Syntax, Semantics and Pragmatics of Conceptual Modelling

Bernhard Thalheim

Computer Science Institute,
Christian-Albrechts-University Kiel
thalheim@is.informatik.uni-kiel.de

Abstract. Models, modelling languages, modelling frameworks and their background have dominated conceptual modelling research and information systems engineering for last four decades. Conceptual models are mediators between the application world and the implementation or system world. Currently conceptual modelling is rather a craft and at the best an art. We target on a science and culture of conceptual modelling. Models are governed by their purpose. They are used by a community of practice and have a function within application cases. Language-based models use a language as a carrier. Therefore, semiotics of models must be systematically developed. This paper thus concentrates on the linguistic foundation of conceptual modelling.

1 Introductory Notions for Conceptual Modelling

Conceptual modelling is a widely applied practice in Computer Science and has led to a large body of knowledge on constructs that might be used for modelling and on methods that might be useful for modelling. It is commonly accepted that database application development is based on conceptual modelling. It is however surprising that only very few publications have been published on a *theory of conceptual modelling*. We continue our research [2, 3, 15–18] and aim in a theory of linguistic foundations for modelling within this paper.

1.1 Goals, Purposes and Deployment Functions of Conceptual Models

Purpose is often defined via intention and mixed with function. *Goal* (or intention or target or aim) is a ternary relation between a current state, envisioned states, and people (community of practice). Typical - sometimes rather abstract - intentions are perception support, explanation and demonstration, preparation to an activity, optimisation, hypothesis verification, construction, control, and substitution.

The *purpose* is a binary relation between intentions and means or instruments for realisation of the intention. The main mean we use is the language. Semiotics is widely intentionally used for modelling; however without paying attention to it.

G. Bouma et al. (Eds.): NLDB 2012, LNCS 7337, pp. 1–10, 2012.

The *deployment function* of a model relates the model purpose to a practice or application cases or application 'game' similar to Wittgenstein's language game (we call it better *deployment case* and is characterised by answering the classical W-questions: how, when, for which/what or why, at what/which (business use case), etc. We add to purpose: application, conventions, custom, exertion, habit, handling, deployment, service, usage, use, and way of using. The model has a role and plays its behaviour within this application game.

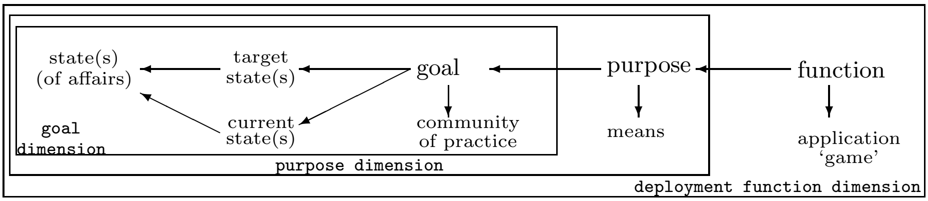

Fig. 1. Distinction of Intention, Purpose,and Deployment Function of a Model

1.2 Abstraction and Conceptual Modelling

Abstraction is one of the most overloaded conceptions in Computer Science and at the same time one of the most under-specified. Abstraction means development of general concepts by abstracting common properties of specific concepts. Using the approach in [14] we develop three dimensions of abstraction:

Structural abstraction is used for highlighting essential, necessary, and general structural elements of the origin. Structural abstraction has three main constituents: *combining structural abstraction* (often called) *classifying structural abstraction* (often called meta(-meta(-meta)) abstraction) combines things of interest into collections that contain these things as elements; *generalising structural abstraction* (often called pattern or templates).

Context abstraction "factors out" repeating, shared or local patterns of things and functionality from individual things. Context abstraction assumes that the surroundings of a things under consideration are commonly assumed by a community of practice or within a culture and focuses on the concept, turning away attention from its surroundings such as the environment and setting. Models use ambiguities, ellipses, metaphors and commonly assumed conceptions.

Behavioural abstraction is used for concentrating of essential and general behavioural elements. We may distinguish between *combining*, *classifying*, and *generalising behavioural abstraction*. Aspect separation, encapsulation, and modularisation are specific techniques.

The opposite of abstracting is detailing. *Refinement* is a specific kind of faithful detailing. Refinement uses the principle of modularisation and information hiding. Developers typically use conceptual models or languages for representing and conceptualising abstractions. Classical forms of abstraction are generalisation, isolation, and idealisation.

We thus may concentrate on three main tasks for abstraction within a community of practice:

- *Choose the right scope* within the application area in dependence on goals.
- *Choose the right focus* at the right level and in the right granularity .
- *Choose the right observation* with the right behaviour and right properties.

