TOWARDS A SEMIOTICS OF MATHEMATICAL TEXT (PART 2*)

PAUL ERNEST

In this paper, the second of three in a series, I define what I mean by semiotic systems, using examples both from the teaching and learning of mathematics and from research mathematics. After locating semiotic systems within their ever-present social contexts of use, I go on to explore more broadly the semiotics of mathematical text. Using the social semiotics of Michael Halliday, I delineate three metafunctions that can usefully be applied to mathematical text. These are the ideational, interpersonal and textual function. I explore the first of these, the ideational textual function, in this paper. The remaining two functions are discussed in the third and final paper in the series.

One of the central characteristics of mathematics, both in school and in research, is the rule based production and use of signs. However, except in degenerate cases, the use of signs and rules is always underpinned by meaning. Even in degenerate cases, such as blind rule following in school mathematics, there are meanings to the signs and processes, it is just that the student involved is not accessing them. Thus signs, rules and meanings are the three components of a semiotic system.

Semiotic systems

The main theoretical tool developed in this paper is that of a semiotic system. The theory of semiotic systems provides a structure for describing the signs (texts) and the transformational rules applied to them in both school and research mathematics. This model includes both the publicly observable features of mathematical activity and the underlying meanings that underpin such activity, especially the rules of textual transformation.

A semiotic system is defined in terms of three components:

- a set of signs,
- a set of rules for sign use and production,
- an underlying meaning structure, incorporating a set of relationships between these signs and rules (see also Ernest 2005, 2006).

Signs

The set of signs comprises both elementary signs and compound signs made up of concatenated sets of signs. These signs constitute abstract *types* (after Peirce), but their *tokens* (*i.e.,* instances) can be spoken or uttered via various media: written, drawn, encoded electronically or represented by any material means.

Semiotic systems can have multimodal sets of signs, such as the semiotic systems of nursery or kindergarten arithmetic. These sets of signs can include a selection of: verbal sounds and spoken words, repetitive bodily movements, arrays of sweets, pebbles, counters, and other objects, including specially designed structural apparatus, sets of marks, icons, pictures, written language numerals and other text, symbolic numerals. Such inscriptions can be represented on the chalkboard, in printed texts and charts, on computer and other ICT displays, and in children's own work on paper. [1]

In contrast, the semiotic system of school algebra at the lower secondary school level has for its signs constants (numerals), variable letters (*x, y, z,* etc.), a 1-place function sign (⁻), 2-place functions signs (+, −, ×, ÷), a 2-place relation sign (=), and punctuation signs (parentheses, comma, full stop). [2] These signs are typically represented as textual inscriptions on the chalkboard, in printed texts or worksheets or in students' written work. In practice, the set of signs changes over the course of schooling. Early on, in the introduction to algebraic notions during the elementary school years, a blank space ' ', empty line '_' or empty box '□' may be used instead of a variable letter. Later, after the introduction of school algebra in secondary school including the signs listed above, further primitive function signs are introduced, including sin, cos, tan, 1- and 2-place function signs (x^c, x^y), etc. At all levels in school algebra, the formal signs may also be supplemented with written language (*e.g.*, English, French, etc.).

Strictly speaking, if we add new primitive signs to a semiotic system we have changed it to a new semiotic system. However, in practice we often act as if we are extending a single semiotic system, or uncovering further parts of a single semiotic system. [3]

Rules

The rules for sign use, combination and production in a semiotic system can be analysed into 3 types, syntactic, semantic and pragmatic, after Morris (1945). The syntactic rules are based only on the signs *qua* signs, such as rules for producing well-formed formulas (WFFs), etc. Thus '2 × 4 = 8 =' is not a WFF whereas '2 × 4 = 8' is one, because '=' is a 2-place relation sign that for syntactic correctness must be combined with 2 well formed term (WFT) signs. The complexity property of WFFs and WFTs described below is also syntactically defined.

* This is the second part of a three-part article. The first part was published in the previous issue of FLM and the final part will be published in the next issue.

Semantic rules concern the dimension of sign interpretation and meaning(s). Thus, for example, deriving '$2x = 4$' from '$2x + 3 = 7$' in a semiotic system incorporating school algebra can be justified in terms of the meanings of the signs '2', '3', '4', '7', 'x', '+' and '='. In terms of significance, the dominant sign in these expressions is the binary equality relation '=' and this has an underpinning informal meaning of *balance* that must be respected to preserve truth. The import of this is that whatever operation is applied to one of the binary relation sign's arguments (one of its 'sides') must also be applied to the other. Another feature at play in this example is an implicit heuristic of simplification. This seeks to reduce the complexity of terms in an equation *en route* to solution. [4]

Pragmatic rules include contingent and rhetorical rules and these are determined purely by social convention. Examples include teacher requirements that students label answers with the prefix *'Ans. ='*, and double underline the answer and this prefix (Ernest, 1993). Likewise, in university and research mathematics the end of a proof is commonly signified with the Halmos bar '∎' (analogous to the classical QED). Such pragmatic rules are socially imposed or agreed conventions which are immaterial to syntactic and semantic correctness.

Semiotic systems in school and research mathematics extend from highly informal systems with rules that are largely implicit or tacit, to at the other extreme, totally formalized systems in which the rules are fully explicit. An example of an informal system is the semiotic system and practice of arithmetical word problems. The symbols are alphanumeric, comprising letters, numbers, words, and elementary arithmetical operations (+, –, =). Most rules are implicit, concerning the translation of written sentences into numerals and operations, and then computing the answer. For example, "Mary has four sweets. Jim has three sweets. How many do they have altogether?" This problem can be translated into "Four and three; Add", or '4 + 3 = _'. But this translation is based on heuristics, such as 'How many altogether' suggests 'add'. This kind of heuristic cannot be turned into an explicit rule or algorithm, as the correctness of this interpretation depends on the context; the sense of the whole set of statements. As has been noted in the literature (Brown & Küchemann, 1976, 1977, 1981), children's attempts to algorthmicize this heuristic leads to incorrect surface rules such as 'more' or 'total' translate to '+'.

