Towards Thought as a Logical Picture of Signs

J.J. Sarbo, S. Hoppenbrouwers and J.I. Farkas
University of Nijmegen, Toernooiveld 1, 6525 ED
Nijmegen, The Netherlands, e-mail: janos@cs.kun.nl

Abstract We are concerned with the problem of summarizing the contents of a coherent text. In this paper we argue that complex units of symbols like sentences, for example, are signs and the meaning of a text arises via their interaction. We introduce a model for the generation of summaries and illustrate its potential by a realistic example.

1   Introduction

When we read a text, in the end, we may have a single thought which is our comprehension of its meaning. Even if the text is long and complex we may be able to summarize its contents by a single sentence. Such a process of summarization could be alternatively called *meaning extraction*.

Traditionally, text summarization is typically based on a statistical or syntactical analysis ([Jon93], [EN98]). We argue that, in virtue of their formal character, such approaches may not be able to find the 'natural' meaning represented by a text. In this paper we introduce an approach which stresses the sign character of language. We will assume that language consists of symbols which are signs, that such signs arise from input stimuli via a cognitive process (making use of existing lexical knowledge) and, finally, that their meaning emerges via mediation.

Recently we introduced such a sign-based model for the syntactic analysis of sentences, which will be extensively referred to in this paper ([FS00], [SF01]). Here, we will argue that a similar model applies if complex units of symbols like sentences themselves are considered as signs. Such a model can explain how complex units may contribute to a text, as signs of thought. We illustrate the proposed approach by a realistic example of a text fragment found in a thesis ([Hui96]).

1.1   Informal Analysis

Our sample text (cf. fig. 1) concerns the definition of the concept of information retrieval system. Before introducing any theory, let us informally analyse the sentences of the given fragment.

The analysis as presented in sect. 4 assumes a *syntactic* analysis of some sort, and therefore we need to provide at least a reasonably detailed analysis at that

label	text	abbreviation
1	There are several document-bases.	*sev, docb*
2	Each document-base contains	*docb*
	different types of information.	*dt-of-info*
3.1	There are various types of users and	*vt-of-users*
3.2	there are vast differences between	
	their information needs.	*vdiff, t-ineeds*
4.1	There are various kinds of search-tasks.	*vk-of-st*
4.2	or stated differently.	
4.3	there are several ways in which a user	*sev-ws*
	can be satisfied with the returned information.	*canb-satf, ret-info*

Fig. 1: Sample text

level. Note, however. that there are certainly other possible analyses than the one presented here, quite probably rendering different semantics. The primary goal of our informal analysis is merely to describe the clusters of language symbols we take to be the input of our further analysis. We assume such an analysis to be strictly based on syntactic properties of the input symbols. The syntactic terminology largely follows a standard work on descriptive English grammar ([QGLS85]). S=subject. V=verb, C=Complement. O=object. A= adverbial.

Though we aim to develop a formal approach to meaning extraction. in the course of the informal pre-analysis presented in this section we tentatively make use of some concepts related to "text grammar" or "naive" (informal) semantics. The hidden agenda of the section is to work towards a potential interpretation of the parts of the text and then relate them to the meaning of the text as a whole.

In (1) above we have so-called existential *there* (S) as the 'grammatical subject'. followed by main verb *be* (V). *document-bases* as the 'notional subject' (O) with the quantifier *several*. *There* is a syntactic place-holder in the regular subjective position. whereas the 'notional subject' is the compound noun taking the object position. *There* and *several document bases* both are of a subjective nature. be it at different levels of analysis.

In (2) we observe: *Each document base* (S): *Each*. which relates to the quantifier *several* in (1): *contains* (V): an objectival noun phrase headed by *information* and modified by genitive *different types (of)*. It is asserted that information is held in each document base, and that this information comes in different types for each document-base. If we assume that no other properties of the document-base are relevant in context. it can be argued that information held by it is in fact what defines document-bases. Therefore. in context *contains* is equivalent to *be*.

The analysis of (3.1) is similar to that of (1): coordinating *and* simply joins the two clauses (3.1) and (3.2). In (3.2) we again have existential *there* (S), *vast differences* heads the object, and the adverbial prepositional phrase makes explicit

138

that those differences hold between the information needs of the users; the possessive pronoun *their* refers to *users* in (3.1).

