

# An Exploration in the Space of Mathematical Knowledge

Andrea Kohlhasse<sup>1</sup> and Michael Kohlhasse<sup>2</sup>

<sup>1</sup> DiMeB, Dept. of Computer Science and Mathematics, University Bremen,  
kohlhasse@informatik.uni-bremen.de

<sup>2</sup> School of Engineering and Science, International University Bremen,  
m.kohlhasse@iu-bremen.de

**Abstract.** Although knowledge is a central topic for MKM there is little explicit discussion on what ‘knowledge’ might actually be. There are specific intuitions about form and content of knowledge, about its structure, and epistemological nature that shape the MKM systems, but a conceptual model is missing.

In this paper we try to rationalize this discussion to give MKM a firmer footing, to start a discussion among MKM researchers and help relate the MKM intuitions and discourses to other communities.

Starting from the observation that many concrete realizations of mathematical knowledge objects are considered equivalent, we propose a conceptual model of the space of (mathematical) knowledge objects graded by levels of abstraction and presentational explicitness and draw conclusions for MKM markup formats.

## 1 Handles on (Mathematical) Knowledge

The concept of ‘knowledge’ is investigated by many scientific disciplines, some take a microscopic, ontological view, some a macroscopic, epistemological view and still others a pragmatic view. The latter seems to be the dominant one in the field of Mathematical Knowledge Management (MKM), but ever so often we find pragmatic limits and have to cross the border. There are multiple ways of looking at mathematical knowledge; for instance there is much discussion about whether we should focus on the essence or the visual appearance of mathematical objects and where to determine the borders between these as they seem to be fluctuating.

In this paper, we start an exploration into the world of mathematical knowledge. Reflections on this mathematical space were inspired in part by an article by SEYMOUR PAPERT, called “An Exploration in the Space of Mathematics Educations” [Pap96]. There, he investigates different math educational approaches, but instead of contrasting them he relates them by interpreting them as axes in an n-dimensional space. Here, we investigate essence/appearance approaches concerning knowledge objects and are interested in the resulting *knowledge space*, hoping that this perspective yields new and unexpected dependencies and relations.

## 1.1 Knowledge and Context

Information theory assumes that the fundamental concepts of data, information, and knowledge are not interchangeable concepts. In particular, the transitive combination of “Lots of available data” and “Information are good data” and “Knowledge is created with information” readily accepted in the Internet Bubble cannot be held.

As data are visually accessible, we need to consider yet another concept: a ‘glyph’ is an arrangement of pixels on a screen (or dots of ink on a sheet of paper) into a recognizable shape. In contrast to the usage of data (which contain something even if we don’t know what), the usage of ‘glyphs’ emphasizes the pure presentation of a single character without any underlying semantics. In order to close in on ‘knowledge’, we want to take a closer look at the meaning of glyphs, data, and information and their relationships and differences based on an established knowledge management model. PROBST ET AL. (see [PRR97]) posit that glyphs, data, information, and last but not least knowledge can be seen as stages of a pipeline that is shown in Figure 1.

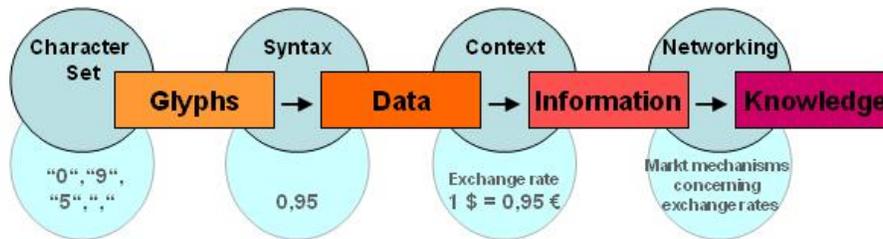


Fig. 1. From Mere Glyphs To Valuable Knowledge

In particular, **glyphs** are just a set of characters or symbols like  $\{0;9;5;\}$  without any structure. A first set of rules imposed on the glyphs — the syntax — yields **data** which can be handled by machines. For obtaining meaning from such data we still need another component: the context. Usually, we discern data from **information** by viewing information as data with a message or data with an intention. DAVENPORT and PRUSAK think of information “as data that makes a difference” [DP98]. Data becomes information if a user can interpret the data in regard to a specific goal (or a local context) e.g. using the string ‘0,95’ as number in an equation concerning exchange rates in our example. In contrast, information becomes **knowledge**, if a user can interpret the information in regard to a global context like understanding the exchange rate equation in the area of specific market behavior with respect to change of exchange rates.

## 1.2 Communities of Practice as Knowledge Context

We described in Section 1.1 how knowledge in the field of Knowledge Management is dealt with. In 1991, BROWN and DUGUID investigated more closely the

global context which transforms information into knowledge (see “The Social Life of Information” [BD00]). In [BD91] they identify LAVE and WENGER’s influential concept of “Community of Practice” [LW91, p. 98] as the *social life of information*, i.e. they link Communities of Practice with organizational learning and hence with knowledge. A **Community of Practice (CoP)** is

“a group of people (e.g. professionals) informally bound to one another through exposure to a common class of problems, common pursuit of solutions, and thereby themselves embodying a store of knowledge.” [WMS02]

In mathematical terms, scientific groups can only build a CoP if their members agree on the validity of certain equivalence relations (which we will call “substance equivalences” in 2.1).

CoPs are considered as the locus of knowledge as opposed to the learner’s mind. The process of obtaining knowledge (learning) is described as “a process that takes place in a participation framework” [LW91, p. 14] where “participation is always based on situated negotiation and renegotiation of meaning in the world” [LW91, p. 51]. So far, the assignment of meaning is done in MKM via semantic annotations, but the necessary agreement on CoP-dependent substance equivalences are not yet paid attention to, even though they seem to play a decisive role in the Mathematical Knowledge Space.

