Semiotic Conceptual Analysis of Part-Whole Relationships in Diagrams

Uta Priss

Fakultät Informatik, Ostfalia University, Wolfenbüttel, Germany
www.upriss.org.uk

Abstract. Proverbially, a picture is worth a 1000 words because it conveys a multitude of concepts, their parts and relationships simultaneously. But it is difficult to precisely describe why and how a picture achieves this. This paper employs Semiotic Conceptual Analysis as a means for providing formal methods for analysing diagrams and other graphical representations, in particular, by modelling part-whole relationships amongst representations and defining quantitative measures and certain qualitative features for comparing different types of representations. The general background for this research is analysing diagrams for teaching purposes.

1 Introduction

A background for this research is mathematics education research. Mathematical concepts often involve several types of representations many of which are diagrammatic. It is well known that mathematical content should be taught using "multiple representations" (Ainsworth 1999). For example, a function can be represented in different formats (as a set, relation, graph, piece of computer code and so on) and using different media (with pen and paper or with a variety of computational tools). Expert mathematicians use and switch between different representations seamlessly often without even being aware of it. But students need to learn when and how to employ representations and how to integrate them into joined-up concepts. For a mathematics teacher the questions arise as to which representations are most effective, how many different types should be used, in which order they should be introduced and so on. While many publications on graphs and diagrams already exist (e.g. Moody (2009)), there is still a need for a development of algorithmic, structural approaches. In this paper we are proposing to use Semiotic Conceptual Analysis (SCA) as a method for analysing and comparing diagrams and other graphical representations. A future goal is to develop computerised tools that support teachers in their decision processes of selecting and structuring teaching materials.

SCA is a formalisation of semiotics based on modelling signs as elements of a triadic relation (Priss 2017). SCA was initially influenced by the semiotics of Charles S. Peirce but its purpose is not philosophy. Instead, SCA can be considered an extension of Formal Concept Analysis (FCA) which is a formalisation of concept hierarchies (Ganter & Wille 1999). Both SCA and FCA are mathematical approaches to modelling data. Therefore signs as defined by SCA and concepts as defined by FCA are abstractions of their philosophical or linguistic counterparts. Once data has been modelled by a user according to SCA or FCA, signs or concepts are deterministically identified by their structural properties. Diagrams and other graphical representations tend to contain parts and are therefore compound signs in SCA. A text containing words which themselves contain letters is also an example of a compound sign. This paper presents a semiotic approach for comparing compound signs, their parts and relationships which applies to many different types of representations without defining grammars or other domain-specific structures. Some of the notions were already introduced in Priss (2022) but are further extended in this paper. All propositions and examples in this paper are new. SCA has previously been applied in a variety of settings (Priss 2019) and has established a vocabulary pertaining to different semiotic topics.

There are other existing theories which consider how semiotic structures are transmitted (e.g. Information Theory) or transformed, such as representation theory, Barwise & Seligman's (1997) information flow theory, Goguen's (1999) algebraic semiotics, theories of formal languages and grammars and others. But such theories tend to focus on proving properties of classes of structures or on certain types of problems. The aim of SCA, however, is to provide means for analysing finite data collected from actual applications. Furthermore, other existing theories tend to use sets as their smallest building blocks in contrast to the use of triadic signs in SCA. Last but not least, there are many methods for analysing graphical representations (e.g. Moody 2009) but they do not tend to build coherent mathematical theories. Questions about diagrams could also be answered by conducting user studies. But that would be a more time consuming approach than a structural analysis with SCA.

Apart from being influenced by Peirce's philosophy, another motivation for SCA was to improve concept lattices developed for natural language data such as the lattices generated from WordNet and Roget's Thesaurus by Priss & Old (2010). The structures emerging from words and their meanings do not automatically resemble intuitive conceptual hierarchies or maintain part-whole relationships because of synonymy and polysemy amongst words and because the relationships within the data are normally created in a somewhat ad-hoc manner. Signs are triadic instead of binary. Thus a semiotic relation can be considered a formal context (in the sense of FCA) where each "cross" is replaced by a set of partial functions that determine under which conditions a cross is relevant. From the view of Triadic FCA, similar contexts are discussed by Ganter & Obiedkov (2004).

The following section repeats SCA core notions and ensures that this paper is self-contained with respect to SCA. Because FCA has been presented many times before at this conference, it is assumed that readers are familiar with FCA. Section 3 describes how compound representamens can be decomposed in a manner that provides measurable features. Sections 4 and 5 apply the theory developed in this paper to two examples of mathematical representations which present typical topics taught in undergraduate mathematics. The paper finishes with a conclusion.

2 Basic SCA Notions

While an FCA *concept* is a pair of a set called *extension* and a set called *intension*, a *sign* in SCA is a triple whose three elements are called *interpretation*, *representamen* and *denotation*. There are no further restrictions placed on the sets of representamens and denotations. But the interpretations must be partial functions from the set of representamens into the set of denotations which means that if an interpretation is defined for a representamen then it must map the representamen onto a unique denotation. If SCA is applied to linguistic data then representamens might be words or lexemes, denotations might be word or lexeme meanings and interpretations might be usage contexts. A set of signs with shared sets of interpretations, representamens and denotations is called a *semiotic relation*. If SCA is applied to diagrams, a user first identifies interpretations that are relevant. Usually different interpretations are applied simultaneously to different parts of a diagram yielding a complex conceptual structure which is then further analysed with SCA and FCA. Compound signs which are defined more precisely below have a denotation as a whole but also denotations via their parts.