An example of a totally formalized system in which the rules are made as fully explicit as possible is given by the formulation of the propositional calculus in Church (1956). Here the rules of deductive inference are specified explicitly and completely. [5] Even in totally formalized semiotic systems, the rules are based on the underlying meaning structure. It is just that the process of distilling these meanings into explicit rules has reached a stage of completeness, and so the rules can be applied in the generation of signs without reference to the underlying meaning structure. The possibility of neglecting meanings in formalized mathematics was noted three centuries ago by Berkeley (1710, p. 59):

> in Algebra, in which, though a particular quantity be marked by each letter, yet to proceed right it is not requisite that in every step each letter suggest to your thoughts that particular quantity it was appointed to stand for.

Thus meaning gives rise to rules with the result that the meaning structure can be neglected (sometimes only temporarily) during the use of a semiotic system. The entire edifice of mechanical calculation upon which digital computing is based depends on this property. Computers apply rules to signs in completely formalized semiotic systems in which there is no possibility of meaning once the computing structure is finally realized by humans.

Historically, the development of semiotic systems involves the distillation of the implicit meanings of sign use, with their constraints and affordances, embodied in meaning structures, into the more explicit system rules. But it should be noted that any translation of rules and meanings from an implicit system (the meaning structure) into a more explicit form (the rules of a semiotic system) is at best a homomorphism and can never be an isomorphism. In translating relatively vague meanings into something more explicit, some elements of meaning must always be lost and further elements must always be added. Refinement and explication of meanings involves the social construction of new meaning and considerable human creativity and ingenuity may be involved.

But, except for artificial examples of totally formalized systems, virtually none of the semiotic systems at play in mathematics education or mathematical research have fully explicit sets of rules. This is because semiotic systems in use, equivalent to Saussure's systems of *parole*, are typically based on implicit or tacit rules inspired by the underlying meaning structure. In semiotic systems, transformations of signs are typically permitted provided that they preserve key meanings in the underlying meaning structure. In school algebra, this is the truth (balance) of equations, and also the relevance of transformations. In arithmetical word problems, this is the arithmetical meanings and values of the inherent arithmetic in the task statements. In propositional calculus it is the truth value of the propositional signs. However, for each of these three types, the rules are in many cases implicit and are very often acquired by learners or practitioners as 'case law' from the social use of semiotic systems.

In research mathematics, especially in the foundations of mathematics, the ideal semiotic system aims to dispense with all uses of the underlying meaning structure, so that the rules for sign use and production can be specified fully explicitly in syntactic terms. However, this can never be fully achieved except in trivial cases because the specification of syntactical rules requires metamathematical formulation. Thus even the specification of the propositional calculus in purely syntactical terms (as discussed above) requires metamathematical constants, variables and expressions, and their deployment depends on an underlying meaning structure. Hilbert (1925), in his contributions to the philosophy of mathematics, freely acknowledged that a meaningful but finitistic metamathematics is ineliminable when mathematical theories are presented fully formally, *i.e.,* with all their rules presented explicitly in syntactical form (Neumann, 1931).

Meaning structure

The underlying meaning structure of a semiotic system is the most elusive and mysterious part, like the hidden bulk of an iceberg. It is the repository of meanings and intuitions concerning the semiotic system that supports its creation, development, and utilitization (Arcavi, 2005). For individuals, it can range from a collection of tenuous ideas and fleeting images (Burton 1999), to something more well-defined, akin to an informal mathematical theory.

The meaning structure of a semiotic system can be described in three (equivalent) ways, as:

- a set of mathematical contents,
- an informal mathematical theory,
- a previously constructed semiotic system.

A meaning structure is a loosely associated set of mathematical contents that can include: signs, concepts, objects, properties, functions, relationships, rules, procedures, methods, heuristics, classifications, problems, examples, ideas, images, metaphors, models, structures, representations, propositions, theorems, arguments, proofs, theories, etc. It is a reservoir of meanings that can be drawn upon in formulating, developing and operating a semiotic system, such as the metaphor "equality is balance" for simple algebra (Sfard, 1994).

An informal mathematical theory can serve as the meaning structure for a semiotic system in which the sets of signs and rules constitute a formal mathematical theory. All formal mathematical theories, I claim, have an underlying informal theory serving as its meaning structure. Lakatos (1978) argues that such an informal theory provides the touchstone for evaluating a formal mathematical theory. The theorems of the informal theory are potential 'heuristic falsifiers' through which the success of the formal theory can be judged, according to whether it captures or contradicts them.

Two different formal theories are generally regarded as equivalent if their signs and rules are inter-translatable, and they both share the same informal mathematical theory as their meaning structure. Thus various formulations of Peano arithmetic, which may vary in their sets of signs or rules, *e.g.,* the choice of 0 or 1 as the first numeral sign, are regarded as equivalent in this way, even though they are distinct semiotic systems.

Additionally, a previously constructed semiotic system can serve as a meaning structure for a new semiotic system. Since the meaning structure of a semiotic system can include signs and rules, this possibility is already inherent in the first of the descriptions given above. It is also possible for more than one existing semiotic system to be drawn upon or combined to make up a meaning structure, and not all of these need to be mathematical systems. The incorporation of entire or elements of non-mathematical semiotic systems into the meaning structure provides the potential for links between mathematics and other human ways of representing our experiences and environment.

This third case sounds circular, but it is not. Rather it is a case of what Peirce terms 'unlimited semiosis', in which the interpretant of a sign is itself a new sign, and so on, *ad infinitum.* For a semiotic system as a whole can be regarded as a sign, with its underlying meaning structure constituting its interpretant (from the triadic, *i.e.,* Peircean, perspective of signs) or its signified (from the dyadic, *i.e.,* neo-Saussurian perspective of signs). Like any other sign, it is linked via its origins, its uses in practice, its meanings and its associations, with other signs, or in this particular case, with other semiotic systems.

