understand the meta-concepts (i.e., the modeling concepts). However, to do so one cannot realistically be expected to look into each individual modeler's semantic idiosyncrasies. Instead, a generalized view on how people with a certain background typically understand the common meta-concepts could be used to infer, to some degree, the outline of their conceptual understanding. Such (stereo)types of modelers could be found by identifying communities of modelers that share similar semantic tendencies for given meta-concepts and analyzing whether they have any shared properties that allow us to treat them similarly. A community in this context is nothing more than a group of people who can be seen to share certain things, in this case their understanding of a modeling language and its (meta-)concepts. As language, or any means of communication, is inherently bound to a community using it (Perelman and Olbrechts-Tyteca 1969) (regardless of whether that community is bound by geography, biology, shared practices and techniques (Wenger and Snyder 2000, Meyerhoff 2008), like-minded people (Alani and Shadbolt 2002), used and shared information (Bishr et al. 1999), cognitive strengths and weaknesses (Wilmont et al. 2012) or simply speech and natural language (Gumperz 2001, Hoppenbrouwers 2003)), it seems safe to assume that there are communities which share a typical way of understanding modeling language concepts. This is not to say that such communities would be completely homogeneous in their semantics, but merely that they show enough overlap to be able to be treated as belonging together during a process which integrates models originating from their members without expecting strong inconsistencies in the final product.

Finding such communities based on, for example, empirical data is not a difficult matter in itself. However, the difficulty lies in going from simply finding communities to understanding them and generalizing them, i.e., being able to predict, on the basis of empirical data or prior experience, that communities of people sharing certain properties will typically use certain semantics. To do so it is necessary to find *markers* – properties that are shared between the members of a community. These *markers* (e.g., dominant modeling language, focus on specific aspects) are needed to be able to postulate that a given modeler, with a given degree of certainty, belongs to some community and thus likely shares this community's typical understanding of a concept.

8.2 The real world situation

Between 2010 and 2012 several collaborative modeling workshops were organized in the context of the Agile Service Development (ASD) project[1] (Lankhorst 2012). With the

[1] The ASD project (`www.novay.nl/okb/projects/agile-service-development/7628`) was a collaborative research initiative focused on methods, techniques and tools for the agile development of business services. The ASD project consortium consisted of Be Informed, BiZZdesign, Everest, IBM, O&i, PGGM,

partners involved in these workshops, who themselves are involved in different kinds of (collaborative) domain modeling (e.g., enterprise modeling, knowledge engineering, systems analysis), we have found that there are a number of common markers that modelers are typically (and often implicitly) grouped by. That is, on the basis of these properties they are often assigned to collaborate on some joint domain modeling task. These properties are, for example, a similar background, education, focus on what aspects to model (e.g., processes, goals), in what sector they do so (e.g., government, health care, telecommunications), and modeling languages used. Thus it seems that in practice, it is assumed that when people have similar backgrounds, use similar modeling languages and methods, etc., they will share a similar enough conceptualization of the involved modeling meta-concepts, and will thus be able to effectively collaborate.

In this chapter we will test whether the premise (that people with similar backgrounds, using similar languages and methods and so on, will share a similar conceptualization of modeling meta-concepts) of this assumption holds, as it is so rarely tested or backed up by empirical data. To do so we hypothesize that commonly used properties (e.g., modeling language used, modeling focus, operating sector) should be reflected in communities that share a similar semantic understanding of the common modeling meta-concepts. To test this we will investigate the personal semantics for practitioners and students alike (whereas other work on finding such communities and their conceptualizations often focuses on analysis of their produced texts or models (Flake et al. 2002, Recker and Dreiling 2007) instead of the modelers themselves). We will then group them by shared semantics and investigate whether they share the expected, or indeed, any amount of properties. If this is found to be so, then the 'naive' grouping procedure commonly used already in practice might have some merit. Furthermore, it could lead to predictive theories that, to a certain degree, predict what (the range of) understanding is that a modeler has for a given concept.

The specific focus of the study presented in this chapter is thus to investigate and test whether this common assumption made in modeling practice can be backed up by empirical investigations. In terms of the types of theories in information systems research described by Gregor (2006), we strive to analyze and describe in detail the modelers' conceptual understandings, and whether that analysis challenges any held assumptions. It is thus out of the scope of this particular study to propose an approach stipulating how to more effectively 'do' the act of modeling, nor is it our intention to describe in elaborate detail how existing methods for particular modeling efforts (e.g., TOGAF, ADM in the context of enterprise modeling) might be adapted to fit with our findings. Instead, we

RuleManagement Group, Voogd & Voogd, CRP Henri Tudor, Radboud University Nijmegen, University Twente, Utrecht University & Utrecht University of Applied Science, TNO and Novay.

will discuss the more fundamental implications our findings have (be they bad or good), and what steps could be taken both by practice and research in order to deal with them.

8.3 Experimental parameters

8.3.1 Limitations

While the amount of participants in each study might seem low compared to other scientific studies with different goals and methodologies, both of our studies are large enough to produce useful results for our purposes. As we will test our hypothesis by attempting to falsify it, we need only counter-examples to the practice of naive grouping we described in the introduction. We are confident that accepting the hypothesis is not unrealistic as it is grounded in empirical observations, and its rejection would also not be a trivial matter. Thus, it is most efficient for a first enquiry into the problem matter to use only as many people as deemed necessary to find a counter-example. Given that the practice we described seems to be widespread, it should thus be found in relatively small samples of participants.

Even though there is more difference (of opinion and interpretation) to be found when it comes to domain concepts than the meta-concepts we focus on in this dissertation, these meta-concepts are more interesting to look at for this study's purposes. As we wish to compare a number of modelers in order to establish whether they can be grouped or not, the concepts we investigate should be shared amongst them. This is definitely the case for the meta-concepts, as they are shared by most languages and methods, whereas the highly specialized domain concepts might not be shared amongst them.

8.3.2 Method

Materials: This community identification study is based on data from the semantic differential studies shown and discussed in Chapters 6.3 and 6.4. They investigated the understanding participants have for the meta-concepts from Table 3.2 (ACTORS, EVENTS, GOALS, PROCESSES, RESOURCES, RESTRICTIONS and RESULTS) on the dimensions from Table 3.3 (natural, human, composed, intentional, necessary, material and vague).

Materials – markers: We use markers to analyze whether groups found in the results of our study reflect commonly used grouping approaches by practitioners. These markers originate from workshop sessions with practitioners and companies as detailed in the introduction. They are the following: what modeling languages and methods people use, what sector they operate in, what the focus of their modeling efforts is, and what kind of stakeholders they interact with during the modeling process.

Participants: The practitioner study as reported on in Chapter 6.3 (n=12, see Table 8.1 for detailed marker information) was carried out in two internationally operating companies that focus on supporting clients in (re)designing organizations and enterprises. Apart from the semantic differential, we explored what modeling languages and methods they use, what sector(s) they operate in, what they model, and what kind of people they mostly interact with in order to see whether these could be used as identifying factors for semantic communities. The student study (n=19 at the original time of this analysis) was an ongoing longitudinal study into the (evolution of) the understanding computing and information systems science students have of modeling concepts. We explored their educational (and where applicable, professional) background, their knowledge of modeling or programming languages and methods, their interests, and career plans. While these students will likely not offer any particularly interesting insight compared to the practitioners, we include them in order to verify whether the phenomena we investigate occur in other groups than just experienced modelers.

Procedure: To find communities of people that shared a certain amount of semantics (i.e., score similarly for given concept-dimension combinations) we initially analyzed the results from the semantic differential studies using repeated bisection clustering. However, we found it not feasible to investigate the existence and borders of communities with this approach, as it was not sensible to *a priori* estimate parameters like optimal cluster size and similarity cutoffs (i.e., how similar people should score in order to be considered part of the same community), given that we had no realistic prior data. For this reason we used principal component analysis (PCA) and the visualizations of it in order to generate a more manageable way of investigating the communities and the rough semantic distance between them. These PCA results and their visualizations (see Figs. 8.1 and 8.2) demonstrate (roughly) the degree to which people share a semantically similar understanding of the investigated concepts and can thus be grouped together. It has to be stressed that this 'unit' of distance is dimensionless and thus should not be used as an objective measure on its own. Instead, it can be seen and used to distinguish groups from groups, while not saying necessarily in detail how objectively far they are from each other. Combining this data with the information we have gathered about the participants (i.e., the markers) we can investigate whether the structure of the found clusters (i.e., semantic communities) reflect what would be expected from the naive grouping commonly performed in practice.

Table 8.1: Participants in the practitioner study and the markers relevant for identifying them as being a particular kind of modeler. 'Proprietary' languages are not publicly available modeling languages or suites, often developed in-house.

No.	Used languages	Sector	Focus	Interacts with
1	Proprietary, RDF, OWL, UML, ERD	Financial, Government	Context, domain knowledge, processes, data	Operational managers, Senior managers, Domain experts
2	Proprietary	Government	Knowledge systems, processes	Managers, domain experts
3	Proprietary	Financial, Government	Knowledge rules, processes, data	Analysts, modelers
4	OWL, UML, BPMN	Government, Public, Healthcare, Finance	Knowledge rules, decisions	Domain experts, project managers, IT engineers, business and enterprise architects
5	UML, Proprietary, Protos	Financial, Government, Non-profit	Application-specific knowledge, process knowledge, knowledge databanks	Domain experts and IT departments
6	Meta-modeling, ontologies, taxonomies	Spatial planning, environment	Processes	Domain experts, analysts, architects
7	Proprietary, UML, Java	Government, spatial planning	Business processes, process structure, supply chain	Domain experts, IT specialists
8	UML, OWL, RDF, Mindmap, Rulespeak, Proprietary	Government, Healthcare	Rules	Business professionals, policymakers, lawyers

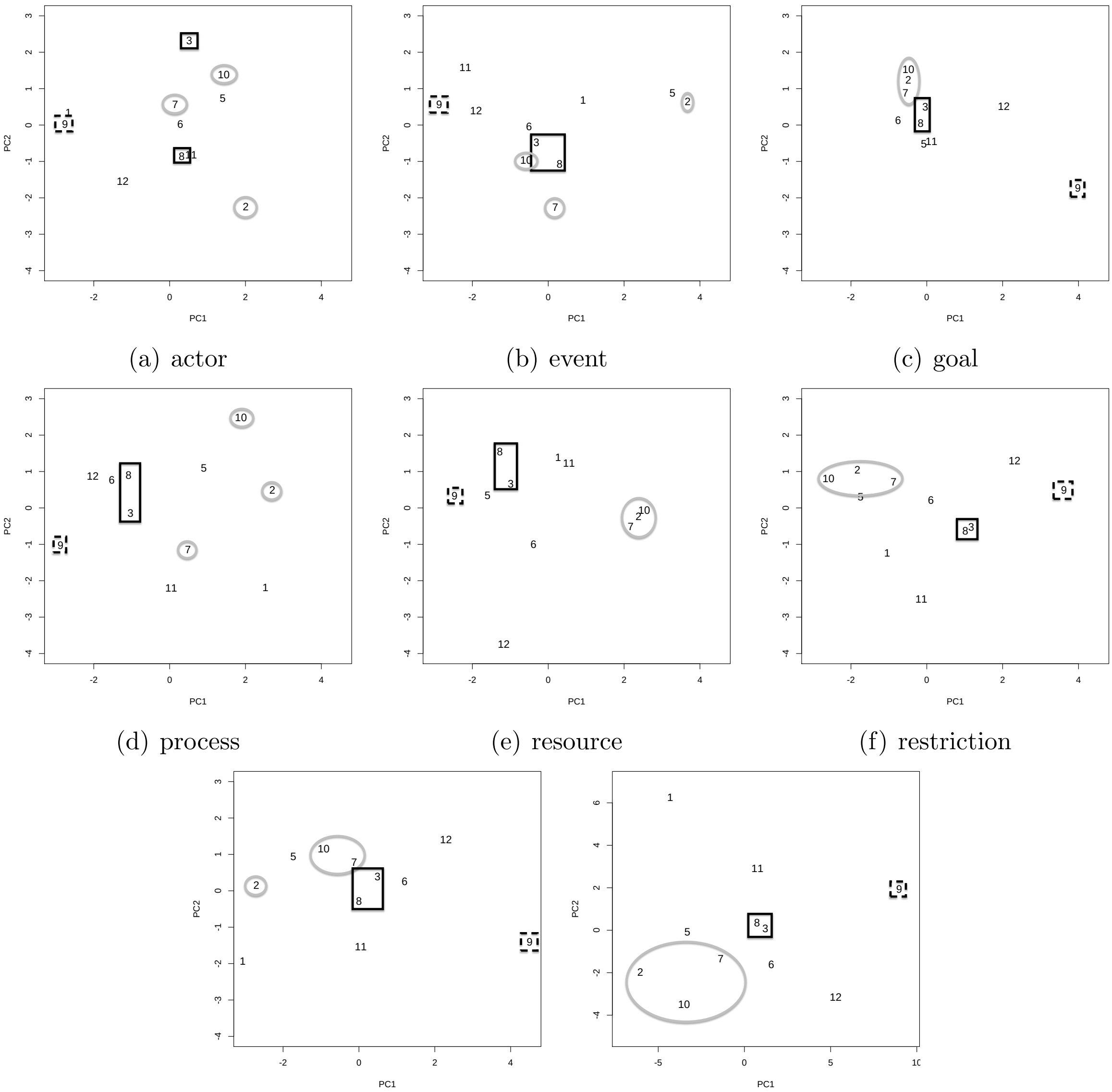
9	Proprietary	Government, Financial	Rules, legislation, policy, processes	Domain experts
10	Proprietary, XML, XSLT	Government, finance	Processes, rules, object definitions for systems	Domain experts, java developers
11	ArchiMate, UML, ORM, ERD, BPMN, Amber, 'improvisational'	–	Enterprise-wide architectures, strategic context, change organization	(senior) line managers, architects, domain experts, process owners
12	ArchiMate 2.0, Amber Architect, Proprietary	Government, Healthcare, Financial, Telecom	Business processes, work processes, instructions, enterprise architecture	Domain experts and managers, 'normal' workers

8.4 Data

Most importantly, the results support the idea that people can be clustered based on their personal semantics. As shown in Figs. 8.1 and 8.2 there are easily detectable clusters (i.e., communities) for most of the investigated meta-concepts, although they vary in terms of their member size and the semantic difference between the members (i.e., the variance within the clusters).

8.5 Discussion

While there are both clusters of people that share a semantic understanding for practitioners and students alike, they do differ somewhat. Internal variance for a number of concepts is greater for students, i.e., the semantics are more 'spread out' (see Table 8.2). This may be explained by practitioners having more exposure to specific interpretations of some concepts, causing a lower spread of measurable semantics. Nonetheless, both practitioners and students are still easily divided into communities based on their semantic differences.

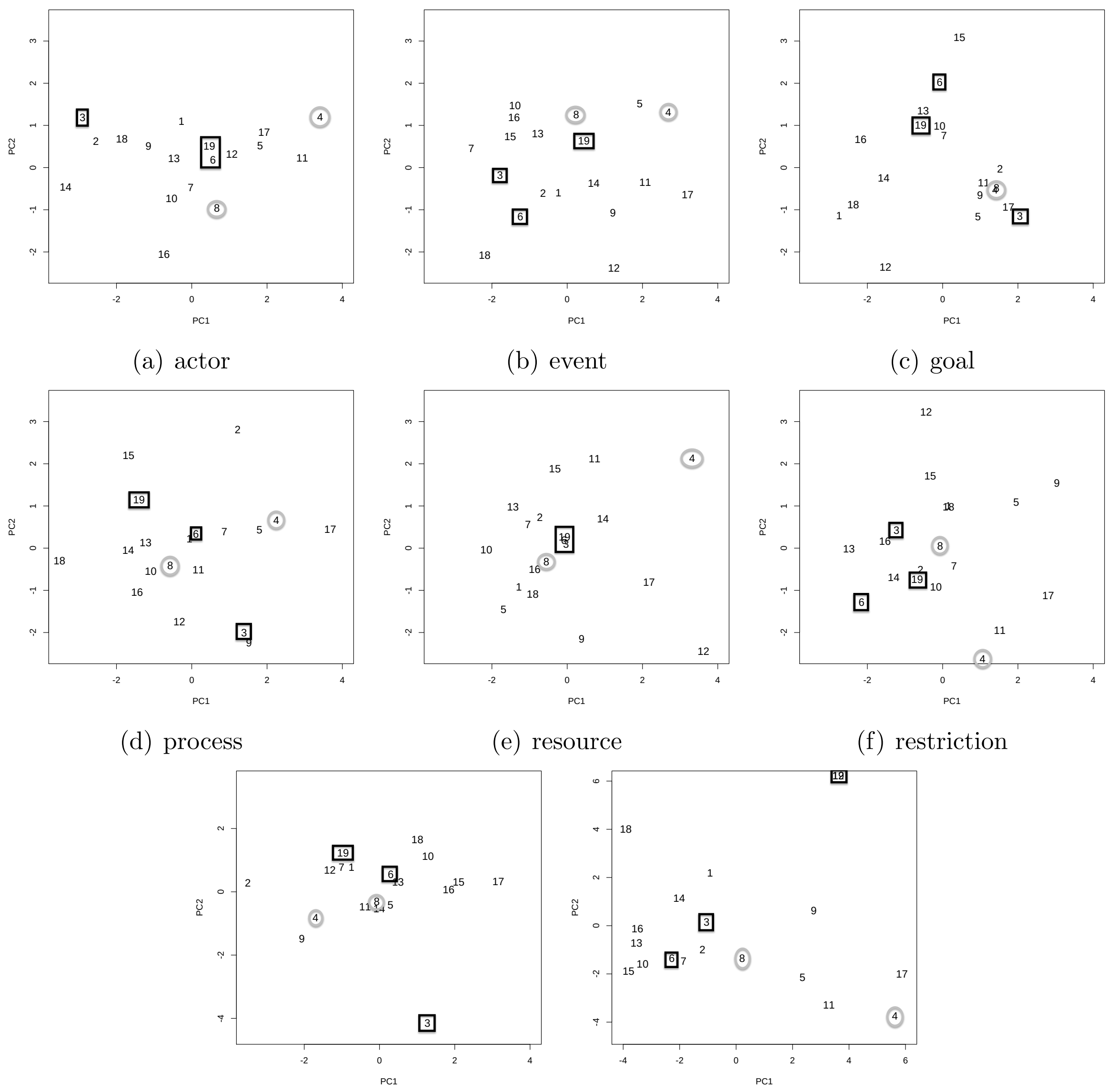


(a) actor (b) event (c) goal

(d) process (e) resource (f) restriction

(g) result (h) overall (averaged)

Figure 8.1: Visualization of the principal component analysis for the investigated meta-concepts (and average overall understanding) in the practitioner sample. The visualizations represent (roughly) the distance between understandings which individual participants have. The further away two participants are on both axes (i.e., horizontal and vertically different co-ordinates), the more different their conceptual understanding has been measured to be. Colored boxes and circles are used to highlight some interesting (potential) groups of participants that are discussed in more detail.



(a) actor (b) event (c) goal

(d) process (e) resource (f) restriction

(g) result (h) overall (averaged)

Figure 8.2: Principal components found in the data of concept-specific understandings for students. The visualizations represent (roughly) the distance between understandings which individual participants have. The further away two participants are on both axes (i.e., horizontal and vertically different coordinates), the more different their conceptual understanding has been measured to be.

8.5.1 Finding communities

To demonstrate the existence and structure of the found communities, we will discuss some of the clusters we found for the understanding practitioners and students have of GOALS, PROCESSES, RESOURCES and RESTRICTIONS. The immediately obvious difference between the practitioners and students is that, where there are clusters to be found

Table 8.2: Comparison of variance for each concept in the investigated data samples. Wilcoxon-testing on the variance to test whether one sample had a lower spread was negative (V=8, p=0.1875). While overall variance is not significantly different, a number of meta-concepts (i.e., ACTOR, EVENT and GOAL) do display potentially interesting disparities.

sample	actor	event	goal	process	resource	restriction	result
practitioner	0.38	0.57	0.68	0.93	0.73	0.94	0.92
student	0.66	0.77	1.01	0.93	0.82	0.83	0.81

amongst the practitioners, they differ mostly on one axis (i.e., component), whereas the students often differ wildly on both axes. Of particular interest to testing our hypothesis are participants 3 & 8, and 2, 7 & 10 from the practitioner data sample. The first community clusters together very closely for their understanding of RESTRICTIONS (and GOALS, albeit to a lesser degree) while they differ only slightly for most other concepts. This means one would expect them to share some real-world properties. Perhaps they are people specialized in goal modeling, or share a typical way of modeling RESTRICTIONS in a formal sense. The second community (participants 2, 7 & 10) cluster together very closely for RESOURCES, fairly close for GOALS and RESTRICTIONS, while being strongly different when it comes to their understanding of PROCESSES.