2 The Notion of the Model

It is common misbelief (e.g., [1] or more generally almost all Computer Science textbooks) that there is no definition of the concept of the model. We consider this claim as the *big misunderstanding* of the science and art of modelling.

A model is simply a material or virtual *artifact* (1) which is called model within a community of practice (2) based on a judgement (3) [7] of appropriateness for representation of other artifacts (things in reality, systems, ...) and serving a *purpose* (4) within this community. We distill thus criteria for artifacts to become a model. We can use on two approaches: abstract properties or we criteria for artifacts.

2.1 Stachowiak, Aristoteles, Galilei and Mahr Properties of Models

Models are often defined through abstract properties they must satisfy [15, 18].

(1) *Mapping* property: Each model has an origin and is based on a mapping from the origin to the artifact.
(2) *Truncation* property: The model lacks some of the ascriptions made to the original and thus functions as an Aristolean model by abstraction of irrelevant.
(3) *Pragmatic* property: The model use is only justified for particular model users, tools of investigation, and period of time.
(4) *Amplification* property: Models use specific extensions which are not observed for the original,
(5) *Distortion* property: Models are developed for improving the physical world or for inclusion of visions of better reality, e.g. for construction via transformation or in Galilean models.
(6) *Idealisation* property: Modelling abstracts from reality by scoping the model to the ideal state of affairs.
(7) *Carrier* property: Models use languages and are thus restricted by the expressive power of these languages.
(8)*Added value* property: Models provide a value or benefit based on their utility, capability and quality characteristics.

(9) *Purpose* property: Models and conceptual models are governed by the purpose. The model preserves the purpose.

The first three properties have been introduced by Stachowiaks [12]. The fourth and fifth property have been introduced by Steinmüller [13]. The seventh property is discussed by Mahr [10]. The sixth, eight and nine properties [18] are however equally if not the most important ones.

2.2 Criteria for Appropriateness of an Artifact to Become a Model

The separation into goal, purpose and deployment function for models provides three main appropriateness criteria:

(1) The *adequacy* of a model defines its *potential* for the goals. Adequacy is given by the similarity of the model with its origin in dependence on its goal, the regularity for the application (within a well-founded system that uses rules for derivation of conclusions), the fruitfulness (or capacity) for goals, and the simplicity of the model through the reduction to the essential and relevant properties in dependence on the goal.
(2) A model is *fit* for its purpose if it is *usable* for the purpose, *suitable* within the given context and for the prescribed purposes, *robust* against small changes in the parameters, *accurate* to the level of precision that is necessary for the purpose, and *compliant* with the funding concepts, application context and meta-model. The model must be *testable* and, if false, it can be disconfirmed by a finite set of observations (finitely testable) and by any of superset of these observation (irrevocably testable).
(3) The *usefulness* for deploying is given by *effectiveness* for complete and accurate satisfaction of the goal, *understandability* for purposeful deployment of the model by users, *learnability* of the model within the deployment stories, *reliability* and a high *degree of precision* of the the model, and *efficiency* of the model for the function of the model within the application. *testability*

Additional criteria are *generality* of the model beside its direct goals and intentions and the extend of *coverage* in the real world for other goals.

2.3 The Hidden Background of Models

The structure and function of a model are based on a correspondence relation between the reality or the augmented and the model. The model can be constructed incrementally. It represents a number of facets of the origin (topology/geometry, state, interaction, causal). The *model pragmatism* is however hidden. It consists of at least three background dimensions:

Founding concepts: A model uses the cultural background within the application area and within a community of practice. Base conceptions (scope, expressions, concept space organisation, quantification/measurement), a namespace/ontology/carrier, a number of definitions (state, intrinsic, object, interaction descriptors and depictors), and a language as cargo [10] characterise these founding concepts.

Application context: The application domain binds the model to some common understanding that is not explicitly defined in the model. Each model has an empirical scope of the model, has specific application-domain driven correspondences, and must satisfy a number of laws and regulations.

Meta-model: Each model has a basement, is restricted by paradigms and theories, has a status in the application; context, displays elements on certain abstraction level and granularity, and uses a scale. It is also prone for paradigmatic evolution within the epistemological profile of community of practice.

These dimensions are partially known in didactics (of modelling) [5]. The three background dimensions drive however the model deployment and development.

3 Language-Backed Modelling

3.1 Language Selection Matters

Languages may however also restrict modelling. This restriction may either be compensated by over-development of language components or by multi-models. The relational database modelling language uses integrity constraint as compensation component for the inadequate expressibility of the language. The Sapir-Whorf hypothesis [19] results in the following principle:

***Principle of linguistic relativity*:** Actors skilled in a language may not have a (deep) understanding of some concepts of other languages. This restriction leads to problematic or inadequate models or limits the representation of things and is not well understood.