Often it will be the case that a newly developed or elaborated semiotic system (for an individual or group of learners) will be more formally and explicitly specified than the previously constructed or utilized semiotic system which serves as its meaning structure. In the teaching and learning of mathematics it is commonly one of the goals of instruction to increase the abstraction, complexity and formality of the semiotic systems to which learners are introduced and inducted, over the course of time. Hence there is a gradient of increased formality and explicitness.

Such processes of increasing formality are evident in institutionalized mathematics teaching at all levels. In school mathematics, the study of number properties and manipulations in numerical calculation precedes and provides the meaning structure for elementary algebra. Operations in number systems not only grow in complexity in the passage from the Natural Numbers, via Integers, Rationals and Algebraic Numbers to Real Numbers, but also each of these transitions takes the semiotic system as the meaning structure for the next semiotic system of number that is sequentially developed. Typically, in university mathematics, the study of concrete structures such as sets and number systems precedes the study of algebraic structures such as group, ring and field theory, and the study of informal or 'naïve' set theory (Halmos, 1974) precedes the study of axiomatic set theory. In each of these examples, the study of a relatively informal semiotic system precedes the study of its relatively more formal counterpart and, I claim, provides a central part of the meaning structure of the subsequently developed, more formal semiotic system.

The theory of semiotic systems provides a model for describing the teaching and learning of mathematics in school. In learning any school mathematics topic in the form of a semiotic system, learners are inducted into a discursive practice involving the signs and rules of that system. Teachers present tasks in the form of signs and present rules for working or transforming the signs for accomplishing the tasks. Most commonly the rules will be exhibited implicitly through worked examples, particular instances of rule applications, rather than explicit rules stated in their full generality. Through observing the examples, working the tasks, and receiving corrective feedback, learners internalize, build and enrich their personal meaning structures corresponding to the semiotic system.

Trying to teach rules explicitly rather than through exemplification can lead to what I term the 'General-Specific paradox' (Ernest, 2006). If a teacher presents a rule explicitly as a general statement, often what is learned is precisely this specific statement, such as a definition or descriptive sentence, rather than what it is meant to embody: the ability to apply the rule to a range of signs. [6] Thus teaching the general leads to learning the specific and in this form it does not lead to increased generality and functional power on the part of the learner. Whereas, if the rule is embodied in specific and exemplified terms, such as in a sequence of relatively concrete examples, the learner can construct and observe the pattern and incorporate it as a rule, possibly implicit, as part of their own appropriated meaning structure. This is how children first acquire the grammatical rules of spoken and written language. Thus the paradox is that general understanding is achieved through concrete particulars, whereas limited and specific responses may be all that results from learning general statements. This resembles the Topaze effect (Brousseau, 1997), according to which the more explicitly the teacher states what it is the learner is intended to learn, the less possible that learning becomes. For the learner is not doing the cognitive work (meaning making) that constitutes learning, but following surface social cues to provide the required sign – the desired response or answer.

The pattern whereby a learner first learns the use of signs through observation and participation in public sign use in discursive practices embodies the well-known dictum of Vygotsky (1978, p. 128) “Every function in the child’s cultural development appears twice, on two levels. First, on the social and later on the psychological level; first between people as an interpsychological category, and then inside the child as an intrapsychological category.” This Vygotskian scheme can be represented as a cyclic pattern for learners’ appropriation of signs and the rules of sign-use through participation in a discursive practice. The pattern is illustrated in Figure 1 (after Ernest, 2005).

The cycle shown in Figure 1 has four phases: appropriation, transformation, publication and conventionalisation, any one of which can be taken as the beginning, for the cycle repeats endlessly. In the figure, Vygotsky’s two levels are shown, first, by the top right quadrant, where the socio-cultural is represented as both public and collective, and second, by the bottom left quadrant, where the (intra)psychological is represented as both individual and private. This latter constitutes the notional space where a learner constructs his or her meaning structure. The other two quadrants are crossing points on the boundary between these two levels and these are the locations where the learner’s semiotic agency is acted out.

In the development of a personal meaning structure, a learner draws on further resources beyond those indicated in Figure 1. These include existing meanings and the meaning structures of other semiotic systems already partly mastered, as well as meta-discussions of sign production and use. The latter are partly included in Figure 1 in terms of social negotiation and critical acceptance.

The model indicates schematically the route through which learners appropriate the rules of sign-use, mostly through observing their exemplification in practice. The earliest uses of a sign or rule can be based on simple imitation, corresponding approximately to Skemp’s (1976) and Mellin-Olsen’s (1981) notion of instrumentalism, because of the performativity involved. Later, after a sequence of related appropriations, performances and conventionalisations, the use of the sign is transformed through the development of personal meanings, including a sense of where and how the sign is to be used acceptably, and a whole nexus of other associations. The successful *appropriation* and *transformation* of a sign, with its nexus of associated meanings and meta-discourse, parallels Skemp’s (1976) notion of ‘relational understanding’. This involves both being able to use the sign correctly, corresponding to conventionally accepted usage within the micro-community of the classroom under the authority of the teacher, and being able to offer a rationale or explanation for the usage.

The next phase is that of *publication,* in which individual learners engage in conversational acts of sign utterance. These utterances can vary from quick, spontaneous verbal, gestural or written responses to a question or other stimulus, through to constructing extended texts elaborated and revised over a period of time, prior to offering them to others. A group of learners can elaborate such texts co-operatively, but such processes subsume several or even

		SOCIAL LOCATION		
		Individual		Collective
	Public	Learner’s public utilization of sign to express personal meaning (Public & Individual)	*Conventionalisation* →	Social (teacher & others) negotiated and conventionalized (via critical acceptance) sign use (Public & Collective)
MANIFESTATION		*Publication* ↑		↓ *Appropriation*
	Private	Learner’s development of personal meanings for sign and its use (Private & Individual)	← *Transformation*	Learner’s own unreflective response to and imitative use of new sign utterance (Private & Collective)

Figure 1: Model of sign appropriation and use in learning

many sub-cycles in which individuals utter signs to others in the group in an extended conversation giving rise to jointly elaborated, negotiated and agreed texts.