### 1.3 The Pragmatics of Mathematical Knowledge

In order to make knowledge amenable to management, it has to be ‘captured’. More specifically, it has to be reified, so that it can be stored, transferred, or retrieved as **knowledge object**. But even if we set aside for the moment the problem of explicit and implicit (tacit) knowledge that is well-studied in learning theories, we have to look at the relationship between a knowledge object and the represented knowledge itself. Therefore, we need to look at the question what knowledge is made of and whether or what part of it ‘exists’. We focus on the philosophy of Mathematics that is concerned with this question and find that it has occupied many famous people like CARNAP, BERNAYS, BENACERRAF, or PUTNAM (see for example “The existence of mathematical objects” in [BP64, pp. 241-311]). An important distinction in this consideration is the one of **substance** and **accidence**<sup>3</sup>. Substance is the unchanging essence of an object, whereas accidence is the object’s appearance. These terms form a *dialectic pair*: even though an object’s substance and appearance can be differentiated, they are inseparable, they form a unity, so that one cannot think of one without the other. The question whether a knowledge object (especially a mathematical object) exists in “being’ or ‘thinking’ ” [Isr79, 7] is mostly irrelevant to mathematicians as long as it can be described. Its answer depends on a person’s underlying ontological belief (see for example BENACERRAF’s essay “What numbers could not be” [BP64, pp. 272-295]). But in real life, mathematicians are pragmatists, they use abstract

<sup>3</sup> This distinction was used by Kant, there are many similar ones, including: essence/appearance (Hegel), matter/form (Aristoteles), content/form (MKM).

objects independent from their existence. Analogously, epistemological issues are pragmatically ignored by (most) MKM systems.

However, the differentiation between content and form found its way into the general MKM discourse. It is consensus in MKM that for a mathematical object we can distinguish its form from its content and express both aspects in markup systems. For instance, the MATHML format [ABC<sup>+</sup>03] has two sub-languages: *presentation*-MATHML describes the two-dimensional layout in an expression like  $\sqrt[3]{x+2}$ , and *content*-MATHML, which can express its functional structure as the application of the cubic root function to a sum with the variable  $x$  and the number 2.

In general, it could be argued, that it makes no difference whether we take the symbol ‘ $\mathbb{R}$ ’ for the real numbers or maybe simply ‘ $R$ ’. We could just call it “different notation”. But do you really believe that your personal selection of symbols is a matter of accident? Especially mathematicians do take great care in this selection out of coherence and consistence reasons [Hei00], but also because they know that different presentations and conceptualizations do have different associations and they make pragmatic use of it. In philosophic terms, we might call this the dialectic character of the substance/accidence aspects of a knowledge object. In many cases, the choice of conceptualization and presentation can make the difference whether a problem is solvable at all; see e.g. [Rob91] for a collection of striking examples.

In the following, if we use the pair **substance/accidence** we want to stress the different perspectives one can take looking at objects. This view is concerned with the relevance and the timeliness of the respective objects. In contrast, if we look at concrete objects, i.e. manifestations of knowledge, we can speak of their **content** and **form**. Here, we can think of content and form as the object’s constitutive elements. They give rise to a knowledge space spanned by substance and accidence, inhabited by knowledge objects with certain “content and form coordinates”.

## 2 A Conceptual Model for Knowledge Spaces

The fundamental observation is that knowledge can only be observed or communicated, if it is in a concrete form, e.g. written down in a book or uttered by a colleague or teacher. For this realization — which we can consider as knowledge *object* — a lot of conceptual and presentational aspects have to be fixed. Some seem to contribute to the meaning of the object, while others are thought of as rather personal choices like the page size of the book that contains the knowledge. In this section, we have a closer look at what the mathematical community deems substantial, yielding substance of knowledge as the totality of traits (which can be modeled as equivalence classes) that constitute the meaning.

We will use the following group definitions as a running example in this paper. It is well-known that groups can alternatively be described in two ways:

**Definition 1** [KM79] A **group<sub>1</sub>** is a set  $G$  together with an associative binary operation  $\circ: G \times G \rightarrow G$ , such that there is a unit element  $e$  for  $\circ$  in  $G$ , and all elements have inverses.

**Definition 2** [Hal59] A **group<sub>2</sub>** is a set  $G$ , together with a (not necessarily associative) binary operation  $/: G \times G \rightarrow G$ , such that  $a/a = b/b$ ,  $a/(b/b) = a$ ,  $(a/a)/(b/c) = c/b$ , and  $(a/c)/(b/c) = a/b$  for all  $a, b, c \in G$ .

For any group<sub>1</sub>  $(G, \circ)$ , we can define a binary operation  $/_{\circ}$  by  $a/_{\circ}b := a \circ b^{-1}$  that shows that  $(G, /_{\circ})$  is a group<sub>2</sub>, and vice versa (using  $a \circ_j b := a/_j b^{-1}$  with  $b_j^{-1} := ((b/b)/b)$ ). So we see that the two definitions are isomorphic (which we could capture as a structure  $\mathcal{G} := (G, \circ, e, \cdot^{-1}, /_{\circ}) = (G, \circ_j, a/a, \cdot_j^{-1}, /)$ ; see [CS98] for a formal account). In Mathematics it is usual to represent a structure like a group simply as the pair  $\mathcal{G}_1 := (G, \circ)$  or a pair  $\mathcal{G}_2 := (G, /)$ , since in a group<sub>1</sub> the unit  $e$  and the inverse operation  $\cdot^{-1}$  are uniquely determined by  $G$  and  $\circ$  (and similarly for a group<sub>2</sub>). So, we can view  $\mathcal{G}$  as the substance of group and  $\mathcal{G}_1$  and  $\mathcal{G}_2$  as its accidents.