The principle of linguistic relativity is not well understood. In [15] we demonstrated via a crossroad example that Petri nets are often not the right tool for representation of behaviour. A similar observation on UML is made by Krogstie [9].

3.2 The Cognitive Insufficiency of the Entity-Relationship Modelling Language

Lakoff introduces six basic schemata of cognitive semantics without stating that this list of schemata is complete.

- The container schema define the distinction between in and out. They have an interior, a boundary and an exterior.
- The part-whole schema define an internal structuring and uses whole, part and configuration as construction units.
- The link schema connects thing of interest. It uses various kinds of links for associating or un-associating things.
- The center-periphery schema is based on some notion of a center. Peripherical elements are not as important than those in the center.
- The source-path-goal schema uses source (or starting point), destination, path, and direction. It allows also to discuss main and side tracks.
- Typical ordering schemata are the up-down, front-back and the linear ordering schema. They use spatial and temporal associations.

We call a modelling language *cognition-complete* if these six schemata can be represented.

The classical ER modelling language suffers from a number of restrictions. It uses the container and the link schemata. It allows to mimic the part-whole schema via special links (called IsA). This work-around is however badly misunderstood. In order to become cognition-complete integrity constraints must be used. Their cognitive complexity is however beyond surveyability of humans. A typical flaw of the classical ER model is the use of monster types that integrate stabile - almost not changing - properties and transient - often changing - properties. Objects are then taken as a whole. Unary relationship types easily resolve this problem if higher-order types are permitted.

Extended ER modelling languages are however also not cognitive complete. The center-periphery schema can only be emulated. The source-path-goal schema can be represented by higher-order relationship types. The part-whole schema is supported by the specialisation via unary relationship types and by generalisation via cluster types. Ordering schemata can be defined using the order types and bulk types.

4 Syntax of Conceptual Models: Structuring and 'Functioning'

Syntax of models is build on deictic context-based rules [20] for construction of complex expressions using a domain-dependent vocabulary and governed by a set of meta-rules for construction (styles, pattern, abstraction).

4.1 Morphology of Conceptual Models and the Form of Elements

Morphology is the science of word form structure. The part of syntax is completely neglected in conceptual modelling. It is however equally important. Elements of a modelling language can be classified according to their categories and roles within a model and according to their specific expression within a model. Expression might similarly ruled by inflection, deviation, and composition. Therefore, techniques like lemmatisation (reduction of words to their base form) and characterisation by the (morpho-syntactic) role within a model.

Based on [11] we distinguish five morphological features: full or partial specification, layering within a model, integrity constraints, cyclic or acyclic structuring, complete set of schemata for cognitive semantics, open or closed context, and kind of data types.

Syntax for models is context-dependent. Constructs are bound by an implicit construction semantics [14] that is an integral component of any language. Models are governed by syntactic rules or explicit and implicit social norms. They are constructed with implicit styles and architectures.

4.2 The Lexicography and the Namespace of Models

Lexicography has developed a number of principles for coding and structuring lexical elements based on the lexicon on the application domain. Most ontological research in Computer Science does not got beyond lexicography and uses a topical annotation of model elements while hoping that every stakeholder has the same interpretation for words such as ‘name’, ‘description’, ‘identifier’ etc. If we consider however more complex entries such as ‘address’ then we detect that such kind of annotation does not work even for one language. The situation becomes far worse if we consider different languages, cultures, or application domains. Then the nightmare of “integration” becomes a challenge.

Models typically use a general and an application-dependent namespace. Moreover, the model is a product of a community of practice with its needs, its common-speak, its specific functions of words, its specific phrases and abbreviations, and its specific vocabulary.

5 Semantics of Conceptual Models

5.1 Kinds of Semantics

Semantics is the study of meaning, i.e. how meaning is constructed, interpreted, clarified, obscured, illustrated, simplified, negotiated, contradicted and paraphrased. It has been treated differently in the scientific community, e.g., in the area of knowledge bases and by database users.

- The scientific community prefers the treatment of ‘always valid’ semantics based on the mathematical logic. A constraint is valid if this is the case in any correct database.
- Database modellers often use a ‘strong’ semantics for several classes of constraints. Cardinality constraints are based on the requirement that databases exist for both cases, for the minimal and for the maximal case.
- Database mining is based on a ‘may be valid’ semantics. A constraint is considered to be a candidate for a valid formula.
- Users usually use a weak ‘in most cases valid’ semantics. They consider a constraint to be valid if this is the usual case.
- Different groups of users use an ‘epistemic’ semantics. For each of the group its set of constraints is valid in their data. Different sets of constraints can even contradict.

