Towards explicit strategies for modeling

S.J.B.A. Hoppenbrouwers, H.A. Proper and Th.P. van der Weide

Institute for Computing and Information Sciences, Radboud University Nijmegen
Toernooiveld 1, 6525 ED Nijmegen, The Netherlands, EU
{S.Hoppenbrouwers, Th.P.vanderWeide, E.Proper}@cs.ru.nl

Abstract. We present an initial framework resulting from our ongoing research concerning *modeling strategies*. Our approach is rooted in a subjectivist, communication-based view on modeling. Under this approach, models are viewed as the result of *modeling dialogues*, which are a specialized sub-type of the diverse conversations that constitute a *system development conversation* at large. By focussing on the *process* of modeling instead of properties of models or modeling languages, we expect, eventually, to be able to better understand and deal with some currently problematic aspects of modeling, in particular *model validation in context*. We sketch plans for an environment for studying modeling conversations and strategies.
As the title suggests, the current status of our research efforts are at the level where we have an initial understanding of the fundamental workings of the playing field. In future research, the actual strategies will make their appearance.

1 Focus on behavior leading to models

Much has been said in modeling and system development literature about "Ways of Modeling" [1] both "formal" and "informal" modeling languages/techniques in system development. The syntax, semantics, verifiability, quality, etc. of models and modeling languages has been extensively studied. In particular in the case of formal modeling, existing literature on this can safely be characterized as being soundly "scientific".

However, though some research has been also done concerning stages in and aspects of "Ways of Working" (i.e. the process or procedure [1]) in modeling [2, 3, 4, among others], the *how* behind the activity of creating models is still mostly art rather than science. First of all, therefore, there is a purely scientific interest in improving our understanding of the details of the modeling process. In addition, such a study might enable us to find ways of improving the modeling process (for example, its quality, efficiency, or effectiveness).

As is widely known among practitioners, a large number of different models for some domain or item may be produced even if one single Way of Modeling (language, formalism) is used. Though some ideas have been formulated on what distinguishes "good models" from "bad models" (typically, by validation or verification *in hindsight*)[5, among others], hardly any material deals in detail with *what to do to get good models*. In addition, some aspects of quality, it seems, can be better achieved through a good modeling process than by just imposing requirements on the end product. This holds in particular (though not exclusively) for matters of validation. In our view, valid models are only attainable by viewing a model *in context*, and therefore in relation to the actors who have created and agreed with some model. The process of modeling thus is a crucial link between the end product and its context.

Perhaps the most practical reason for studying which behavior leads to good models concerns human resources. Generally, few experts are available who are capable of, or willing to, perform high-quality modeling. In any case, this process takes much time and effort (especially formal modeling). However, an increasing number of high quality models (even "light formalizations") are required in system development (e.g. formalized ontologies, business rules, requirements). Anything that helps guide modeling behavior and support the process would be helpful.

2 Communicative perspective on modeling

It makes good sense to consider (the requirements and functionality of) both Ways of Modeling and Ways of Working on the basis of a fundamental Way of Thinking [1] that is an optimally fit paradigm for understanding the nature and purpose of modeling in system development. In this respect, we embrace a subjectivist, situational, and above all *communication-based* view on modeling [4]. We are not alone in this [6, 7, 2]. Essentially, we view modeling as a *knowledge transformation*: the 'knowledge state' of participants in the process is changed. Based on [4], and partly echoing [8], we identify three dimensions for the knowledge states of the development community: (1) topic (what knowledge items are about), (2) level of sharing (awareness, agreement, commitment), and (3) level of explicitness (formality, quantifiability, executability, comprehensibility, completeness). Each modeling conversation is presumed to have some *knowledge goal*: a knowledge state which it aims to achieve (or maintain). In achieving a knowledge goal, a (sub)-conversation will follow one or more *modeling strategies*.