Fig. 1 displays an example of a tiny semiotic relation with 4 signs. The sign s_1 has '15' as its representamen, '15 minutes' as its denotation and i_1('15') = '15 minutes' using some interpretation i_1. In this example the representamens and denotations are strings, but representamens can also be pictures and denotations can be concepts from an FCA concept lattice. Representamens are usually specified based on an equivalence relation. In Fig. 1, 'quarter' and 'QUARTER' might be considered equivalent for an SCA analysis. According to the following definition, the only requirement for these 4 triples is that their interpretations must be partial functions which means that s_2 and s_3 must use different interpretations.

$s_1 = (i_1$, '15', '15 minutes')
$s_2 = (i_1$, 'quarter', '25 cents')
$s_3 = (i_2$, 'QUARTER', '15 minutes')
$s_4 = (i_2$, '25', 'a natural number with value 25')

Fig. 1. Example of a tiny semiotic relation with 4 signs

Definition 1. *For a set R (called* representamens*), a set D (called* denotations*) and a set I of partial functions $i : R \nrightarrow D$ (called* interpretations*), a* semiotic relation *S is a relation $S \subseteq I \times R \times D$. A relation instance $(i, r, d) \in S$ with $i(r) = d$ is called a* sign. *For a semiotic relation, an equivalence relation $\approx_R$ on R, an equivalence relation $\approx_I$ on I, and a tolerance relation $\sim_D$ on D are defined.*

Equivalent representamens are usually considered indistinguishable for SCA purposes and must be mapped onto the same denotation by an interpretation. But writing '=' instead of '$\approx_R$' would be problematic because in most usage contexts '=' refers to denotational equality. For example, '$x = 5$' means having the same denotation and not that x and 5 are equivalent representamens (because, as representamens, letters are most likely not equivalent to numbers). Each interpretation i_n might correspond to a time t_n, location l_n and user u_n. Interpretations can then be considered equivalent if they pertain to the same time and location. Denotations can be equal, such as the denotations of s_1 and s_3 in Fig. 1. But in many cases the relation $\sim_D$ only presents a weak similarity represented via a tolerance relation which is reflexive and symmetric but need not be transitive. In Fig. 1 the denotations of s_2 and s_4 might be considered similar to each other. The denotation of s_2 might also be similar to '26 cents' which might not be similar to the denotation of s_4. The following definition transfers linguistic notions to their more abstract semiotic counterparts:

Definition 2. *For a semiotic relation, two signs* (i_1, r_1, d_1) *and* (i_2, r_2, d_2) *are:*

- synonyms $\Leftrightarrow d_1 \sim_D d_2$,
- polysemous $\Leftrightarrow r_1 \approx_R r_2$ *and* $d_1 \sim_D d_2$,
- homographs $\Leftrightarrow r_1 \approx_R r_2$ *and* $d_1 \nsim_D d_2$.
- ambiguously polysemous $\Leftrightarrow$ *polysemous and* $i_1 \not\approx_I i_2$,
- simultaneously polysemous $\Leftrightarrow$ *polysemous and* $i_1 \approx_I i_2$.

It follows from the definition that every sign is synonymous and polysemous to itself and that polysemy is a special kind of synonymy. In Fig. 1, the signs s_1 and s_3 are synonyms. Whether s_2 and s_3 are polysemous depends on $\sim_D$. In general, modelling data with SCA always depends on decisions made by a user of SCA. Homographs occur if a representamen is used with totally unrelated denotations. Usually, homographs can be avoided by renaming representamens. For example, the word "lead" could be disambiguated into "lead (verb)" and "lead (metal)". As mentioned before, interpretations may be considered equivalent if they pertain to the same time and location. Thus ambiguous polysemy occurs if a representamen is used with slightly different denotations in different usage contexts whereas simultaneous polysemy affects mostly compound signs which have many denotations simultaneously.

Similar but not identical to Peirce's distinction between icon, index and symbol, three types of interpretations can be distinguished: interpretations are *arbitrary* (following Saussure's notion of 'arbitraire') if the assignment between denotations and representamens is conventional without any detectable pattern. Interpretations are *algorithmic* if the relationship between representamens and denotations follows some pattern. For example, if a further sign with representamen '16' is added to Fig. 1 and the information is provided that it is built in the same manner as s_1 then an algorithm can determine that its denotation is '16 minutes'. Last but not least, interpretations are *observational*[1] if the denotations can be derived directly from the representamens using some general background knowledge. For example, if 'quarter' is represented graphically, it may be possible to determine its meaning just from its shape and usage context. The distinction between the three types of interpretations is not mathematically precise but again depends on judgement of a user of SCA.

3 Decomposing Compound Representamens

Triadic signs are more efficient than binary representations because n interpretations and m representamens can represent $n \times m$ denotations. If representamens are compounds such as pairs, then a sign user can express up to $n \times m^2$ denotations while only having to memorise a total of $n + m$ interpretations and representamens. Thus forming compound representamens increases the number of denotations that can be expressed with a fixed vocabulary.

[1] In analogy to Stapleton et al.'s (2017) notion of 'observability'.

Definition 3. *For a set R of representamens with a partial order $(R, \leq)$, a representamen r is a* compound representamen *if* $\exists r_p \in R$ *with* $r_p < r$ *and a* part representamen *if* $\exists r_w \in R$ *with* $r < r_w$.