Semantics is currently one of the most overused notions in modern computer science literature. Its understanding spans from synonyms for structuring or synonyms for structuring on the basis of words to precise defined semantics. This partial misuse results in a mismatch of languages, in neglecting formal foundations, and in brute-force definitions of the meaning of syntactic constructions.

Semantics of models uses also commonsense, intended and acceptable meanings, various kinds of quantifications, (logical) entailment, deduction, induction, and abduction.

5.2 The Lexicology and the Namespace of Models

Ontologies are becoming very popular in Computer Science research. Philosophy developed now a rather restrained usage of ontologies. Lexicology [4] - as a part of philology - or semasiology is based on semantic relations in the vocabulary of a language. Lexicology of models studies elements of models and their meaning, relations between these elements, sub-models and the namespace in the application domain. Classical linguistic relations such as homonym, antonym, paronym, synonym, polysemy, hyponym, etc. are used for stereotyped semantics in the namespace.

Models combine two different kinds of meaning in the namespace: referential meaning establishes an interdependence between elements and the origin ('what'); functional meaning is based on the function of an element in the model ('how'). The referential meaning is well investigated and uses the triangle between element, concept and referent. The functional meaning relates elements in a model to the model context, to the application context, and to the function of this element within the model. It thus complements the referential meaning. Additionally model lexicology use the intext (within the model), the general, the part-of-model, and the differential (homonym-separating) meaning. Further, we need to handle the change of meaning for legacy models.

6 Pragmatics of Conceptual Models

6.1 General Pragmatics of Modelling in a Community of Practice

Pragmatics for modelling is the study how languages are used for intended deployment functions in dependence on the purposes and goals within a community of practice. Functions, purposes and goals are ruling the structure and function of the model. We distinguish the descriptive-explanatory and persuasive-normative functions of a model. Models are used for (1) acting (2) within a community, especially the modeller and have (3) different truth or more generally quality [8]. We may distinguish between *far-side* and *near-side* pragmatics separating the 'why' from the 'what' side of a model. General pragmatics allows to describe the overall intentions within the community and strategies for intention discovery.

We may distinguish between the development and deployment modes. The first mode starts with abstraction and mapping and then turns to the representation. The second mode inverses the first mode. Based on this distinction we infer a number of basic principles of modelling pragmatics similar to Hausser [6]: model surface compositionality (methodological principle), model presentation order's strict linearity relative to space (empirical principle), model interpretation and production analysed as cognitive processes (ontological principle), reference modelled in terms of matching an model's meaning with context (functional principle).

Models must be methodologically valid, support subjective deductive (paradeductive) inference with an open world interpretation, allow context-dependent reasoning (implicature), provide means for collaborative interaction, weaken connectives and quantifications, and integrate deductive, inductive, abductive and paradeductive reasoning.

6.2 Visualisation or the 'Phonetics' of Conceptual Models

Phonology is the science of language sounds. *Phonetics* investigates the articulatory, acoustic, and auditive process of speech. It is not traditionally considered not considered to be a part of a grammar. But it is equally important for practical language deployment. Phonology of models is concerned with the ways in which intentions can be conveyed using conventional and non-conventional resources. The modeller uses reference for transferring the specific point of view that has been used for modelling.

Visualisation is the 'phonetics' in modelling. It is based on three principles:

Principles of visual communication are based on three constituents: *Vision*, *cognition*, and *processing and memorizing characteristics.* We may use specific visual features such as contrast, visual analogies, presentation dramaturgy, reading direction, visual closeness, symmetric presentation and space and movement.

Principles of visual cognition refer to *ordering*, *effect* delivery, and *visualisation*. We base those on *model organisation*, *model economy*, *skills of users*, and *standards*.

Principles of visual design are based on *optical vicinity* , *similarity* , *closeness*, *symmetry*, *conciseness*, *reading direction*.

These principles help to organise the model in a way that correspond to human perception.

7 Summary

Models are artifacts that can be specified within a (W^4+$W^{17}H$)-frame that is based on the classical rhetorical frame introduced by Hermagoras of Temnos[1]. Models are primarily characterised by W^4: wherefore (purpose), whereof (origin), wherewith (carrier, e.g., language), and worthiness ((surplus) value). Secondary characterisation $W^{17}H$ is given by:

- user or stakeholder characteristics: by whom, to whom, whichever;
- characteristics imposed by the application domain: wherein, where, for what, wherefrom, whence, what;
- purpose characteristics characterising the solution: how, why, whereto, when, for which reason; and
- additional context characteristics: whereat, whereabout, whither, when.