The clearest aspect of knowledge transformations that can be observed is the exchange of particular statements (in formal or informal language, verbal or graphical) in order to achieve a knowledge goal. This is why we view "statements" (with some function, stated in some language) as the most appropriate level at which we can study what goes on in IS development and modeling. Two basic sets of statements are to be distinguished here: the statements that are a record of the modeling conversation, and the statements that constitute the knowledge that is agreed upon by all participants. The former set is a recording of the modeling conversation, the second essentially constitutes the model. Importantly, the statements are *grounded* in their (social) context: the people that stated them and agreed with them [9].

Modeling, then, to us is a process of gathering and refining statements: a sequence of questions and answers. Consequently, we reject the view of modeling

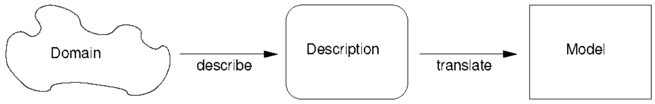

Fig. 1. A linear view on modeling

as a *linear* process of first describing a (relevant aspect-of/abstraction-from a) domain and then producing a model corresponding to this description. This is illustrated in Figure 1. In this linear process, the resulting model essentially is a *translation* from the (natural language) statements forming the description of the domain to a to a (formal) model [9]. Even though this linear approach can indeed be seen as *a* modeling strategy, we take the view that modeling strategies are needed that take a more iterative approach to modeling. This is illustrated in Figure 2. Here, the descriptions from Figure 1 are replaced by questions and answers. The collected answers could be re-constructed into a description, which should essentially be a verbalization of the model model.

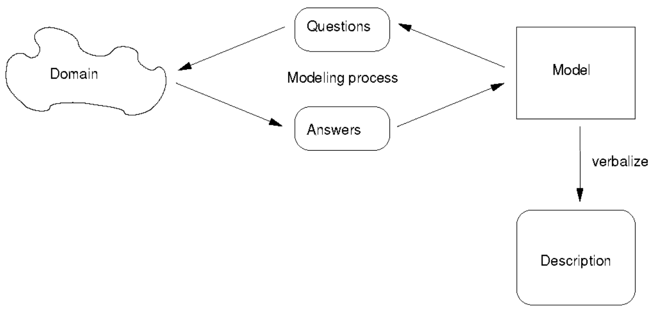

Fig. 2. An iterative view on modeling

From a generic system development perspective, modeling is – at least in theory – to be integrated with less typically "model-oriented stages" in the process, e.g. explorative requirements gathering and negotiation, obtaining commitment for models, and so on. We expect that some strategies that are useful for modeling can also be fruitfully applied outside "pure modeling". This perspective reflects the wish to apply *rational principles* (communication-oriented or otherwise) in system development where relevant and feasible [10].

3 Strategies and conversations for modeling

As a starting point for our exploration of modeling behavior we take (aspects of) the detailed ORM Conceptual Schema Design Procedure (CSDP; its Way of Working) [2]. Guidelines for modeling from other methods are also useful input

[11, 7, 12, among others]. Our ultimate goal is to distill modeling strategies that are essentially modeling language independent.

Our view of a modeling process is as follows. For particular contexts and particular knowledge goals, Ways of Modeling are selected [4] *as well as* Ways of Working: (sets of) strategies matching the Ways of Modeling; the behavioral view on modeling. Situationally selected strategies thus guide an actual, executed modeling conversation, yet the individual strategies are deployed only when the conversation requires it. Which strategy is to be applied exactly when depends on the course of the conversation, and cannot be predicted –only triggered. Strategies typically consist of a *precondition* (which matches an "initial knowledge state" in the (sub-)conversation), a *course of action* (which may involve deployment of other strategies), and a *postcondition* (which should match a knowledge goal in the (sub-)conversation).

Modeling conversations are, of course, executed by *participants*. Each participant shares knowledge statements with the others, trying to achieve agreement on them. However, the knowledge goals of participants, as well as their competencies [13], may be different in nature, in particular with respect to the level of explicitness they strive for. In an often used view on the ORM modeling process, this difference in goals and competencies is reflected in two participant roles engaging in a modeling dialogue: the Domain Expert (DE) and the System Analyst (SA). The former focuses on achieving completeness and validity of the model with respect to the domain it reflects, the latter focuses on satisfying the demands posed on explicitness by the modeling language/formalism.