One could expect this to mean that they have some shared focus on RESOURCES, either through a language they use (e.g., value-exchange or deployment languages) which are often strongly connected to GOALS (as either requiring them, or resulting in their creation). Conversely, one would not necessarily expect there to be much overlap between the participants with regard to PROCESSES, as they are grouped with a wide spread.

For the students, there are several potentially interesting communities to look at. Participants 4 & 8 differ strongly for several concepts (e.g., their strong differentiation on two components for RESOURCES, and for PROCESSES and RESTRICTIONS), but they have an almost exactly similar understanding of GOALS. One would expect that some kind of property shared between them might be used to identify other participants that cluster together for GOALS, but not necessarily share other understandings. Participants 3, 6 & 19 also cluster together closely for one concept – RESOURCES – but differ on their understanding of the other investigated concepts. As such, if (some) experience in the form of having used specific programming and modeling languages is correlated to their conceptual understanding, one would expect to find some reflection of that in the clusterings of these students.

8.5.2 Identifying communities

However, when we add the information we have about the participants (see Tables 8.3 and 8.4) to these clusters, we run into some problems. It is often the case that communities do not share (many) pertinent properties, or when they do, there are other communities with the same properties that are far removed from them in terms of their conceptual understanding. For instance, consider participants 2, 7 & 10 (highlighted with a gray oval) from the practitioner data sample. While they share some properties, (e.g. operating in the same sector, having some amount of focus on PROCESSES, and interacting with domain experts), when we look at other communities it is not as simple to use this combination of properties to uniquely identify them. For instance, participants 3 & 8 (highlighted with a black rectangle) cluster together closely in their own right, but do share some overlapping properties (both operate in the government sector). Thus, merely looking at the sector a modeler operates in cannot be enough to identify them.

Table 8.3: Comparison of some practitioners based on investigated properties. The proprietary language is an in-house language used by one of the involved companies.

No.	Used languages	Sector	Focus	Interacts with
3	Proprietary	Financial, Government	Knowledge rules, processes, data	Analists, modelers
8	UML, OWL, RDF, Mindmap, Rulespeak, Proprietary	Government, Healthcare	Rules	Business professionals, policymakers, lawyers
2	Proprietary	Government	Knowledge systems, processes	Managers, domain experts
7	Proprietary, UML, Java	Government, spatial planning	Business processes, process structure	Domain experts, IT specialists
10	Proprietary, XML, XSLT	Government, finance	Processes, rules, object definitions for systems	Domain experts, java developers
9	Proprietary	Government, Financial	Rules, legislation, policy, processes	Domain experts

Another interesting observation is the fact that while participants 2, 7 & 10 cluster together closely for a number of concepts (e.g., GOALS, RESOURCES and RESTRICTIONS), they do not appear to have a similar understanding of what constitutes a PROCESS, even though they all share a strong focus on modeling processes. Looking at the combination of sector and focus is not enough either, as under these conditions participant 8 and 10 should be grouped closer together because they both have a focus on rules. When we finally look at the combination of sector, focus and interaction we have a somewhat higher chance of uniquely identifying communities, although there are still counter-examples. Participant 9 (highlighted with a gray rectangle), for example, shares all the properties with participants 2, 7 & 10, but is conceptually far removed from all others. The dataset shows a similar trend for most other participants, providing both examples and counterexamples for most of these property combinations, making it generally very difficult, if not impossible to identify communities.

We face the same challenge in the student data sample, although even more pronounced on an almost individual level. There are participants that share the same properties while having wildly varying conceptual understandings. There seems to be some differentiation on whether participants have prior experience, but even then this sole property does not have enough discriminatory power. Take for example participants 4 & 8 (highlighted with a black rectangle) and participants 3, 6 & 19 (highlighted with a gray oval). Both these communities cluster closely together for a specific concept, but then differ on other concepts. One could expect this has to do with a small amount of properties differing between them, which is the case, as there is consistently a participant with some prior experience in programming and scripting languages amongst them. However, if this property really is the differentiating factor, one would expect that on the other concepts the participants with prior experience (4 & 6) would be further removed from other participants than the ones without experience are, which is simply not the case. It thus seems rather difficult to link these properties to the communities and their structure.

This challenge could be explained by a number of things. First and foremost would be a plain lack in the amount of properties (or their granularity, as might be the case in the student data sample) to identify communities by, while it is also possible that the investigated concepts were not at the right abstraction level (i.e., either too specific or too vague), or that the investigated concepts were simply not the concepts people use to model. We will discuss each of these possibilities.

The simplest explanation is that the properties we attempt to identify communities by are not the right (i.e., properly discriminating) ones. It is possible (especially for the student data sample) that some of the properties are not necessarily the wrong ones, but that they are not discriminative enough. For example, knowing what modeling languages someone

Table 8.4: Comparison of some students based on investigated properties. Profiles are standardized packages of coursework students took during secondary education, nature being natural sciences, technology a focus on physics and health a focus on biology.

No.	Study	Profile	Prior experience
4	Computing Science	Nature & Technology & Health	Some programming and scripting experience
8	Computing Science	Nature & Technology	None
3	Information Systems	Nature & Technology	None
6	Computing Science	Nature & Technology	Programming experience
19	Information Systems	Nature & Health	None

uses could be described in more detail because the language could have multiple versions that are in use, and it is possible (indeed quite likely) that a language used is not the same as the 'official' language. However, this line of reasoning is problematic for two reasons. The first being that these are properties that are used by practitioners to (naively) group modelers together, the second that there is no clear-cut way to identify reasonable other properties that are correlated to the modeling practice. If these properties are not useful, we would have to reject the hypothesis on the grounds of them being a 'bad fit' for grouping people. Other properties that could be thought of could include reflections of the cultural background of modelers. However, these are less likely to be of influence in our specific case as the Enterprise Modelers we investigate are all set in a Western European context and there is little cultural diversity (or granularity, as might be the case in the student data sample) in this sense.

Another explanation could be that the meta-concepts we chose are not at the right abstraction level (i.e., concept width), meaning that they are either too vague or specific. For example, some modelers could typically think on near-instantiation level while others think more vaguely. If concepts are very specific, one would expect to find differences much faster (as the distance between people's conceptual understanding can be expected to be larger), which thus makes it *easier* to find communities. If they are (too) vague though, people would not differ much because there are not enough properties to differ on in the first place. However, the way we set up our observations rules out the vagueness possibility, as participants were given a semantic priming task before the semantic differential task of each concept. What we investigated was thus their most typical specific understanding of a concept. For this reason it is unlikely that the abstraction level of the concepts was the cause of the challenge of identifying the communities.

Finally, the most obvious explanation could be a flaw in our preliminary work, namely that we did not select the right meta-concepts, irrespective of their abstraction level. Considering the meta-concepts were derived from an analysis of conceptual modeling languages and methods used for many aspects of enterprises, and that there simply does not seem a way to do without most of them, we find it very unlikely this is the case. The unlikely option that what we investigated was not actually the modeling concept, but something else entirely (i.e., someone considering their favorite Hollywood actors over a conceptual modeling interpretation of ACTOR) can also be ruled out as the priming task in our observation rules out this possibility. It is therefore unlikely that these potential issues affected our analysis, leading us to conclude that the identification of communities of modelers based on the investigated properties is not feasible.

While we had admittedly hoped that these observations would yield a positive result to the hypothesis, the lack of support we have shown means that a theory of predicting how modelers understand the key concepts they use, and thus what the additional 'implicit' semantics of a model could be (as alluded to in the introduction) is likely not feasible. Nonetheless, the observations do help to systematically clarify that these different personal understandings exist, can be measured, and might be correlated to communication and modeling breakdown due to unawareness of linguistic prejudice.

8.5.3 Consequences

If we wanted to simply discount the possibility of these properties being good ways to identify communities that share a semantic understanding of some concepts with, we would now be done. But as noted in the introduction, the rejection of this hypothesis carries with it certain consequences, especially as these properties *are* being used to identify communities and group people together in practice (e.g., the earlier discussed workshops within the ASD project Lankhorst (2012)). Our findings are thus of direct relevance to groups like model facilitators and enterprise modelers as they can use these kind of findings to support them in determining good and effective modeling strategies (e.g., by having more of an insight into the basic 'kinds' of modelers, being more aware of common differences). More generally, the consequences of our findings are not simply that we should stop grouping modelers together in a naive fashion, but that we should strive to gain a better understanding of why we do so, what else we might do in its stead, and what avenues of research should be explored to deal with the consequences.

Like the studies presented in Chapter 6, this study thus again re-iterates the need that before actually modeling a domain, whether with modelers or stakeholders, it would be prudent to discuss the understandings the involved parties have of the (meta-)concepts to be used. This should not be relegated to a purely abstract discussion of the types (meta-concepts) used in the modeling language, but should rather focus on exploring what in the universe of discourse needs to be modeled, and as a result, what types are needed for this. As a consequence, one can focus on elaborating how the people involved understand those meta-concepts. For example, when modeling a specific universe of discourse which entails the necessity to model rules and the way they affect people, it can easily be derived that some meta-concept for rules or restrictions is necessary. We can then move towards a discussion concerning what kind of properties this meta-concept should *at least* be able to distinguish between, e.g., that some rules are logical conditions that cannot be violated (alethic), while some other ones are moral conditions that can, but ought not be violated (deontic). As a result of having done this, we now know before actually starting to model what the modeling language as such should accommodate.

This study furthermore stresses the point that a 'model' should not just be regarded solely in terms of its graphical or textual representation. Instead, we need to understand that the actual model underlying whatever form it is represented in contains more information than the representation itself. This includes, for example, the personal understanding that people have of the concepts and meta-concepts, and in particular, the (joint) understanding of the meta-concepts used by the model. To ensure that one does not leave out these personal understandings and their possible effects during the model creation and use, a number of practices can be applied during the modeling process.

However, we should not focus exclusively on attempting to solve the issue by looking only at the conceptual aspects. Attempts to understand more clearly the reasons and challenges in the modeling process as discussed could be undertaken in, for example by understanding why people become part of a community (in our case, of shared semantics) could help to deal with their conceptualization processes by understanding more clearly how the group dynamics affect them. Several drivers (e.g., economical, political and cultural) could drive people to become part of such communities and have received attention already (see e.g., Huang et al. (2002) or recent work in EA by Niemietz (2013)), and is a worthwhile angle of investigation to extend our understanding of such group dynamics.

To summarize, we have shown that the often implicit assumption that "people have strongly comparable semantics for the common modeling meta-concepts if they share an expertise in certain sectors, modeling focus or used languages" cannot be backed up by our empirical investigation. While not an exhaustive disproof of the hypothesis by any means, it casts enough doubt on it that it would be a considerate practice for those involved in (collaborative) modeling efforts to be more careful and double-check their assumptions when modeling together with, or using models from, others practitioners.

8.6 Summary & outlook

In this chapter we have shown a way to discover communities that share personal semantics through analysis of psychometric data and discussed the difficulties in identifying them through shared properties between their members. On the basis of this we have rejected the hypothesis that modelers with certain shared properties (such as used languages, background, focus, etc.) can be easily grouped together and expected to share a similar understanding of the investigated meta-concepts. While this offers support for one of our main research questions, "*Can particular (kinds of) conceptual understandings of certain meta-concepts be attributed to a community?*", it more importantly gives a negative answer to the follow up question of whether those attributions can be used to predict a person's typical understanding of a language construct simply by finding their community membership. Furthermore, we have discussed the consequences of these findings for the modeling process and elaborated on what avenues of research might prove fruitful with these consequences.

Part IV

Closing

CHAPTER 9

Conclusions and Further Research

Abstract. This chapter discusses the conclusions reached in this thesis and provides directions for further research. First, a summary of the answers for each of the research questions is given and briefly discussed, which substantiates the contributions that this work has made. Second, some aspects that can be studied in more detail and other issues still left open are described as potential avenues for further research.

9.1 Answers, summarized

While the main findings and conclusions of each study are discussed in detail in their own Chapters, we will briefly summarize the answers to each of the individual research questions here.

9.1.1 Fundamental part

1. What are the meta-concepts (i.e., language constructs) shared between modeling languages used in enterprise modeling? We derived a number of meta-concepts common to conceptual modeling languages through an analysis of modeling language notations. These meta-concepts are ACTOR, EVENT, GOAL, PROCESS, RESOURCE, RESTRICTION, and RESULT. Chapter 3.3 goes into more detail on the derivation process and the meaning of these meta-concepts and what exemplars fit into them. Furthermore, we also derived a number of dimensions which can be used to distinguish the different interpretations people have of these meta-concepts (e.g., for person x an ACTOR is a *natural, human* thing, while for person y it is a *non-natural, non-human, material* thing). These dimensions can be found in Table 3.3.

2. What is the conceptual understanding that people have of these meta-concepts? The conceptual understanding that people have of these constructs widely diverges from person to person. In Chapter 6 we have presented the results of two semantic differential studies that measured the conceptual understanding enterprise modelers and computing/information science students have of the meta-concepts. It did so by finding how well the derived dimensions related to each meta-concept (i.e., whether a RESOURCE is a material thing, a human thing, and so on). The main finding of both studies was that there is a wide range of different conceptual understandings held from person to person. However, some aspects do seem to be shared between people, as well as over the course of the longitudinal study. These are for example that the meta-concepts are conceptualized as (having to be) non-vague things, and many of the meta-concepts being of an intentional nature (e.g., goals are formulated for a reason, resources are specifically listed for some purpose, results are intentionally created). Further details can be found in the discussion of the study among practitioners (see Section 6.3), and the study among students (see Section 6.4). However, even though there are some common aspects of the conceptual understanding of the meta-concepts, we should keep in mind the strong individual differences between the way people see each meta-concept, and attempt to understand how these individuals involved in a modeling effort understand the used meta-concepts.

2.1 To what extent do the categories for these meta-concepts more resemble a graded or discrete structure (e.g., artifactual or natural)? The categories for the meta-concepts are likely of a graded nature (that is, some of their members are considered better members than others), as these categories reflect artifactual benchmark categories and contain many members (i.e., terms) that are judged to be partial members. As a result, the validity of instantiations of models containing these meta-concepts should be considered to be a graded affair as well. While a model can have several valid instantiations in a given domain, some of them might be more conceptually valid than others. The data used to support this conclusion is given and discussed in Section 4.3.2.

2.2 To what extent do conceptual understandings differ on key points between different people (e.g., different sets of features, typicality)? With the feature listing study presented in Chapter 5 we found that there is a wide variety of different typical views held about what the meta-concepts are to people. As the participants listed their most typical features for the meta-concepts, it is easier to pinpoint what is shared between people's conceptual understandings, instead of pinpointing what is different. Outside of these conserved features shared by many people there are strong differences from person to person in how they see each meta-concept, and thus also how they differ in their conceptual understanding. As discussed in Section 5.4, we can see for the meta-concepts that there are conserved feature aspects such as e.g., the ability and capability of an actor to act, events being triggers for things, resources being identified by their nature (e.g., material or immaterial).

3. Does a person's conceptual understanding of a meta-concept change over time? A person's conceptual understanding can be observed to change over time, as we have discussed for the longitudinal semantic differential study performed with students in Section 6.4. On the one hand, most of the measured conceptual understandings changed over time, even if only slightly, during the course of the study. However, the way in which they changed (sometimes moving back and forth from a specific understanding) was difficult to characterize, to such a degree that it is difficult to attribute these changes to any external factors. On the other hand, the 'average' conceptual understanding taken of all participants is more stable, with no particularly significant (or systematic) changes occurring over time for the general trend. As can be seen in Fig. 6.2, for most meta-concepts the dimensions *natural, human*, and *vague* remain negative, while for the dimensions *necessary* and *intentional* they stay positive. Thus, taken on average there does not seem to be a significant switch of polarity occurring over time.

3.1 If so, does training in particular languages or techniques affect these base conceptual understandings? There do not seem to be any significant correlations between the way the the conceptual understandings we measured changed, and the kinds of educational stimuli that the students received (i.e., the courses they took, modeling language and techniques they learned). Thus, while training in a particular language or technique might affect the base conceptual understanding someone has of the meta-concepts we study, it would be difficult to pinpoint what the effect would be. Because of this, it is difficult to know what to teach people in order to affect their conceptual understanding in a meaningful way (e.g., ensuring people trained in a particular value modeling technique conceptualize resources in the same way).

9.1.2 Practical part

4. To what extent do modeling languages allow for people to express their particular conceptual understanding of meta-concepts? As detailed in Chapter 7.1, the degree to which existing modeling languages allow one to express their particular conceptual understandings varies. We used the meta-concepts and dimensions to investigate whether combinations of them were accommodated by a modeling language's specification (e.g., "Is it possible to model an actor element as a human being?"), and did so for a number of meta-concept and dimension combinations. Of these, around half of the meta-concept and dimension combinations we investigated were not accommodated by any investigated modeling language. Thus, it does not seem to be the case that modeling languages are inherently semantically expressive, even for those conceptual distinctions that we have found to be particularly salient among enterprise modelers. Furthermore, some common ways of conceptualizing things were found to have no support at all, such as making an explicit distinction between the necessity of goals exhibited with Clegg and Barker (1994)'s MoSCoW technique. However, nuances to this should also be added, as accommodating each possible different conceptualization of a meta-concept would increase the complexity of a modeling language to such a degree that it would become nearly impossible to use the language in a practical setting. These findings and issues are discussed in more detail in Section 7.3.1.

4.1 Do modeling languages make a distinction between explicitly or implicitly allowing them to do so? From language to language there are differences as to whether those conceptual distinctions that are accommodated by the notational elements from the language are done so explicitly (i.e., having their own distinct notational elements), or implicitly. By implicit accommodation we mean the case where language constructs are defined in such a way that they assume a particular conceptual understanding (e.g., actor elements are human beings), while this interpretation is not necessarily clear from the notation itself. In such a case it can be difficult to infer the exact semantics of a model without knowledge of (or access to) the specification of the language. Some languages used variations in the visual elements for their meta-concepts which denoted these differences as well, such as the ArchiMate notation using stick puppets for actor elements to denote they represent people.

4.2 Can particular (kinds of) conceptual understandings of certain meta-concepts be used to infer which modeling language or notation is best suited for a specific modeler? While it is possible to look at individual people and their conceptual understanding of the common meta-concepts to infer which language accommodates most of what is important to them, a modeling language is rarely ever used by one person. Given that we have discussed the strong amount of differences between how people understand the meta-concepts, it is thus difficult to choose a 'best' modeling language in terms of how many important conceptual distinctions are accommodated. Instead, one can opt to use a language which allows for the explicit modeling of these distinctions (e.g., by virtue of its notation allowing people to easily discriminate between differences, such as with the stick puppet example used in ArchiMate). Doing so, and avoiding languages for which the semantics include many specific interpretations of meta-concepts (e.g., actors are human things, resources are material things) ensures that for a modeling effort with many different people involved the different conceptual understandings they have can all be captured and more importantly, explicitly presented to one another.

5. Can particular (kinds of) conceptual understandings of certain meta-concepts be attributed to a community (of discourse, practice, etc.)? In Chapter 8 we discussed the results of a study that aimed to discover communities with shared conceptual understandings, and then, based on their members' real-world properties (e.g., what these like-minded people share in terms of professional background, focus, used modeling languages) investigated whether predicting community membership based on such real-world properties was feasible. While this study found that it was possible to find groups of people that shared similar conceptual understandings, attributing them to a particular set of real-world properties (e.g, a community of discourse or practice) was more difficult. The individuals grouped together in the clusters often did not have a significant overlap of real-world properties, or in the cases where that did happen, there were individuals outside of the cluster who shared these same properties and should have thus been in the same cluster, if the conceptual understanding of a meta-concept was to be attributed to these properties.