Modelling is the art, the systematics and the technology of model (re)development and model application. It uses model activities and techniques. This paper is going to be extended by more specific aspects of the modelling art in the context of semiotics and linguistics.

[1] The rhetor Hermagoras of Temnos, as quoted in pseudo-Augustine's De Rhetorica defined seven "circumstances" as the loci of an issue: Quis, quid, quando, ubi, cur, quem ad modum, quibus adminiculis(W^7: Who, what, when, where, why, in what way, by what means). See also Cicero, Thomas Aquinas, and Qunitillian's loci argumentorum as a frame without quesrioning. The Zachman frame uses an over-simplification of this frame.

References

1. Agassi, J.: Why there is no theory of models. In: Niiniluoto, I., Herfel, W.E., Krajewsky, W., Wojcicki, R. (eds.) Theories and Models in Scientific Processes, Amsterdam, Atlanta, pp. 17–26 (1995)
2. Dahanayake, A., Thalheim, B.: Towards a Framework for Emergent Modeling. In: Trujillo, J., Dobbie, G., Kangassalo, H., Hartmann, S., Kirchberg, M., Rossi, M., Reinhartz-Berger, I., Zimányi, E., Frasincar, F. (eds.) ER 2010. LNCS, vol. 6413, pp. 128–137. Springer, Heidelberg (2010)
3. Dahanayake, A., Thalheim, B.: Enriching Conceptual Modelling Practices through Design Science. In: Halpin, T., Nurcan, S., Krogstie, J., Soffer, P., Proper, E., Schmidt, R., Bider, I. (eds.) BPMDS 2011 and EMMSAD 2011. LNBIP, vol. 81, pp. 497–510. Springer, Heidelberg (2011)
4. Ginsburg, R.S., Khidekei, S.S., Knyazeva, G.Y., Sankin, A.A.: A course in modern English lexicology, 2nd edn. Vysschaja Schkola, Moscov (1979) (in Russian)
5. Halloun, I.A.: Modeling Theory in Science Education. Springer, Berlin (2006)
6. Hausser, R.: Foundations of computational linguistics - human-computer communication in natural language, 2nd edn. Springer (2001)
7. Kaschek, R.: Konzeptionelle Modellierung. PhD thesis, University Klagenfurt, Habilitationsschrift (2003)
8. Korta, K., Perry, J.: Critical Pragmatics. Cambridge University Press, Cambridge (2011)
9. Krogstie, J.: Quality of uml. In: Encyclopedia of Information Science and Technology (IV), pp. 2387–2391 (2005)
10. Mahr, B.: Information science and the logic of models. Software and System Modeling 8(3), 365–383 (2009)
11. Ritchey, T.: Outline for a morphology of modelling methods - Contribution to a general theory of modelling
12. Stachowiak, H.: Modell. In: Seiffert, H., Radnitzky, G. (eds.) Handlexikon zur Wissenschaftstheorie, pp. 219–222. Deutscher Taschenbuch Verlag GmbH & Co. KG, München (1992)
13. Steinmüller, W.: Informationstechnologie und Gesellschaft: Einführung in die Angewandte Informatik. Wissenschaftliche Buchgesellschaft, Darmstadt (1993)
14. Thalheim, B.: Entity-relationship modeling – Foundations of database technology. Springer, Berlin (2000)
15. Thalheim, B.: Towards a theory of conceptual modelling. Journal of Universal Computer Science 16(20), 3102–3137 (2010), `http://www.jucs.org/jucs_16_20/towards_a_theory_of`
16. Thalheim, B.: The art of conceptual modelling. In: Proc. EJC 2011, Tallinn, pp. 203–222 (2011)
17. Thalheim, B.: The Science of Conceptual Modelling. In: Hameurlain, A., Liddle, S.W., Schewe, K.-D., Zhou, X. (eds.) DEXA 2011, Part I. LNCS, vol. 6860, pp. 12–26. Springer, Heidelberg (2011)
18. Thalheim, B.: The theory of conceptual models, the theory of conceptual modelling and foundations of conceptual modelling. In: The Handbook of Conceptual Modeling: Its Usage and Its Challenges, ch.12, pp. 543–578. Springer, Berlin (2011)
19. Whorf, B.L.: Lost generation theories of mind, language, and religion. Popular Culture Association, University Microfilms International, Ann Arbor, Mich. (1980)
20. Wojcicki, R.: Wyklady z logiki z elementami teorii wiedzy. Wyd. Naukowe SCHOLAR (2003) (in polish)