(4.1) is analyzed similar to (1). (4.2) is an explicit statement at text grammatical level, indicating that (4.1) is paraphrased in (4.3). Interestingly, we are presented with two different expressions of what must essentially be the same meaning. It is to be expected that the two phrasings contribute to the general meaning of the complete text by presenting the same essential meaning by means of two quite different expressions.

In (4.3), not unlike in (2), the *several ways* are not only complemented by *in which...*, but because no other information is relevant in context concerning those *ways*, the complement defines *ways*. In context, the complement is arguably more meaningful than *several ways*, and in any case is linked to it tightly. From a semantic point of view, the prepositional construction *in which ...* can be seen as similar to an explicit predication using *be*.

Interestingly, a definite article is used in *the returned information*, supposedly referring to *returned information* already understood to be there by the reader, whereas no such returned information has been explicitly mentioned in the text. Closest comes the *types of information* in (2); *returned* is never explicitly mentioned, only hinted at in (4.1) (assuming that tasks have something to do with the returning of information). In fact, it could be argued that, given that (4.1) already introduces *search tasks*, the explicit introduction of *returned information* in the text is the most important semantic contribution of (4.3).

Having identified clusters of symbols, we now turn to the question why these clusters, given the particular order in which they occur, do indeed define something meaningful that amounts to a definition of *information retrieval system*.

2 A Semiotic Model of Language

In our analysis of signs we follow the principles of Peirce's semiotics ([Pei31], [Tej88]). Accordingly, a sign signifies its object to an agent in some sense, which is called the interpretant of the sign. The irreducible relation of sign, object and interpretant (each of which is a sign, recursively) is called the *triadic relation* of sign. We assume that the ground for any sign is a *contrast* in the 'real' world. How is such a contrast cognitively observed?

2.1 Cognitive Basis

Following cognition theory ([Har87]), the recognition of any sign must begin with the sensation of the physical input. Physical stimuli enter the human receiver via the senses which continuously transform the raw data into internal sensation. The output of the senses, a bio-electric signal, is processed by the brain in percepts. The generation of such a percept is triggered by a change in the input, typically, or by
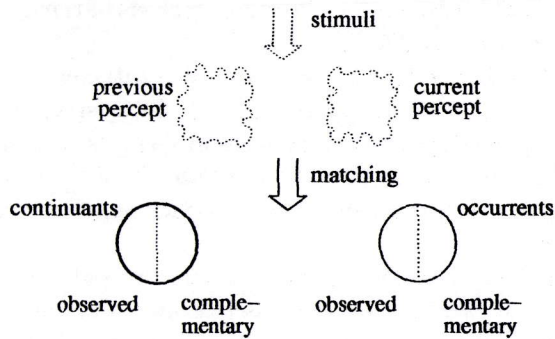
Fig. 2: Cognitive model of sign recognition

the duration of some sampling time, e.g. in the case of visual perception.

The brain compares the current percept with the previous one, and this enables it to distinguish between two sorts of input qualities: one, which was there and remained there, something stable, which we will call a *continuant*; and another, which was not there, but is there now (or vice versa), something changing, which we will call an *occurrent*. The collections of continuants and occurrents, which are inherently related to each other, form the basis for our perception of a phenomenon as a sign. We also assume that, by means of *selective attention*, we recognize in these collections coherent sets of qualities: the qualities of the *observed* and those of the *complementary* part of the phenomenon. We will refer to these sets collectively as the *input*. We will assume that these sets are the *primary* signs we observe: qualities which are signs. Such signs, which are called in Peirce's terminology a *qualisign*, are special signs for which we have no denotation. Although qualisigns are coherent, by definition, we experience them as independent signs.

In [FS00] we have shown that the above model allows for qualisigns to be interpreted as Boolean logical functions. If the above collections are represented as sets (in the mathematical sense), we can define our 'universe' as the union of the sets of qualities of the qualisigns. Notice that these sets are *coherently* related to each other, contrary to formal logic in which the universe is an *arbitrary* set. If the sets are finite, the universe is a 'closed' world by definition. A schematic diagram of our cognitive model of signs and their logical representation is depicted in fig. 2.