5.1 Can this be used to predict a person's typical conceptual understanding of a meta-concept by finding their community membership? Given that we found that particular kinds of conceptual understandings of meta-concepts cannot be attributed easily to some combination of real-world properties, we also rejected the idea of predicting conceptual understandings of an individual based solely on their community membership.

9.2 Proposals for further research

9.2.1 Group and context differentiation

Most of the studies we have performed in this thesis used non-specialized groups of participants. That is, we sought out people with experience in enterprise or conceptual modeling, people working in conceptual modeling, students studying relevant topics and so on. Doing so led to many interesting findings, but makes them general in terms of what kind of 'conceptual modeler' they are valid for. It can be expected that within different domains (e.g., telecom, government, finance, law) there are particular ways of looking at the world, just as well as one could still expect that modelers focused on some particular aspect (e.g., business processes, value exchanges, architecture, rules) might have a particular way of looking at the world.

For several of the things we investigated (i.e., category structure, feature listing, conceptual understandings) it might be interesting to repeat these studies with more strongly differentiated groups of people. Doing so could lead to interesting insights into what different domains and aspects look at, and what is important to them. On the other hand, given the findings from our last study into detecting communities with shared personal semantics, and the difficulty we had in attempting to find coherent groups of similar people (e.g., sharing real-world properties) that also had similar personal semantics, these group and context differentiated studies might also not necessarily yield results specific to their aspect. However, even if that is the case, more data which either corroborates earlier findings (for example, that of the meta-concept categories being of a graded nature), or expands and builds further on them (for example, by eliciting more lists of features for meta-concepts) would still be valuable enough to justify the research effort.

The studies we have performed on modeling concepts and terminology were performed in isolation. That is, we investigated how people thought about and conceptualize modeling concepts and terminology, but presented these concepts and terms out of their logical context (i.e., in an actual model). While this has obvious advantages, such as that of easily managing the complexity of stimuli and having some control over not introducing unwanted information, it can have an effect on the results (or the amount of it). The more context that is available to a person about a given concept, the more specific their understanding will likely be. Thus, if we were to do, say, a feature listing study where we did not simply ask what someone thought of some term with a small description, but provided them with a number of (visual) models and examples of the use of the concept we wish to study, more detailed information could potentially be elicited.

9.2.2 Quantitative determination of feature typicality

The feature listing study we have performed has a limited amount of quantitative information about the typicality of the elicited features. By counting the amount of times participants listed a (similar) feature an occurrence score was computed. However, this cannot be used to compare how typical each feature is, as simply assuming that features listed more often are more typical is incorrect. People might not have listed features because they did not know about them, regardless of how typical they might be for them. Incorporating an explicit quantitative survey in which we investigate how typical each feature elicited earlier on is (e.g., on a Likert scale) would result in a dataset from which the relative typicality of these features could be computed. With such data gathered, a ranking list of what kind of conceptual distinctions must, should and ideally would be supported by a modeling language can then be systematically produced. Performing such work for different specialized groups (e.g., business process modelers, goal modelers) as argued above, and repeating it on regular intervals should lead to a situation where we do not only have useful methods to ensure that a modeling language is conceptually complete, but that they can be based on tried relevant personal insights as well.

9.2.3 Determination of important distinctions

One of the more important things that should also be done is to ensure a more detailed insight into what distinctions are conceptually relevant and important to actual users of modeling languages (i.e., *modelers*) and a created model's end-users (i.e., *stakeholders*). We should focus on doing empirical work based on finding out what entities (and with which properties they manifest themselves) are vital to modelers and stakeholders for the typical domains they model, and in doing so determine what the conceptual needs are for domain-specific languages for particular domains (e.g., that goal modeling languages need to at least explicitly distinguish between hard and soft goals, goals that have to be achieved versus goals that ought to be achieved, and so on).

In this thesis we have derived the meta-concepts and dimensions through an analysis of modeling languages. However, more dimensions should be found by investigating what conceptual distinctions are important to modelers. This could be done with more in-depth qualitative work, exploring the aspect that a modeler works on, and trying to elicit as much information about what is important to them in this domain. Interviews with a number of enterprise modelers could yield a large amount of text to analyze for such potential conceptual distinctions, which can then be tested for their importance in simpler surveys, for instance by asking a large group of modelers how important some dimension is for their focus.

Finally, it is important to be aware of the constantly changing conceptual landscapes that modelers and stakeholders operate in. The concepts that become important to modelers and stakeholders can appear and change much more quickly. For this reason alone we should also focus our efforts on a repeating empirical effort to elicit such conceptual needs.

9.3 Closing remarks

In this thesis we have made a number of contributions (see Section 1.5.3) to the state of knowledge on how individual people conceptualize modeling language constructs (i.e., meta-concepts), and what implications these different personal understandings have for both the process of modeling (e.g., more explicitly focusing on the clarification of meta-concepts during communication about models, using modeling languages, grouping modelers), as well as the things used for it (e.g., what kind of formalizations are necessary to capture particular kinds of understandings, what modeling languages accommodate some conceptual understandings). First, we have derived a number of common meta-concepts shared between modeling languages used in enterprise modeling, and a number of semantically distinct dimensions on which to distinguish between different conceptual understandings of these meta-concepts. With these inputs we have performed several studies that have shown, amongst others, that the category structure for these meta-concepts is of a graded nature, what typical features for these meta-concepts are, what typical conceptual understandings are and how these affect to what degree a model's instantiation is conceptually valid, that these conceptual understandings are not always accommodated by modeling languages, and finally, that grouping people together based on similar conceptual understandings is no simple task.

These contributions do not directly solve the larger issue of it being difficult to *a posteriori* create an integrated conceptual model of an enterprise and ensure that it is reflective of all the involved people's conceptual understandings. However, the presented insights into how people understand the meta-concepts, and how we can measure and make them explicit are valuable for modeling efforts in practice to make them more aware of the fact that people have different personal semantics. Our main hope is that with the findings presented in this thesis more people in practice become aware of how different people can understand these core modeling meta-concepts. If they as well *acknowledge* that people can have different, valid understandings of the world and their modeling languages, and that the only way to ensure correct models is to constantly talk to one another about the way we understand not only our domain, but the meta-concepts we use to model as well, the quality and consistency of models and their integrationss should be, even if only slightly, more guaranteed.

References

B. Adelson. Comparing natural and abstract categories: A case study from computer science. *Cognitive Science*, 9(4):417 – 430, 1985.

X. Aimé, F. Furst, P. Kuntz, and F. Trichet. Conceptual and Lexical Prototypicality Gradients Dedicated to Ontology Personalisation. In R. Meersman and Z. Tari, editors, *On the Move to Meaningful Internet Systems: OTM 2008*, volume 5332 of *Lecture Notes in Computer Science*, pages 1423–1439. Springer Berlin Heidelberg, 2008.

X. Aimé, F. Furst, P. Kuntz, and F. Trichet. Ontology Personalization: An Approach Based on Conceptual Prototypicality. In L. C. et al., editor, *Advances in Web and Network Technologies, and Information Management*, volume 5731 of *LNCS*, pages 198–209. Springer Berlin Heidelberg, 2009.

H. Alani and N. Shadbolt. Identifying communities of practice: Analysing ontologies as networks to support community recognition. In *Proceedings of the 2002 IFIP World Computer Congress*, 2002.

M. B. Almeida. A proposal to evaluate ontology content. *Applied Ontology*, 4(3-4):245–265, August 2009.

J. Altarriba and L. M. Bauer. The distinctiveness of emotion concepts: A comparison between emotion, abstract, and concrete words. *The American journal of psychology*, pages 389–410, 2004.

V. Anaya, G. Berio, and M. Verdecho. Evaluating quality of enterprise modelling languages: The ueml solution. In R. J. Gonalves, J. P. Mueller, K. Mertins, and M. Zelm, editors, *Enterprise Interoperability II*, pages 237–240. Springer London, 2007.

V. Anaya, G. Berio, M. Harzallah, P. Heymans, R. Matulevicius, A. L. Opdahl, H. Panetto, and M. J. Verdecho. The unified enterprise modelling language–overview and further work. *Computers in Industry*, 61(2):99 – 111, 2010. Integration and Information in Networked Enterprises.

F. Ashby and W. Maddox. Relations between prototype, exemplar, and decision bound models of categorization. *Journal of Mathematical Psychology*, 37(3):372 – 400, 1993.

C. P. Ayala, C. Cares, J. P. Carvallo, G. Grau, M. Haya, G. Salazar, X. Franch, E. Mayol, and C. Quer. A comparative analysis of i*-based agent-oriented modeling languages. In *SEKE'05, Taipei, Taiwan*, pages 43–50, 2005.

J. Barjis, G. Kolfschoten, and A. Verbraeck. Collaborative enterprise modeling. In Proper et al., editor, *Advances in Enterprise Engineering II*, volume 28 of *LNBIP*, pages 50–62. Springer Berlin Heidelberg, 2009.

R. Barr and L. Caplan. Category representations and their implications for category structure. *Memory & Cognition*, 15(5):397–418, 1987.

R. Bidarra, E. Van Den Berg, and W. Bronsvoort. Collaborative modeling with features. In *Proceedings of DET*, volume 1, page 2001, 2001.

Y. Bishr, W. Kuhn, and M. Rdwan. Probing the concepts of information communities - a first step towards semantic interoperability. *Interoperating Geographic Information Systems*, 495:55–70, 1999.

P. Bloom. Theory of artifact categorization. *Cognition*, 66(1):87 – 93, 1998.

C. Brandt and F. Hermann. Conformance analysis of organizational models: a new enterprise modeling framework using algebraic graph transformation. *International Journal of Information System Modeling and Design (IJISMD)*, 4(1):42–78, 2012.

R. Breu, U. Hinkel, C. Hofmann, C. Klein, B. Paech, B. Rumpe, and V. Thurner. Towards a formalization of the Unified Modeling Language. In M. Aksit et al., editor, *ECOOP'97 Object-Oriented Programming*, volume 1241 of *LNCS*, pages 344–366. Springer Berlin Heidelberg, 1997. 10.1007/BFb0053386.

K. Britz, J. Heidema, and T. Meyer. Modelling object typicality in description logics. In A. Nicholson and X. Li, editors, *AI 2009: Advances in Artificial Intelligence*, volume 5866 of *LNCS*, pages 506–516. Springer Berlin Heidelberg, 2009.

A. Burton-Jones, Y. Wand, and R. Weber. Guidelines for empirical evaluations of conceptual modeling grammars. *Journal of the Association for Information Systems*, 10 (6):495–532, 2009.

Y. Cai and H.-F. Leung. A formal model of fuzzy ontology with property hierarchy and object membership. In Li et al., editor, *ER*, volume 5231 of *LNCS*, pages 69–82. Springer Berlin, 2008.

R. Chaffin. Associations to unfamiliar words: Learning the meanings of new words. *Memory & cognition*, 25(2):203–226, 1997.

P. P.-S. Chen. The entity-relationship model—toward a unified view of data. *ACM Trans. Database Syst.*, 1(1):9–36, Mar. 1976.

W. Chen and R. Hirschheim. A paradigmatic and methodological examination of information systems research from 1991 to 2001. *Information Systems Journal*, 14(3):197–235, 2004.

R. Clarke, A. Burton-Jones, and R. Weber. Improving the semantics of conceptual-modeling grammars: A new perspective on an old problem. In *Proceedings of the Thirty Fourth International Conference on Information Systems: ICIS 2013*, pages 1–17, 2013.

D. Clegg and R. Barker. *Case method fast-track: a RAD approach*. Addison-Wesley Longman Publishing Co., Inc., 1994.

T. Colburn and G. Shute. Abstraction in computer science. *Minds and Machines*, 17(2): 169–184, 2007.

N. P. Dalal, M. Kamath, W. J. Kolarik, and E. Sivaraman. Toward an integrated framework for modeling enterprise processes. *Communications of the ACM*, 47(3):83–87, 2004.

A. Dardenne, A. van Lamsweerde, and S. Fickas. Goal-directed requirements acquisition. *Sci. Comput. Program.*, 20:3–50, April 1993.

I. Davies, P. Green, M. Rosemann, M. Indulska, and S. Gallo. How do practitioners use conceptual modeling in practice? *Data & Knowledge Engineering*, 58(3):358 – 380, 2006.

D. Dean, R. Orwig, and D. Vogel. Facilitation methods for collaborative modeling tools. *Group Decision and Negotiation*, 9:109–128, 2000.

D. Delen, N. P. Dalal, and P. C. Benjamin. Integrated modeling: the key to holistic understanding of the enterprise. *Communications of the ACM*, 48(4):107–112, 2005.

F. J. Di Vesta and W. Dick. The Test-Retest Reliability of Children's Ratings on the Semantic Differential. *Educational and Psychological Measurement*, 26(3):605–616, 1966.

R. Dieng and S. Hug. Comparison of personal ontologies represented through conceptual graphs. In *Proceedings of the European Conference on Artificial Intelligence (ECAI)*, pages 341–345. Citeseer, 1998.

G. Diesendruck and S. Gelman. Domain differences in absolute judgments of category membership: Evidence for an essentialist account of categorization. *Psychonomic bulletin & review*, 6(2):338–346, 1999.

R. M. Dijkman, M. Dumas, and C. Ouyang. Semantics and analysis of business process models in BPMN. *Information and Software Technology*, 50(12):1281 – 1294, 2008a.

R. M. Dijkman, D. A. C. Quartel, and M. J. van Sinderen. Consistency in multi-viewpoint design of enterprise information systems. *Inf. Softw. Technol.*, 50(7-8):737–752, 2008b.

Z. Estes. Domain differences in the structure of artifactual and natural categories. *Memory & cognition*, 31(2):199–214, 2003.

Z. Estes. Confidence and gradedness in semantic categorization: Definitely somewhat artifactual, maybe absolutely natural. *Psychonomic bulletin & review*, 11(6):1041–1047, 2004.

D. Ferrariolo, J. Cugini, and R. Kuhn. Role-based access control (RBAC): Features and motivations. In *Proceedings of the 11th Annual Computer Security Applications Conference*. IEEE Computer Society Press, 1995.

P. Fettke and P. Loos. Ontological evaluation of reference models using the bunge-wand-weber models. In *AMCIS*, volume 384, pages 2944–2955, 2003.

P. Fettke, C. Houy, A.-L. Vella, and P. Loos. Towards the reconstruction and evaluation of conceptual model quality discourses – methodical framework and application in the context of model understandability. In Bider et al., editor, *Enterprise, Business-Process and Information Systems Modeling*, volume 113 of *LNBIP*, pages 406–421. Springer Berlin Heidelberg, 2012.

J. R. Firth. A Synopsis of Linguistic Theory, 1930-1955. *Studies in Linguistic Analysis*, pages 1–32, 1957.

G. Flake, S. Lawrence, C. Giles, and F. Coetzee. Self-organization and identification of web communities. *Computer*, 35(3):66–70, 2002.

J. A. Fodor. *Concepts: Where cognitive science went wrong.* Clarendon Press/Oxford University Press, 1998.

M. S. Fox and M. Gruninger. Enterprise modeling. *AI magazine*, 19(3):109, 1998.

R. France and B. Rumpe. Domain specific modeling. *Software and Systems Modeling*, 4 (1):1–3, 2005.

F. Francesconi, F. Dalpiaz, and J. Mylopoulos. TBIM: A language for modeling and reasoning about business plans. Technical Report DISI-13-020, University of Trento. Department of Information Engineering and Computer Science, May 2013.

U. Frank. Multi-perspective enterprise modeling (MEMO) - conceptual framework and modeling languages. In *Proceedings of the 35th Annual Hawaii International Conference on System Sciences*, volume 3, pages 1258 – 1267. IEEE Computer Society, 2002.

U. Frank, D. Heise, H. Kattenstroth, D. Ferguson, E. Hadar, and M. Waschke. ITML: A Domain-Specific Modeling Language for Supporting Business Driven IT Management. In *Proc. of the 9th OOPSLA workshop on DSM*, 2009.

P. Frederiks and T. van der Weide. Information modeling: the process and the required competencies of its participants. *Data & Knowledge Engineering*, 58(1):4–20, 2006.

M. Freund, J.-P. Descles, A. Pascu, and J. Cardot. Typicality, contextual inferences and object determination logic. In *FLAIRS*, volume 4, pages 491–495, 2004.

L. Gabora, E. Rosch, and D. Aerts. Toward an Ecological Theory of Concepts. *Ecological Psychology*, 20(1):84–116, 2008.

A. Gaillard, I. Urdapilleta, O. Houix, and C. Manetta. Effects of task and category membership on representation stability. *Psicológica*, 32(1):31–48, 2011.

Y. Gal and A. Pfeffer. A language for modeling agents' decision making processes in games. In *IFAAMAS'03*, pages 265–272. ACM, 2003.

R. Garland. A Comparison of Three Forms of the Semantic Differential. *Marketing Bulletin*, 1:19–24, 1990.

D. Geeraerts. *Theories of Lexical Semantics.* Oxford University Press, 2010.

A. Gemino and Y. Wand. A framework for empirical evaluation of conceptual modeling techniques. *Requirements Engineering*, 9(4):248–260, 2004.

N. Genon, D. Amyot, and P. Heymans. Analysing the cognitive effectiveness of the UCM visual notation. In F. A. Kraemer and P. Herrmann, editors, *System Analysis and Modeling: About Models*, volume 6598 of *Lecture Notes in Computer Science*, pages 221–240. Springer Berlin Heidelberg, 2011a.

N. Genon, P. Heymans, and D. Amyot. Analysing the cognitive effectiveness of the BPMN 2.0 visual notation. In B. Malloy, S. Staab, and M. Brand, editors, *Software Language Engineering*, volume 6563 of *Lecture Notes in Computer Science*, pages 377–396. Springer Berlin Heidelberg, 2011b.

L. Giordano, V. Gliozzi, N. Olivetti, and G. Pozzato. Reasoning about typicality in preferential description logics. In Hoelldobler et al., editor, *Logics in Artificial Intelligence*, volume 5293 of *LNCS*, pages 192–205. Springer Berlin Heidelberg, 2008.

F. Giunchiglia, J. Mylopoulos, and A. Perini. The TROPOS software development methodology: processes, models and diagrams. In *Agent-Oriented Software Engineering III*, pages 162–173. Springer, 2003.

J. Gordijn and J. Akkermans. Value-based requirements engineering: Exploring innovative e-commerce ideas. *Requirements engineering*, 8(2):114–134, 2003.

J. Gordijn, E. Yu, and B. van der Raadt. e-Service Design Using i* and e3value Modeling. *IEEE Software*, 23:26–33, 2006.

G. Grau, J. Horkoff, E. Yu, and S. Abdulhadi. i* guide 3.0. Internet, August 2007.

N. Green and C. Bate. *Lost in Translation*. Evolved Technologist, 2007.

S. Gregor. The nature of theory in information systems. *MIS Quarterly*, 30(3):611–642, 2006.

T. R. Gruber. A translation approach to portable ontology specifications. *Knowledge Acquisition*, 5:199–220, 1993.

N. Guarino. Formal ontology, conceptual analysis and knowledge representation. *International Journal of Human-Computer Studies*, 43:625 – 640, 1995.

N. Guarino. Formal Ontology and Information Systems. In *Proceedings of Formal Ontology and Information Systems*, pages 3–15. IOS Press, 1998.

N. Guarino, M. Carrara, and P. Giaretta. Formalizing ontological commitments. In *Proceedings of the twelfth national conference on Artificial intelligence (vol. 1)*, AAAI '94, pages 560–567, Menlo Park, CA, USA, 1994. American Association for Artificial Intelligence.

G. Guizzardi and G. Wagner. A unified foundational ontology and some applications of it in business modeling. In *CAiSE Workshops (3)*, pages 129–143, 2004.

G. Guizzardi and G. Wagner. Using the unified foundational ontology (ufo) as a foundation for general conceptual modeling languages. In *Theory and Applications of Ontology: Computer Applications*, pages 175–196. Springer, 2010.