The cycle is completed through the process of *conventionalisation*, in which signs are uttered within the classroom conversation and are subjected to attention, critique, negotiation, reformulation and acceptance or rejection, primarily by the teacher. Teacher approval is normally the final arbiter of acceptance, because of the power and authority relations in the classroom. Typically the conventionalized sign utterance that is accepted will satisfy three criteria:

Relevance. The sign or text is perceived to be a relevant response or putative solution (or an intermediary stage to one) to a recognised (*i.e.*, sanctioned) starting sign that has the role of a task, question or exercise. This task might be teacher imposed or otherwise shared and authorized.

Justification. The mode and steps in the derivation of the sign from the authorized starting point will normally be exhibited as a semiotic transformation of signs, employing acceptable rules or sign transformations within the semiotic system, or justified meta-linguistically. [7]

Form. Both the signs and their transformations will normally exhibit teacher-acceptable form, thus conforming to the rhetoric of the semiotic system involved as realized and defined in that classroom. This system could be that of spoken verbal comments, drawn and labelled diagrams, numerical calculations, algebraic derivations, or some combination of these or other sign types (Ernest 1993).

An overall schematic model of a semiotic system within its social context of use is given in Figure 2. This model summarizes the three parts that make up a semiotic system: the signs, rules and meaning structure. The signs can be elementary or composite, and indeed sequences of signs will often be all that is explicitly exhibited in the operation of the semiotic system. Many rules will often be implicit, exhibited only via specific applications as sign transformations, as well as explicitly presented rules. For learners the meaning structure will be developed as the semiotic system is utilized, although learners will also draw on other, partially mastered semiotic systems, as well as their general repertoire of meanings, including some of the items listed, from signs and concepts to the relevant informal mathematical theory.

The use of semiotic systems always takes place within a social context. Within social settings there are persons and their roles, positions, power relations, and relations with social institutions such as schools. An important dimension of social understanding as it relates to semiotic systems is the concept of school learning task and the aims, goals, and purposes of schoolwork that is presupposed by operating semiotic systems in school settings. The transformation of signs in semiotic systems is directional, and the understanding of directionality in general is socially acquired in a variety of social settings including home and school. [8] Ultimately, directionality in activities results from directions given by a person in some powerful role. Conversation is a major channel for the communication of such matters (Ernest, 1998). Where it concerns working matters pertaining to the semiotic system in use, *i.e.,* the signs produced, the rules employed, and the accomplishment of mathematical tasks, I term this meta-discussion of the semiotic system. This plays an important role in correcting and shaping sign production and rule deployment, as well as enabling the development of the public meaning structure within the particular social context.

The semiotics of mathematical text

Mathematical text is a peculiar form of text on account of its subject matter. Mathematical text does not refer to the embodied world of our experiences, with its frameworks of space and time. Instead it is understood to refer to a different realm, one in which time stands still or does not exist. I have earlier critiqued the myth that this other realm is a Platonic universe beyond our social and material world (Ernest, 2008). Instead, mathematics refers to a semiotic space, a socially constructed realm of signs and meanings. Human beings are sign-using and sign-making creatures, and most humans can participate in the semiotic space of mathematics, even if only to a limited extent. However, this participation involves the deployment of semiotic resources and tools and the assumption of certain identities. Entry to this semiotic space is via the texts of mathematics, but it is important to assert that it is not a timeless zone (Mazur, 2004). While there is no universal timepiece ticking away in semiotic space, nevertheless individual and group engagement in mathematical activity is always over time (Mason, Drury, & Bills, 2007). [9] What this means is that accessing mathematical texts always has a sequential nature. Sequential development is the semiotic or logical analogue of time.

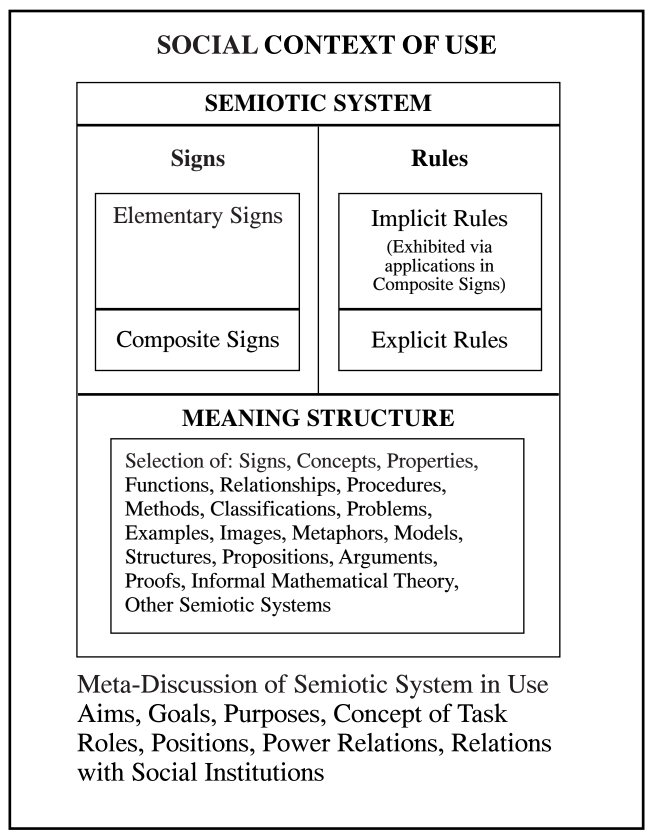


Figure 2. Model of semiotic system within its social context of use

However, reorderings of mathematical texts are always possible, and are often affected. This is similar to reorderings in the construction and editing of films and fictional texts, as well as of academic texts like the one you are currently reading. Nevertheless, at any given stage, any text is made up of sequentially ordered contiguous signs. This sequential contiguity is the analogue of time and is often enacted through the use of space. [10] Any array of semi-permanent material signs – that is, marks on a page rather than sounds in the air – is ordered by spatial conventions of access into a beginning, middle and end. In any form of representation, there is always an ordering present and this structures the access and role of readers.