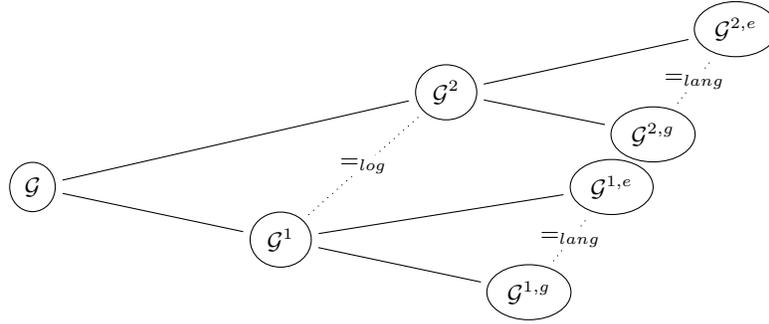
Mathematicians frequently speak of  $\mathcal{G}_1$  and  $\mathcal{G}_2$  as different *representations*. Note, that there often is a mix-up between the terms ‘presentation’ and ‘representation’. Principally, ‘**presentation**’ is used to describe an explicit realization whereas ‘*representation*’ is used to describe an implicit formalization<sup>4</sup>. In the example, instead of the usage of the symbols  $\{(G, \circ); e; \cdot^{-1}\}$  just as well the symbols  $\{(S, +); 1; -\}$  respectively could have been used in the presentation. In order to avoid confusions and for the purposes of the discussion in this paper we prefer to phrase these *representations* as “**conceptualizations**” to mark them off their “presentations”. The term ‘**representation**’ is therefore freed and serves as superordinated expression for conceptualizations as well as presentations.

## 2.1 Substance Equivalence

In Mathematics and in the natural sciences it is customary to consider presentational aspects like the (natural) language to be irrelevant for the meaning of a mathematical text. In this view, any document can be translated to any natural language without loss of meaning. As we have seen in the groups example above, Mathematics knows an even stronger equality notion — isomorphism. These distinct notions of equality of representations give rise to equivalence relations like  $=_{lang}$  or  $=_{log}$  which we will call **substance equivalences**.

In particular, we can consider the relationship model of  $\mathcal{G}$  as knowledge reification along equivalence relations (Figure 2). The oval nodes are knowledge objects, whereas the various edges in the triangular graph signify the relations

<sup>4</sup> In German, this connection is exemplified in the language itself: presentation translates to ‘Darstellung’ whereas *representation* translates to ‘Darstellungsweise’, i.e. the mode of presentation. This corresponds to the frequent usage of the term ‘presentation’ in combination with the preposition ‘for’ in contrast to ‘*representation*’-usage concentrating on the ‘of’-object. In other words, presentation is targeted with respect to the potential audience, whereas *representation* is focused on the content and its structure.



**Fig. 2.** Knowledge Reification in Mathematics

between the objects. Starting the description in a bottom-up way from right to left, the nodes  $\mathcal{G}^{i,*}$  where  $* \in \{e, g\}$  and  $1 \leq i \leq 2$  stand for concrete variants of Definition group<sub>*i*</sub> in English (*e*) and German (*g*). As Mathematics considers translations between natural languages to be meaning-preserving, these are considered to be “conceptually equal” (see the dotted line between them). This allows us to obtain the knowledge objects  $\mathcal{G}^i$  for the two conceptualizations of groups discussed above, which we model as the equivalence classes of  $\mathcal{G}^{i,*}$ :  $\mathcal{G}^i := [\mathcal{G}^{i,e}]_{=lang} = [\mathcal{G}^{i,g}]_{=lang}$ . These are logically equivalent, so they give rise to another knowledge object given as the equivalence class of all logically equivalent conceptualizations of groups, which we denote with  $\mathcal{G} := [\mathcal{G}^1]_{=log}$ <sup>5</sup> (that corresponds to the structure  $\mathcal{G}$  above).

We can see the diagram in Figure 2 as a visualization of the passage from implicitly represented knowledge objects on the left to explicitly presented ones on the right, making representation choices along the way.

## 2.2 Substance Equivalences and MKMarkup Formats

The substance equivalences we discussed above are generally accepted in Mathematics. If we look at other disciplines, e.g. in the Arts or Humanities, the assumption that representations can be translated without loss of substance would be highly controversial; there are literary texts (e.g. poems) that are considered “un-translatable”. Hence, the diagram in Figure 2 would look completely different for e.g. literary science<sup>6</sup>. As Communities of Practice are marked by their collective value judgments about knowledge, we argue that substance equivalences are the defining characteristics of CoPs. Moreover, we have to assume that at least some are (implicitly) inscribed into the representation formats used by a CoP. But exactly how are the mathematical substance equivalences inscribed

<sup>5</sup> Note, that the equivalence class construction of  $\mathcal{G}$  is independent of the order of intermediate layers, i.e. the equivalence relations commute and therefore it is done modulo the transitive closure of  $=log \cup =lang$ .

<sup>6</sup> We suspect that it would be a left-right-inverted (dual) version of the triangle for Mathematics, but leave the investigation of this to further research.

into the MKM formats? We differentiate between the well-known formula, statement, and theory level of mathematical knowledge objects and give examples in several MKMarkup formats.

We will start our analysis with the simplest case: the *formula level* and have a look at various MKM formats. In  $\text{T}_{\text{E}}\text{X}/\text{L}^{\text{A}}\text{T}_{\text{E}}\text{X}$ , we can specify the exact sizes, colors, or fonts of the glyphs that make up the two-dimensional layout of a mathematical formula. Since we are given handles how to specify all these, we have to assume that these parameters matter, and therefore that the format does not inscribe equivalence of formulae where the glyphs differ in size, color, or font. In presentation  $\text{MATHML}$ , the specification of these traits is not possible in the prime vocabulary, but relegated to a CSS style system (which allows the specifications to be overridden in the client by standard means) which we take as a hint that stronger substance equivalences are in effect than in  $\text{T}_{\text{E}}\text{X}/\text{L}^{\text{A}}\text{T}_{\text{E}}\text{X}$ .  $\text{OPENMATH}$  [BCC<sup>+</sup>04] is of course the most radical in the substance equalities it assumes. It is impossible to specify the presentation of an  $\text{OPENMATH}$  object, as this format is geared towards communication of mathematical objects between systems. Communication with humans will be done via  $\text{OPENMATH}$  editors and presentation systems; which are free to choose any presentation suitable. Obviously, any two presentational variants e.g.  $\binom{n}{k}$ ,  ${}_nC^k$ ,  $C_k^n$ , and  $C_n^k$  are substance equivalent, since they all mean the same:  $\frac{n!}{k!(n-k)!}$  (see [Koh05b] for a discussion).