Modelling part-whole relationships can be challenging (cf. Priss 1998). In contrast to a mathematical partial order, a part-whole relationship might contain multiple copies of the same part, such as a sentence with multiple copies of a word. Furthermore, different types of part-whole relationships might be involved. For example, words as parts of sentences present a different type of part-hood compared to letters as parts of sentences which establish pronunciation instead of meaning. Last but not least, identifying parts within a compound is also a matter of interpretation and the role of each part is determined by an interpretation. Therefore a focus of the modelling of part-whole relationships in SCA is on interpretations. The pairs of part representamens and their interpretations form a multiset where multiple copies of elements can exist. In this paper, duplicate elements of a multiset are differentiated by a natural number as an index but the index 0 is usually omitted. The interpretations in I_π are partial functions $i : R \twoheadrightarrow D$.

Definition 4. *For $(R, \leq)$, a set $I_\pi \subseteq I$ of interpretations and a set $R_\phi \subseteq R$, a* decomposition *is a function* $\phi(r) := \{(i_p, r_q)_n \mid i_p \in I_\pi, r_q < r, r_q \in R_\phi, n \in \mathbb{N}_0\}$ *where the multiset contains exactly one pair for each occurrence of an $r_q \in R_\phi$ in r. A function $\phi_\star(r)$ is defined analogously as a set without multiple occurrences. The corresponding sets of decompositions are denoted by Φ and $\Phi_\star$. The sets of interpretations and representamens, respectively, occurring in $\phi(r)$ are denoted by $\phi(r)_{|I}$ and $\phi(r)_{|R}$.*

Because $i_p \in I_\pi$ is an interpretation, each element (i_p, r_q) of $\phi(r)$ constitutes a sign $(i_p, r_q, i_p(r_q))$ itself. Table 1 shows examples of representamens decomposed according to Def. 4. The number 74 is decomposed into a part '7' with an interpretation i_{d10} 'multiply by 10' and a part '4' that is multiplied by 1. The denotation of '7' together with i_{d10} is 70. The algorithm for how the parts are put together (e.g. adding 70+4) is not explicitly represented. Different decompositions can exist for a representamen. For instance, the representamen 74 can also be decomposed into a '7' interpreted as a decimal position and a '4' in a unary position. As a Roman numeral, LXXIV contains 5 parts, 4 of which are added and one is subtracted. The numeral X occurs twice with the same interpretation.

Proposition 1. *For $(R, \leq)$ with a decomposition $\phi \in \Phi$:*

a) $(r \in R_\phi$ *and* $r < r_1) \iff r \in \phi(r_1)_{|R}$
b) $\phi(r_1)_{|R} \subseteq R_\phi$
c) $r_1 = r_2 \Rightarrow \phi(r_1) = \phi(r_2)$
d) $r_1 \leq r_2 \Rightarrow \phi(r_1)_{|R} \subseteq \phi(r_2)_{|R}$

Proof: a) and b) follow from Def. 4. c) is true because ϕ is a function. d) follows from a) because $r \in \phi(r_1)_{|R} \Rightarrow r < r_1 \Rightarrow r < r_2$ and $r < r_2 \Rightarrow r \in \phi(r_2)_{|R}$.

Proposition 1c) and d) only apply if the same R_ϕ is involved. For example, in Table 1, (3,5)< {(3,5),(4,6)} and $\phi_g((3,5)) \subseteq \phi_g(\{(3,5),(4,6)\})$ but $\phi_g((3,5)) \nsubseteq \phi_{g1}(\{(3,5),(4,6)\})$. Def. 4 does not force a representamen to always be used with the same interpretation or that representamens are physically non-overlapping.

Table 1. Examples of decompositions (representamens in bold without quotes)

$\phi_g(\mathbf{74}) = \{(i_{d10},\mathbf{7}), (i_{d1},\mathbf{4})\}$
$\phi_p(\mathbf{74}) = \{(i_{p1},\mathbf{7}), (i_{p2},\mathbf{4})\}$
$\phi_g(\mathbf{LXXIV}) = \{(i_{add},\mathbf{L}), (i_{add},\mathbf{X}), (i_{add},\mathbf{X})_1, (i_{sub},\mathbf{I}), (i_{add},\mathbf{V})\}$
$\phi_p(\mathbf{LXXIV}) = \{(i_{p1},\mathbf{L}), (i_{p2},\mathbf{X}), (i_{p3},\mathbf{X}), (i_{p4},\mathbf{I}), (i_{p5},\mathbf{V})\}$
$\phi_g(\mathbf{\{(3,5),(4,6)\}}) = \{(i_{le},\mathbf{3}), (i_{ri},\mathbf{5}), (i_{le},\mathbf{4}), (i_{ri},\mathbf{6}), (i_m,\mathbf{\{}), (i_m,\mathbf{(}), (i_m,\mathbf{(})_1, (i_m,\mathbf{,}),$
$\quad (i_m,\mathbf{,})_1, (i_m,\mathbf{,})_2, (i_m,\mathbf{)}), (i_m,\mathbf{)})_1, (i_m,\mathbf{\}})\}$
$\phi_p(\mathbf{\{(3,5),(4,6)\}}) = \{(i_{p1.1},\mathbf{3}), (i_{p1.2},\mathbf{5}), (i_{p2.1},\mathbf{4}), (i_{p2.2},\mathbf{6}), (i_m,\mathbf{\{}), (i_m,\mathbf{(}), ...(i_m,\mathbf{\}})\}$
$\phi_{g1}(\mathbf{\{(3,5),(4,6)\}}) = (i_e,\mathbf{(3,5)}), (i_e,\mathbf{(4,6)}), (i_m,\mathbf{\{}), (i_m,\mathbf{\}}), (i_m,\mathbf{,})\}$
$\phi_g(\mathbf{(3,5)}) = \{(i_{le},\mathbf{3}), (i_{ri},\mathbf{5}), (i_m,\mathbf{(}), (i_m,\mathbf{,}), (i_m,\mathbf{)})\}$
with $R_{\phi_g} = R_{\phi_p} = \{1, 2, ..., 9, I, V, X, L, C, \{, \}, (,), ...\}$ and $R_{\phi_{g1}} = \{(1, 1), (1, 2), ...\{, \}, ...\}$