G. Guizzardi, H. Herre, and G. Wagner. On the general ontological foundations of conceptual modeling. In S. Spaccapietra, S. March, and Y. Kambayashi, editors, *Conceptual Modeling (ER) 2002*, volume 2503 of *Lecture Notes in Computer Science*, pages 65–78. Springer Berlin Heidelberg, 2003.

J. Gumperz. *Linguistic anthropology: A reader*, chapter The speech community, pages 66–74. Wiley-Blackwell, 2001.

T. Halpin. ORM 2. In *On the Move to Meaningful Internet Systems 2005: OTM 2005 Workshops*, pages 676–687. Springer, 2005.

J. A. Hampton. Similarity-based categorization and fuzziness of natural categories. *Cognition*, 65(2-3):137 – 165, 1998.

J. A. Hampton, D. Dubois, and W. Yeh. Effects of classification context on categorization in natural categories. *Memory & Cognition*, 34(7):1431–1443, 2006.

S. Harnad. To cognize is to categorize: Cognition is categorization. In C. Lefebvre and H. Cohen, editors, *Handbook of Categorization*. Elsevier, December 2005.

Z. S. Harris. Distributional structure. *Word*, 10:146–162, 1954.

B. Henderson-Sellers. UML - the Good, the Bad or the Ugly? Perspectives from a panel of experts. *Software and System Modeling*, 4(1):4–13, 2005.

B. Henderson-Sellers. Random thoughts on multi-level conceptual modelling. In *The evolution of conceptual modeling*, pages 93–116. Springer, 2011.

P. Heymans, G. Saval, and G. Dallons. A template-based analysis of grl. *Advanced Topics in Database Research, Volume V*, 5:124, 2006.

S. Hofer. Instances over algorithms: A different approach to business process modeling. In P. Johannesson, J. Krogstie, and A. Opdahl, editors, *The Practice of Enterprise Modeling*, volume 92 of *Lecture Notes in Business Information Processing*, pages 25–37. Springer Berlin Heidelberg, 2011.

S. J. B. A. Hoppenbrouwers. *Freezing language : conceptualisation processes across ICT-supported organisations*. PhD thesis, Radboud University Nijmegen, 2003.

S. J. B. A. Hoppenbrouwers, H. A. Proper, and T. P. van der Weide. Formal modelling as a grounded conversation. In G. L. M. Goldkuhl and S. Haraldson, editors, *Proceedings of the 10th International Working Conference on the Language Action Perspective on Communication Modelling (LAP05)*, pages 139–155. Linköping : Linköpings Universitet and Högskolan i Boras, 2005.

S. J. B. A. Hoppenbrouwers, L. Lindeman, and H. A. Proper. Capturing modeling processes–towards the modial modeling laboratory. In *On the Move to Meaningful Internet Systems 2006: OTM 2006 Workshops*, pages 1242–1252. Springer, 2006.

S. J. B. A. Hoppenbrouwers, H. Weigand, and E. Rouwette. Setting rules of play for collaborative modeling. *International Journal of e-Collaboration (IJeC)*, 5(4):37–52, 2009.

T. Horgan and M. Potrč. Blobjectivism and indirect correspondence. *Facta Philosophica*, 2(2):249–270, 2000.

C. Houy, P. Fettke, and P. Loos. Understanding understandability of conceptual models - what are we actually talking about? In Atzeni et al., editor, *Conceptual Modeling*, volume 7532 of *LNCS*, pages 64–77. Springer, 2012.

J. C. Huang, S. Newell, and R. D. Galliers. The impact of organizational sub-cultures on the implementation of component-based development: A case study of an international investment bank. In *European Conference on Information Systems (ECIS 2002)*, pages 219–228, 2002.

IEEE. Systems and software engineering – architecture description. *ISO/IEC/IEEE 42010:2011(E) (Revision of ISO/IEC 42010:2007 and IEEE Std 1471-2000)*, pages 1 –46, 1 2011.

D. Jardine. Concepts and terminology for the conceptual schema and the information base. *Computers and Standards*, 3(1):3 – 17, 1984.

H. Jonkers, R. van Buuren, F. Arbab, F. de Boer, M. Bonsangue, H. Bosma, H. ter Doest, L. Groenewegen, J. G. Scholten, S. J. B. A. Hoppenbrouwers, M.-E. Iacob, W. Janssen, M. M. Lankhorst, D. van Leeuwen, H. A. Proper, A. Stam, L. van der Torre, and G. Veldhuijzen van Zanten. Towards a Language for Coherent Enterprise Architecture Descriptions. In *EDOC '03: Proceedings of the 7th International Conference on Enterprise Distributed Object Computing*, page 28, Washington, DC, USA, 2003. IEEE Computer Society.

J. Kaidalova, U. Seigerroth, T. Kaczmarek, and N. Shilov. Practical challenges of enterprise modeling in the light of business and it alignment. In Sandkuhl et al., editor, *The Practice of Enterprise Modeling*, volume 134 of *LNBIP*, pages 31–45. Springer Berlin Heidelberg, 2012.

H. Kaindl, S. Brinkkemper, J. A. Bubenko Jr, B. Farbey, S. J. Greenspan, C. L. Heitmeyer, J. C. S. d. P. Leitea, N. R. Mead, J. Mylopoulos, and J. Siddiqi. Requirements engineering and technology transfer: Obstacles, incentives and improvement agenda. *Requirements Engineering*, 7(3):113–123, 2002.

C. W. Kalish. Essentialism and graded membership in animal and artifact categories. *Memory & Cognition*, 23(3):335–353, 1995.

A. Katifori, C. Vassilakis, I. Daradimos, G. Lepouras, Y. Ioannidis, A. Dix, A. Poggi, and T. Catarci. Personal ontology creation and visualization for a personal interaction management system. In *Workshop on Personal Information Management CHI2008*, volume 2008, 2008.

V. Katifori, A. Poggi, M. Scannapieco, T. Catarci, and Y. Ioannidis. OntoPIM: how to rely on a personal ontology for Personal Information Management. In *Proceedings of the 1st Workshop on The Semantic Desktop*. Citeseer, 2005.

E. Kindler. On the semantics of EPCs: A framework for resolving the vicious circle. In *Business Process Management*, pages 82–97. Springer, 2004.

J. Krogstie, O. I. Lindland, and G. Sindre. Defining quality aspects for conceptual models. In *Proceedings of the IFIP international working conference on Information system concepts: Towards a consolidation of views*, pages 216–231, London, UK, UK, 1995. Chapman & Hall, Ltd.

H. Kühn, F. Bayer, S. Junginger, and D. Karagiannis. Enterprise Model Integration. In *E-Commerce and Web Technologies*, volume 2738 of *Lecture Notes in Computer Science*, pages 379–392. Springer Berlin Heidelberg, 2003.

M. Lacher and G. Groh. Facilitating the exchange of explicit knowledge through ontology mappings. In *Proceedings of the 14th International FLAIRS conference*, pages 305–309, 2001.

M. M. Lankhorst. Enterprise architecture modelling–the issue of integration. *Advanced Engineering Informatics*, 18(4):205 – 216, 2004.

M. M. Lankhorst, editor. *Agile Service Development: Combining Adaptive Methods and Flexible Solutions*. Springer, 2012.

M. M. Lankhorst, H. A. Proper, and H. Jonkers. The anatomy of the archimate language. *International Journal of Information System Modeling and Design (IJISMD)*, 1(1):1–32, 2010.

S. Laurence and E. Margolis. *Concepts: core readings*, chapter Concepts and cognitive science, pages 3–81. MIT Press Cambridge, MA, 1999.

I.-E. Liao, S.-C. Liao, K.-F. Kao, and I.-F. Harn. A Personal Ontology Model for Library Recommendation System. In S. Sugimoto, J. Hunter, A. Rauber, and A. Morishima, editors, *Digital Libraries: Achievements, Challenges and Opportunities*, volume 4312 of *Lecture Notes in Computer Science*, pages 173–182. Springer Berlin Heidelberg, 2006.

O. I. Lindland, G. Sindre, and A. Solvberg. Understanding quality in conceptual modeling. *Software, IEEE*, 11(2):42–49, 1994.

L. Liu and E. Yu. From requirements to architectural design - using goals and scenarios. In *Software Requirements to Architectures Workshop (STRAW)*. 2001.

L. López, X. Franch, and J. Marco. Making explicit some implicit i* language decisions. In *Conceptual Modeling–ER 2011*, pages 62–77. Springer, 2011.

A. R. Luria and O. S. Vinogradova. An objective investigation of the dynamics of semantic systems. *British Journal of Psychology*, 50(2):89, 1959.

A. R. Luria and O. S. Vinogradova. Семантическаја структура слова *(Semantic structure of the word)*, chapter Обективное исследованны динамики семантических систем (Objective studies of the dynamics of semantic systems), pages 27–63. 1971. (in Russian).

E. Machery. *Doing without concepts*. Oxford University Press Oxford, 2009.

A. Majid, M. van Staden, J. Boster, and M. Bowerman. Event categorization: A cross-linguistic perspective. In K. Forbus, D. Gentner, and T. Tegier, editors, *Proceedings of the 26th Annual Meeting of the Cognitive Science Society*, pages 885–890. Mahwah, NJ : Erlbaum, 2004.

B. C. Malt. Features and beliefs in the mental representation of categories. *Journal of Memory and Language*, 29(3):289 – 315, 1990.

B. C. Malt and E. C. Johnson. Do artifact concepts have cores? *Journal of Memory and Language*, 31(2):195 – 217, 1992.

B. C. Malt, S. Gennari, M. Imai, E. Ameel, N. Tsuda, and A. Majid. Talking about walking. *Psychological Science*, 19(3):232, 2008.

B. C. Malt, E. Ameel, S. Gennari, M. Imai, and A. Majid. Do words reveal concepts? In *Proceedings of the 33rd Annual Conference of the Cognitive Science Society*, pages 519–524, 2011.

K. McRae, G. S. Cree, M. S. Seidenberg, and C. McNorgan. Semantic feature production norms for a large set of living and nonliving things. *Behavior research methods*, 37(4): 547–559, 2005.

M. Mernik, J. Heering, and A. M. Sloane. When and how to develop domain-specific languages. *ACM computing surveys (CSUR)*, 37(4):316–344, 2005.

M. Meyerhoff. *Communities of Practice*, pages 526–548. Blackwell Publishing Ltd, 2008.

D. L. Moody. Theoretical and practical issues in evaluating the quality of conceptual models: current state and future directions. *Data & Knowledge Engineering*, 55(3): 243–276, 2005.

D. L. Moody. The physics of notations: Toward a scientific basis for constructing visual notations in software engineering. *IEEE Transactions on Software Engineering*, 35: 756–779, 2009.

D. L. Moody. The “physics” of notations: A scientific approach to designing visual notations in software engineering. In *Proceedings of the 32nd ACM/IEEE International Conference on Software Engineering*, pages 485–486, New York, NY, USA, 2010. ACM.

D. L. Moody and J. van Hillegersberg. Evaluating the visual syntax of UML: An analysis of the cognitive effectiveness of the uml family of diagrams. *Lecture Notes in Computer Science*, 5452:16–34, 2009.

D. L. Moody, P. Heymans, and R. Matuleviaius. Visual syntax does matter: Improving the cognitive effectiveness of the i* visual notation. *Requirements Engineering*, 15(2): 141–175, 2010.

J. Mylopoulos, A. Borgida, M. Jarke, and M. Koubarakis. Telos: Representing knowledge about information systems. *ACM Trans. Inf. Syst.*, 8(4):325–362, Oct. 1990.

F. Neuhaus, E. Florescu, A. Galton, M. Gruninger, N. Guarino, L. Obrst, A. Sanchez, A. Vizedom, P. Yim, and B. Smith. Creating the ontologists of the future. *Applied Ontology*, 6(1):91–98, Jan. 2011.

H. Niemietz. The impact of organisational subcultures on the enterprise architecture process. In *Proceedings of the International Conference Information Systems 2013*, pages 397–400, 2013.

I. Niles and A. Pease. Towards a standard upper ontology. In *Proceedings of the international conference on Formal Ontology in Information Systems-Volume 2001*, pages 2–9. ACM, 2001.

Object Management Group. Business Process Model and Notation (BPMN) FTF Beta 1 for Version 2.0. Internet, 2010a.

Object Management Group. Business motivation model (bmm), version 1.1. Internet, 2010b.

Object Management Group. Decision model and notation (dmn), version 1.0 - beta 1. Internet, 2014a.

Object Management Group. Documents associated with value delivery metamodel 1.0 - beta 1. Internet, 2014b.

A. L. Opdahl and G. Berio. Interoperable language and model management using the UEML approach. In *Proceedings of the 2006 International Workshop on Global Integrated Model Management*, pages 35–42, New York, NY, USA, 2006.

C. E. Osgood, G. J. Suci, and P. Tannenbaum. *The Measurement of Meaning*. University of Illinois Press, Urbana, 1957.

S. Patig. Modeling deployment of enterprise applications. In Wil Aalst et al., editor, *Information Systems Evolution*, volume 72 of *LNBIP*, pages 253–266. Springer Berlin Heidelberg, 2011.

C. Perelman and L. Olbrechts-Tyteca. *The New Rhetoric: A Treatise on Argumentation*. University of Notre Dame Press, June 1969.

J. P. Peter. Reliability: A review of psychometric basics and recent marketing practices. *Journal of Marketing Research*, 16(1):6–17, 1979.

S. Pinker. *The stuff of thought: Language as a window into human nature*. Viking Press, 2007.

G. Plataniotis, S. de Kinderen, and H. A. Proper. EA Anamnesis: Towards an approach for Enterprise Architecture rationalization. In S. Printing, editor, *Proceedings of the The 12th Workshop on Domain-Specific Modeling (DSM12)*. ACM DL, 2012.

G. Poels, A. Maes, F. Gailly, and R. Paemeleire. Measuring the perceived semantic quality of information models. In J. Akoka, S. Liddle, I.-Y. Song, M. Bertolotto, I. Comyn-Wattiau, W.-J. van den Heuvel, M. Kolp, J. Trujillo, C. Kop, and H. Mayr, editors, *Perspectives in Conceptual Modeling*, volume 3770 of *Lecture Notes in Computer Science*, pages 376–385. Springer Berlin Heidelberg, 2005.

H. A. Proper and D. Greefhorst. The roles of principles in enterprise architecture. In *Trends in Enterprise Architecture Research*, volume 70 of *LNBIP*, pages 57–70. Springer, 2010.

J. Pucher and R. Buehler. Making cycling irresistible: lessons from the Netherlands, Denmark and Germany. *Transport Reviews*, 28(4):495–528, 2008.

D. A. Quartel, W. Engelsman, H. Jonkers, and M. Van Sinderen. A goal-oriented requirements modelling language for enterprise architecture. In *EDOC'09*, pages 3–13. IEEE, 2009.

J. C. Recker. Opportunities and constraints: the current struggle with BPMN. *Business Process Management Journal*, 16(1):181–201, 2010.

J. C. Recker and A. Dreiling. Does it matter which process modelling language we teach or use? An experimental study on understanding process modelling languages without formal education. In *18th Australasian Conference on Information Systems*, pages 356–366, Toowoomba, Australia, 2007.

G. Regev. Systems and Business Process Modeling: are we Confusing the Map with the Territory? Keynote speech presented for the joint BPMDS and EMMSAD working conferences co-located with CAiSE2014, Thessaloniki, Greece, June 2014.

M. Renger, G. L. Kolfschoten, and G.-J. Vreede. Challenges in collaborative modeling: A literature review. In Jan Dietz et al., editor, *Advances in Enterprise Engineering I*, volume 10 of *Lecture Notes in Business Information Processing*, pages 61–77. Springer Berlin Heidelberg, 2008.

P. Rittgen. Collaborative modeling – a design science approach. In *System Sciences, 2009. HICSS'09. 42nd Hawaii International Conference on System Sciences*, pages 1–10. IEEE, 2009.

S. Robinson. Conceptual modeling for simulation: issues and research requirements. In *Proceedings of the 38th conference on Winter simulation*, WSC '06, pages 792–800. Winter Simulation Conference, 2006.

E. Rosch. Natural categories. *Cognitive Psychology*, 4(3):328–350, 1973.

E. Rosch. Reclaiming concepts. *Journal of Consciousness Studies*, 6:61–77(17), 1999.

E. Rosch and C. B. Mervis. Family resemblances: Studies in the internal structure of categories. *Cognitive Psychology*, 7(4):573 – 605, 1975.

E. Rosch, C. B. Mervis, W. D. Gray, D. M. Johnson, and P. Boyes-Braem. Basic objects in natural categories. *Cognitive Psychology*, 8(3):382 – 439, 1976.

M. Rosemann, A. Hjalmarsson, M. Lind, and J. C. Recker. Four facets of a process modeling facilitator. In D. Galletta and T.-P. Liang, editors, *32nd International Conference on Information Systems*, pages 1–16, Shanghai International Convention Center, Shanghai, December 2011. AIS.

M. Rospocher, C. Ghidini, L. Serafini, B. Kump, V. Pammer, S. Lindstaedt, A. Faatz, and T. Ley. Collaborative enterprise integrated modelling. In *5th International Workshop on Semantic Web Applications and Perspectives (SWAP 2008)*, 2008.

B. H. Ross and V. S. Makin. Prototype versus exemplar models in cognition. *The nature of cognition*, pages 205–241, 1999.

E. Rouwette, S. Hoppenbrouwers, and B. Dangerfield. Collaborative systems modeling and group model building: a useful combination. In *26th International Conference of the System Dynamics Society*, 2008.

W. Ruts, S. De Deyne, E. Ameel, W. Vanpaemel, T. Verbeemen, and G. Storms. Dutch norm data for 13 semantic categories and 338 exemplars. *Behavior Research Methods*, 36:506–515, 2004.

A. Santos, S. E. Chaigneau, W. K. Simmons, and L. W. Barsalou. Property generation reflects word association and situated simulation. *Language and Cognition*, 3(1):83–119, 2011.

A.-W. Scheer and M. Nüttgens. ARIS Architecture and Reference Models for Business Process Management. In *Business Process Management*, volume 1806 of *LNCS*, pages 301–304. Springer Berlin Heidelberg, 2000.

J. Scott. Business Capability Maps – The missing link between business strategy and IT action. *Architecture & Governance*, 5(9):1–4, 2009.

K. Siau and X. Tan. Improving the quality of conceptual modeling using cognitive mapping techniques. *Data & Knowledge Engineering*, 55(3):343 – 365, 2005.

E. E. Smith. Theories of semantic memory. *Handbook of learning and cognitive processes*, 6:1–56, 1978.

C. Sonnenberg, C. Huemer, B. Hofreiter, D. Mayrhofer, and A. Braccini. The rea-DSL: a domain specific modeling language for business models. In *CAiSE'11*, pages 252–266. Springer, 2011.

J. Sowa. Ontology, metadata, and semiotics. In B. Ganter and G. Mineau, editors, *Conceptual Structures: Logical, Linguistic, and Computational Issues*, volume 1867 of *Lecture Notes in Computer Science*, pages 55–81. Springer Berlin Heidelberg, 2000.

J. Sowa. The Role of Logic and Ontology in Language and Reasoning. In *Theory and Applications of Ontology: Philosophical Perspectives*, pages 231–263. Springer Berlin Heidelberg, 2010.

D. Ssebuggwawo, S. J. B. A. Hoppenbrouwers, and H. A. Proper. Interactions, goals and rules in a collaborative modelling session. In A. Persson and J. Stirna, editors, *The Practice of Enterprise Modeling*, volume 39 of *Lecture Notes in Business Information Processing*, pages 54–68. Springer Berlin Heidelberg, 2009.

G. Stahl. A model of collaborative knowledge-building. In *Proc. of 4th Int. Conference of the Learning Sciences (ICLS 2000), Ann Arbor, MI*, pages 70–77, 2000.

J. Stirna, A. Persson, and K. Sandkuhl. Participative enterprise modeling: Experiences and recommendations. In Krogstie et al., editor, *Advanced Information Systems Engineering*, volume 4495 of *LNCS*, pages 546–560. Springer Berlin Heidelberg, 2007.

G. Storms, P. D. Boeck, and W. Ruts. Prototype and exemplar-based information in natural language categories. *Journal of Memory and Language*, 42(1):51 – 73, 2000.