At the *level of mathematical statements*, where e.g. our groups example is located, things are more complicated. We have already seen that, here, issues like the (natural) language employed in a definition, or the conceptualization play a major role. This leads us to another way, in which substance equivalences can be inscribed into MKMarkup formats. For instance, our own  $\text{OMDOC}$  format [Koh05a] has an explicit concept of language variants e.g. in the **definition** element (which represents a definition such as the one for  $\text{group}_1$ ): it can incorporate a multilingual collection of **CMP** elements that contain definitional text fragments that are explicitly considered language variants of each other. So, the substance equivalence  $=_{lang}$  from our example in Figure 2 is inscribed into  $\text{OMDOC}$ . We can see that the substance equivalence  $=_{log}$  is inscribed into  $\text{OMDOC}$  as well. It is provided by the **alternative** element, which in our example would allow to phrase the definition of  $\text{group}_2$  as an alternative definition to  $\text{group}_1$  as long as we have proofs for the equivalence. In this situation,  $\text{OMDOC}$  only provides one concept for a group, a clear sign that  $=_{log}$  is assumed in  $\text{OMDOC}$ .

At the *level of theories*,  $\text{OMDOC}$  has still another way of inscribing substance equalities into the format as it supports *theory morphisms*, i.e. structures that allow to prove that one theory is included in (or even isomorphic to) another modulo a variety of translations. In particular, isomorphic theories are considered as logically interchangeable (even if they are pragmatically different), another materialization of  $=_{log}$ .

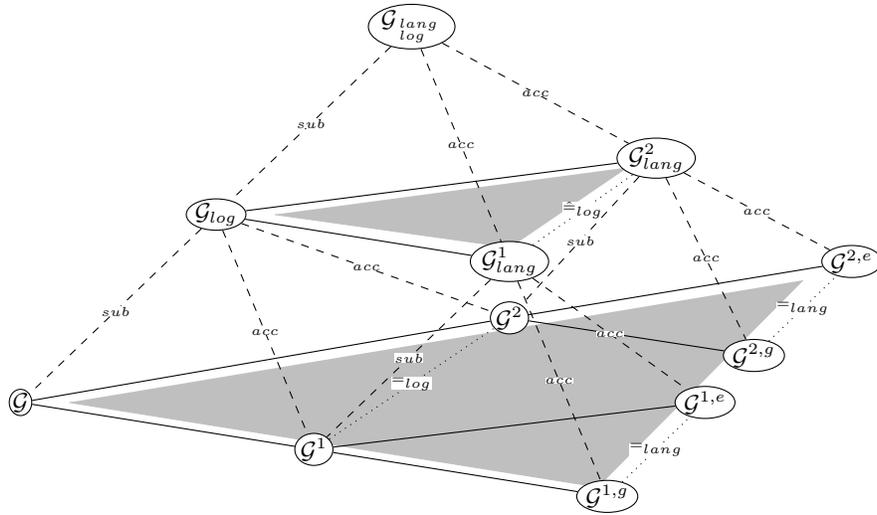
### 2.3 The Conceptual Model of MKS

We will now take a look at how the reified knowledge (text fragments marked up in an MKM format) fit into a conceptual model of the **Mathematical Knowl-**

**edge Space (MKS).** We develop the intuition for MKS by constructing the MKS for groups. Its generalization we leave to the gentle reader.

It is a central observation, that — even though we may actually *want* to write down an abstract object like  $\mathcal{G}^1$  or even  $\mathcal{G}$  in Figure 2 — we only *can* write down a leaf. Given the discussion in the last section, we have to assume that when we express mathematical knowledge in an MKM format, we actually write down a **markup pair** consisting of a concrete realization and the assumed substance equivalence relation. For instance, to formulate the conceptualization group<sub>1</sub>, we can either type the markup pair  $\langle \mathcal{G}^{1,e}, =_{lang} \rangle$  or  $\langle \mathcal{G}^{1,g}, =_{lang} \rangle$ . Note, that these markup pairs contain enough information to reconstruct  $\mathcal{G}^1$  as  $[\mathcal{G}^{1,e}]_{=_{lang}}$  or  $[\mathcal{G}^{1,g}]_{=_{lang}}$ . We can consider  $\mathcal{G}^1$  as their substance and  $\mathcal{G}^{1,*}$  as their accidents.

We can lift  $=_{lang}$  to an **markup equivalence relation**  $\hat{=}_{lang}$  by setting  $\langle x, R \rangle \hat{=}_{lang} \langle y, R \rangle$  iff  $x =_{lang} y$ . As  $\mathcal{G}^{1,g} =_{lang} \mathcal{G}^{1,e}$  the pairs  $\langle \mathcal{G}^{1,e}, =_{lang} \rangle$  and  $\langle \mathcal{G}^{1,g}, =_{lang} \rangle$  are  $\hat{=}_{lang}$ -equivalent, giving rise to an equivalence class  $\mathcal{G}^1_{lang}$ , which we consider to be the “*language-independent markup object for a group*”. For example,  $\mathcal{G}^1_{lang}$  is naturally represented by the multilingual **definition** element in OMDOC.