To explore the semiotics of mathematical text including the roles of readers and writers it is necessary to look at the language of mathematics. Here it is important to look beyond mathematics to other disciplines that have analysed language and text without the ideological presuppositions often present in discussions about mathematics (Ernest, 2008). An obvious area to turn to is linguistics, whose subject matter is language and text.

The linguistic theorist Halliday (1985) has developed the theory of systemic functional grammar, which provides an illuminating tool for this analysis. In some of his publications, he and his colleagues have examined the linguistics of mathematical and scientific text (Halliday and Martin, 1993, UNESCO, 1974). Halliday distinguishes three overlapping metafunctions of text in use, and these can usefully be applied to mathematical text. These are the ideational, interpersonal and textual functions.

1. The *ideational or experiential function* concerns the contents of the universe of discourse referred to, the subject matter of the text, the propositional content. This includes the processes described and the objects or subject matters involved in the process. Morgan (1998, p. 78) relates this to mathematical questions such as "What does this mathematical text suggest mathematics is about? How is the mathematics brought about? What role do human mathematicians play in this?"

2. The *interpersonal function* concerns the position of the speaker, the interaction between speaker and addressees, and their social and personal relations. The related mathematical questions suggested by Morgan (1998, p. 78) include "Who are the author and the reader of this mathematical text? What is their relationship to each other and to the knowledge constructed in the text?"

3. The *textual function* is about how the text is created and structured, and how it uses signs, and so on. Morgan (1998, p. 78) relates this to mathematical questions such as "What is the mathematical text attempting to do? Tell a story? Describe a process? Prove?"

In mathematical text, as in the general theory, these three functions all overlap, and cannot be treated as disjoint (Morgan, 1998). Thus, for example, the ideational function of representing the universe of discourse overlaps with the textual function of binding linguistic elements together into broader texts. Further, the intended audience cannot be divorced from these issues. All treat the semiotics of mathematics, where subject matter and form are indissolubly bound up together, as is the intended social function of the text.

The ideational function

In exploring the ideational function of mathematical text the following questions arise. What is the propositional content of mathematical text, and what is mathematics/ mathematical text about? What objects and processes are described? The answer, from a semiotic perspective, is that mathematics is about mathematical signs and the operations applied to them (Radford, 2002). The meaning of the signs of mathematics resides primarily in their uses and functions. Some signs, such as numerals, appear to represent objects – numbers in the case of numerals. Some signs represent operations on signs and, through them, appear to represent processes applied to objects. Thus the 2-place addition operation sign '+' is applied to numerals and thus appears to act on numbers (objects). However, the classification of signs as representing objects or processes is relative. It depends on the function that is foregrounded and the perspective to which it gives rise. In fact all mathematical signs represent both objects and processes. [11] A sign represents an object when viewed or used as a unified entity in itself, or as a single signifying entity. A sign represents a process when viewed in terms of its parts or in its actions on, or relations with, other signs. The coordination or structuring of parts into a whole, even when apparently static, is a process or relation. For example, the number 3 is on the one hand, a basic mathematical object. But on the other hand, in Peano arithmetic it is formally defined as the successor of 2, and ultimately defined only in terms of the two primitive signs S and 0 (3 = def S2 = SS1 = SSS0). This analysis, which incidentally mirrors both historical and psychological developments of number (Ernest, 2006), reveals the number 3 as constituted by a process.

Mathematics is constituted by semiotic systems, and by being made up of signs and sign-based activities it appears to be at one remove from its subject matter, the objects and processes of mathematics. For signs, by their nature, are always distinct from the objects they signify. [12] Thus one might say that if the objects and processes of mathematics are to be found anywhere, it is in the underpinning meaning structures of semiotic systems, for this is the domain to which the signs within a semiotic system refer or 'point'. At first blush, all appears well and good, until we remember what makes up the meaning structures, notably the objects and processes of mathematics and their constitutive signs, concepts, functions, and so on, up to and including other semiotic systems. But this then raises again the question of the nature of the objects, processes and other entities that make up the meaning structures. What is this domain to which the signs of mathematics refer? It is nothing but the semiotic realm of mathematics itself, for mathematical signs and texts do not represent or refer to some reality beyond our material world, as I argued in Part 1 (Ernest, 2008). Nor do they represent our material world itself. What mathematical texts and signs refer to is other mathematical signs, and these are cultural objects, not material objects. The vast array of mathematical semiotic systems created, communicated and sustained by human activity both make up the sign systems of mathematics and the subject matter to which these signs refer. Far from being a vicious, self-contradictory cycle, this is a virtuous cycle that creates and brings into being an ever growing universe of mathematical signs, objects and semiotic systems.

Just as human imagination brings into being the countless characters, situations and worlds of fiction, so too it brings into being the unending realms of mathematical objects, including all of the constituents of the meaning structures of semiotic systems. Unlike the former, which are usually the product of individual imaginations, the latter are socially constructed. For it is not enough for just one mathematician to assert that mathematical theories and objects exist. They need to be shared and accepted before they can be said to be properly mathematical, as indicated in Figure 1. [13]

Thus the signs and texts in a mathematical semiotic system refer both to the signs of the semiotic system itself, and to objects and processes in the meaning structure of the system, which are themselves sign-based. The meaning structure draws on signs and meanings created in other semiotic systems, but also grows in meaning and complexity as the semiotic system develops and creates its own domain of signs during its development and use. [14]

Mathematics is a special subject matter in that its signs refer only to other signs and sign-based processes and operations. It is also unique in the depth and complexity of its sign formation operations whereby sign-based processes are condensed and reified into objects, and so transformed and constituted as further sign objects. [15] In most of the semiotic systems in mathematics these operations create a potentially endless supply of signs of increasing structural complexity and abstraction. Mathematical theories sometimes encompass such unending processes as a whole and in one bound turn them into new, yet more complex signs. By these processes, the semiotic systems of mathematics can represent and incorporate infinite objects and processes.