2.2 Classification of Sign

In his semiotic theory, Peirce defined an ingenious classification of signs. In his view, the most complete signs are the icon, index, and symbol which represent their object on the basis of, respectively, *similarity, causality* and *arbitrary consensus*. Besides this taxonomy, Peirce also distinguishes signs, respectively, according to the categorical status of the sign, and according to the relationship between object
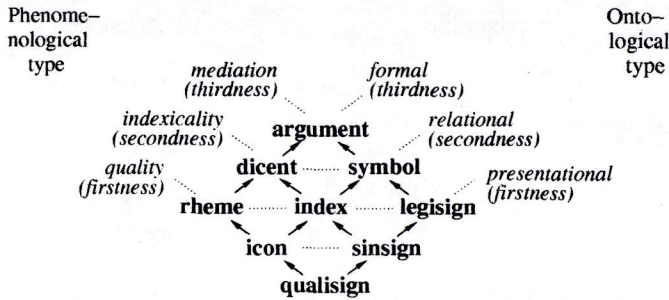
Fig. 3: Peirce's classification of signs

and interpretant. From a categorical perspective. signs can be qualisigns, sinsigns or legisigns, which correspond, respectively, to firstness, secondness and thirdness. In other words, a sign can be a *quality*, an *actual event*, or a *rule*. Seen from the perspective of the relationship between object and interpretant, a sign may be a rheme, a dicent or an argument. In other words a sign may signify a *qualitative possibility*, an *actual existence*, or a *proposition*. Thus we obtain nine kinds of sign which may be arranged in a matrix as shown in fig. 3 (the meaning of the horizontal lines and directed edges will be explained later). Although Peirce also defined more complex systems of signs. we hold that his 'simple' classification is the most practical ([DFS99],[FS99]). We argue that his signs can be identified in different symbol phenomena like Boolean logic. syntax. and syllogistic logic. The first two of these have been discussed in [FS00] and [SF01], the last one is the subject of the current paper.

2.3 A Model of Signs

How do complex signs emerge? We argue that semiosis (the generation of the interpretant) can be defined as a process in which trichotomic relations emerge recursively, revealing gradually more accurate and clear *approximations* of the full richness of a sign of the observed phenomenon. Accordingly, the proposition of the input as a sign arises from the input qualisigns via a number of other signs. In this process, icon, index and symbol signs function as sign (in the sense of the triadic relation) whereas other signs function as (the signs of) their object. It turns out ([FS00]) that semiosis can be defined as the *interaction* of *adjacent* signs (cf. the horizontal lines in fig. 3). For example, an argument sign can emerge from the interaction of a dicent and symbol sign.

In virtue of the fast and continuous nature of cognition, we will assume that the signs of a phenomenon are not recognized isolatedly, but only as 'temporary' signs which are approximations of the final proposition. We argue that such signs are *re*-presentations of the qualisigns, and their types are identical to the classes defined by Peirce. The recognition process we have in mind can be illustrated by
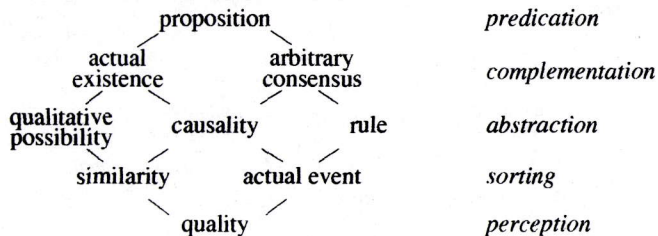
141

proposition   *predication*
          actual   arbitrary
        existence   consensus   *complementation*
    qualitative   causality   rule   *abstraction*
    possibility
          similarity   actual event   *sorting*
                    quality   *perception*

Fig. 4: Aspects of signs and stages of recognition

the perception of a motion picture. In that process, a series of pictures are input which are not recognized isolatedly, but which are necessary for observing motion as a change between the first and last picture of such a series.