F. Tan and M. Hunter. The repertory grid technique: A method for the study of cognition in information systems. *MIS Quarterly*, pages 39–57, 2002.

The Open Group. *ArchiMate 1.0 Specification.* Van Haren Publishing, 2009.

The Open Group. *ArchiMate 2.0 Specification.* Van Haren Publishing, 2012.

M. Uschold. Making the case for ontology. *Applied Ontology*, 6(4), 2011.

W. M. van der Aalst. Formalization and verification of event-driven process chains. *Information and Software technology*, 41(10):639–650, 1999.

T. Verhagen and S. Meents. A framework for developing semantic differentials in is research: Assessing the meaning of electronic marketplace quality (EMQ). Serie Research Memoranda 0016, VU University Amsterdam, Faculty of Economics, Business Administration and Econometrics, 2007.

F. B. Vernadat. UEML: towards a unified enterprise modelling language. *International Journal of Production Research*, 40:4309–4321, 2002a.

F. B. Vernadat. Enterprise modeling and integration (EMI): Current status and research perspectives. *Annual Reviews in Control*, 26(1):15 – 25, 2002b.

J. von Neumann. *Theory Of Games And Economic Behavior*. Princeton University Press, 1944.

Y. Wand and R. Weber. Mario bunge's ontology as a formal foundation for information systems concepts. *Studies on Mario Bunge's Treatise, Rodopi, Atlanta*, pages 123–149, 1990.

E. Wenger and W. Snyder. Communities of practice: The organizational frontier. *Harvard business review*, 78(1):139–146, 2000.

R. J. Wieringa. Conceptual modeling in social and physical contexts, April 2008. URL `http://doc.utwente.nl/64804/`.

R. J. Wieringa. Real-world semantics of conceptual models. In R. Kaschek and L. Delcambre, editors, *The Evolution of Conceptual Modeling*, volume 6520 of *Lecture Notes in Computer Science*, pages 1–20. Springer Berlin Heidelberg, 2011.

R. J. Wieringa and J. M. G. Heerkens. The methodological soundness of requirements engineering papers: a conceptual framework and two case studies. *Requirements Engineering*, 11(4):295–307, 2006.

I. Wilmont, S. Brinkkemper, I. van de Weerd, and S. J. B. A. Hoppenbrouwers. Exploring intuitive modelling behaviour. In I. Bider, T. Halpin, J. Krogstie, S. Nurcan, H. A. Proper, R. Schmidt, and R. Ukor, editors, *Enterprise, Business-Process and Information Systems Modeling*, volume 50 of *Lecture Notes in Business Information Processing*, pages 301–313. Springer Berlin Heidelberg, 2010.

I. Wilmont, E. Barendsen, S. J. B. A. Hoppenbrouwers, and S. Hengeveld. Abstract reasoning in collaborative modeling. In R. Sprague, editor, *Proceedings of the 45th Annual Hawaii International Conference on System Sciences*, pages 170–179. Los Alamos : IEEE Computer Society, 2012.

I. Wilmont, S. Hengeveld, E. Barendsen, and S. J. B. A. Hoppenbrouwers. Cognitive mechanisms of conceptual modelling. In W. Ng, V. Storey, and J. Trujillo, editors, *Conceptual Modeling*, volume 8217 of *Lecture Notes in Computer Science*, pages 74–87. Springer Berlin Heidelberg, 2013.

P. Wolff and K. J. Holmes. Linguistic relativity. *Wiley Interdisciplinary Reviews: Cognitive Science*, 2(3):253–265, 2011.

L. Wu and L. W. Barsalou. Perceptual simulation in conceptual combination: Evidence from property generation. *Acta Psychologica*, 132(2):173 – 189, 2009. Spatial working memory and imagery: From eye movements to grounded cognition.

B. Wyssusek and H. Klaus. On the foundation of the ontological foundation of conceptual modeling grammars: the construction of the bunge-wand-weber ontology. In J. Castro and E. Teniente, editors, *Proceedings of the CAiSE '05 Workshops*, volume 2, pages 583–593, 2005.

H. Yang and J. Callan. Learning the distance metric in a personal ontology. In *Proceeding of the 2nd International Workshop on Ontologies and Information Systems for the Semantic Web*, ONISW '08, pages 17–24, New York, NY, USA, 2008. ACM.

C. A. Yeung and H.-F. Leung. Ontology with likeliness and typicality of objects in concepts. In D. W. Embley, A. Olive, and S. Ram, editors, *Conceptual Modeling - ER 2006*, volume 4215 of *Lecture Notes in Computer Science*, pages 98–111. Springer Berlin Heidelberg, 2006.

C. A. Yeung and H.-F. Leung. A formal model of ontology for handling fuzzy membership and typicality of instances. *Comput. J.*, 53(3):316–341, 2010.

Summary

The focus of this thesis is the personal semantics people have of concepts used to model: *meta-concepts*. That means, how they typically view those concepts, how they categorize them, what their semantic structure is, and so on. Our reason for investigating this, and the objective we hope to achieve, is to make practitioners and academics more aware that there can be significant differences in the way people interpret meta-concepts. It is not our intention to propose yet another method or way of doing things to deal with these differences, or create yet another series of pictures and models, but instead focus on the knowledge questions that underlie them. In doing so we hope to stimulate other researchers in the field of conceptual modeling to perform more empirical work, grounding their designs and statements in real world observations, and for practitioners to be able to use such insights in their daily practice.

We first introduce the issues to do with personal semantics in our specific research context: Enterprise Modeling (EM). The goal of EM is to create models that correctly and coherently describe the structure and function of all interconnected aspects of an enterprise, such as, for example, its processes, design decision making, deployment and management of IT artifacts, goals & motivations, value exchanges. These models can then be used to reason about the enterprise's structure and behavior, and when and where relevant, propose changes to it. As the models describing such different aspects are typically written in different modeling languages, often by different people, personal differences in the way people interpret not only the concepts from the domain, but also the concepts used for modeling themselves can lead to a host of problems. To elaborate on this we go into more detail on the conceptual modeling done for such aspects, what concepts are from a cognitive point of view, and how personal differences affect them. We conclude by discussing

the research objectives, the questions we aim to answer, and the contributions we hope to make.

As we perform several empirical studies to investigate personal semantics, in Chapter 2 we will go into the way we will perform these studies and how we use them to answer the research questions. We then give an overview of the methods and how we have adapted them to be relevant to the context of (conceptual) modeling, and where available show an example of experimental implementation. Amongst the methods we have used are semantic differentials, which we have implemented according to existing quality criteria for use in Information Systems research, term categorization experiments adopted from psycholinguistic studies for several features such as typicality and centrality, and feature listing studies.

Because it is important to know *what* to investigate the personal semantics of, we will use Chapter 3 to discuss the framing of our empirical work. This pertains to the meta-concepts we investigate (e.g., ACTORS, GOALS, RESOURCES), on what dimensions we investigate them (e.g., their physical status, composition, intentionality), and what elements of concept structure are relevant to look at. We give an overview of the analysis we have performed on a number of (specifications of) conceptual modeling languages and notations that are used to represent, or reason about different aspects of enterprises. Resulting from this is a high level categorization of their core terms which we use throughout the course of our research. On the basis of this we also deduced a number of discriminatory dimensions on which terms commonly differ. We elaborate on the materials used for the semantic differentials, that is, the terms we piloted, selected and ended up using to describe the found dimensions by.

On basis of the selected meta-concepts, we discuss the results of several fundamental studies in Chapters 4 through 6. Combined, they present an overview of the personal semantics that people *have* of the meta-concepts and give us more insight into their structure. The main findings discussed are that many of the meta-concepts are likely of a graded nature. That is, some things are considered to be better members than others, which can have an effect on the semantics of models and their derivatives. Given that these meta-concepts are judged in a graded way, instantiations of such models likely should be too, leading to different degrees of conceptual validity for different instantiations of the same model. Furthermore, there are many kinds of features that people attribute to these meta-concepts, as well as these features being of different kinds (e.g., alethic, deontic, temporal, mereological). Some features re-occur quite often, and are thus likely to be shared between many people. However, many features unique to individuals can be found as well. Such features can be used to figure out how people identify meta-concepts, and also what aspects are important to them. Finally, the conceptual understanding that people have of these meta-concepts can be differentiated on different semantic dimensions. Measuring them in such a way shows that there is often a strong variety in these kinds of understandings. Some understandings are shared between many people such as most meta-concepts being considered to have to be non-vague, but people are, like the earlier findings from the feature listing study corroborates, unique individuals that cannot be easily generalized. Furthermore, these understandings change over time as well, although it is difficult to find systematic patterns in how the cores of such personal semantics changes, or what affects them.

Using the findings from the prior studies we investigate to what degree existing modeling languages accommodate conceptual distinctions. That is, whether they allow for the modeling of specific interpretations of a meta-concept (for example, a *human* actor), and whether they do so explicitly by providing clear distinct notational elements, or implicitly by assuming a particular interpretation in the specification of the language. In Chapter 7 we discuss in more detail whether modeling languages should optimally be capable of doing so for all possible conceptual distinctions, and what kind of common distinctions are lacking support, as there are still several potentially relevant conceptual distinctions that are either left unaccommodated or left implicit (e.g., explicit encoding of the necessity of goals, regardless of their satisfaction criteria). Furthermore, we argue that we should keep up to date with the conceptual needs of modelers and stakeholders in practice in order to ensure that our modeling languages (and notations) are supportive of their needs.

Afterwards, we turn our attention towards discovering groups of people that share personal semantics through analysis of their conceptual understandings. Apart from merely finding such groups, it would be useful to be able to predict what kind of personal semantics people have based on some real-world properties, such as what kind of background they have, and what languages and techniques they are experienced in. In our experience many modeling efforts assume such a link between real-world properties and the personal semantics of their modelers, which we put to the test in Chapter 8. Based on data from semantic differential studies we have used clustering methods to find such groups, and attempted to predict a conceptual modeler's personal semantics based on real-world properties. However, in our studies we were unable to find support for this, casting doubt on whether this assumption from practice that people with similar real-world properties have similar personal semantics is valid.

Samenvatting

De focus van dit proefschrift is de persoonlijke semantiek die mensen hebben van concepten die ze gebruiken om te modelleren: *meta-concepten*. Daarmee bedoelen we, hoe ze deze concepten zien, hoe ze ze categoriseren, wat hun semantische structuur is, enzovoort. Onze reden voor dit onderzoek, en wat ermee hopen te bereiken, is om andere onderzoekers en de praktijk meer bewust te maken van het feit dat er significante verschillen kunnen zijn in de manier waarop mensen meta-concepten interpreteren. Het is niet onze bedoeling om een zoveelste methode voor te stellen, of enkel een lading plaatjes en modellen te maken, maar om onszelf te richten op de onderliggende kennisvragen bij deze kwestie. Hiermee hopen we andere onderzoekers in het gebied van conceptueel modelleren te stimuleren om ook meer empirisch werk uit te voeren, hun ontwerpen en uitspraken in observaties van de echte wereld te aarden, en dat mensen uit de praktijk zulke inzichten kunnen gebruiken in hun dagelijks werk.

We introduceren eerst enkele kwesties gerelateerd aan persoonlijke semantiek in onze specifieke onderzoekscontext: Enterprise Modellering (EM). Het doel van EM is om modellen te maken die een correct en samenhangend overzicht geven van de structuur en de functie van alle verschillende aspecten van een Enterprise (bijvoorbeeld zijn processen, ontwerpbesluitsvorming, inzet en management van IT artefacten, doelen & motivaties, waardeuitwisselingen). Deze modellen kunnen vervolgens gebruikt worden om te redeneren over de structuur en het gedrag van de Enterprise, en waar relevant, om suggesties te geven voor aanpassingen aan die structuur. Omdat de modellen van de verschillende aspecten vaak in verschillende modelleertalen geschreven zijn, vaak ook nog door verschillende mensen, kunnen persoonlijke verschillen in hoe mensen de domein- en modelleerconcepten interpreteren tot een lading aan problemen leiden. We gaan hierom

meer in op het conceptuele modellering wat gedaan wordt voor zulke aspecten, wat de 'concepten' die we modelleren zijn van een cognitief gezichtspunt, en hoe persoonlijke verschillen die concepten beïnvloeden. We concluderen door onze onderzoeksobjectieven, onderzoeksvragen, en de hopelijke bijdragen te beschrijven.

Omdat we meerdere empirische studies uitvoeren om persoonlijke semantiek te onderzoeken, gaan we in Hoofdstuk 2 in op de manier hoe we deze studies uitvoeren, en hoe we ze gebruiken om onze onderzoeksvragen te beantwoorden. Vervolgens geven we een overzicht van de methoden en hoe we ze aangepast hebben om relevant te zijn in onze context van (conceptueel) modelleren. Waar mogelijk geven we ook (een voorbeeld van) een experimentele implementatie. De methoden die we gebruikt hebben zijn onder andere semantische differentialen (geïmplementeerd aan de hand van bestaande kwaliteitscriteria voor het gebruik in Informatie Systemen onderzoek), term categorisatie experimenten (voor meerdere eigenschappen zoals bijvoorbeeld typicaliteit en centraliteit), en eigenschapselicitatie studies.

Omdat het belangrijk is om te weten *waarvan* we de persoonlijke semantiek onderzoeken, bediscussïeren we in Hoofdstuk 3 de inkadering van ons empirisch werk. Dit heeft betrekking op de meta-concepten die we zullen bestuderen (bijvoorbeeld UITVOERDERS, DOELEN, BRONNEN), op de dimensies waar we ze op onderzoeken (bijvoorbeeld hun fysieke status, compositie, intentionaliteit), en welke elementen van conceptstructuur relevant zijn om naar te kijken. We geven een overzicht van de analyse die we uitgevoerd hebben over een aantal (specificaties van) modelleertalen en notaties die gebruikt worden voor verschillende aspecten van Enterprises. Het resultaat hiervan is een categorisatie van de belangrijkste termen die we voor de rest van ons onderzoek gebruiken. Op basis van deze categorisatie hebben we ook een aantal onderscheidende dimensies afgeleid waarop deze termen onderscheiden kunnen worden.

Op basis van de geselecteerde meta-concepten bediscussïeren we de resultaten van meerdere fundamentele studies in Hoofdstuk 4 tot en met 6. Samengenomen geven deze hoofdstukken een overzicht van de persoonlijke semantiek die mensen *hebben* van de meta-concepten en geven ze ons meer inzicht in hun structuur. De voornaamste vondsten zijn dat veel van de meta-concepten van een gegradeerde aard zijn (daarmee bedoelen we dat sommige dingen betere voorbeelden van een meta-concept zijn dan anderen), wat een effect can hebben op de semantiek van modellen en hun derivaten. Hierdoor moeten instantiaties van modellen waarschijnlijk ook op een gegradeerde manier bekeken worden wat betreft hun validiteit, waardoor er verschillende niveaus van conceptuele validiteit ontstaan voor meerdere instantiaties van hetzelfde model. Bovendien zijn er vele soorten eigenschappen die mensen toeschrijven aan de meta-concepten, en zijn deze eigenschappen van meerdere soorten (bijvoorbeeld aletisch, deontisch, tijdelijk en mereologisch).

Sommige eigenschappen komen vaak voor, en worden dus waarschijnlijk gedeeld door veel mensen. Echter, veel eigenschappen zijn ook uniek voor bepaalde individuen. Zulke eigenschappen kunnen gebruikt worden om erachter te komen hoe mensen meta-concepten identificeren, en welke aspecten belangrijk voor ze zijn. Ten slotte kan de manier waarop mensen de meta-concepten zien gedifferentïeerd worden op meerdere semantische dimensies. Door ze zodanig te meten blijkt dat er vaak een sterke variatie is in de manier hoe mensen deze meta-concepten zien. Sommige van deze zienswijzen zijn gedeeld door meerdere mensen (bijvoorbeeld dat veel meta-concepten gezien worden als dingen die niet vaag moeten zijn), maar mensen blijken over het algemeen toch unieke individueën te zijn die niet makkelijk gegeneraliseerd kunnen worden. De manier waarop mensen deze meta-concepten zien verandert ook over tijd, alhoewel het moeilijk is om hier systematische patronen in te vinden, of te achterhalen wat deze veranderingen beïnvloedt.

Met behulp van de bevindingen uit de voorgaande studies onderzoeken we in welke mate bestaande modelleringstalen conceptuele onderscheidingen ondersteunen. Dat wil zeggen of ze het modelleren van specifieke interpretaties van een meta-concept (bijvoorbeeld een *menselijke* uitvoerder) ondersteunen, en of ze dit expliciet doen door middel van specifieke elementen in hun notatie, of impliciet door een bepaalde interpretatie in hun specificatie aan te nemen. In Hoofdstuk 7 bespreken we dit in meer detail, door in te gaan op de vraag of modelleertalen alle mogelijke conceptuele onderscheidingen zouden moeten ondersteunen, wat van voorkomende onderscheidingen vaak niet ondersteund worden. Verder pleiten we dat we constant moeten blijven onderzoeken wat de conceptuele benodigdheden zijn van modelleurs en hun belanghebbenden uit de praktijk om ervoor te zorgen dat onze modelleertalen (en notaties) deze benodigdheden ondersteunen.

Hierna richten we onze aandacht op het ontdekken van groepen mensen die persoonlijke semantiek delen door middel van analyse van hun zienswijzen. Behalve zulke groepen te vinden, zou het handig zijn om te kunnen voorspellen wat van persoonlijke semantiek mensen hebben, bijvoorbeeld op basis van eigenschappen uit de echte wereld zoals hun achtergrond, kennis van talen, enzovoort. In onze ervaring wordt dit in de praktijk vaak aangenomen dat er zo een dergelijke link is tussen deze eigenschappen en persoonlijke semantiek. In Hoofdstuk 8 stellen we deze aanname op de proef. Gebaseerd op gegevens van de semantische differentiaal studies hebben we clustermethodens gebruikt om zulke groepen te vinden, en vervolgens geprobeerd om de persoonlijke semantiek van een modelleur te voorspellen gebaseerd op eigenschappen uit de echte wereld. Echter, in onze studies konden we hier geen steun voor vinden, wat een twijfel oproept over deze aanname in de praktijk.

Curriculum Vitae

Dirk van der Linden is a PhD candidate in Information Science at the Radboud University Nijmegen, working with prof. Henderik Proper at the CRP Henri Tudor in Luxembourg. His doctoral work is concerned with the personal understanding people have of concepts used by modeling languages. Before starting his doctoral research he worked as an educational developer at the Radboud University Nijmegen, developing and implementing IT-supported education methods and tools such as simulator software, collaborative work/learning environments and multi-media infrastructures. He holds an MSc (cum laude) in Information Science from Radboud University Nijmegen, and a BASc. in Bioinformatics from HAN University of Applied Sciences. His research is supported by the Fonds Nationale de Recherche Luxembourg. His home on the web can be found at `www.dirkvanderlinden.eu`.

EE Team Dissertation Series

2015-1 D.J.T. van der Linden, *Personal semantics of meta-concepts in conceptual modeling languages*, Radboud University Nijmegen, Nijmegen, the Netherlands, February 13, 2014

2014-2 C. Feltus, *Aligning Access Rights to Governance Needs with the Responsibility MetaModel (ReMMo) in the Frame of Enterprise Architecture*, University of Namur, Namur, Belgium, March 11, 2014

2014-1 F. Tulinayo, *Combining System Dynamics with a Domain Modeling Method, 2014, Radboud University Nijmegen*, Nijmegen, the Netherlands, January 27, 2014

2013-2 C. Brandt, *An Enterprise Modeling Framework for Banks using Algebraic Graph Transformation*, Technische Universität Berlin, Berlin, Germany, December 20, 2013

2013-1 R. Wagter, *Enterprise Coherence*, Radboud University Nijmegen, Nijmegen, the Netherlands, November 19, 2013

2012-2 A. Nakakawa, *A Collaborative Process for Enterprise Architecture Creation*, Radboud University Nijmegen, Nijmegen, the Netherlands, November 21, 2012

2012-1 D. Ssebuggwawo, *Analysis and Evaluation of Modelling Processes*, Radboud University Nijmegen, Nijmegen, the Netherlands, November 21, 2012

2009-2 S. Overbeek, *Bridging Supply and Demand for Knowledge Intensive Tasks*, Radboud University Nijmegen, Nijmegen, the Netherlands, April 24, 2009

2009-1 J. Nabukenya, *Improving the Quality of Organizational Policy Making using Collaboration Engineering*, Radboud University Nijmegen, Nijmegen, the Netherlands, March 3, 2009

2006-1 B. van Gils, *Aptness on the Web*, 2006, Radboud University Nijmegen, the Netherlands, March 3, 2006

2004-1 S.J.B.A. Hoppenbrouwers, *Freezing Language – Conceptualisation Processes across ICT-Supported Organizations*, Radboud University Nijmegen, Nijmegen, the Netherlands, December 10, 2004

Part V

Appendices

APPENDIX A

Categorization Data

The categorization data is presented in the following tables. They contain, respectively, the number of terms that are full, partial or non-members, followed by the proportion of full, partial and non-members.