The social function of semiotic systems is primarily to represent and to solve mathematical problems and to work mathematical tasks. These operations require the transformation of texts employing sign-based processes following the rules of the semiotic system. Explicit and implicit rules are the key operative mechanisms and principles through which new signs are formed and composite texts are constructed and elaborated. The overt function of these rules is to provide a technology for the transformation of mathematical signs in a goal directed way. [16] That is, a means for bringing the signs closer to some desired (and sometimes locally defined) canonical state, and in so doing preserving key invariants within the meaning structure. The two best known types of transformations, corresponding to two dominant problem solving activity types, are numerical calculation and proof.

Calculations are sequences of numerical terms that, through a sequence of numerical value-preserving transformations, lead to the derivation of new terms that are relatively simpler [17] and increasingly approach one of the canonical forms of numerals. For example, the standard canonical form for numerals is the decimal place value representation, which can be represented:

$\sum_0^\infty 10^n \times k_n,$ for $n \in \mathrm{N}$, where $k_n = 0$ except for a finite number of values of n, where $0 < k_n < 10$.

The key characteristic of this representation is that, for any numerical value, it is unique. [18] Because of this property many calculational tasks, and the sequences constructed in accomplishing the tasks, have unique endpoints (answers), which is a special and defining characteristic of school mathematics. In a semiotic system encompassing arithmetical calculations, there are an infinite number of terms that can be transformed 1-1 onto any given numerical value. [19] Thus, for any answer, there is an unlimited corresponding set of tasks.

This discussion illustrates how the rules for the transformations of terms in numerical calculations are based upon the principle of preservation of numerical value. There are also rules used in inequality transformations that, for example, do not increase (they preserve identically or decrease) the numerical values. Similarly, deductive proof sequences draw upon rules for the transformations of propositional signs that do not decrease the truth values attached to them. As in these cases, the rules of the semiotic systems typically originate in or refer to certain interpretations or valuations of the signs in the underlying meaning structure and preserve these attributes as invariants.

In addition to providing a technology of sign transformations, semiotic systems have a further function in providing representations of mathematical problems. Mathematical texts provide users with a language for representing structural aspects of a range of situations. However, situations are never given directly, they are always mediated by a second semiotic system, acting as part of the meaning structure of the mathematical semiotic system in which the new texts are constructed. Typically such subsidiary semiotic systems are less formal, perhaps more intuitive systems than that in which the 'modeling' text is constructed for application and use. The wide range of objects that can be drawn upon in constituting the meaning structure of a mathematical semiotic system includes existing semiotic systems, some drawn from outside of mathematics. Thus the structural modeling function enables the construction of texts that, on the one hand enable the application of mathematics to extra-mathematical situations, and on the other hand, serve as starting points for transformations within the mathematical semiotic system. It enables semiotic systems to both 'look outwards' via modeling and application functions and to 'look inwards' towards the sign transformation functions.

The modeling function utilizing the relation between a primary mathematical semiotic system and a subsidiary semiotic system incorporated informally within the meaning structure of the former can serve a range of functions. It can be the location for plans or schematic representations to be implemented in the primary semiotic system. As a subsidiary domain of signification and meaning, it can be used to formulate informal representations serving as plans or outline ideas, to be used to guide sign constructions and transformations in the primary semiotic system. For example, a mathematical proof is often based on an informal proof idea (Rotman, 1988). Such an idea, represented in a subsidiary semiotic system, serves as the plan or schematic outline of the proof transformation constructed in the primary semiotic system as a sequence of signs. This can be generalized because quite often, rather than being generated 'from scratch' in semiotic systems, transformational sequences are implemented 'translations', in some loose sense, of schematic ideas or plans developed outside the semiotic system. Typically such outline ideas are generated and represented in a subsidiary semiotic system incorporated in the meaning structure.

This concludes the treatment of the ideational function of mathematical text. The interpersonal and textual functions are treated in the third and final paper in this series, although it is evident that they have already entered into the above discussion of the ideational function.

Notes

[1] Unlike some researchers, *e.g.,* Duval (1995), Hayfa (2006), I do not see the need to create separate semiotic systems for different types of sign/register/mathematical topic. I prefer a simple all-encompassing definition of a semiotic system because it is more flexible and offers more generality than multiple system types. Although there are complexities involved in coordinating different registers within one system, especially for learners, the ultimate educational goal is for unity to triumph over difference.

[2] Technically I should put, *e.g.,* '=' for =, in this account, but instead I am following common usage to allow = to stand ambiguously for both a 2-place relation sign in the object language and for the metalinguistic sign ('=') that names it in the metalanguage, where my discussion takes place.

[3] Within limits this practice is justifiable, as we can embed a 'substructure' (a structure in which each of the constituents sets of a semiotic system is a subset of the corresponding sets within a greater 'superstructure') conservatively, within the said superstructure. However, the enlargement of a semiotic system, such as extending the semiotic system of natural numbers to that of integers can lead to new metatheoretical results (*e.g.,* multiplication can produce products less than either multiplicand) contradictory to the state of affairs in the system of natural numbers, leading to epistemological obstacles (Ernest, 2006). Arzarello (2006) extends the notion of a semiotic system to that of a semiotic bundle. This comprises a collection of semiotic systems and a set of relationships between the systems. This notion would allow a natural way of treating families of semiotic systems related by the addition of further signs, rules or meanings, as discussed in the text.