According to our model ([FS00]), sign recognition proceeds in stages. Each stage is associated with a different kind of re-presentation of the qualisigns (cf. fig. 4). In the first stage the observed qualisigns are grouped by *sorting* them according to the aspects of part (icon) and whole (sinsign). Notice that any 'part' of the input is similar to it, and that the input as a 'whole' is a simultaneous occurrence of the parts, an actual event (which happens now). The qualisigns of a phenomenon are different signs, but which have a common origin. Therefore, in the second stage, icon and sinsign are compared with each other yielding the 'abstract' re-presentations of the qualisigns: the observed qualisigns independent from each other (rheme), how they are related to the complementary signs (index), and the law-like relation of their simultaneous occurrence (legisign). The index, which is alternatively called the *context*, is a pointer, linking and *converting* between part and whole. In the third stage, the abstract signs are completed by the information of the complementary ones. This yields the signs of the actual relations of the parts of the observed phenomenon (dicent), and the property characterizing them as a whole (symbol). Dicent and symbol are also called subject and predicate. Finally, the completed signs are merged to a single proposition (argument) about the final meaning of the relation between the observed qualisigns.

The above model has been applied for the derivation of a classification and ontology of Boolean logical signs. Currently we have shown that a *sequential* version of this model amounts to a model of *language* signs (morphological and syntactical). In such a model, input signs which are symbols appear one after the other as qualisigns, and complex signs arise via interaction in a *bottom-up* fashion (cf. the directed edges in fig. 3). Because the universes of such qualisigns may be independent of each other, syntactic symbols can only interact (or *bind*) if they are compatible. Whether two symbols are, or are not compatible follows from their relational properties. Such properties, which are sets of qualities, are defined class-wise. The sign of a binding is a representation of the union of the qualities of the constituent symbols, i.e of their qualisigns, recursively.

Due to their sequential nature, language symbol interactions have also degenerate forms. One of them is *accumulation*, in which an existing sign is combined with another sign of the same type. Such an interaction assigns the same meaning to both constituents thereby rendering them indistinguishable. The other one is *coercion*, in which a new sign is generated for the denotation of an existing sign. Coercion applies if the signs, which are to interact, are incapable of accumulation or binding ([SF01]).

3   Syllogistic Logic

We will argue that syllogistic signs can be modeled analogously to Boolean logical and language signs. Accordingly, in this section we will refer by a sign class to the classification of syllogistic signs. Although syllogism is the simplest form of predicate logic, its importance in human cognition is generally acknowledged.

The doctrine of syllogisms goes back to Aristotle ([Boc61]) who identified three syllogistic schemes. The differences between the schemes lie in the representation of the common terms (cf. fig. 5). In scheme 1 they are subject and predicate, respectively, in the first and second premise. In scheme 2 they are subject, and in scheme 3 predicate, in both premises. Syllogistic logic is related to semantics and reasoning. In what follows, we will briefly introduce semantic signs. Because of space, in this paper we will restrict ourselves to the description of those aspects of semantics which are needed for our sample example in sect. 4. The signs of reasoning –abduction, induction and deduction– are not specified in this paper.

3.1   Semantics

Semantics is concerned with the qualities of symbols from the categorical perspective of firstness. We argue that semantic signs arise via a re-analysis of the syntactic sign interactions. We assume that the semantic qualities of a sign can be specified in terms of the most complete triad of signs: icon, index and symbol. For example, verbs can denote existence (e.g. *be*), state (e.g. *sleep*), and transition (e.g. *give*); nouns can be expressive of a thing (e.g. *apple*), reference (e.g. the name *John*), and something symbolic (e.g. agent *John*). Such a triad can be interpreted as a 'meaning' ordering: icon<index<symbol. For example, an iconic sign of a triad is less meaningful than an indexical one of the same triad.

Although the qualities of a symbol are lexically defined with respect to its function as a sign in the sense of the trichotomic relation, a symbol can also function as object. We require that in every symbol interaction, those qualities are considered which correspond to the category aspect of the *sign*. For example, in a rheme and index interaction the qualities referring to the category of secondness are taken from *both* the rheme and the index sign. Accordingly, though syntactic and semantic signs are isomorphic by definition, their sign recognition processes are different.