One might distinguish *positional* decompositions and interpretations which describe the exact or relative position of parts within compounds, for example using coordinates, and *grammatical* decompositions and interpretations which determine some type or category for each representamen. For images, a bitmap decomposes an image positionally by providing locations (as interpretations) for each colour pixel (as representamens). A grammatical decomposition of an image is a vector graphical representation. In Table 1, interpretations indexed with *p*... and decomposition ϕ_p are positional. Both grammatical and positional decompositions may contain an interpretation i_m which indicates punctuation marks. A part together with a grammatical interpretation tends to generate a sign that is useful even outside the context of a compound, parts with positional interpretations usually do not. For example, a sign for '70' (extracted from '74') might be useful, but a sign for '7 in the decimal position' is probably not of interest. The next definition discusses how to reconstruct a partial order amongst representamens from decompositions.

Definition 5. *For* $(R, \leq)$ *with a set* Φ *of decompositions,* $\phi \in \Phi$ *is called* reversible *if for all* $r_1, r_2 \in R$ *with non-empty* $\phi(r_1)$, $r_1 \leq r_2 \iff \phi(r_1) \subseteq \phi(r_2)$.
A set $\Phi_1 \subseteq \Phi$ *is called* reversible *if for all* $r_1, r_2 \in R$ *the following two conditions hold:*
$(\phi(r_1) \neq \emptyset$ *for all* $\phi \in \Phi_1) \Rightarrow (r_1 \leq r_2 \iff (\forall \phi_m \in \Phi_1)\ \phi_m(r_1) \subseteq \phi_m(r_2))$,
$(\exists \phi \in \Phi_1$ *with* $\phi(r_1) = \emptyset) \Rightarrow (r_1 \leq r_2 \iff (\exists \phi_n \in \Phi_1)\ \ r_1 \in \phi_n(r_2)_{|R})$.
Analogous definitions apply to a reversible $\phi_\star$ *and a reversible* $\Phi_{\star 1}$.

The condition involving ϕ_m means that all ϕ_m must agree in order for $r_1 \leq r_2$ to hold. The condition involving ϕ_n builds on Proposition 1a. A reversible ϕ ensures that some upper part of $(R, \leq)$, a reversible Φ_1 that the complete $(R, \leq)$ can be constructed from the decompositions. For Table 1, $\phi_g((3,5)) \subseteq \phi_g(\{(3,5),(4,6)\})$ but also $\phi_g((3,6)) \subseteq \phi_g(\{(3,5),(4,6)\})$. Therefore, ϕ_g is not reversible and by itself insufficient to determine $(R, \leq)$. But $\Phi_1 = \{g, g1\}$ is a reversible set of decompositions because $\phi_{g1}((3,6)) = \emptyset$ and only $(3, 5) \in \phi_{g1}(\{(3,5),(4,6)\})_{|R}$. As another example, a decomposition of a text into letters is not reversible but in combination with further decompositions corresponding to parse trees (words, parts of sentences, paragraphs and so on) a set of decompositions could be reversible. Restoring a representamen from its parts (if ϕ is bijective) also requires some rules with respect to how the parts fit together. SCA assumes that such rules are stored somewhere but they are not discussed explicitly by

SCA. The rules for restoring from a bijective positional decomposition tend to be simpler than from a grammatical decomposition. Grammatical interpretations also provide positional information but in a less deterministic manner. For example, i_{sub} states that a Roman numeral is to the left of the next larger numeral but if several 'X's occur, their order is not determined (and not relevant). For natural languages some grammatical categories determine locations (such as subject-predicate-object in English) but some are flexible. Thus both positional and grammatical interpretations have some advantages and disadvantages.

Proposition 2. *If ϕ is reversible and $\phi(r_1) \neq \emptyset$, then $r_1 = r_2 \iff \phi(r_1) = \phi(r_2)$*

Proof: '$\Rightarrow$' because ϕ is a function. '$\Leftarrow$' $\phi(r_1) = \phi(r_2) \Rightarrow \phi(r_1) \subseteq \phi(r_2)$ and $\phi(r_2) \subseteq \phi(r_1) \Rightarrow r_1 \leq r_2$ and $r_2 \leq r_1$.

Thus a reversible ϕ means that the parts contain sufficient information to identify a compound. Each Φ generates a formal context as described in the next definition, but contexts use sets instead of multisets.

Definition 6. *For $(R, \leq)$ with a set Φ of decompositions, a* decomposition context *is a formal context $(\bigcup_{\phi,r} \phi_\star(r), R, J)$ with $(i_1, r_1) J r_2 \iff r_1 = r_2$ or $(\exists \phi_\star \in \Phi_\star)$ $(i_1, r_1) \in \phi_\star(r_2)$. The corresponding lattices are called* decomposition lattices.

Fig. 2 shows an example of a decomposition lattice for ϕ_g from Table 1. In this case, the lattice order preserves the partial order amongst representamens. But that would not be the case if a representamen $(3, 6)$ was added because ϕ_g is not reversible.