[4] The complexity C of an expression (term or formula) is defined inductively in terms of its syntactic structure. The complexity of an atomic expression t, denoted $C(t) =_{\text{def}} 1$. Given a set of k expressions $t_1, t_2, \ldots, t_k$, the maximum complexity of which is n, and a k-place function or relation symbol F, the complexity of the expression $Ft_1t_2\ldots t_k$, denoted $C(Ft_1t_2\ldots t_k) =_{\text{def}} n+1$. Complexity is used as a measure of the impact of transformational rules on a term constituting one of the arguments ('sides') of an equation. When a legitimate (*i.e.,* rule following) transformation of the equation achieves a reduction of the complexity of the equation or its terms, normally, the task is closer to completion. The simplification heuristic motivates the use of rules to reduce the complexity of terms or expressions, and it is normally implicated throughout the solution of algebraic and arithmetical equations, for task goals are typically based on maximal simplification of texts.

[5] But note that the rules specifying, for example, which axioms may be inserted into deductive sequences must either be given as metalinguistic schemas, *e.g.,* $P\to(Q\to P)$, where P and Q are metalinguistic variables ranging over and replaceable consistently by any well formed formulas, or as an infinite set of all possible instances of replacements in this expression within the object language (*i.e.,* the signs of the semiotic system), or where there is a single instance included as a privileged sign (axiom), *e.g.,* $p\to(q\to p)$, where p and q are elementary propositional variables, coupled with a metalinguistic rule of replacement for Q by $Q(p/q)$, whereby all instances of p occurring in Q are replaced by instances of q in $Q(p/q)$, such that if (and only if) Q is true or assertable, so is $Q(p/q)$.

[6] This also applies to any general item of knowledge that is applicable in multiple and novel situations, such as a mathematical concept, rule, generalised relation, skill or strategy.

[7] Sign transformations do not always mean the replacement of one or more parts of a compound sign by different parts, with the retention of the unreplaced parts. It may involve the construction of a wholly new sign in the sequence. For example, axiom use in a logical proof can involve the insertion of a new sign with no components shared or overlapping with the previous step.

[8] Note that the simplification heuristic described above plays a central role in operationalizing directionality in mathematical tasks. That is, a significant part of the appropriation of directionality is associated with the implicit understanding of the simplification heuristic as a technique for goal-directed activity.

[9] Note that there is no universal time in empirical space either, as Einstein's Relativity Theory tells us. However, there are standardized, socially agreed regional time conventions in the local physical world in a way that there are not in mathematical semiotic space.

[10] Spatial or sequential contiguity also correspond with the syntagmatic and metonymic axes referred to in linguistics; whereas equivalence and other substitution relations correspond with the paradigmatic and metaphoric axes. In my discussion I assume sequential access to all texts. Although mathematical texts include diagrams and other objects not reducible to the linear (Rotman, 1995) this does not compromise the overall sequential access of the reader.

[11] Several authors have remarked on the dual nature of mathematical objects from philosophical, psychological and sociological perspectives, and stressed the role of reification in the construction of objects from processes in mathematics (Ernest, 1991, 1998; Machover, 1983; Radford, 2002; Restivo, 1992; Sfard, 1987, 1994).

[12] This follows from both de Saussurian and Peircean theories of semiotics, as well as theories of the sign (*e.g.,* Morris, 1952). Even in the extreme case of formal mathematical systems, in which it is possible to have a sign signifying itself, such as in Henkin's (1959) proof of the completeness of the first-order functional calculus, this involves two wholly distinct roles for the sign in different domains, serving as signifier in the metalanguage (the mention-function) and signified in the object language (the use-function).

[13] Perhaps this is also true of the objects of fiction. Maybe they do not exist in any meaningful way until they have been both described in fiction and interpreted by a reader.

[14] Semiotic systems have a dual nature as is illustrated in Figure 1, both public/collective and private/individual. Culturally semiotic systems grow and develop historically as their creators develop them as human sociocultural artefacts. For educational purposes these are reconfigured, recontextualized, and presented to learners for them to meet on the interpsychological plane and internalize and appropriate on the intrapsychological plane (Vygotsky 1978). Through such processes, learners reconstruct private/individual meaning structures, although their sign utterances and rule applications are primarily public.

[15] Elsewhere I argue for the importance and prevalence of the reification in the semiotics and philosophy of mathematics (Ernest, 1991, 1998). See also Note 11.

[16] The rules embody, in operative form, the structural meanings of objects, relations and processes in the meaning structure of the semiotic system.

[17] By 'simpler' I mean having reduced complexity, as discussed in Note 4.

[18] My illustration only concerns the numerals representing natural numbers, although with further elaborations it can be extended to other domains of number including Rational and Real numbers, bearing in mind the complexities introduced by infinite decimal representations.

[19] I leave aside the problem, which encompasses almost the whole history of mathematics, of numerical terms signifying objects that are not part of the semiotic system in which the terms have been constructed, *e.g.*, given a semiotic system for N, if $m > n \in$ N and, then $n - m \notin$ N. The generation of such non-elements is often a motivating factor in the extension of semiotic or structural systems in the history of mathematics to new systems to include them.

References

Arcavi, A. (2005) 'Developing and using symbol sense in mathematics', *For the Learning of Mathematics* **25**(2), 42-47.

Arzarello, F. (2006) 'Semiosis as a multimodal process', *Revista Latinoamericana de Investigación en Matemática Educativa* **9** (Número Especial), 267-300.

Berkeley, G. (1710, reprinted 1962) *The principles of human knowledge.* Glasgow: W. Collins.

Brousseau, G. (1997, translated and edited by N. Balacheff, M. Cooper, R. Sutherland and V. Warfield) *Theory of didactical situations in mathematics: didactique des mathematiques, 1970–1990,* Dordrecht, Kluwer Academic Publishers.

Brown, M. and Küchemann, D. E. (1976) 'Is it an add miss? Part 1', *Mathematics In School* **5**(5), 15–17.

Brown, M. and Küchemann, D. E. (1977) 'Is it an add miss? Part 2', *Mathematics In School* **8**(1), 9–10.

Brown, M. and Küchemann, D. E. (1981) 'Is it an add miss? Part 3', *Mathematics In School* **10**(1), 13–15.

Burton, L. (1999) 'Why is intuition so important to mathematicians but missing from mathematics education?', *For the Learning of Mathematics* **19**(3), 27–32.