A.1 Actor

Table A.1: Categorization data for actor

	full	partial	non	p.full	p.partial	p.non
unit	17	13	10	0.43	0.33	0.25
requirement unit	10	9	21	0.25	0.23	0.53
actor	38	1	1	0.95	0.03	0.03
role	15	4	21	0.38	0.10	0.53
collaboration	1	5	34	0.03	0.13	0.85
player	36	4	-	0.90	0.10	-
infrastructure component	10	12	18	0.25	0.30	0.45
organizational component	14	11	15	0.35	0.28	0.38
device	17	10	13	0.43	0.25	0.33
application software	16	10	14	0.40	0.25	0.35
organizational unit	22	9	9	0.55	0.23	0.23
position	6	5	29	0.15	0.13	0.73

perspective	1	3	36	0.03	0.08	0.90
market segment	6	8	26	0.15	0.20	0.65
hardware role	12	9	19	0.30	0.23	0.48
software role	11	6	23	0.28	0.15	0.58
hardware	13	12	15	0.33	0.30	0.38
software	15	12	13	0.38	0.30	0.33
organizational role	14	6	20	0.35	0.15	0.50
environment agent	28	9	3	0.70	0.23	0.08
software agent	28	11	1	0.70	0.28	0.03

A.2 Event

Table A.2: Categorization data for event

	full	partial	non	p.full	p.partial	p.non
event	40	-	-	1.00	-	-
behavior	11	8	21	0.28	0.20	0.53
function	7	7	26	0.18	0.18	0.65
interaction	24	6	10	0.60	0.15	0.25
activity	22	7	11	0.55	0.18	0.28
task	12	8	20	0.30	0.20	0.50
business rule	1	-	39	0.03	-	0.98
service task	11	10	19	0.28	0.25	0.48
transaction	24	9	7	0.60	0.23	0.18
start event	36	3	1	0.90	0.08	0.03
intermediate event	34	4	2	0.85	0.10	0.05
value activity	10	11	19	0.25	0.28	0.48
value interface	4	4	32	0.10	0.10	0.80
value offering	10	5	25	0.25	0.13	0.63
connection	21	5	14	0.53	0.13	0.35
move	32	3	5	0.80	0.08	0.13
contribution	11	8	21	0.28	0.20	0.53
correlation	4	8	28	0.10	0.20	0.70
dependency	1	4	35	0.03	0.10	0.88
means-ends	5	3	32	0.13	0.08	0.80
control	12	8	20	0.30	0.20	0.50
goal refinement	12	4	24	0.30	0.10	0.60

monitor	8	7	25	0.20	0.18	0.63
operation	20	7	13	0.50	0.18	0.33
operationalisation	15	6	19	0.38	0.15	0.48
performance	2	4	34	0.05	0.10	0.85
operation	20	9	11	0.50	0.23	0.28
initiatives	15	8	17	0.38	0.20	0.43

A.3 Goal

Table A.3: Categorization data for goal

	full	partial	non	p.full	p.partial	p.non
goal	40	-	-	1.00	-	-
hard-goal	33	5	2	0.83	0.13	0.05
soft-goal	32	8	-	0.80	0.20	-
business goal	39	1	-	0.98	0.03	-
achieve goal	30	6	4	0.75	0.15	0.10
assignment	8	10	22	0.20	0.25	0.55
avoid goal	23	5	12	0.58	0.13	0.30
cease goal	30	5	5	0.75	0.13	0.13
expectation	6	16	18	0.15	0.40	0.45
maintain goal	30	3	7	0.75	0.08	0.18
requirement	14	12	14	0.35	0.30	0.35
consumer needs	20	10	10	0.50	0.25	0.25
belief	6	8	26	0.15	0.20	0.65
value	11	5	24	0.28	0.13	0.60
target	30	6	4	0.75	0.15	0.10

A.4 Process

Table A.4: Categorization data for process

	full	partial	non	p.full	p.partial	p.non
organizational service	23	11	6	0.58	0.28	0.15
infrastructure service	22	9	9	0.55	0.23	0.23
information service	22	10	8	0.55	0.25	0.20

other service	20	11	9	0.50	0.28	0.23
service	21	13	6	0.53	0.33	0.15
IT service	21	11	8	0.53	0.28	0.20
process	40	-	-	1.00	-	-
sub flow	23	9	8	0.58	0.23	0.20
business flow	27	6	7	0.68	0.15	0.18
process flow	23	9	8	0.58	0.23	0.20
business process	38	1	1	0.95	0.03	0.03
dependency path	7	8	25	0.18	0.20	0.63
game	22	9	9	0.55	0.23	0.23
task	17	10	13	0.43	0.25	0.33

A.5 Resource

Table A.5: Categorization data for resource

	full	partial	non	p.full	p.partial	p.non
artifact	18	10	12	0.45	0.25	0.30
location	14	9	17	0.35	0.23	0.43
hardware	32	4	4	0.80	0.10	0.10
cpu	30	5	5	0.75	0.13	0.13
hd	19	6	15	0.48	0.15	0.38
memory	34	4	2	0.85	0.10	0.05
software	31	6	3	0.78	0.15	0.08
value	16	8	16	0.40	0.20	0.40
data object	25	10	5	0.63	0.25	0.13
business object	17	13	10	0.43	0.33	0.25
object	24	12	4	0.60	0.30	0.10
node	10	9	21	0.25	0.23	0.53
network	29	6	5	0.73	0.15	0.13
network device	28	8	4	0.70	0.20	0.10
representation	6	7	27	0.15	0.18	0.68
meaning	4	5	31	0.10	0.13	0.78
device	31	6	3	0.78	0.15	0.08
computer hardware	33	5	2	0.83	0.13	0.05
machine resource	34	3	3	0.85	0.08	0.08
environmental data	26	4	10	0.65	0.10	0.25

data input	21	9	10	0.53	0.23	0.25
input	21	8	11	0.53	0.20	0.28
value object	19	10	11	0.48	0.25	0.28
information	32	5	3	0.80	0.13	0.08
resource	39	1	-	0.98	0.03	0.00
content	21	8	11	0.53	0.20	0.28
value port	10	8	22	0.25	0.20	0.55

A.6 Restriction

Table A.6: Categorization data for restriction

	full	partial	non	p.full	p.partial	p.non
interaction	9	7	24	0.23	0.18	0.60
contract	28	6	6	0.70	0.15	0.15
interface	11	10	19	0.28	0.25	0.48
message	7	3	30	0.18	0.08	0.75
catching	7	10	23	0.18	0.25	0.58
throwing	7	10	23	0.18	0.25	0.58
boundary	31	4	5	0.78	0.10	0.13
rule	30	5	5	0.75	0.13	0.13
decomposition	8	8	24	0.20	0.20	0.60
belief	15	9	16	0.38	0.23	0.40
priority	26	8	6	0.65	0.20	0.15
license	26	9	5	0.65	0.23	0.13
boundary condition	34	2	4	0.85	0.05	0.10
conflict	15	10	15	0.38	0.25	0.38
domain property	24	6	10	0.60	0.15	0.25
permission	27	4	9	0.68	0.10	0.23
value	9	5	26	0.23	0.13	0.65
policy	31	6	3	0.78	0.15	0.08
trust	19	8	13	0.48	0.20	0.33
non-interrupting	16	8	16	0.40	0.20	0.40
interrupting	18	8	14	0.45	0.20	0.35
strategy	16	12	12	0.40	0.30	0.30
measure	9	9	22	0.23	0.23	0.55
strategic objective	16	14	10	0.40	0.35	0.25

A.7 Result

Table A.7: Categorization data for result

	full	partial	non	p.full	p.partial	p.non
product	33	5	2	0.83	0.13	0.05
human output	32	5	3	0.80	0.13	0.08
material output	33	3	4	0.83	0.08	0.10
service output	35	3	2	0.88	0.08	0.05
data output	35	5	-	0.88	0.13	0.00
outcome	37	-	3	0.93	0.00	0.08
end event	16	10	14	0.40	0.25	0.35
payoff	24	11	5	0.60	0.28	0.13

APPENDIX B

Feature Listing Data

B.1 Actor

has a name, has permissions, has a role, is a role, has skills, has capabilities, can perform an action, performs an action, stops an action, relaunches an action, schedules an action, cancels an action, belongs to a group, participates in a process, has a responsibility, can execute a task, can invoke a service, is assigned to a role, performs a task, performs an activity, has an assigned goal, uses resources, is responsible for tasks, initiates actions, is a responsible agent, can execute actions, can monitor the execution of actions, is capable of proactive action, is socially capable, is capable of assuming different roles, can perform actions, can receive information, can sense, is proactive, is deliberative, is an automatic thing, has responsibilities, can perform tasks, has access to a resource, can control a resource, can contribute to the fulfillment of goals, is composed of parts, performs actions, is human, is technical, can react to events, has to work under constraints, has rights, has access to resources, can fulfill multiple roles, carries knowledge, carries culture, is an active entity, is physical, displays behavior, can be human, can be a machine, carries out work, can carry responsibility, can receive responsibility, has an identity, has wishes, has desires, has competence, can act autonomously, executes work, requests work, is independent, uses services, uses active structures, is a person, is an organization, is responsible for a business function, is responsible for a business process, can trigger actions, administrates vision and principles, realizes vision and principles, realizes a goal, carries out actions, is non-human, is responsible, is accountable, is autonomous, is identifiable, is a computer, is a machine, is part of a business function, receives triggers, receives service requests, delivers a desired or agreed upon service, participates in a collaboration, is a company, is a person with a role in a company, has a concern, is a stakeholder, has a contractual relation, has relation to other actors, is a thing, does things, fulfills a role, has properties, acts in a context, is someone active in a context, is responsible in a context, has influence on a system, is reactive, can act, ensures the achievement of a result, is active, reacts to actions, makes decisions, always has an input, always has an output, has a quality level, has a role of a person, has a role of a machine, acts, can be a department, can be an organization, can be a user, can be a system, can be a consumer of services, can offer services, has tasks, is absolute, has an intention, has an objective, has a goal, relates to a use case, relates to a scenario, relates to a function

B.2 Event

has a trigger, can be caught by something, has a related objective, is related to an actor, has a description, started on some time, started by some thing, has a precondition, has a postcondition, has a duration, has a cost, is recursive, has a cause, has consequences, has a beginning, has an end, is triggered by something, can be temporal, has a name, can trigger an action, defines a transition from some state to some state, occurs in a process, has to be monitored, triggers a task, triggers a process, is atomic, is finite, is short, can cause a change, can be recognized by an actor, is unique, is identifiable, is instantaneous, is an occurrence, has some phenomena, is composed, is a change of state, occurs during an interval in time, is limited in time, can be planned, can be spontaneous, happens, triggers a business process, triggers a function, triggers a transformation, involves at least one subject, is executed by at least one subject, can be observed, results from a change, does not have a duration, has conditions, impacts other things, has a source, is a process end, is measurable, has relevance to business, triggers something, is the change of the state of an object, is related to a process, is related to a result, can trigger actions, is a fact, does not happen, is repeatable, can trigger a reaction, is predictable, occurs in the real world, can be anticipated upon, is a trigger, results in a case, is generated by a source, may be notified to listeners, is reported by an actor, report of an event is received by an actor, has a type, has a frequency, occurs at a specific time, triggers an action, has properties, influences a process flow, can disrupt a process, is caused by something, causes something, is unpredictable, can trigger a process, can occur spontaneously, has a timespan, triggers change in a situation, is spontaneous, can be caused by an action, is immutable, has an external cause, has an internal cause, occurs in the business layer, can trigger an activity, can be triggered, has a date, has a time

B.3 Goal

has a time, has a result, has a related objective, is related to an actor, has a description, has a precondition, has a postcondition, has a cost, is defined by someone, has a name, has a type, has a relationship with a subgoal, has a relationship with an actor, is based on a need, is based on an objective, has to be reached during a process, is subjective, is relevant to someone, may be quantifiable, is agreed upon, is formulated according to the S.M.A.R.T. standard, is a desired state, is not currently the case, is an intention of an actor, has properties, has to be fulfilled, can be refined into a subgoal, can conflict with a goal, can be satisfied, has a property, is desired by an actor, is an objective, is clear, has a limited scope, can be reached, is challenging, sensibility of which is attested by human ACTORS, can be supported by a person, is temporally valid, is a guideline, can be described quantitative, can be described qualitative, is formulated to be achieved in the middle-long-term, is delimited, has a related stakeholder, describes a desired outcome, implies change, is the S.M.A.R.T. formulation of an aim, has a related actor, can be achieved, has a condition, has a purpose, gives direction, is thoughtful, is measurable, is owned by a stakeholder, should address a driver, becomes an evaluation criteria for future accomplishments, can be abstract, can be concrete, is future oriented, describes a desired state, describes a desired change, has value for an actor, can be quantitative, can be qualitative, has a source, is a desired future state, supports an organization, is formulated to support the strategy, fits within a broader framework, represents the directions given by something, is realizable within some time, is desired, is a future state, can be formulated according to the S.M.A.R.T. standard, has a timespan, has a KPI, is on some governance level, is tangible, is the desired result of an action, has to be achieved, is a reason for the existence of something, is a desired result, gives direction to an actor, describes a situation, is in the future, can be achieved in multiple ways, is planned, requires action to be achieved, makes strategy concrete, requires controlled actions to be achieved, is a state, has implications, can have subgoals, can have a strategy, can be broken down into parts

B.4 Process

can be an instance of a method, is composite, groups tasks, has an endpoint, is executed atomically, is created by a modeler, has an input, has an output, is defined through its components, is executable, is pauseable, is stoppable, is rewindable, is formally defined, is written in a modeling language, is bounded, is executed for a specific purpose, is a succession of actions, can be a chemical reaction, can be a corporate process, has steps, has a timeline, has related actors, has input, has output, is a set of related actions, reaches a goal, may involve roles, uses resources, has operating conditions, has a trigger event, may trigger another process, may depend on another process, can be owned, can be measured, has a trigger, is intangible, is composed of some tasks, has a sequence, has concurrency, is abstract, is detailed, has a start, has an end, has steps which are connected, can be conscious, can be passive, is specific to a problem, is specific to a solution, is not generic, is specific

B.5 Resource

has features, has characteristics, can be available, needs permission for use, is used by an actor, has a related objective, has a description, has an owner, has a value, is composed, may be decomposed, is consumable, has an expiration date, satisfies a precondition, is material, is immaterial, is a person, is finite, is countable, has a name, has a label, has a type, can be human, can be material, is used for something, carries data, carries technology, carries knowledge, is used for actions, is used for behavior, has utility, has operational utility, has financial utility, has a benefit to something, can be abstract information, can be a representation of information, can be decomposed, can be accessed, can be controlled, can be software, can be hardware, is active, is passive, may be the realization of an EA artifact, classifies a resource, is scalable, is scarce, is exchangeable, is used by something, is valuable, has to be protected, has a capacity, has costs, is in a specific form, has a source, has a target audience, has a creation date, has a modification date, has a version, access is controlled, has a volume, is core, is non-core, is skilled, has structure, has boundaries, requires management, can be scarce, is a means to execute actions, is a means to initiate behavior, is transformed into a product, is consumed in a process, has a life cycle, is useful for some goal, has a price, is of a specific kind, exists in itself, has an availability, is human, is money, is input, is needed to realize output, is a means to perform an activity, can be a machine, is a quantity of human capacity, is a machine, is some amount of machine capacity, is needed for an activity, is optional, is required, has properties, is needed to achieve a goal, has availability, is difficult to manage, is a means of support for an actor, can have alternatives, is required for an activity, costs money, has a quality level, uses something, is an employee, is used up, has to be refilled, can be quantified, can be measured

B.6 Restriction

is defined, is verifiable, has a type, has an owner, restricts something, limits something, is valid from some time, is valid until some time, is invalidated by someone, is cancelled by someone, has conditions, may have consequences, is defined by rules, has a name, is part of a policy, limits an actor's privileges, prevents an action, sets a precondition for an action, sets a postcondition for an action, sets a condition for an action, limits the permissible value range for an object, can be expressed numerically, is a regulation, is a limitation, is natural, is artificial, limits possible solutions, originates outside of its applicable area, is owned, needs to be discussed, has to be made explicit, should be formulated according to S.M.A.R.T., is a contribution of an actor to a result, is common, is measurable, is quantitative, can prevent an action, can limit possible solutions, has a source, is not allowed, is forbidden, has a source which is authoritative, has an operation domain, needs to be met during some period of time, can be temporal, can be determined by something, restricts the way something can be done, is a consequence of a higher principle, limits the behavior of an actor, has a motivation, violation may result in some punishment, has a level of generality, something might be compliant to a restriction, limits choices, may be represented as business rules, is a forbidden choice, is stipulated by others, is stipulated by the environment, limits a resource, can be absolute, limits an activity, is temporal, is situational, can influence a goal, can influence an event, is dogmatic behavior, is concrete, limits choice of scenarios, limits time, limits budget, limits available knowledge, is specific, is applicable to something, supports the realization of a goal, can have a quantitative value

B.7 Result

has a state, has data, has a related objective, has a related event, has a postcondition, has a precondition, satisfies a precondition, is measurable, has a name, is the output of a service, is the outcome of an action, meets a goal, is produced by an actor, is observable, can have an effect on something, is the consequence of an event, is output, is an outcome, can be aimed for, follows an action, is achieved by specific actions, can be described quantitatively, can be described qualitatively, is temporal, has value for someone, is delimited, is complete, has a quality, has related process, has a description, has a related actor, has a purpose, is the result of behavior, can affect a structure, is valued by a stakeholder, can be evidence of success, can be quantitative, can be qualitative, is achieved, is not achieved, exists after its achieving, is clearly described, is material, is immaterial, is a planned outcome of a process, can be part of a KPI, achievement can be monitored, is realized, is desired, achieving it can be a goal, has a type, is based on input values, is a product, is a service, has a form, has content, has volume, has a frequency, is a KPI, results after a process step, results after a process, can be composed, required to result from something, results from an action, enables other actions, results from an actor, is clear, can be input for an actor, contributes to a goal, has a property, has a benefit, has a value, results from a clear process, has a related value, has costs, is related to an activity, is a state of something, can be part of a process, associated with a process

APPENDIX C

Semantic Differential Data

C.1 Adjectives used in the semantic differentials

natural: natuurlijk, organisch, ontworpen, gemaakt, intrinsiek
human: menselijk, heeft gevoelens, is zelfbewust, heeft verlangens, is een machine
composed: unit, heeft onderdelen, atomair, geheel, totaal
necessary: noodzakelijk, essentieel, nodig, fundamenteel, moet opgevolgd worden
intentional: opzettelijk, spontaan, willekeurig, toevallig, geformuleerd
material: solide, fysiek, stoffelijk, echt, aanraakbaar
vague: exact, concreet, gedefinieerd, voorspelbaar, dubbelzinning

C.2 Raw data

The data given here are the raw, non-normalized results of the average adjective score for each meta-concept dimension combination. Rows represent respectively the meta-concepts ACTOR, EVENT, GOAL, PROCESS, RESOURCE, RESTRICTION, RESULT. The columns represent respectively the dimensions *natural, human, composed, necessary, intentional, material, vague.*