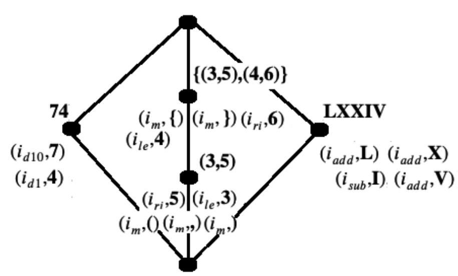


Fig. 2. A decomposition lattice for ϕ_g from Table 1

A non-reversible ϕ will lose some information about representamens. Furthermore because lattices are constructed using sets they omit information about multiple occurrences of parts. Therefore in general, a decomposition lattice generates its own partial order amongst representamens which need not correspond to $(R, \leq)$. It is therefore of interest to specify under which conditions $(R, \leq)$ is maintained in a decomposition lattice.

Proposition 3. *For $(R, \leq)$ with a reversible $\phi_\star \in \Phi$ and $(\bigcup_r \phi_\star(r), R, J)$ as a decomposition context, it follows that $(\forall r \in R)$ $r' = \phi_\star(r)$ and $r_1 \leq r_2 \iff r_1' \leq r_2'$ for all r_1 and r_2 with non-empty $\phi_\star(r_1)$.*

Proof: $r_1 \leq r_2 \iff \phi_\star(r_1) \subseteq \phi_\star(r_2) \iff r'_1 \leq r'_2$.

Thus the attribute order of a decomposition lattice for a reversible $\phi_\star$ preserves at least an upper part of $(R, \leq)$. A similar proposition can be stated for a reversible $\Phi_\star$. Below, measures for comparing semiotic relations based on their compound representamens are defined. It can be beneficial to employ a small set of part representamens (for example an alphabet of letters) to form a large set of compound representamens (for example words and texts). This often involves ambiguous polysemy of representamens because they are used with different interpretations. For example a digit can be used in the unary, decimal, centesimal and so on position of a decimal number. Measures can therefore be defined based on the length, number of interpretations, reuse of representamens and so on of decompositions. Def. 7 only shows some examples. Many similar measures can be defined.

Definition 7. *For a semiotic relation S with a reversible ϕ, measures can be defined:*

a) $len_\phi(r) := |\phi(r)|$ *as the* length *of* r.
b) $\mu_\phi(r) := |\phi(r)_{|I}|$ *as the number of different interpretations.*
c) $rpt_\phi(r_p) := max_{(i,r)}(n \mid (i, r_p)_n \in \phi(r))$ *as the maximal repetition of any* (i, r_p).

These can be extended to measures for semiotic relations considering maximal, minimal or average values. For example, $\mu_{max}(S) := max_r(\mu_\phi(r))$, $\mu_{min}(S) := min_r(\mu_\phi(r))$ *and* $\mu_{avg}(S) := avg_r(\mu_\phi(r))$.

For these measures a low value is preferable. The measures should be balanced against each other. A semiotic relation that has low values for one measure need not have low values for any other measures. The signs of the decimal and Roman numeral representation of '74' are synonyms, but their measures are different: len_ϕ('74') $= 2$, len_ϕ('LXXIV') $= 5$, μ_ϕ('74') $= \mu_\phi$('LXXIV') $= 2$, rpt_ϕ('X') $= 1$ and $rpt_\phi() = 0$ for all digits of a decimal number. But len_ϕ('100') $= 3 > 1 = len_\phi$('C'). Therefore, in some cases, the measure for decimal numbers is better than for Roman numerals and vice versa.

So far compound representamens and signs established by their parts have been discussed, but not yet signs which have compound representamens. Because signs are triadic, a partial order for any of the three components can contribute to a partial order amongst the signs. This section suggests that an *extended denotation* of a compound sign consists of its own polysemous meanings as well as the set of meanings attributed to its parts.

Definition 8. *For a semiotic relation S with $(R, \leq)$ and a set Φ of decompositions, a sign (i, r, d) is called a* compound sign *if* $\exists \phi \in \Phi$ *with* $\phi(r) \neq \emptyset$.

Requiring $\sim_D$ to be an equivalence relation as in the next definition, partitions the set of denotations so that, if no homographs exist, each representamen is mapped into a partition via its interpretations. For natural language words, such conditions are too strong unless the partitions of $\sim_D$ are large and general, for example, denoting a content domain. For signs representing mathematical content, individual denotations can be very succinct so that each denotation correlates with a clearly definable concept.

Definition 9. *For a semiotic relation S with $(R, \leq)$ and a set Φ of decompositions, without homographs, for a single equivalence class of $\approx_I$ and where $\sim_D$ is an equivalence relation, a compound sign $s := (i, r, d)$ has an* extended denotation $D(s) := \{d_1 \mid (\exists\phi)(\exists(i_1, r_1) \in \phi(r))\ i_1(r_1) = d_1, d \sim_D d_1\} \cup \{d_1 \mid (\exists(i_1, r, d_1) \in S)\ d \sim_D d_1\}$. *The extended denotation of a set $S_1 \subseteq S$ is $D(S_1) := \bigcup_{s \in S_1} D(s)$.*

Thus compound representamens lead to compound signs which are simultaneously polysemous and have potentially large sets of extended denotations. Def. 9 provides a formal explanation for why a picture might be "worth a 1000 words".