Church, A. (1956) *Introduction to mathematical logic, volume 1,* Princeton, NJ, Princeton University Press.

Duval, R. (1995) *Sémiosis et pensée humaine,* Bern, Peter Lang.

Ernest, P. (1991) *The philosophy of mathematics education,* London, Falmer Press

Ernest, P. (1993) 'Mathematical activity and rhetoric: towards a social constructivist account', in Hirabayashi, I., Nohda, N., Shigematsu, K. and Lin, F. L. (eds), *Proceedings of 17th Meeting of the International Group for the Psychology of Mathematics Education,* Tsukuba, Japan, University of Tsukuba, **2**, pp. 238–245.

Ernest, P. (1998) *Social constructivism as a philosophy of mathematics,* Albany, NY, SUNY Press.

Ernest, P. (2005) 'Agency and creativity in the semiotics of learning mathematics', in Hoffmann, M., Lenhard, J. and Seeger, F. (eds), *Activity and sign – grounding mathematics education,* New York, Springer, pp. 23–34.

Ernest, P. (2006) 'A semiotic perspective of mathematical activity: the case of number', *Educational Studies in Mathematics* **61**, 67–101.

Ernest, P. (2008) 'Towards a semiotics of mathematical text (part 1)', *For the Learning of Mathematics* **27**(2), 2–8.

Halliday, M. A. K. (1985) *An introduction to functional grammar,* London, Edward Arnold.

Halliday, M. A. K. and Martin, J. R. (1993) *Writing science: literacy and discursive power,* London, Falmer Press.

Halmos, P. R. (1974) *Naïve set theory,* New York, Springer Verlag.

Hayfa, N. (2006) 'Impact of language on conceptualization of the vector', *For the Learning of Mathematics* **26** (2), 36–40.

Henkin, L. (1949) 'The completeness of the first-order functional calculus', *Journal of Symbolic Logic* **14**, 159–166. Reprinted in Hintikka, J. (ed.) (1969) *The philosophy of mathematics,* Oxford, Oxford University Press, pp. 42–50.

Hilbert, D. (1925) 'On the infinite', *Mathematische Annalen*, **95**, 161–190. Translation in Benacerraf, P. and Putnam, H. (eds) (1964), *Philosophy of mathematics: selected readings*, Englewood Cliffs, NJ, Prentice-Hall, pp. 134–151.

Lakatos, I. (1978) *The methodology of scientific research programmes (philosophical papers volume 2)*, Cambridge, UK, Cambridge University Press.

Machover, M. (1983) 'Towards a new philosophy of mathematics', *British Journal for the Philosophy of Science* **34**, 1–11.

Mason, J., Drury, H. and Bills, E. (2007) 'Explorations in the zone of proximal awareness', in Watson, J. and Beswick, K. (eds), *Mathematics: essential research, essential practice. Proceedings of the 30th annual conference of the Mathematics Education Research Group of Australasia,* Adelaide, MERGA, **1**, pp. 42–58.

Mazur, B. (2004) 'On the absence of time in mathematics', *For the Learning of Mathematics* **24**(3), 18–20.

Mellin-Olsen, S. (1981) 'Instrumentalism as an educational concept', *Educational Studies in Mathematics* **12**, 351–367.

Morgan, C. (1998) *Writing mathematically: the discourse of investigation,* London, The Falmer Press.

Morris, C. (1945) *Foundations of the theory of signs,* Chicago, University of Chicago Press.

Neumann, J. von (1931) 'The formalist foundations of mathematics', *Erkenntnis,* 91–121. Translation in Benacerraf, P. and Putnam, H. (eds) (1964), Philosophy of mathematics: selected readings, Cambridge, UK, University Press, 50–54.

Radford, L. (2002) 'The seen, the spoken and the written: a semiotic approach to the problem of objectification of mathematical knowledge', *For the Learning of Mathematics* **22**(2), 14–23.

Restivo, S. (1992) *Mathematics in society and history,* Dordrecht, Kluwer.

Rotman, B. (1988) 'Towards a semiotics of mathematics', *Semiotica* **72**(1/2), 1–35.

Sfard, A. (1987) 'Two conceptions of mathematical notions: operational and structural', in Bergeron, J. C., Herscovics, N. and Kieran, C. (eds), *Proceedings of the 11th annual meeting of the International Group for the Psychology of Mathematics Education,* Montreal, University of Montreal, **3**, pp. 162–169.

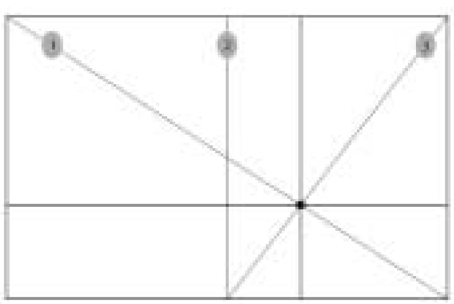
Sfard, A. (1994) 'Reification as the birth of metaphor', *For the Learning of Mathematics* **14**(1), 44–55.

Skemp, R. R. (1976) 'Relational understanding and instrumental understanding', *Mathematics Teaching* **77**, 20–26.

UNESCO (1974) *Interactions between linguistics and mathematics education,* UNESCO, Paris.

Vygotsky, L. S. (1978) *Mind in society: the development of the higher psychological processes,* Cole, M., John-Steiner, V., Scribner, S., Souberman, E. (eds), Cambridge, MA, Harvard University Press.

What fraction is it?



Take a rectangular piece of paper. Place it on the table in front of you, landscape. (1) Fold it across a main diagonal. (2) Unfold it, and fold it in half parallel to the shorter side so that you have two equal rectangles on your left and right. (3) Unfold again and fold the main diagonal of one of the smaller rectangles.

You should see the pair of diagonals crossing at a point. I say that a line through this point parallel to either side of the original paper, divides the paper exactly into thirds (shortways or lengthways). If you can prove this, show that it is (theoretically) possible to fold any rational fraction of the paper.

[with thanks to Leo Rogers]