The approach described is largely driven by what could be seen as the mother of all conversation strategies: *uncertainty reduction* [14] with respect to achieving the knowledge goals in modeling and, ultimately, system development. Two chief types of uncertainty reduction can be distinguished: *epistemic* and *linguistic* [15]. Both are covered by a specific set of strategies, but some overlap is expected. Ways of *dealing with* uncertainty [16, 9] can also be related to strategy categories: *gather more information*, *make assumptions*, *choose between alternatives*, or *temporarily ignore uncertainty*. Furthermore, generic strategy categories may relate to *topicality*, (levels of) *agreement*, and (levels of) *explicitness* (see section 2). Many more types and sub-types of strategy can be thought of. We have only just begun exploring this area.

Ideally, the participants work towards mutually agreed set of statements in a *controlled language*: natural language statements (or possibly, some agreed-on graphical format) with a clear and simple syntax that reduces the danger of ambiguity and confusion. This is in fact another high-level, generically applied conversation strategy that is crucial in reducing linguistic uncertainty. In this manner, the demands posed by the formalism underlying the way of modeling can be fulfilled without discussing the syntax or semantics of the formalism as such. In addition, interpretation of terms used in the conversation may be explored in dedicated sub-conversations leading to more statements expressing the meaning of the term, to the point of satisfaction of all participants [17].

Whether or not these statements are to become part of the model is a modeling decision.

4 Sketch of a framework for modeling modeling dialogues

The backbone of our framework is a "dialogue grammar", which is used to structure the sequence in which the actions can take place, and some further restrictions imposed on it [18]. In the preliminary grammar, six straightforward dialogue actions are distinguished, most of them dealing with statements:

- Propose(a, s) Actor a proposes statement s. It does not become part of the common model until every actor accepts it.
- Withdraw(a, s) Actor a withdraws statement s. Withdraw is the opposite of a propose.
- Accept(a, s) Actor a accepts statement s as a valid statement; it may eventually become part of its internal model M_a. A statement can only be accepted after it is proposed.
- Reject(a, s) Actor a rejects statement s, because a finds s unacceptable even for further consideration. Reject is the counterpart of accept.
- Ask(a, q) Actor a asks question q, to be answered by some actor. Queries can be withdrawn or answered.
- Answer(a, q, s) Actor a answers question q with statement s; an answer functions as a special Propose.

As a mere illustration of our approach to modeling dialogues and strategies, consider this very simple example, that reflects a possible strategy in ORM-like modeling: we might call it the "delayed specificity strategy". It comes in two flavors: one taking the "gather more information" approach, the other the "make assumptions" approach. A more elaborate, formalized analysis of this strategy can be found in [18]. The ORM formalism requires relational structures to be *specific*: enough information about entities must be provided, including of what *type* an entity is. Roughly, the strategy is formulated as follows:

Precondition There is a non-specific statement. It is required that the SA can identify such statements.

Course of action A Solve the non-specificity by asking the domain expert for missing information.

Course of action B Alternatively, solve the non-specificity by assuming the missing information. This requires the SA sub skill of conceiving and verbalizing plausible information.

Postcondition The non-specificity is resolved.

The strategy could lead to the following dialogue, which reflects a step-wise approach to the gathering of the information needed:

```
propose DE    John lives in Nijmegen.
    ask SA    What kind of thing is John?
```

```
    propose DE    John is a person.
ask/propose SA    Do we distinguish John from other persons
                  by means of his name?
     accept DE,SA Yes [we distinguish John from other persons
                  by means of his name].
ask/propose SA    Do you agree that John is a person with
                  name John?
     accept DE,SA Yes [John is a person with name John].
```