4 An Application to Euler Diagrams

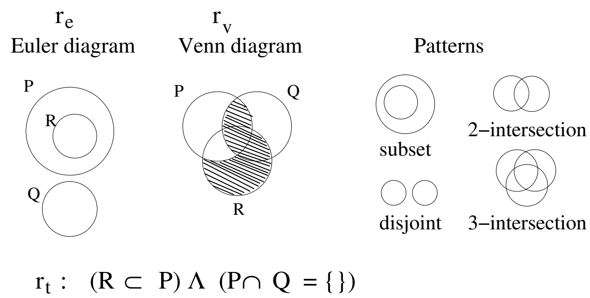


Fig. 3. Euler diagram, Venn diagram and decomposition patterns

This section applies the SCA theory described so far to an example of diagrams frequently used in teaching mathematics. The left hand side of Fig. 3 shows an example (referred to in this section as r_e) of an Euler diagram from Stapleton et al. (2017) for the statement $(R \subset P) \wedge (P \cap Q = \emptyset)$ which is referred to as r_t. In Stapleton's terminology r_e provides a "free ride" or "observations" because it also shows that $R \cap Q = \emptyset$ contrary to r_t from which $R \cap Q = \emptyset$ can be concluded but not observed. Using SCA, r_t can be decomposed into a conjunction: $\phi_{conj}(r_t) = \{(i_{term},$'$R \subset P$'$)$, $(i_{term},$'$P \cap Q = \emptyset$'$)$, $(i_{op},$'$\wedge$'$)\}$. The representamen r_e can be decomposed into zones: $\phi_{zones}(r_e) = \{(i_{Q\overline{PR}},\text{zn}), (i_{P\overline{RQ}},\text{zn}), (i_{PR\overline{Q}},\text{zn}), (i_{\overline{PQR}},\text{zn})\}$, but other decompositions of r_e are also possible.

Fig. 4 displays a decomposition lattice for ϕ_{zones} for representamens of well-formed Euler diagrams for 3 sets. Euler diagrams are considered well-formed if they do not contain triple points, adjacent lines, disconnected zones and so on (Flower, Fish, & Howse 2008). For many configurations of sets, it is not possible to draw a well-formed Euler diagram. Contrary to what might be intuitively expected of a partial order amongst representamens, the partial order in Fig. 4 does not consist of adding or deleting graphical elements. Instead a representamen r_1 is beneath a r_2 if r_2 can be obtained from r_1 by moving (including possibly enlarging) curves in a manner that adds zones without losing zones. Thus apart from replicating previously known partial orders, decomposition lattices can potentially provide tacit information a user was not previously aware of.

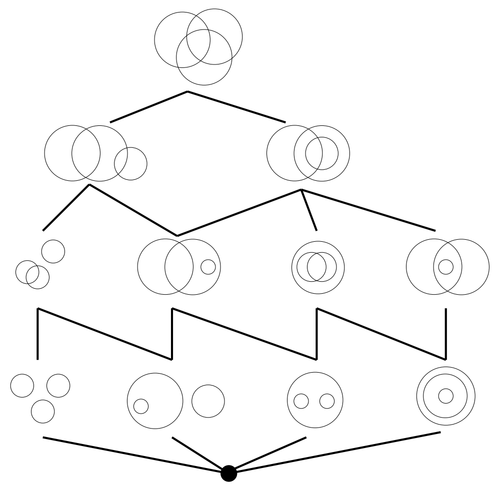


Fig. 4. A decomposition lattice for well-formed Euler diagrams for 3 sets

The patterns on the right hand side of Fig. 3 show different possibilities for two sets to relate to each other using well-formed diagrams (disjoint, intersect or subset). For n sets only the intersection of all of them is new, all other intersections affect $n-1$ sets. Thus there are n+1 patterns for n sets with respect to well-formed Euler diagrams. Considering such patterns, a decomposition of r_e is: $\phi_{pattern}(r_e) = \{(i_{RP}$,subs), $(i_{PQ}$,disj), $(i_{RQ}$,disj)$\}$. According to Def. 4, a decomposition has a pair for each occurrence of a part. Therefore $\{(i_{RP}$,subs), $(i_{PQ}$,disj)$\}$ would not be a decomposition of r_e. A further decomposition that only focuses on circles might be: $\phi_{circ}(r_e) = \{(i_P$,'○'), $(i_Q$,'○'), $(i_R$,'○')$\}$. Each decomposition of r_e highlights another possibility for reading the diagram. Other decompositions could be defined for r_t, but compared with r_e there are fewer possibilities and the extended denotation of a sign with r_t as a representamen would be smaller than a sign with r_e as a representamen.

Table 2. Comparing semiotic relations for representing n sets

	$\|R\|$	$\|R_\phi\|$	$\mu_{max}(S)$	polysemy (for R_ϕ)	reversible
$\phi_{conj}(r_t)$	2^{2^n}	2^n	1	no	yes
$\phi_{zones}(r_e)$	2^{2^n}	1	2^n	yes	yes
$\phi_{pattern}(r_e)$	2^{2^n}	$n+1$	2^n	yes	yes
$\phi_{circ}(r_e)$	2^{2^n}	1	n	yes	no

Whether or not observations can actually be seen by users is a different question. If diagrams are too small, large or complex, users may not be able to make every possible observation. Students, teachers, experts, people with a visual disability or dyslexia will have different skills for visually parsing representamens. With respect to patterns, it is a modelling question whether patterns with more than two sets should also be considered because they are visually much more difficult to detect. In any case, Euler and Venn diagrams for more than 3 sets can become difficult to visually parse.