**Fig. 3.** The Mathematical Knowledge Space for Groups

In Figure 3 we picture the substance and accidence relations **sub** and **acc** resp. with dashed lines, interpreting the triangle from Figure 2 as the base face of a tetrahedral graph and positioning  $\mathcal{G}^1_{lang}$  in the first level above it. An analogous construction yields  $\mathcal{G}^2_{lang}$ . This gives us license to construct a knowledge object  $\mathcal{G}_{log}$  as the equivalence class of the  $\mathcal{G}^i_{lang}$  modulo  $\hat{=}_{log}$ , on the first level just as we did for the lower level in Section 2.1. Note, that

$$\mathcal{G}_{log} := [\mathcal{G}^1_{lang}]_{\hat{=}_{log}} = \left[ \left[ \langle \mathcal{G}^{1,e}, =_{lang} \rangle \right]_{\hat{=}_{lang}} \right]_{\hat{=}_{log}} \equiv \left[ \left[ [\mathcal{G}^{1,e}]_{\hat{=}_{lang}}, =_{lang} \right] \right]_{\hat{=}_{log}}$$

The right-hand side of this is again an equivalence class of markup pairs, so we can consider  $\mathcal{G}$  as the substance of  $\mathcal{G}_{log}$  and the  $\mathcal{G}^i$  as its accidences, making  $\mathcal{G}_{log}$  the “*conceptualization-independent conceptualizations of group*”. In particular, all the relations in the lower part of Figure 3 commute. Note, that just as in the lower level, the objects become more explicit from left to right. Finally, we complete the picture by iterating the construction to obtain a knowledge object  $\mathcal{G}_{log}^{lang}$  for the group that is “*independent of everything ‘relevant’*”, where the relevance is determined by the knowledge object’s author’s CoP, in our example the conceptualizations with respect to *log* and *lang*.

In particular, *we obtain a knowledge object that no longer contains anything that the given Community of Practice deems substance-irrelevant.*

### 3 Interpretations of MKS

In Section 2.1 we have presented a model of the reification of knowledge based on substance equivalences. We can interpret Figure 2 — i.e. the *base face of the MKS tetrahedron* — as the perspective of an *author* who writes down her knowledge with an audience in mind. Naturally, her membership in a Community of Practice (see 2.2) determines the employed *implicit* substance equivalences.

In 2.3, we completed this picture by extending the analysis with an account of markup processes resp. *markup formats*, yielding the mathematical knowledge space in Figure 3. Here, we can interpret the *right face of the MKS tetrahedron* (i.e. the triangle area between  $\mathcal{G}_{log}^{lang}$ ,  $\mathcal{G}^{1,g}$ , and  $\mathcal{G}^{2,e}$ ) as the markup process, starting out with concrete materialization of knowledge, ending with a knowledge object in a markup format with *explicit* or *inscribed* substance equivalences.

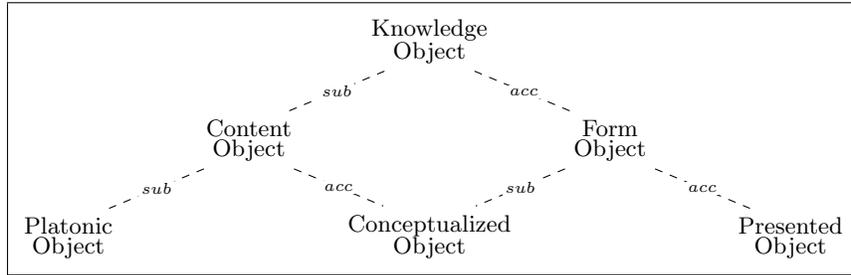
#### 3.1 MKS and the Content/Form Distinction

Let us now consider the *front face of the MKS tetrahedron* (i.e. the triangle area between  $\mathcal{G}_{log}^{lang}$ ,  $\mathcal{G}$ , and  $\mathcal{G}^{1,g}$ )<sup>7</sup>. Starting at the top with  $\mathcal{G}_{log}^{lang}$  which we call the **Knowledge Object**, we can distinguish its content from its form arriving at what we call the “**Form Object**” and the “**Content Object**” — which can be recurrently subjected to the same analysis (see Figure 4 for the resulting view of the front face of the MKS). With the substance perspective on the Content Object we arrive at what we call the “**Platonic Object**”<sup>8</sup>. Successively looking down the substance branch of the tree, we arrive at more and more fundamental, abstract objects. In particular, these are increasingly liberated from their conceptualization as well as presentation. In contrast, looking down the accidence

<sup>7</sup> Note, that the front face of the MKS tetrahedron is the only surface conceptually left as the back face’s interpretation is analogous as it is just a variant.

<sup>8</sup> The existence of such an object is not discussed, since either ontological assumption has no consequences for the conceptual model. As soon as we start reifying implicit knowledge (independent from the underlying ontology) we have to choose a form which in turn materializes the object.

branch we arrive at more and more concrete and tangible objects. In detail, the accidense view on the Content Object leads to its conceptualization level (the “**Conceptualized Object**”), where we have a representation of the content in which certain decisions of how to think about it have been taken (e.g. do we want to think about a group as an object, where the associative operation  $\circ$  is the primary concept or not?).



**Fig. 4.** The Analysis Triangle of a Knowledge Object

Now, let us look at the accidense aspect of the Form Object. As it becomes more and more concrete, we are led to a presentation level and therefore to the concrete “**Presented Object**”. The substance perspective on the Form Object reveals again a conceptualization level, which by our analysis above is the Conceptualized Object. Let us clarify this with the group example: if we want to talk about what ‘the group’ really is (i.e. the Platonic Object) we have to decide on a representation (otherwise communication is impossible). This selection determines which of the above definitions will be applied. In other words, the choice of the definition fixes the conceptualization of a group. The MKM community seems to concentrate on conceptualizations as *semantic* representations (accidense of the content = substance of the form).

Interestingly, so far capturing knowledge has always aimed at those knowledge objects that are “independent of everything” and *not* at the Platonic Objects themselves (possibly because we mistook them for the same).