C.2.1 Practitioners

PARTICIPANT 1

1.60	1.40	2.60	2.00	2.60	2.00	3.00
3.00	3.40	2.60	1.40	2.40	3.40	4.20
3.60	3.00	2.00	1.00	1.00	2.60	5.00
3.40	3.80	1.00	1.00	1.00	2.60	5.00
2.60	3.00	3.00	2.00	2.20	2.00	4.40
3.80	3.40	3.40	1.00	1.40	2.60	5.00
4.00	4.20	1.40	1.00	1.00	2.60	5.00

PARTICIPANT 2

2.60	3.60	2.40	3.40	2.00	3.40	3.40
4.20	4.20	2.00	1.60	1.20	3.60	4.60
3.40	4.40	3.20	2.00	2.20	3.40	3.40
4.20	4.20	2.20	1.40	1.20	4.00	4.80
4.20	4.20	2.60	2.60	1.20	3.00	4.40
4.20	4.40	3.40	2.00	1.20	4.00	5.00
4.20	4.40	2.20	1.60	1.20	3.60	4.80

PARTICIPANT 3

3.00	2.20	3.00	2.00	2.00	1.80	4.20
3.60	3.40	3.00	2.00	3.40	3.80	3.20
3.20	3.60	3.40	1.60	1.80	3.60	3.40
3.20	3.20	2.40	2.20	2.80	3.60	2.00
3.20	2.80	3.00	2.40	3.00	2.60	2.80
3.00	3.40	3.20	2.80	2.00	3.00	3.60
3.60	3.60	3.20	2.60	2.00	3.40	3.80

PARTICIPANT 4

3.40	2.80	2.60	2.40	2.00	2.20	3.80
3.00	3.40	2.80	2.00	2.80	3.60	3.20
3.00	2.80	3.00	3.00	3.20	2.40	3.40
2.60	3.40	3.00	3.00	3.20	3.20	3.40
3.00	3.00	3.00	3.00	3.00	3.00	3.00
3.00	3.00	3.00	3.00	3.00	3.00	3.00
3.00	3.00	3.00	3.00	3.00	3.00	3.00

PARTICIPANT 5

3.00	2.40	2.40	1.40	1.80	3.00	3.80
4.00	3.40	2.00	1.00	2.00	4.20	4.60
3.40	2.60	3.00	1.40	2.60	4.00	3.60
4.20	3.40	3.20	1.60	2.00	4.20	3.80
2.80	3.00	3.00	2.20	3.20	3.40	2.60
4.20	3.80	3.20	1.80	1.00	3.60	5.00
3.80	4.20	2.60	1.80	1.40	4.20	4.40

PARTICIPANT 6

3.00	2.80	3.00	2.00	2.60	2.80	3.60
3.60	3.20	3.00	3.40	2.60	3.20	3.80
3.20	3.60	2.80	2.00	2.60	3.40	4.60
3.60	3.00	3.40	3.80	3.00	3.40	3.80
3.20	3.40	3.60	3.60	2.60	2.60	3.60
3.80	3.60	2.80	3.00	2.60	3.60	4.20
3.20	3.40	3.20	3.00	2.60	3.40	4.00

PARTICIPANT 7

2.80	2.60	3.40	1.80	2.20	3.00	3.60
3.20	4.60	3.20	2.80	2.80	4.00	3.60
3.20	4.00	3.20	2.20	1.80	3.20	4.00
3.00	3.80	1.40	2.00	2.00	3.20	3.00
4.00	4.80	3.20	2.60	1.80	2.20	4.20
4.40	4.40	3.40	2.00	2.00	3.80	3.80
3.40	4.20	3.40	2.40	1.60	3.60	4.00

PARTICIPANT 8

2.40	3.40	2.80	1.60	2.40	3.00	3.40
3.00	3.80	2.40	2.40	3.20	4.20	3.40
3.20	3.80	2.60	1.60	2.80	3.80	3.20
3.40	3.40	2.60	3.00	3.00	4.20	2.40
2.60	2.00	3.20	2.20	2.80	2.00	3.40
3.20	4.00	3.00	2.20	3.00	3.40	3.00
3.40	4.20	3.00	1.60	2.00	3.00	3.60

PARTICIPANT 9

2.40	2.20	3.60	2.00	3.20	2.20	2.80
2.80	2.80	3.40	2.40	3.60	2.60	2.60
2.40	1.80	3.60	1.60	3.60	3.20	2.20
2.60	2.20	3.40	2.60	3.40	2.80	2.60
2.60	2.80	3.80	2.00	3.40	3.00	1.80
2.40	2.20	3.80	2.40	3.40	3.60	2.80
2.60	2.40	3.60	3.80	3.80	2.00	3.20

PARTICIPANT 10

3.20	2.60	2.80	2.80	1.60	1.80	3.80
2.60	3.60	3.00	2.40	3.20	3.80	3.80
3.20	3.40	3.20	2.40	1.60	4.40	4.20
4.60	4.20	3.20	2.20	1.40	5.00	3.80
4.60	4.20	3.80	1.40	1.60	2.00	4.40
4.60	4.20	2.80	2.00	1.00	4.20	5.00
4.40	4.20	2.60	3.00	2.00	3.60	4.00

PARTICIPANT 11

3.20	2.80	2.60	1.80	2.60	3.20	3.20
2.60	2.40	2.80	2.40	3.00	2.60	2.80
3.20	2.80	3.00	2.40	1.80	2.80	3.40
3.00	2.80	2.20	1.80	1.80	2.20	4.40
3.20	3.40	2.60	1.40	2.40	2.80	4.00
3.60	4.40	2.80	2.00	2.60	2.00	3.80
3.60	3.40	2.80	1.60	1.80	2.40	4.00

PARTICIPANT 12

2.60	2.60	2.60	2.60	2.80	2.40	2.40
3.20	3.00	3.00	3.20	3.00	2.80	2.80
3.00	3.00	3.40	3.20	3.00	2.80	3.00
3.20	3.40	3.80	3.60	3.80	3.40	3.60
3.60	3.80	4.20	3.80	4.00	4.40	4.40
3.60	3.60	3.80	3.80	3.80	3.40	3.60
3.60	3.60	3.80	3.80	3.80	3.40	3.60

MEDIAN OF ALL

2.80	2.60	2.80	2.00	2.40	2.80	3.40
3.20	3.40	3.00	2.40	3.00	3.60	3.60
3.20	3.40	3.20	2.00	2.20	3.40	3.40
3.40	3.40	2.60	2.20	2.00	3.40	3.80
3.20	3.40	3.20	2.20	2.60	2.60	4.00
3.80	3.80	3.20	2.00	2.00	3.60	3.80
3.60	4.20	3.00	2.40	2.00	3.40	4.00

SD OF ALL

0.14	0.21	0.14	0.14	0.21	0.28	0.21
0.28	0.28	0.21	0.42	0.28	0.35	0.49
0.07	0.35	0.14	0.28	0.28	0.21	0.21
0.21	0.28	0.49	0.49	0.64	0.42	0.64
0.35	0.35	0.21	0.21	0.35	0.28	0.28
0.28	0.28	0.14	0.21	0.49	0.14	0.42
0.14	0.28	0.28	0.57	0.42	0.21	0.28

C.2.2 Students

Phase 1

PARTICIPANT 1

3.40	3.60	2.60	2.20	2.80	3.00	3.40
2.60	3.60	3.20	3.60	3.40	3.40	2.60
4.20	3.60	2.60	3.40	1.40	3.60	4.00
3.60	4.60	2.40	2.00	1.60	3.20	4.00
2.60	4.20	3.00	2.40	2.60	2.20	3.00
3.80	3.80	3.20	3.00	1.60	2.80	3.80
3.80	4.00	2.60	2.20	2.20	2.00	4.40

PARTICIPANT 2

3.80	3.40	2.80	1.80	2.40	2.40	4.00
3.20	3.80	2.40	2.60	2.00	3.60	3.60
3.00	3.40	3.00	1.40	2.80	2.60	4.60
2.60	3.80	2.40	1.80	2.20	2.00	3.80
3.00	3.60	3.20	2.20	2.00	2.40	4.00
3.60	3.40	2.80	1.80	2.20	3.20	3.60
3.80	3.40	2.20	2.20	1.80	2.20	4.00

PARTICIPANT 3

3.60	4.20	2.60	2.20	1.20	2.80	4.00
4.00	4.80	3.20	1.20	2.80	3.20	5.00
4.00	4.20	1.80	1.20	2.00	3.60	3.40
4.00	4.40	2.60	1.80	1.40	3.60	4.40
2.40	3.80	3.60	2.60	2.20	3.60	2.00
3.80	4.00	3.60	4.60	1.00	4.20	4.20
3.20	3.80	1.40	2.40	2.80	3.00	3.20

PARTICIPANT 4

3.40	3.80	2.60	1.80	2.60	2.80	4.00
3.80	3.80	4.00	2.80	3.00	4.00	2.80
3.00	4.00	3.20	2.20	1.80	2.80	3.80
3.60	4.20	2.20	2.20	1.60	3.40	3.20
3.20	4.20	2.40	2.20	3.00	3.80	3.40
3.00	3.60	3.00	2.40	3.20	2.00	4.00
2.80	3.80	2.80	2.80	2.00	2.00	4.00

```
PARTICIPANT 5
4.00 4.20 1.80 1.60 1.20 4.00 4.20
4.40 4.80 2.60 1.40 2.40 3.40 5.00
4.20 3.80 1.80 1.20 2.80 2.60 4.20
4.20 4.80 1.60 1.20 2.00 3.60 4.40
4.20 4.80 1.20 1.40 1.80 3.00 5.00
4.20 4.00 1.80 1.40 1.60 3.00 4.20
4.20 4.60 1.60 2.00 2.40 3.20 4.00

PARTICIPANT 6
2.20 2.00 3.20 2.40 3.20 2.20 2.80
4.00 4.00 3.00 2.40 2.60 4.20 3.80
3.80 3.80 3.00 2.40 2.40 4.00 2.80
3.00 3.60 2.60 2.40 2.20 3.00 2.80
4.00 4.60 2.80 2.40 2.40 2.20 3.80
3.00 3.40 2.60 2.20 2.20 3.20 3.60
3.40 4.00 2.40 2.20 2.60 3.40 3.60

PARTICIPANT 7
3.00 3.00 2.80 2.40 2.80 3.00 3.80
3.00 4.20 3.00 3.00 3.00 2.60 3.00
3.60 4.20 2.80 2.40 2.20 2.80 3.60
4.00 4.20 2.40 3.00 2.80 2.80 3.40
2.80 3.00 3.20 2.60 3.00 2.40 3.20
3.40 3.80 3.00 2.80 2.00 2.80 4.20
3.20 3.00 2.80 2.60 2.60 2.60 3.20

PARTICIPANT 8
3.80 4.20 3.40 3.80 2.60 3.60 4.20
3.40 3.20 4.40 1.60 3.00 2.00 3.60
3.20 3.60 4.00 2.20 1.40 2.00 4.20
2.60 3.40 1.40 2.40 1.80 2.60 3.60
2.40 4.00 2.00 1.80 1.60 1.40 3.20
3.40 3.20 4.20 3.00 1.80 4.00 3.40
2.00 3.60 3.00 3.00 2.20 2.60 2.60

MEDIAN OF ALL
3.50 3.70 2.70 2.20 2.60 2.90 4.00
3.60 3.90 3.10 2.50 2.90 3.40 3.60
3.70 3.80 2.90 2.20 2.10 2.80 3.90
3.60 4.20 2.40 2.10 1.90 3.10 3.70
2.90 4.10 2.90 2.30 2.30 2.40 3.30
3.50 3.70 3.00 2.60 1.90 3.10 3.90
3.30 3.80 2.50 2.30 2.30 2.60 3.80

SD OF ALL
0.21 0.35 0.07 0.21 0.14 0.21 0.14
0.28 0.21 0.21 0.49 0.14 0.28 0.49
0.35 0.14 0.21 0.35 0.35 0.35 0.21
0.35 0.28 0.14 0.21 0.21 0.28 0.28
0.28 0.28 0.28 0.14 0.28 0.28 0.28
0.21 0.21 0.21 0.28 0.21 0.21 0.21
0.35 0.14 0.21 0.14 0.21 0.35 0.28
```

Phase 2

```
PARTICIPANT 1
3.00 3.00 3.60 3.00 4.00 3.80 2.60
2.40 2.00 3.60 3.60 3.40 3.00 2.00
3.80 4.00 2.80 1.80 1.00 2.80 4.60
3.20 3.60 2.60 2.20 2.00 3.80 3.00
4.00 4.60 2.80 3.00 2.00 2.20 4.00
4.40 4.40 3.00 2.00 1.00 3.40 4.60
3.80 3.80 2.20 2.40 1.20 2.60 4.60

PARTICIPANT 2
3.40 4.00 3.40 2.40 2.60 2.20 4.00
3.40 4.20 3.00 2.60 2.40 3.40 3.80
3.80 4.40 2.40 2.00 1.80 2.60 3.60
3.00 4.00 2.80 2.00 2.60 2.60 3.20
3.00 4.20 3.20 2.00 2.40 2.60 3.60
3.60 4.00 3.40 2.00 2.20 3.40 3.40
3.60 4.00 2.40 2.20 2.40 2.20 4.00

PARTICIPANT 3
2.20 1.20 2.00 1.60 2.60 1.20 3.00
4.00 4.20 2.20 2.60 2.80 3.40 3.00
3.80 3.60 1.00 2.00 1.00 2.20 4.40
3.80 4.00 3.40 2.00 2.00 3.80 3.60
2.00 4.20 3.20 2.40 3.40 4.20 3.40
4.60 4.20 3.60 2.20 1.00 3.60 4.40
4.00 4.20 3.40 3.60 1.60 3.40 4.40

PARTICIPANT 4
2.40 2.80 2.80 2.00 2.60 3.00 2.80
3.40 4.00 4.60 3.20 3.60 4.40 3.20
3.20 3.00 2.40 1.80 1.60 2.20 4.00
3.00 2.80 2.60 2.00 2.00 4.20 3.00
2.80 4.40 3.00 2.40 2.00 1.60 4.00
3.20 4.00 3.00 2.40 3.40 3.40 4.40
2.80 3.40 2.60 3.00 3.20 2.80 4.00
```

```
PARTICIPANT 5
3.20 2.80 2.40 1.80 2.00 1.80 4.80
3.60 3.60 2.80 2.00 1.80 2.60 4.60
4.40 4.20 2.40 2.60 3.60 3.40 3.60
2.80 2.80 2.20 1.60 2.00 2.20 5.00
3.00 3.40 2.00 1.20 3.40 1.20 4.00
3.40 4.00 2.60 2.20 1.60 2.00 4.40
3.60 4.00 2.40 2.20 3.00 2.40 3.80
```

```
PARTICIPANT 6
1.60 1.00 3.00 2.80 3.20 1.60 3.00
3.80 3.80 3.00 2.20 2.60 3.40 3.80
4.00 3.80 3.00 2.20 2.80 3.40 4.00
3.80 3.40 3.00 2.80 2.40 3.60 3.80
3.40 4.80 2.40 2.00 2.40 2.60 3.60
3.80 4.20 2.40 2.00 2.00 3.20 3.40
3.80 4.20 2.60 2.20 1.60 2.80 3.20
```

```
PARTICIPANT 7
3.00 3.00 3.20 3.00 3.40 3.00 3.00
3.00 3.00 3.00 3.20 3.00 3.20 2.80
3.00 3.00 3.00 3.00 2.20 3.20 2.80
3.40 3.00 2.60 3.00 2.80 3.00 3.20
3.20 3.40 2.80 3.00 2.80 3.00 3.40
3.00 3.00 3.00 2.60 3.00 3.00 3.00
3.20 3.40 2.80 3.00 2.80 2.80 3.00
```

```
PARTICIPANT 8
3.20 2.00 3.60 1.60 3.40 2.40 2.60
3.60 4.60 3.80 4.00 2.40 3.20 3.60
2.60 3.20 2.00 1.80 1.40 2.00 4.20
3.40 3.40 3.60 3.00 2.60 4.20 2.60
3.00 3.20 1.80 1.80 3.40 1.40 2.00
2.80 3.60 1.80 3.00 1.60 1.80 3.80
3.80 3.60 1.20 2.40 1.60 2.40 3.60
```

```
MEDIAN OF ALL
3.00 2.80 3.10 2.20 2.90 2.30 3.00
3.50 3.90 3.00 2.90 2.70 3.30 3.40
3.80 3.70 2.40 2.00 1.70 2.70 4.00
3.30 3.40 2.70 2.10 2.20 3.70 3.20
3.00 4.20 2.80 2.20 2.60 2.40 3.60
3.50 4.00 3.00 2.20 1.80 3.30 4.10
3.70 3.90 2.50 2.40 2.00 2.70 3.90
```

```
SD OF ALL
0.21 0.35 0.28 0.42 0.28 0.49 0.21
0.14 0.21 0.28 0.35 0.21 0.07 0.28
0.28 0.35 0.28 0.14 0.42 0.35 0.28
0.21 0.35 0.14 0.21 0.14 0.35 0.21
0.14 0.35 0.28 0.21 0.42 0.49 0.21
0.28 0.14 0.28 0.14 0.42 0.14 0.28
0.07 0.21 0.14 0.14 0.42 0.14 0.28
```

Phase 3

```
PARTICIPANT 1
3.00 3.20 2.60 1.80 2.00 2.40 4.00
3.60 4.00 2.60 3.00 2.40 3.80 4.00
4.00 4.00 2.60 2.00 1.60 3.40 4.00
3.60 4.00 2.20 2.20 2.80 3.40 3.60
4.00 4.00 2.60 2.60 2.20 2.40 4.00
3.80 4.00 3.00 1.80 1.60 2.80 4.60
3.60 4.00 2.20 2.80 3.00 2.80 4.00
```

```
PARTICIPANT 2
3.00 2.20 3.00 2.20 2.80 2.00 3.80
3.40 2.80 2.40 2.20 3.00 2.60 3.80
3.20 3.60 2.60 1.40 2.40 2.20 4.00
2.80 3.40 2.60 2.00 3.00 2.20 3.80
2.40 2.80 3.00 2.00 2.60 2.20 3.60
3.00 3.60 3.00 2.20 2.00 2.40 3.80
3.00 3.80 2.40 2.00 2.60 2.00 3.80
```

```
PARTICIPANT 3
2.40 2.40 3.60 2.00 2.60 2.00 3.80
4.00 4.40 4.00 2.00 2.20 4.00 4.00
4.20 3.80 2.40 2.60 1.00 3.40 3.80
3.60 3.60 3.80 2.00 2.00 4.40 3.40
4.40 4.20 2.80 1.80 1.80 3.20 3.60
4.00 4.00 3.80 3.40 1.80 3.60 4.20
4.00 4.20 2.60 3.60 2.40 2.40 3.40
```

```
PARTICIPANT 4
2.40 2.00 2.20 2.00 2.80 2.20 3.40
3.20 3.60 4.20 3.20 3.20 4.40 4.00
4.20 3.60 1.40 1.60 1.60 3.60 4.40
2.60 3.00 2.00 2.20 2.00 2.80 3.00
2.60 3.40 3.40 2.20 2.00 1.80 4.00
3.60 3.20 4.20 1.80 1.20 3.40 4.20
2.60 3.80 2.40 2.20 1.80 2.40 4.00
```

```
PARTICIPANT 5
3.00 2.00 3.00 1.80 2.80 2.40 3.00
3.20 5.00 3.40 2.80 2.60 4.00 2.80
3.20 3.80 2.40 1.80 3.00 3.00 2.40
2.60 4.00 3.60 2.20 3.40 2.40 2.60
2.60 3.60 2.40 2.40 3.00 2.40 2.40
2.80 3.40 3.60 2.60 3.00 3.20 2.40
3.00 4.00 2.40 2.20 3.20 2.20 2.40

PARTICIPANT 6
3.20 3.00 2.40 2.20 3.20 2.80 3.00
3.40 4.20 2.80 2.60 2.80 3.80 3.60
3.60 4.00 2.80 2.60 2.20 3.60 3.60
3.80 4.20 2.40 2.80 2.00 3.80 4.00
3.60 4.40 3.00 2.00 2.60 3.40 3.80
3.40 3.80 3.00 2.00 2.20 3.40 3.80
3.80 4.20 2.80 2.00 2.60 3.00 3.40

PARTICIPANT 7
3.20 3.00 3.40 3.00 2.20 2.20 3.80
3.20 4.20 2.80 2.20 3.00 4.20 3.20
3.20 4.20 2.80 1.80 1.80 2.80 3.80
3.00 4.20 2.20 2.80 2.80 3.00 3.40
3.00 4.40 4.00 2.20 1.80 2.00 3.80
3.00 4.20 4.00 2.00 1.60 3.60 3.60
3.40 3.80 2.60 2.40 2.60 2.60 3.40

PARTICIPANT 8
2.60 2.40 3.00 2.00 2.80 2.20 3.20
4.20 3.80 3.80 2.80 3.00 4.00 3.40
3.40 3.60 1.60 2.00 1.80 3.60 4.40
3.60 4.00 1.60 2.80 1.40 3.00 3.00
3.20 3.80 3.00 2.40 1.80 1.60 3.60
3.00 3.40 3.20 2.20 1.40 3.00 4.20
3.40 4.20 3.80 2.40 2.00 4.00 3.80

MEDIAN OF ALL
3.00 2.40 3.00 2.00 2.80 2.20 3.60
3.40 4.10 3.10 2.70 2.90 4.00 3.70
3.50 3.80 2.50 1.90 1.80 3.40 3.90
3.30 4.00 2.30 2.20 2.40 3.00 3.40
3.10 3.90 3.00 2.20 2.10 2.30 3.70
3.20 3.70 3.40 2.10 1.70 3.30 4.00
3.40 4.00 2.50 2.30 2.60 2.50 3.60

SD OF ALL
0.14 0.28 0.28 0.14 0.07 0.14 0.21
0.14 0.21 0.42 0.28 0.14 0.14 0.21
0.21 0.14 0.14 0.14 0.21 0.14 0.14
0.28 0.14 0.21 0.14 0.28 0.35 0.28
0.35 0.28 0.21 0.14 0.21 0.28 0.07
0.21 0.21 0.28 0.14 0.21 0.21 0.14
0.28 0.14 0.07 0.14 0.21 0.21 0.14
```