143

The linking and conversion potential of the index can be identified in our model of semantic signs, as follows. We allow for two index signs to unify, either if they are equivalent (semantically), or, if they are converses of each other. The latter relation holds if the two signs refer to semantic qualisigns which are *lexical* counterparts. By definition, two qualisigns are *counterparts* if they represent, respectively, the continuant and occurrent collections of the observed part of the same phenomenon. For example, 'search' and 'way' can be related in such a way. Indeed, if we observe 'search' to occur, there may be present some 'way', in which the search is carried out, for example. Another example of a pair of counterparts is 'need' (continuant) and 'task' (occurrent). We require that the information about counterparts is lexically given.

In our approach to meaning extraction we will assume that in a symbol interaction a semantically less meaningful sign can be ignored (or *removed*) potentially. However, such a removal must also respect the sign's referential properties. By keeping track of the sets of qualities of the different qualisigns we can find out which qualities are not referred to in a later interaction. Clearly, such qualities can be removed without affecting the meaning of later signs. For example, the sign 'various' can be removed from 'various types' if the number of 'types' is not referred to in the rest of the text. Although we may only remove a qualisign when it turns out that it has no relevance for a later sign, it will be assumed that such 'knowledge' is always available. Finally, we mention that anaphoric references will be assumed to be solved via the unification of qualities. Also quantification is considered a semantic sign, in the generation of which the role of index symbols is central. In so far as the index is, amongst others, the sign of the complementary qualities of the observed part ([FS00]), it can complete a rheme or a legisign with the qualities of their referential meaning. Although such a meaning can be represented in different ways via conversion (e.g. existentially or universally), this aspect is not discussed in the paper. We will assume that the sentences of the example of fig. 1 can all be represented as universally quantified premises.

In the rest of the paper we will concentrate on syllogisms. We will assume that the input sentences (which are syntactic signs) are first recognised as semantic signs. Such signs will be considered as qualisigns in a subsequent syllogistic sign generation. We will assume that all earlier representations of such signs are available (e.g. their syntactic interactions, semantic qualities etc.). Because a sentence arises by merging subject (S) and predicate (P), a proposition can be interpreted in two different ways, either as a characterisation of S by P, or the other way round. For example, the sentence 'this stove is black' can be comprehended as a characterisation of 'this stove' by 'being black', or of 'to-be-black' by 'this stove'.

3.2 Syllogism

From a categorical perspective, syllogisms correspond to secondness. We argue that a classification of syllogistic logical signs is isomorphic to the *sequential* model of
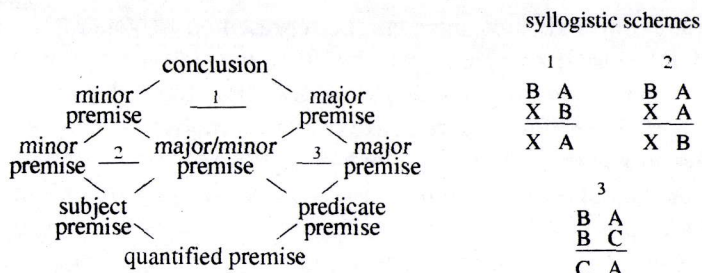
syllogistic schemes

conclusion

minor premise — 1 — major premise

minor premise — 2 — major/minor premise — 3 — major premise

subject premise — predicate premise

quantified premise

1	2
B A	B A
X B	X A
X A	X B

3

B A
B C
C A

Fig. 5: Classification of syllogistic signs and syllogisms

signs. Accordingly. input premises enter one after the other as qualisign, and signs arise bottom-up via syllogism (cf. binding): accumulation corresponds to unification. Because a syllogism requires a common term which may be represented differently in the two premises. a syllogism may need unification as well. Also unification can be represented as a premise: due to its bi-directional nature. such a premise has two different forms which can be used alternatively.

A Peircean specification of syllogistic signs follows directly from the differences between the three schemes (cf. fig. 5: the label of a horizontal line refers to the corresponding scheme). Because a sentence is an assertion about the subject or the predicate. and in the sequential model each sign has a single type. a premise can be represented as either a continuant or an occurrent syllogistic qualisign. From this it follows. that such a premise can be coerced to an icon ('subject premise'). or a sinsign ('predicate premise'). Such a sign. then. can become a rheme or a legisign. or an index. respectively. a minor or major premise of a syllogism. In accordance with the meaning of the schemes 2 and 3. such a rheme and legisign, respectively. refer to a possible subject and predicate of the entire text as a phenomenon. This follows from their function in the realisation of the major term (respectively. X and C) of the conclusion. in both schemes. The meaning of an indexical premise (if there is any) is conform to the complementary meaning of the index. Such a sign can contribute to a more complete sign of the subject or the predicate of the entire text.