For each "accept", both DE and SA are registered as agreeing with the statement. In two cases, the action of asking and proposing is collapsed in one statement. As a result of this small sub-dialogue, the statement "John is a person with name John" is added to the separate set of "agreed statements", on the basis of which a complete formalization can later be compiled (preferably automatically). In addition, by means of the "defoleating strategy" (i.e. removing instances from the statements, which is a typical step in "fact based modeling"), a type-level generalization can be derived (weeding out the instance "John"), suggested, and confirmed:

```
ask/propose SA    Do persons generally have names?
     accept DE,SA Yes [persons generally have names].
```

5 Planned further research

The core goal of our research-in-progress is to study, describe, validate, and eventually *improve* strategies for (formal) modeling in a system development context. As the title suggests, the current status of our research efforts are at the level where we have an initial understanding of the fundamental workings of the playing field. In future research, the actual strategies will make their appearance. The only way of achieving satisfying results in an effort like ours is to prominently include *empirical data and validation* in the study. Our initial ideas concerning modeling and system development as a conversation (as sketched in this paper) are grounded in literature, our own experiences, and interviews with experienced modelers [4]. However, this is clearly insufficient. We plan, therefore, to create an *experimentation environment* for controlled study of modeling conversations. Progressively, we intend to study:

- Which strategies are "naturally" used by (various types of) participants in modeling conversations;
- How modeling conversations/strategies relate to the grounding and validity of the resulting models;
- How modelers can be guided in using particular strategies;
- How (sets of) strategies can be improved/optimized, in line with particular contexts, goals, and/or modeling languages.

The research approach we intend to follow can be classified as "action research" [19, 20]. This entails that our work will progress (evolutionary) through two major stages (taken from [21]):

Diagnostic stage – *This stage involves a collaborative analysis of a situation by the researcher and the subjects of the research. Theories are formulated concerning the nature of the research domain.*

Therapeutic stage – *This involves collaborative change experiments. In this stage changes are introduced and the effects are studied [22].*

Having passed the diagnostic stage at an exploratory level, we are currently preparing to start a more in-depth diagnostic study of actual modeling dialogues and strategies. However, we expect the study to move into the initial therapeutic stage before long: the moment the guiding of the dialogue is introduced. Various diagnostic and therapeutic issues will arise and be dealt with iteratively. Importantly, participants in the experiments will be "imported" from real life contexts as much as possible, so their knowledge and knowledge goals are familiar to them and optimally realistic, even if the modeling environment is, inevitably, not.

The environment for recording and guiding modeling dialogues will, at least initially, use verbal questions-and-answers in a workable controlled language. The design metaphor used is that of a "chatbox" (distantly related to the well known "telephone heuristic" in modeling), with an "automated modeling agent" (SA) as a participant in the chat. Creating an effective environment will be a study in itself. We intend to make opportunistic use of existing agent interaction and dialogue paradigms, and NLP techniques. We will start off by investigating information and domain modeling in ORM. Planned areas of application to be studied (i.e. contexts) are information modeling, pre-negotiation ontology construction, and requirements modeling.

References

1. Seligmann, P., Wijers, G., Sol, H.: Analyzing the structure of I.S. methodologies, an alternative approach. In Maes, R., ed.: Proceedings of the First Dutch Conference on Information Systems, Amersfoort, The Netherlands, EU (1989).
2. Halpin, T.: Information Modeling and Relational Databases, From Conceptual Analysis to Logical Design. Morgan Kaufman, San Mateo, California, USA (2001). ISBN 1-55860-672-6
3. Bleeker, A., Proper, H., Hoppenbrouwers, S.: The role of concept management in system development – a practical and a theoretical perspective. In Grabis, J., Persson, A., Stirna, J., eds.: Forum proceedings of the 16th Conference on Advanced Information Systems 2004 (CAiSE 2004), Riga, Latvia, EU, Faculty of Computer Science and Information Technology, Riga Technical University, Riga, Latvia, EU (2004) 73–82. ISBN 9984-9767-0-X
4. Hoppenbrouwers, S., Proper, H., Weide, T.v.d.: Understanding the requirements on modelling techniques. Technical report, Radboud University Nijmegen Institute for Computing and Information Science (2005) Accepted for publication in: Proceedings of the The 17th Conference on Advanced Information Systems Engineering (CAiSE'05), 13-17 June 2005, Porto, Portugal.
5. Krogstie, J.: A semiotic approach to quality in requirements specifications. In Liu, K., ed.: Organizational semiotics: evolving a science of information systems. IFIP TC8/WG8.1 working conference on organizational semiotics, July 23-25, 2001, Montreal, Quebec, Canada, Amsterdam: Kluwer (2002) 231–24.