Phase 4

```
PARTICIPANT 1
3.00 3.20 2.60 1.60 2.20 3.40 3.80
3.60 4.20 2.00 2.40 2.40 3.60 3.60
4.40 4.00 2.00 1.60 1.20 3.80 4.60
3.60 4.40 2.40 2.00 1.40 3.40 4.20
3.80 4.20 2.40 2.60 2.20 3.40 4.20
4.60 4.20 3.00 1.80 1.00 3.80 4.80
4.00 4.20 2.20 2.40 1.20 2.80 4.60

PARTICIPANT 2
2.80 2.60 2.60 1.60 2.40 2.20 3.80
2.80 3.40 2.80 2.20 3.40 3.00 3.40
3.00 3.00 2.40 1.80 2.20 2.00 4.00
3.00 3.60 2.40 2.00 2.40 2.00 4.00
2.40 3.40 2.80 2.20 2.60 2.20 3.00
3.20 3.40 2.80 2.00 2.20 2.20 3.60
3.60 3.80 2.20 2.20 2.40 2.00 4.00

PARTICIPANT 3
2.80 2.00 2.40 2.00 1.80 2.00 3.20
3.60 4.20 4.20 1.60 1.80 4.40 4.40
3.80 4.20 1.60 2.80 1.00 3.60 4.40
3.80 4.60 4.20 2.00 1.60 4.80 3.80
4.00 4.80 4.20 1.40 1.20 2.00 4.00
3.60 4.20 4.00 2.40 1.40 3.20 3.80
3.40 4.20 2.20 2.40 1.20 3.80 3.80

PARTICIPANT 4
2.80 1.80 2.80 2.00 3.20 2.00 3.60
2.80 3.60 3.60 3.40 3.40 3.80 4.00
2.60 2.80 3.80 2.00 1.60 2.20 3.60
3.00 3.60 1.60 2.00 2.00 2.40 3.60
2.60 3.00 3.60 2.00 3.20 1.80 3.20
2.40 4.00 4.20 2.00 1.80 2.80 4.20
3.00 3.60 2.80 2.20 2.20 2.00 4.00
```

PARTICIPANT 5

3.80	3.40	2.40	1.40	1.80	2.00	4.60
4.80	4.60	3.00	2.40	1.80	2.60	4.00
4.00	4.20	2.20	2.20	1.40	3.20	4.00
5.00	4.60	1.60	1.00	1.40	3.60	4.80
4.00	5.00	2.40	1.40	1.80	2.20	5.00
4.20	3.80	2.20	2.00	1.60	3.00	4.80
4.20	3.60	1.80	1.60	1.20	1.80	3.80

PARTICIPANT 6

3.00	2.80	2.80	2.40	2.60	2.80	2.60
3.40	4.00	2.80	2.00	2.80	3.80	3.40
3.80	3.80	2.80	2.00	2.20	3.40	4.00
3.40	3.80	2.60	2.00	2.20	2.80	3.60
3.00	4.00	2.80	1.60	2.20	2.80	3.60
3.00	4.00	3.00	1.80	2.20	3.00	4.20
3.40	4.00	2.80	1.60	2.00	2.80	4.20

PARTICIPANT 7

3.00	3.20	3.40	2.60	2.20	2.80	4.00
3.80	3.40	3.40	2.60	2.40	3.40	3.80
3.60	4.20	3.40	2.00	1.60	2.80	4.20
3.20	4.60	2.40	2.20	2.00	3.20	3.80
2.80	3.20	3.00	1.60	2.20	2.60	3.60
2.80	3.00	3.80	1.80	2.80	3.00	3.80
3.40	3.00	2.40	1.80	2.20	2.40	3.80

PARTICIPANT 8

2.60	2.40	3.20	4.00	3.20	2.00	2.60
4.00	4.00	3.60	4.20	3.00	3.60	4.00
3.00	3.60	1.80	1.60	1.40	2.40	3.40
3.00	3.60	1.60	2.80	1.80	2.60	3.20
2.60	3.20	2.20	2.20	2.20	1.80	3.00
3.00	3.20	2.00	2.40	1.40	2.80	3.60
3.80	4.00	1.60	3.00	2.00	2.80	4.20

MEDIAN OF ALL

2.90	2.70	2.70	2.00	2.30	2.10	3.70
3.60	4.00	3.20	2.40	2.60	3.60	3.90
3.70	3.90	2.30	2.00	1.50	3.00	4.00
3.30	4.10	2.40	2.00	1.90	3.00	3.80
2.90	3.70	2.80	1.80	2.20	2.20	3.60
3.10	3.90	3.00	2.00	1.70	3.00	4.00
3.50	3.90	2.20	2.20	2.00	2.60	4.00

SD OF ALL

0.07	0.35	0.14	0.28	0.28	0.07	0.28
0.21	0.21	0.28	0.21	0.42	0.14	0.14
0.35	0.21	0.35	0.14	0.14	0.42	0.21
0.21	0.35	0.35	0.00	0.21	0.35	0.14
0.28	0.35	0.28	0.28	0.14	0.28	0.35
0.28	0.21	0.57	0.14	0.28	0.14	0.21
0.14	0.21	0.21	0.21	0.21	0.28	0.14

Phase 5

PARTICIPANT 1

3.40	3.20	2.60	2.40	1.40	3.20	4.40
3.40	4.20	2.20	2.40	2.60	4.00	4.00
4.60	3.60	2.00	1.60	1.20	5.00	5.00
3.40	4.40	2.60	2.20	2.00	3.40	4.40
4.00	4.00	2.60	2.40	1.60	3.00	4.20
4.40	4.20	3.00	2.00	1.20	3.80	4.60
3.80	4.20	1.80	1.80	1.60	2.80	5.00

PARTICIPANT 2

2.80	2.20	2.40	2.40	2.60	2.00	4.00
3.20	3.40	3.00	2.00	2.80	2.60	3.80
3.60	4.40	2.40	2.00	2.00	1.80	4.20
3.20	3.60	3.00	2.20	3.00	2.20	3.60
2.80	3.60	3.20	2.00	2.60	2.60	3.00
3.40	3.60	3.40	2.20	2.40	2.00	3.40
3.80	3.80	2.40	2.20	2.60	2.00	3.80

PARTICIPANT 3

4.20	4.20	3.00	3.80	2.20	4.20	4.00
3.80	4.00	3.80	2.80	3.60	3.80	3.40
3.60	4.00	3.60	4.20	1.60	3.40	4.00
3.20	4.20	3.00	2.00	1.80	3.80	3.20
4.00	4.40	3.80	2.40	1.80	2.00	3.60
3.40	4.20	3.80	3.60	2.00	3.40	4.40
3.60	4.20	3.40	3.00	2.60	2.80	3.40

PARTICIPANT 4

3.40	1.60	2.80	2.40	3.20	2.00	3.60
3.60	4.20	3.40	3.40	3.60	3.40	3.60
3.20	3.00	2.40	1.60	1.40	3.20	3.80
3.00	4.00	2.40	1.60	1.60	2.60	3.60
2.20	2.80	2.60	2.60	3.00	2.00	3.60
2.60	2.60	2.80	2.60	2.00	3.20	4.00
2.80	3.80	2.40	2.20	2.40	2.40	4.20

PARTICIPANT 5

4.60	2.40	2.00	1.40	1.60	2.80	4.60
4.00	4.60	2.60	1.00	1.20	2.20	4.80
4.40	4.20	2.00	1.20	1.60	3.00	4.60
2.80	2.80	3.00	1.20	3.60	4.00	2.60
4.60	4.40	1.80	1.00	1.00	1.60	5.00
3.80	3.20	2.60	2.00	1.00	2.00	4.40
3.40	4.20	3.60	1.00	1.80	2.60	4.80

PARTICIPANT 6

2.80	3.20	3.00	3.20	3.20	3.20	3.00
3.20	3.80	3.00	2.80	3.60	3.40	3.40
4.00	4.20	2.80	2.00	2.00	3.80	3.40
3.60	2.80	3.00	2.60	2.40	3.60	3.80
3.80	4.00	2.80	2.20	2.40	2.60	3.00
3.60	4.00	3.20	2.40	1.60	4.00	4.40
3.40	4.40	2.80	1.40	1.60	3.00	3.40

PARTICIPANT 7

3.00	3.00	2.80	3.00	2.60	2.60	3.00
3.00	4.20	3.20	2.80	2.60	3.60	3.20
3.40	3.80	3.20	2.20	2.20	2.80	3.80
3.60	3.60	3.60	3.00	2.40	3.60	3.40
3.20	3.00	3.60	1.60	2.40	2.40	3.60
3.60	3.60	2.80	2.80	2.40	2.80	3.80
3.20	3.60	3.00	2.40	2.00	2.80	3.60

PARTICIPANT 8

2.40	2.40	4.00	3.00	3.60	1.80	2.60
3.40	4.20	4.60	3.20	3.20	3.20	4.20
2.80	3.20	2.00	1.80	1.60	3.00	4.40
3.80	3.40	4.00	2.40	2.40	4.20	3.20
2.80	3.60	2.40	2.40	2.40	1.20	2.40
3.80	3.20	3.60	1.80	2.00	3.60	4.60
3.40	3.80	2.40	2.40	1.60	2.20	3.40

MEDIAN OF ALL

3.20	2.70	2.80	2.70	2.60	2.70	3.80
3.40	4.20	3.10	2.80	3.00	3.40	3.70
3.60	3.90	2.40	1.90	1.60	3.10	4.10
3.30	3.60	3.00	2.20	2.40	3.60	3.50
3.50	3.80	2.70	2.30	2.40	2.20	3.60
3.60	3.60	3.10	2.30	2.00	3.30	4.40
3.40	4.00	2.60	2.20	1.90	2.70	3.70

SD OF ALL

0.28	0.35	0.14	0.21	0.42	0.42	0.49
0.14	0.07	0.28	0.28	0.35	0.21	0.21
0.28	0.21	0.28	0.21	0.21	0.21	0.21
0.21	0.35	0.14	0.21	0.35	0.21	0.21
0.42	0.28	0.28	0.14	0.28	0.28	0.42
0.14	0.28	0.21	0.21	0.28	0.35	0.14
0.14	0.14	0.21	0.21	0.21	0.14	0.21

C.3 Detailed polarity scores

Table C.1: Polarities of responses in phase 1 of the longitudinal study.

pol.	nat	hum	com	nec	int	mat	vag	avg.
				actor				
neg	40%	50%	0%	10%	0%	20%	60%	26%
neu	50%	40%	90%	20%	70%	60%	40%	53%
pos	10%	10%	10%	70%	30%	20%	0%	21%
				event				
neg	40%	70%	20%	10%	0%	30%	50%	31%
neu	60%	30%	70%	50%	80%	60%	50%	57%
pos	0%	0%	10%	40%	20%	10%	0%	11%
				goal				
neg	50%	70%	10%	0%	0%	30%	60%	31%
neu	50%	30%	70%	30%	40%	60%	40%	46%
pos	0%	0%	20%	70%	60%	10%	0%	23%
				process				
neg	50%	70%	0%	0%	0%	20%	50%	27%
neu	50%	30%	40%	30%	30%	70%	50%	43%
pos	0%	0%	60%	70%	70%	10%	0%	30%
				resource				
neg	20%	70%	10%	0%	0%	20%	30%	21%
neu	60%	30%	60%	40%	50%	30%	60%	47%
pos	20%	0%	30%	60%	50%	50%	10%	31%
				restriction				
neg	40%	50%	20%	10%	0%	20%	70%	30%
neu	60%	50%	70%	50%	30%	70%	30%	51%
pos	0%	0%	10%	40%	70%	10%	0%	19%
				result				
neg	30%	60%	0%	0%	0%	0%	50%	20%
neu	60%	40%	60%	50%	50%	70%	50%	54%
pos	10%	0%	40%	50%	50%	30%	0%	26%

Table C.2: Polarities of responses in phase 2 of the longitudinal study.

pol.	nat	hum	com	nec	int	mat	vag	avg.
				actor				
neg	0%	10%	20%	0%	10%	10%	20%	10%
neu	70%	60%	60%	50%	80%	40%	80%	63%
pos	30%	30%	20%	50%	10%	50%	0%	27%
				event				
neg	40%	60%	30%	20%	10%	10%	40%	30%
neu	50%	30%	60%	60%	60%	90%	50%	57%
pos	10%	10%	10%	20%	30%	0%	10%	13%
				goal				
neg	50%	50%	0%	0%	10%	0%	70%	26%
neu	50%	50%	50%	40%	30%	70%	30%	46%
pos	0%	0%	50%	60%	60%	30%	0%	29%
				process				
neg	20%	30%	10%	0%	0%	50%	30%	20%
neu	80%	70%	80%	50%	50%	40%	70%	63%
pos	0%	0%	10%	50%	50%	10%	0%	17%
				resource				
neg	10%	50%	0%	0%	0%	10%	50%	17%
neu	80%	50%	70%	40%	60%	50%	40%	56%
pos	10%	0%	30%	60%	40%	40%	10%	27%
				restriction				
neg	40%	70%	10%	0%	0%	10%	50%	26%
neu	60%	30%	70%	40%	40%	70%	50%	51%
pos	0%	0%	20%	60%	60%	20%	0%	23%
				result				
neg	60%	60%	0%	10%	0%	0%	60%	27%
neu	40%	40%	60%	40%	50%	70%	40%	49%
pos	0%	0%	40%	50%	50%	30%	0%	24%

Table C.3: Polarities of responses in phase 3 of the longitudinal study.

pol.	nat	hum	com	nec	int	mat	vag	avg.
				actor				
neg	0%	0%	10%	0%	0%	0%	40%	7%
neu	80%	50%	70%	30%	80%	30%	60%	57%
pos	20%	50%	20%	70%	20%	70%	0%	36%
				event				
neg	30%	70%	30%	0%	0%	70%	50%	36%
neu	70%	30%	60%	70%	80%	30%	50%	56%
pos	0%	0%	10%	30%	20%	0%	0%	9%
				goal				
neg	40%	80%	0%	0%	0%	30%	70%	31%
neu	60%	20%	60%	40%	30%	60%	20%	41%
pos	0%	0%	40%	60%	70%	10%	10%	27%
				process				
neg	40%	60%	20%	0%	0%	20%	30%	24%
neu	60%	40%	30%	50%	60%	60%	70%	53%
pos	0%	0%	50%	50%	40%	20%	0%	23%
				resource				
neg	30%	60%	10%	0%	0%	0%	70%	24%
neu	60%	40%	80%	30%	50%	40%	20%	46%
pos	10%	0%	10%	70%	50%	60%	10%	30%
				restriction				
neg	30%	50%	40%	0%	0%	20%	70%	30%
neu	70%	50%	60%	40%	30%	70%	20%	49%
pos	0%	0%	0%	60%	70%	10%	10%	21%
				result				
neg	30%	80%	10%	10%	0%	10%	40%	26%
neu	70%	20%	50%	30%	70%	50%	50%	49%
pos	0%	0%	40%	60%	30%	40%	10%	26%

Table C.4: Polarities of responses in phase 4 of the longitudinal study.

pol.	nat	hum	com	nec	int	mat	vag	avg.
				actor				
neg	10%	0%	0%	10%	0%	0%	50%	10%
neu	90%	70%	80%	30%	50%	50%	50%	60%
pos	0%	30%	20%	60%	50%	50%	0%	30%
				event				
neg	50%	60%	30%	10%	0%	50%	60%	37%
neu	50%	40%	60%	40%	60%	50%	40%	49%
pos	0%	0%	10%	50%	40%	0%	0%	14%
				goal				
neg	50%	60%	10%	0%	0%	20%	70%	30%
neu	50%	40%	40%	30%	20%	50%	30%	37%
pos	0%	0%	50%	70%	80%	30%	0%	33%
				process				
neg	30%	80%	10%	0%	0%	20%	70%	30%
neu	70%	20%	30%	30%	20%	60%	30%	37%
pos	0%	0%	60%	70%	80%	20%	0%	33%
				resource				
neg	30%	40%	20%	0%	0%	0%	50%	20%
neu	60%	60%	50%	30%	40%	50%	50%	49%
pos	10%	0%	30%	70%	60%	50%	0%	31%
				restriction				
neg	30%	50%	30%	0%	0%	10%	80%	29%
neu	60%	50%	50%	20%	30%	80%	20%	44%
pos	10%	0%	20%	80%	70%	10%	0%	27%
				result				
neg	40%	70%	0%	0%	0%	10%	80%	29%
neu	60%	30%	40%	30%	20%	50%	20%	36%
pos	0%	0%	60%	70%	80%	40%	0%	36%

Table C.5: Polarities of responses in phase 5 of the longitudinal study.

pol.	nat	hum	com	nec	int	mat	vag	avg.
				actor				
neg	20%	10%	10%	10%	10%	10%	50%	17%
neu	70%	50%	70%	50%	60%	60%	50%	59%
pos	10%	40%	20%	40%	30%	30%	0%	24%
				event				
neg	30%	70%	20%	0%	30%	30%	50%	33%
neu	70%	30%	70%	70%	60%	60%	50%	59%
pos	0%	0%	10%	30%	10%	10%	0%	9%
				goal				
neg	50%	60%	10%	10%	0%	20%	70%	31%
neu	50%	40%	40%	20%	20%	70%	30%	39%
pos	0%	0%	50%	70%	80%	10%	0%	30%
				process				
neg	30%	50%	20%	0%	10%	50%	40%	29%
neu	70%	50%	70%	40%	30%	40%	60%	51%
pos	0%	0%	10%	60%	60%	10%	0%	20%
				resource				
neg	40%	60%	20%	0%	0%	0%	50%	24%
neu	50%	40%	60%	30%	40%	50%	40%	44%
pos	10%	0%	20%	70%	60%	50%	10%	31%
				restriction				
neg	50%	50%	20%	10%	0%	30%	70%	33%
neu	50%	50%	80%	40%	20%	50%	30%	46%
pos	0%	0%	0%	50%	80%	20%	0%	21%
				result				
neg	30%	80%	10%	0%	0%	0%	50%	24%
neu	70%	20%	50%	30%	40%	70%	50%	47%
pos	0%	0%	40%	70%	60%	30%	0%	29%