Because in our model all sets are finite. premises and conclusions can always be *universally* quantified. Clearly such a quantification requires that the universe of the signs involved is restricted to the qualities defined by all unifications due to earlier syllogisms. The consequences of this are twofold. First, it simplifies the generation of syllogistic signs: second, it implies the need for the representation of sets of qualities referring to the "context" (in the traditional sense) in which a premise holds.

We mentioned that in our model the input premises are not recognised isolatedly, but only as approximations of the (single) argument sign of the phenomenon representing the entire text. Because, contrary to syntactic symbols, premises possess a

meaning which is finished, such signs can be represented (degenerately) as a sign of *any* class. Which class it eventually will be, may depend on a hypothesis. Another consequence of the above property of premises is that the order of the input signs is *less* strict then, for example, in the case of syntax (note that the temporal aspect of signs is beyond the scope of this paper).

Although approximations are unfinished signs, such signs *are* premises, therefore we allow for an approximation to be represented as a finished sign in order to be stored in the (long term) memory. When required, such signs can be recalled and represented as an approximation sign. We assume that the memory is *random* accessible.

4   Sample Analysis

In this section we analyse the premises of our sample text, and generate their summary stepwise. We will denote a premise as $<Q,A,B>$ where Q stands for the qualifier and A and B denote the minor and major terms, respectively (notice that B is always preceded by *is*). We will refer to the sentences by the labels and abbreviations introduced in fig. 1. In some cases, symbols which are removed, syntactically or semantically, are not omitted for reasons of legibility. In a syllogism, however, such sign will be considered as not being present.

With respect to their semantic value, we will assume that there are *iconic* (or degenerate indexical) signs: *docb, needs, info, types, diff, vast, various, several, be, between*; and *indexical* signs: *of-users, dt-of, their, vk-of*. We allow for a symbol, '*x of y*', to be alternatively represented as *(x of) y*, or *x (of y)*, or *(x) (of) (y)* where symbols enclosed in parentheses refer to a single sign. If any of *x* or *y* is removed, *of* can be omitted as well.

1   $<ALL,docb,is>$

The 'grammatical' subject *there* can be removed syntactically (also in other cases); *several* can be omitted ('quantificationally'), its meaning is incorporated in the quantification sign. We assume that (1) is represented as an icon, which hypothesis is supported by the iconic meaning of *docb* and *is* (semantically).

2   $<ALL,dt-of-info,is docb>$

We argue that *contain* can, in context, be semantically interpreted as a form of *be*, hence the sentence can be paraphrased as shown: *each* is incorporated in the quantifier. Because we represented (1) as an icon, we may assume that (2) will become a sinsign. Indeed, the current premise is about some 'change' (an appearing new fact), which makes its interpretation as an 'actual event' possible. Because the remaining premises are not directly related to those analyzed so far, we will assume that (1) and (2) are coerced to a dicent and a symbol sign which, then, are merged syllogistically.

146

1   *<ALL,docb,is>*
2   *<ALL,dt-of-info,is docb>*
⇒   *<ALL,dt-of-info,is>*

The unification of the two occurrences of the common term amounts to a tautology which is removed. By now we know that 'there are *dt-of-info*', which could be a summary of the text analyzed so far. Because there are more sentences to come, we will assume that the current summary as a sign is saved in the memory for later use.

3.1   *<ALL,vt-of-users,is>*

The degenerate indexical sign *various* can be removed from the subject term, semantically, for reasons mentioned earlier.

3.2   *<ALL,vdiff,is t-ineeds>*

Semantically, the iconic *vast* and the degenerate indexical *between* can be omitted; *their* and *users* are unified. Again, we will initially assume that the current premises are icon and sinsign which become a dicent and a symbol sign via coercions.