### 3.2 MKS and the MKM User

Now we want to look at the MKS from the perspective of the recipient of knowledge, i.e. *the user* or learner who starts with the concrete materialization of knowledge like a certain document. The user heads for the knowledge itself — the Platonic Object — which is a Knowledge Object’s author’s point of departure. A reader has to differentiate between the potential content and the concrete form of a document. Depending on her personal choice what content and what form is, she understands and builds up her own knowledge. In contrast to the sender of knowledge, who knows the used equivalence relations (and more) and actively chooses the representation of content, the recipient of knowledge has to infer the applicable equivalence relations.

We claim that the user perspective is already present in the analysis triangle that we have studied in the last section: let us look at a student confronted with a book. It contains the knowledge in its final presented representation (*Presented Object*), but the student is aiming at an understanding of the underlying substance (*Platonic Object*). In order to decide what the content or the form is in the Presented Object, the student has to envisage a *Knowledge Object*, i.e. a potential model of the real knowledge to be learned. From this hypothetical Knowledge Object she can infer the *Content Object* and the *Form Object*. This dramatically reduces the search space of possible interpretations of the Presented Object to the presentations of the Form Object. Here, “understanding” means that the student is able to distinguish between the content of the Form Object (*Conceptualized Object*) and the Presented Object as its form.

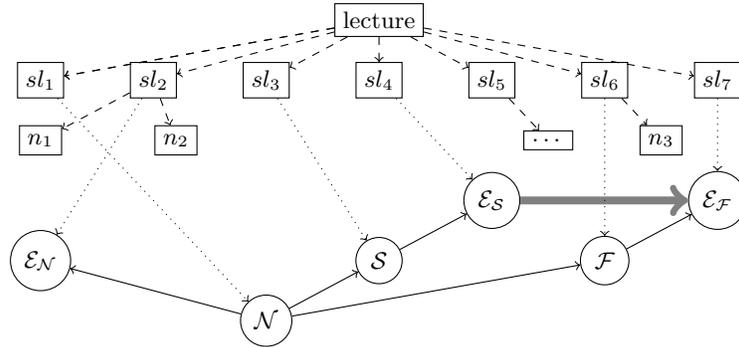
Again, interestingly, the user generally is thought of as either modeling the Platonic Object (e.g. in case of a lecture) or the Knowledge Object (e.g. in case of an MKM system), whereas we conjecture that the user is building a Conceptualized Object as approximation of the Platonic Object. Taking this seriously might help to understand how MKM systems need to be positioned in a learning cycle.

### 3.3 MKS and Narratives

In the discussion of knowledge/document markup formats on the level of theories, it is always difficult to decide what to mark up; the underlying knowledge or the structure of the document that conveys it. Note, that the underlying structures depend on the choice of conceptualization and therefore can be quite different.

Take for instance a didactically enhanced document that introduces a new concept by first presenting a naive, reduced approximation  $\mathcal{N}$  of the real theory  $\mathcal{F}$ , only to show an example  $\mathcal{E}_{\mathcal{N}}$  of where this is insufficient. Then the document proposes a first (straw-man) solution  $\mathcal{S}$ , and shows an example  $\mathcal{E}_{\mathcal{S}}$  of why this does not work. Based on the information gleaned from this failed attempt, the document builds the eventual version  $\mathcal{F}$  of the concept and demonstrates that this works on  $\mathcal{E}_{\mathcal{F}}$ . Let us visualize the narrative- and content structure in Figure 5. The structure with the solid lines and boxes at the bottom of the diagram represents the content structure, where the boxes  $\mathcal{N}$ ,  $\mathcal{E}_{\mathcal{N}}$ ,  $\mathcal{S}$ ,  $\mathcal{E}_{\mathcal{S}}$ ,  $\mathcal{F}$ , and  $\mathcal{E}_{\mathcal{F}}$  signify theories for the content of the respective concepts and examples. The arrows mark the conceptual dependency structure, e.g. theory  $\mathcal{F}$  impacts theory  $\mathcal{N}$ . The top part of the diagram with the dashed lines stands for the narrative structure, where the arrows mark up the document structure. For instance, the slides  $sl_i$  are grouped into a lecture. The dashed lines between the two documents are pointers into the content structure. In the example in Figure 5, the second slide of “lecture” presents the first example: the text fragment  $n_1$  links the content  $\mathcal{E}_{\mathcal{N}}$ , which is referenced from the content structure to slide 1. The fragment  $n_2$  might say something like “this did not work in the current situation, so we have to extend the conceptualization...”.

If we look carefully, we can see that the lower level of the diagram represents the content of the knowledge (structured by the inherent semantic relations of



**Fig. 5.** An Introduction of a Concept via Straw-Man

the objects involved), and the upper part the form (structured, so that humans are motivated to concern themselves with the material, understand why some definitions are stated in just this way, and get the new information in easily digestible portions). For instance, the OMDOC format [Koh05a] contains theory-level content- and presentation markup infrastructure for these aspects. The *theory-level content markup* contains the constitutive representations structured by OMDOC theories and their semantic relations (e.g. inheritance), and the *narrative markup* contains the document structure (e.g. that of a course divided into lectures and further into slides), motivating narrative, course-specific information (“When is the final exam?”), etc.

Just as for content-based systems on the formula level, there are now MKM systems that generate presentation markup from content markup, based on general presentation principles, also on this level. For instance, the ACTIVEMATH system [SBF<sup>+</sup>00] generates a simple narrative structure (the presentation; called a personalized book) from the underlying content structure (given in OMDOC) and a user model. However, a systematic analysis as we have attempted for the formula and statement levels above yielding the MKS is still missing. We do not even have a good understanding what the substance equivalences (and consequently the markup primitives) at the theory level might be. We conjecture that a thorough understanding of the substance/accidence aspects of the theory level (and a theory-level MKS) could eventually lead to a new generation of MKM systems, that can dynamically play with the content/form distinction to the benefit of the individual user.