6. Nijssen, G., Halpin, T.: Conceptual Schema and Relational Database Design: a fact oriented approach. Prentice-Hall, Sydney, Australia (1989). ASIN 0131672630
7. Embley, D., Kurtz, B., Woodfield, S.: Object-Oriented Systems Analysis – A model-driven approach. Yourdon Press, Englewood Cliffs, New Jersey, USA (1992). ASIN 0136299733
8. Pohl, K.: The three dimensions of requirements engineering: a framework and its applications. Information Systems **19** (1994) 243–258. ISSN 0306-4379
9. Hoppenbrouwers, S., Proper, H., Weide, T.v.d.: Formal modelling as a grounded conversation. Technical report, Institute for Computing and Information Science, Radboud University Nijmegen (2005) Submitted to the 10th Anniversary International Working Conference on the Language-Action Perspective on Communication Modelling; Kiruna, Sweden, June 19-20 2005.
10. Veldhuijzen van Zanten, G., Hoppenbrouwers, S., Proper, H.: System Development as a Rational Communicative Process. Journal of Systemics, Cybernetics and Informatics **2** (2004).
`http://www.iiisci.org/Journal/sci/pdfs/P492036.pdf`
11. Jonkers, H., Lankhorst, M., Buuren, R.v., Hoppenbrouwers, S., Bonsangue, M., Torre, L.v.d., Proper, H.: Enterprise Architecture at Work : Modelling, Communication and Analysis. Springer, Berlin, Germany, EU (2005) Lankhorst, M.M., editor. ISBN 35-402-4371-2
12. Booch, G.: Object-oriented Analysis and Design. 2nd edn. New York: Benjamin/Cummings (1994).
13. Frederiks, P., Weide, T.v.d.: Information modeling: the process and the required competencies of its participants. Data & Knowledge Engineering (2004) To appear in a special issue on the NLDB 2004 conference.
14. Hoppenbrouwers, S., Proper, H., Weide, T.v.d.: Dealing with uncertainty in information modeling. Technical report, Institute for Computing and Information Science, Radboud University Nijmegen (2005).
15. Regan, H., Hope, B., Ferson, S.: Analysis and portrayal of uncertainty in a food web exposure model. Human and Ecological Risk Assessment **8** (2002) 1757–1777.
16. Lipshitz, R., Strauss, O.: Coping with uncertainty: a naturalistic decision-making analysis. Organizational Behaviour and Human Decision Processes **2** (1997) 152–154.
17. Hoppenbrouwers, S.: Freezing Language; Conceptualisation processes in ICT supported organisations. PhD thesis, University of Nijmegen, Nijmegen, The Netherlands, EU (2003). ISBN 90-9017318-8
18. Bosman, S., Weide, T.v.d.: Towards formalization of the information modeling dialog. Technical report, Computing Science Institute, University of Nijmegen, Nijmegen, The Netherlands (2004).
19. Avison, D., Lau, F., Meyers, M., Nielsen, P.: Action research. Communications of the ACM **42** (1999) 94–97.
20. Baskerville, R., Wood-Harper, A.T.: A critical perspective on action research as a method for information systems research. Journal of Information Technology **11** (1996) 235–246.
21. Baskerville, R.: Investigating Information Systems with Action Research. Communications of the Association for Information Systems **2** (1999).
22. Blum, F.: Action research – a scientific approach? Philosophy of Science **22** (1955) 1–7.