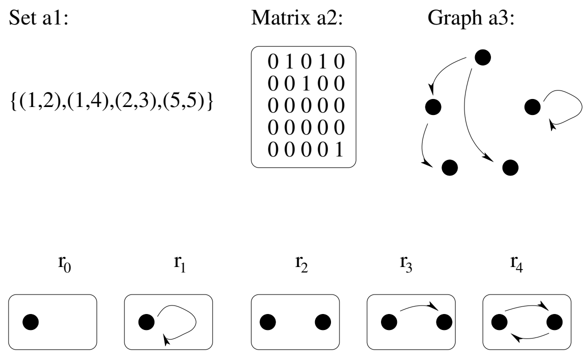
Table 2 summarises some measures for the decompositions of Euler diagrams. In this case all interpretations are algorithmic. Thus users only need to learn the principle of how to read Euler diagrams in order to determine all denotations. There is no repetition in the diagrams because each circle refers to a different set. The decomposition ϕ_{zones} might be positional or grammatical for Euler diagrams but is positional if it is used to shade the zones of a previously drawn Venn diagram (as r_v in Fig. 3). The decomposition ϕ_{circ} can be made positional if the interpretations include vector coordinates for the circles. The parts of r_t are not polysemous because each term has exactly one meaning. Graphical elements in r_e are polysemous because they are independent of the actual labels. For example, the same representamen (a circle) is used for every set and thus creates as many polysemous signs as there are sets. If parts are polysemous, then $|R_\phi|$ can be smaller. For ϕ_{zones} and ϕ_{circ} the polysemy is inefficiently high because $|R_\phi| = 1$.

5 An Application to Diagrams for Binary Relations

Apart from graphical representations of set theory, a further example which is also relevant for teaching introductory mathematics, is the visualisation of binary relations. The top half of Fig. 5 shows a binary relation as a set, as a matrix which contains a '1' if the row and column element are a pair of the relation and as a graph which has a node for each element and an arrow for each pair of the relation. For matrices it should be assumed that the rows and columns are sorted in a fixed sequence. The middle of Fig. 5 shows five representamens that are patterns that can be observed in a graph diagram. For a single node in a graph, the node either has no arc to itself as in r_0 or one arc to itself as in r_1. Any pair of two nodes corresponds to a representamen r_2, r_3 or r_4. The bottom half of Fig. 5 displays decompositions for the three representamens depicted in the top.

In the following, a_1 is used for a representamen of type "set", a_2 for type "matrix" and a_3 for type "graph". Table 3 calculates measures for semiotic relations containing only one of the three types for a set with $m = 4$ elements. Semiotic relations for a_1 and a_2 contain a maximum of 2^{16} representamens because each pair of elements either exists or does not exist. The number of different graphs for binary relations of 4 elements is 3044 and can be obtained from the sequence Number A000595 in the On-Line Encyclopedia of Integer Sequences[2]. Therefore a semiotic relation for a_3 has much fewer representamens than semiotic relations for a_1 and a_2. One could argue that the rows and columns of a matrix are also not labelled. But while it is feasible to observe whether two unlabelled graphs are isomorphic, it is difficult for larger matrices to determine whether one matrix is the result of permutation of the other one by just visually comparing the matrices. Each compound representamen of type a_3 polysemously represents different binary relations if the nodes are not labelled. The table shows that lower values for one measure usually correlate with larger values for other measures. But overall, a_3 outperforms the other types of representamens.

[2] http://oeis.org/A000595



$\phi_g(a_1) = \{(i_e, `(1,2)'), (i_e, `(1,4)'), (i_e, `(2,3)'), (i_e, `(5,5)'), (i_m, `\{'), (i_m, `\}'), (i_m, `,')_2\}$
$\phi_p(a_2) = \{(i_{p1.2}, `1'), (i_{p2.3}, `1'), (i_{p1.4}, `1'), (i_{p5.5}, `1'), (i_{p1.1}, `0'), ..., (i_{p4.5}, `0')\}$
$\phi_p(a_3) = \{(i_{p1.2}, r_3), (i_{p2.3}, r_3), (i_{p1.4}, r_3), (i_{p5.5}, r_1)\}$
$\phi_g(a_3) = \{(i_{pair}, r_3)_2, (i_{ind}, r_1)\}$

Fig. 5. Different representamens with decompositions for binary relations

Table 3. Comparing semiotic relations for binary relations with $m = 4$ elements

of type	$\|R\|$	a_i polysemous	$\|R_\phi\|$	$len_{Max}(S)$	$\mu_{Max}(S)$	$rpt > 0$
a_1	$2^{m^2} = 65536$	no	$m^2 + 3 = 19$	$2m^2 - 1 + 2 = 33$	2	yes (“,”)
a_2	$2^{m^2} = 65536$	no	2	m^2=16	m^2=16	no
a_3 with ϕ_p	3044	yes	5	$4 + 6$	$4 + 6$	no
a_3 with ϕ_g	3044	yes	5	$4 + 6$	2	yes

Table 4 compares semiotic relations of binary relations as sets, matrices or graphs with respect to whether properties of binary relations (as defined in the table) can be observed from them. As mentioned before, an application of SCA always involves judgement. In particular, whether some property can be observed or can only be computed may depend on who is observing a representamen. It would be possible, however, to experimentally measure the time and accuracy which users need to determine properties of binary relations from a representamen and use that as an indication for whether users observe or calculate. A testable hypothesis is that observation would be faster and more accurate. In our judgement, for sets (a_1) properties always need to be calculated by looking at each individual pair and evaluating whether the defining condition is fulfilled or not. For matrices (a_2) the first five properties can be observed. Transitivity cannot be observed. Semiconnex and connex need to be checked for each pair. For graphs (a_3) all properties apart from transitivity can be observed from the part representamens $r_0, ..., r_4$ together with the quantifiers ‘no’ and ‘all’. It might be possible to describe how transitivity can be observed, but that is not as easy and would still need to be checked for many triples of elements.