## 4 Consequences for Mathematical Knowledge Management

Let us now speculate about the consequences of the suggested conceptual model for the field of mathematical knowledge management and the knowledge representation formalisms employed there.

The first consequence is that we have to extend the MKM representation formats with an explicit representation of substance equivalence relations. First steps into this direction have already been taken in [MKB04,KBHL<sup>+</sup>03], where a mathematical knowledge base was extended by a concept of “*variants*” to model language, verbosity, formalism, and partially also versioning variants. Our model here suggests that this approach needs to be systematized and explicit representations for ‘higher-level’ objects need to be introduced. Furthermore, we have to take stock of the various (sub)-relations of the substance equivalences. In [KA03], we have studied how (rather low-level, technical) substance equivalences interact with distribution and versioning of mathematical knowledge and documents. We will have to extend this to the semantic substance equivalences discussed here; [Hut04] goes first steps into this direction.

Another obvious consequence is that we will have to model *Communities of Practice* together with the mathematical knowledge in order to make the CoP-dependencies explicit. However, it is not directly obvious how to model CoPs and their relations to each other. There are large CoPs, like the CoP shared by all the **STEM** fields<sup>9</sup>, and small ones whose members agree on particular mathematical objects and differ on others. For instance, it is a matter of CoP in Mathematics whether you accept the law of excluded middle or the axiom of choice. Such CoP differences can already be modeled in MKM formalisms that have a notion of theories that are ordered by inclusion or inheritance. But inclusion of accepted theories is not the only relation between CoPs. For instance, there are two CoPs in theoretical physics, one standardizing the Ricci tensor to twice the other. To model the equivalence (modulo renormalization), we need rich theory structure with theory morphisms like the ones assumed in [Far00,Koh05a]. The differing group conceptualizations can be handled in the same way.

But as we have seen, the influence of the CoP reaches much farther: Even *the set of substance equivalences is determined by the CoP*. Currently, the assumptions about these seem to be hardwired into the representation formats discussed and utilized in MKM. Depending on how extensive such influences turn out to be, we may have to make representation formats parametric to be able to model such influences explicitly. This also begs the question, whether future MKM knowledge bases will be CoP-specific (severely limiting their usefulness) or whether we will be able to have CoP-spanning knowledge bases. In the latter case, we could annotate documents with new kind of metadata, e.g. the CoP (or CoPs) of the author, or of the intended audience. This has been studied informally by projects that attempt large-scale multi-disciplinary knowledge collections like the Connexions project (see [CNX05,HBK03]) and turned out to be of central relevance for integration, navigation, and quality assurance.

## 5 Conclusion and Further Work

We have explored the space of (mathematical) knowledge MKS spanned by the substance/accidence distinction (with some philosophical excursions) that trig-

---

<sup>9</sup> STEM — Sciences, Technology, Engineering and Mathematics

gers content/form coordinates of a knowledge object which are discussed in various communities (like Artificial Intelligence, Logic & Foundations of Mathematics, document markup, or MKM). We have extended or contrasted the discussion there with cognitions in other scientific disciplines, specifically Knowledge Management and Social Sciences. In the former, the relevance of the context of knowledge is understood, whereas in the latter, the context of knowledge is studied and identified as Community of Practice. Based on this broadened view, we propose a more fine-grained, multi-layered model of the content/form distinction, which explains the apparent multi-faceted nature of the content/form debate. This model is based on two assumptions :

- (i) the existence of a CoP-determined set of substance equivalences that identify the substance of a knowledge object by equating accidental representation commitments, and
- (ii) the dialectic property of the substance and accident aspects of a knowledge object

In our view, assumption (i) is very natural in the field of Mathematics and neighboring disciplines, and the CoP-dependency is often neglected, since it seems to hold for the large CoP shared by the STEM disciplines. This CoP also delineates the applicability of MKM techniques (which currently seem to hard-wire the substance equivalences into the representation formats) to the STEM fields. Turned positively, we conjecture that

*MKM techniques can go wherever the substance equivalences of Mathematics hold!*

On the technological side, our fine-grained knowledge space and its CoP-dependence open whole areas of applications. CoP-information would allow to personalize presentations that are generated from content without assuming total knowledge about the user's preferences. Knowledge about the substance equivalences will (in principle) allow automatic translation (generation of variants tailored to the user and situation).

We hope that the musings presented in this paper will be taken up by others and contribute to a consensus about the foundations of mathematical knowledge, so that we can better manage it. In particular, we have to leave to further investigations the tasks of coming up with a content-oriented model of CoPs and their interrelations (we have only been able to motivate the necessity of this and identify some guiding questions in Section 4) and that of fully exploring the consequences for Mathematical Knowledge Management.

Unfortunately, the model of the knowledge space we have presented opens up as many questions as it helps answer, e.g. what does the knowledge space look like where other forms of substance equivalences are involved. For instance, in the Arts and Humanities, a similar model might be applicable, only that form of an artifact or representation is considered its substance, e.g. in a poem, whereas the meaning is considered its accident. After all, we 'interpret a poem' giving it multiple possible meanings. This suggests the existence of a tetrahedral

knowledge space that is ‘dual’ to the one depicted in Figure 3. This could help solve the riddle that knowledge is considered “objective” in some communities and “subjective” by others.