Students need to learn to switch between the different representamen types depending on a task. For example, an empty relation cannot be represented using a represen-

Table 4. Observability of properties of binary relations

property	definition	as set (a_1)	as matrix (a_2)	as graph (a_3)
reflexive	$(\forall a \in A)\, aRa$	(compute)	filled diagonal	all r_1
irreflexive	$(\forall a \in A)\, \neg aRa$	(compute)	empty diagonal	no r_1
symmetric	$(\forall a, b \in A)\, aRb \rightarrow bRa$	(compute)	symmetric	no r_3
asymmetric	$(\forall a, b \in A)\, aRb \rightarrow \neg(bRa)$	(compute)	asymmetric	no r_1, no r_4
antisymmetric	$(\forall a, b \in A)\, aRb$ and $bRa \rightarrow a = b$	(compute)	asym. or diag.	no r_4
transitive	$(\forall a, b, c \in A)\, aRb$ and $bRc \rightarrow aRc$	(compute)	(compute)	(compute)
semiconnex	$(\forall a \neq b \in A)\, aRb$ or bRa	(compute)	(compute)	no r_2
connex	$(\forall a, b \in A)\, aRb$ or bRa	(compute)	(compute)	all r_1, no r_2

tamen of type a_3. A common misconception that students have[3] is that relationships between properties can be observed (instead of computed) from a_1. For example, students might think that the properties semiconnex and connex are closely connected to the property of symmetry because the definitions ‘look similar’. In that case students are incorrectly applying an interpretation of observation to parts of logical formulas mostly because they have not yet sufficiently learned to read quantifiers and logical operators.

6 Conclusion

A purpose of SCA is to make tacit knowledge explicit. The different methods presented in this paper support this purpose. Middendorf & Pace (2004) develop an educational method of *decoding the disciplines* that helps educators to discover tacit knowledge and skills in their discipline so that they can anticipate misconceptions and difficulties that students might encounter. An analysis with SCA as presented in this paper can support such a process of decoding the disciplines by first determining what kinds of interpretations teachers use when they read and write mathematical representations and by then formally recording these as decompositions. Finally, these decompositions can be further analysed with SCA and FCA in order to detect implicit structures using some form of conceptual exploration.

The application of SCA can be quite technical as presented in the previous section. It is not intended that every teacher conducts her or his own analysis with SCA, but instead that such analyses and results are shared. A conclusion for the two examples presented in this paper is that positional interpretations are usually preferred for formal modelling and software implementations because they are easily restorable and allow a translation between different representations. For human users, however, graphical representations are often more advantageous because they have a high simultaneous polysemy which means that many statements can be observed from them simultaneously. But the patterns involved in making observations from diagrams are not necessarily self-explanatory and need to be explicitly taught. Many aspects of part-whole relationships amongst diagrams have been addressed in this paper, but many are also still open for further research.

[3] Based on personal teaching experience.

References

1. Ainsworth, S. (1999). The functions of multiple representations. Computers & Education, 33, 2-3, p. 131-152.
2. Barwise, J.; Seligman, J. (1997). Information Flow. The Logic of Distributed Systems. Cambridge University Press.
3. Flower, J.; Fish, A.; & Howse, J. (2008). Euler diagram generation. Journal of Visual Languages & Computing, 19, 6, p. 675-694.
4. Ganter, B.; Wille, R. (1999). Formal Concept Analysis. Mathematical Foundations. Berlin-Heidelberg-New York: Springer.
5. Ganter, B.; Obiedkov, S. (2004). Implications in triadic formal contexts. In International Conference on Conceptual Structures, Springer, p. 186-195.
6. Goguen, J. (1999). An introduction to algebraic semiotics, with application to user interface design. Computation for metaphors, analogy, and agents. Springer, p. 242-291.
7. Middendorf, J.; Pace, D. (2004). Decoding the disciplines: A model for helping students learn disciplinary ways of thinking. New directions for teaching and learning, 98, p. 1-12.
8. Moody, Daniel (2009). The 'physics' of notations: toward a scientific basis for constructing visual notations in software engineering. IEEE Transactions on software engineering, 35, 6, p. 756-779.
9. Priss, Uta (1998). The Formalization of WordNet by Methods of Relational Concept Analysis. In Fellbaum, Christiane (ed.), WordNet: An Electronic Lexical Database and Some of its Applications, MIT press, p. 179-196.
10. Priss, Uta; Old, L. John (2010). Concept Neighbourhoods in Lexical Databases. In Kwuida; Sertkaya (eds.), Proceedings of the 8th International Conference on Formal Concept Analysis, ICFCA'10, Springer Verlag, LNCS 5986, p. 283-295.
11. Priss, Uta (2017). Semiotic-Conceptual Analysis: A Proposal. International Journal of General Systems, Vol. 46, 5, p. 569-585.
12. Priss, Uta (2019). Applying Semiotic-Conceptual Analysis to Mathematical Language. In Alam, Sotropa, Endres (eds.), Graph-Based Representation and Reasoning, Proceedings of ICCS'19, Springer Verlag, LNCS 11530, p. 248-256.
13. Priss, Uta (2022). A Semiotic Perspective on Polysemy. Annals of Mathematics and Artificial Intelligence. Available at https://doi.org/10.1007/s10472-022-09795-1
14. Stapleton, G.; Jamnik, M.; Shimojima, A. (2017). What makes an effective representation of information: a formal account of observational advantages. Journal of Logic, Language and Information, 26, 2, p. 143-177.