## References

- [ABC<sup>+</sup>03] Ron Ausbrooks, Stephen Buswell, David Carlisle, Stphane Dalmas, Stan Devitt, Angel Diaz, Max Froumentin, Roger Hunter, Patrick Ion, Michael Kohlhase, Robert Miner, Nico Poppelier, Bruce Smith, Neil Soiffer, Robert Sutor, and Stephen Watt. Mathematical Markup Language (MathML) version 2.0 (second edition). W3c recommendation, World Wide Web Consortium, 2003. Available at <http://www.w3.org/TR/MathML2>.
- [BCC<sup>+</sup>04] Stephen Buswell, Olga Caprotti, David P. Carlisle, Michael C. Dewar, Marc Gaetano, and Michael Kohlhase. The Open Math standard, version 2.0. Technical report, The Open Math Society, 2004. <http://www.openmath.org/standard/om20>.
- [BD91] John Seely Brown and Paul Duguid. Organizational Learning and Communities of Practice:Toward a Unified View of working, Learning and Innovation. *Organization Science*, 2(1):40–57, 1991.
- [BD00] John Seely Brown and Paul Duguid. *The Social Life of Information*. Harvard Business School Press, 2000.
- [BP64] Paul Benacerraf and Hilary Putnam, editors. *Philosophy of Mathematics: Selected Readings*. Cambridge University Press, 2nd edition 1983 edition, 1964.
- [CNX05] CONNEXIONS. Project home page at <http://cnx.rice.edu/>, seen January 2005.
- [CS98] Lassaad Cheikhrouhou and Volker Sorge. Planning equivalence proofs. In *Workshop on Using AI Methods in Deduction at CADE-15*, 1998.
- [DP98] Thomas H. Davenport and Laurence Prusak. *Working Knowledge*. Harvard Business School Press, 2000 edition, 1998.
- [Far00] William Farmer. An infrastructure for intertheory reasoning. In David McAllester, editor, *Automated Deduction – CADE-17*, number 1831 in LNAI, pages 115–131. Springer Verlag, 2000.
- [Hal59] Marshall Hall. *The Theory of Groups*. The Macmillan Company, New York, 1959.
- [HBK03] Geneva Henry, Richard G. Baraniuk, and Christopher Keltz. The connexions project: Promoting open sharing of knowledge for education. In *Syllabus, Technology for Higher Education*, 2003.
- [Hei00] Bettina Heintz. *die innenwelt der mathematik. zur kultur und praxis einer beweisenden disziplin*. Springer-Verlag Wien, 2000.
- [Hut04] Dieter Hutter. Towards a generic management of change. In Christoph Benzmüller and Wolfgang Windsteiger, editors, *Computer-Supported Mathematical Theory Development*, number 04-14 in RISC Report Series, pages 7–18. RISC Institute, University of Linz, 2004. Proceedings of the first “Workshop on Computer-Supported Mathematical Theory Development” held in the frame of IJCAR’04 in Cork, Ireland, July 5, 2004. ISBN 3-902276-04-5. Available at <http://www.risc.uni-linz.ac.at/about/conferences/IJCAR-WS7/>.

- [Isr79] Joachim Israel. *The Language of Dialectics and the Dialectics of Language*. Munksgaard, Humanities Press, USA, 1 edition, 1979. (American) ISBN 0 391 01000X.
- [KA03] Michael Kohlhase and Romeo Anghelache. Towards collaborative content management and version control for structured mathematical knowledge. In Andrea Asperti, Bruno Buchberber, and James Harold Davenport, editors, *Mathematical Knowledge Management, MKM'03*, number 2594 in LNCS, pages 147–161. Springer Verlag, 2003.
- [KBHL<sup>+</sup>03] B. Krieg-Brückner, D. Hutter, A. Lindow, C. Lüth, A. Mahnke, E. Melis, P. Meier, A. Poetzsch-Heffter, M. Roggenbach, G. Russell, J.-G. Smaus, and M. Wirsing. Multimedia instruction in safe and secure systems. In *Recent Trends in Algebraic Development Techniques*, volume 2755 of LNCS, pages 82–117. Springer Verlag, 2003.
- [KM79] M. I. Kargapolov and J. I. Merzljakov. *Fundamentals of the Theory of Groups*. Graduate Texts in Mathematics. Springer Verlag, 1979.
- [Koh05a] Michael Kohlhase. OMDOC an open markup format for mathematical documents (version 1.2), 2005. Manuscript, <http://www.mathweb.org/omdoc/omdoc1.2.ps> to appear in Springer LNAI.
- [Koh05b] Michael Kohlhase. Semantic markup for T<sub>E</sub>X/L<sup>A</sup>T<sub>E</sub>X. Manuscript, available at <http://kwarc.eecs.iu-bremen.de/software/stex>, 2005.
- [LW91] Jean Lave and Etienne Wenger. *Situated Learning: Legitimate Peripheral Participation (Learning in Doing: Social, Cognitive and Computational Perspectives S.)*. Cambridge University Press, 1991.
- [MKB04] Armin Mahnke and Bernd Krieg-Brückner. Literate ontology development. In Robert Meersman, Zahir Tari, and Angelo Corsaro et al., editors, *On the Move to Meaningful Internet Systems 2004: OTM 2004 Workshops*, number 3292 in LNCS, pages 753–757. Springer Verlag, 2004.
- [Pap96] Seymour Papert. An Exploration in the Space of Mathematics Educations. *International Journal of Computers for Mathematical Learning*, 1(1):95–123, 1996.
- [PRR97] G. Probst, St. Raub, and Kai Romhardt. *Wissen managen*. Gabler Verlag, 4 (2003) edition, 1997.
- [Rob91] J. A. Robinson. Formal and informal proofs. In R. S. Boyer, editor, *Automated Reasoning: Essays in Honor of Woody Bledsoe*, pages 267–282. Kluwer, London, 1991.
- [SBF<sup>+</sup>00] Jörg Siekmann, Christoph Benz Müller, Armin Fiedler, Andreas Franke, George Goguadze, Helmut Horacek, Michael Kohlhase, Paul Libbrecht, Andreas Meier, Erica Melis, Martin Pollet, Volker Sorge, Carsten Ullrich, and Jürgen Zimmer. Adaptive course generation and presentation. In P. Brusilovski, editor, *Proceedings of the Fifth International Conference on Intelligent Tutoring Systems—Workshop W2: Adaptive and Intelligent Web-Based Education Systems*, pages 54–61, Montreal, 2000.
- [WMS02] E. Wenger, R.A. McDermott, and W. Snyder. *Cultivating of Communities of Practice*. Harvard Business School Press, 2002.