    <ALL,vt-of-users,is>
    <ALL,vdiff,is t-ineeds>
⇒   *<SOME,t-ineeds,is>*

By restricting its universe, the conclusion can be represented as a universally quantified premise. It will be assumed that this sign becomes an index, via degeneration. The two forms of the common term are *vt-of-users* and *vdiff*. Because, semantically, *types* is iconic, but *of users* is indexical, the existence of a common term requires that *types* and *diff* unify. We assume that these symbols are lexical counterparts.

At this point we can fetch the earlier sign, *<ALL,dt-of-info,is>*, and accumulate it with the current one, *<ALL,t-ineeds,is>*. There are indexical signs in both premises, respectively, *dt-of* and *their*, but which have different references. The sign *ineeds* is a syntactic accumulation of the iconic nouns, *info* and *needs*. Clearly, *dt-of-info* and *t-ineeds* have a common iconic part (*info*), their unification can be expressed as *<ALL,t-ineeds,is dt-of-info>* which is universally quantified via restricting the universe of its terms. The resulting premise is stored in memory. Our analysis so far illustrates that there is a relation between *users* and *docb* via *t-ineeds*.

4.1   *<ALL,vk-of-st,is>*
4.3   *<ALL,user-canb-satf-with-ret-info,is sev-ws>*

In (4.1) the degenerate indexical sign *various* can be removed, semantically. Notice that (4.2) is a coordination sign, syntactically, which is an explicit statement about the different types of (4.1) and (4.3). We may use this information in raising the 'right' hypotheses for these sentences as syllogistic signs. In (4.3), *in which* can be removed syntactically.

By assuming that (4.1) is finally recognized as a dicent, and (4.3) as a symbol sign

(which is motivated by the presence of the modal auxiliary *can*), the final argument sign emerges from their interaction as follows.

      <ALL,vk-of-st,is>
      <ALL,user-canb-satf-with-ret-info,is sev-ws>
$\Rightarrow$ *<ALL,user-canb-satf-with-ret-info,is>*

The common terms are *vk-of-st* and *sev-ws*. Here, *vk-of* is indexical; *sev* can be removed semantically. The existence of a common term requires that *st* and *ws* unify. Such unification is possible, because *st* is a syntactic accumulation of *search* and *task*, and *search* and *ways* are lexical counterparts.

If, as we argue, there are *vk-of-st* and *sev-ws* which are equivalent, why then are *vk-of-st* and *user-canb-satf-with-ret-info* related?

We may observe that the latter term is a sentence which arises from the dicent *user* and the symbol *canb-satf-with-ret-info*. Here, also the predicate is a complex sign, in which *with-ret-info* is a context sign, syntactically. Because *with-ret-info*, which is sign in the sense of the trichotomic relation, points in the direction of its object, the legisign *canb-satf*, and selects its actual meaning, it is sufficient to show that there is a relation in which *vk-of-st* and *with-ret-info* unify. Indeed, *with* can be removed (syntactically), and *st* and *ret-info* unify, because *returned* can be a lexical counterpart of *search*. From this it follows that *vk-of-st* is syllogistically connected with the predicate, via *with-ret-info*.

We represent the above unification as an indexical premise and merge it with the most recent memory sign (a rheme). The resulting sign is represented as an index.

      <ALL,vk-of-st,is sev-ws>
      <ALL,t-ineeds,is dt-of-info>
$\Rightarrow$ *<SOME,dt-of-info,is sev-ws>*

Here, it is required that *vk-of-st* and *t-ineeds* unify. Because the references of *vk* and *their* are different, it follows that we have to unify *search-task* and *info-needs*. Such a unification is possible, because both signs are accumulations, and *task* and *need* are lexical counterparts.

The relation of the syllogistic signs of our text as a phenomenon is depicted in fig. 6. In this graph, the terms of a premise are connected by a boldface solid line. Semantic relations are represented by a (normal) solid line. Signs which are semantically equivalent, or are converses of each other, are connected by a dotted and dashed line, respectively. Unification is indicated by a pair of arrows.

4.1 Towards a Summarization

A summary of our text is generated as follows. We start from the graph of fig. 6, find a path between the subject (*user*) and the predicate (*canb-satf-with-ret-info*) of the entire text as a phenomenon, via context signs (i.e. all other signs), and generate a syntactic representation for this path as a sentence. Clearly, all signs involved in such a path can contribute to the meaning of a summary (semantically).
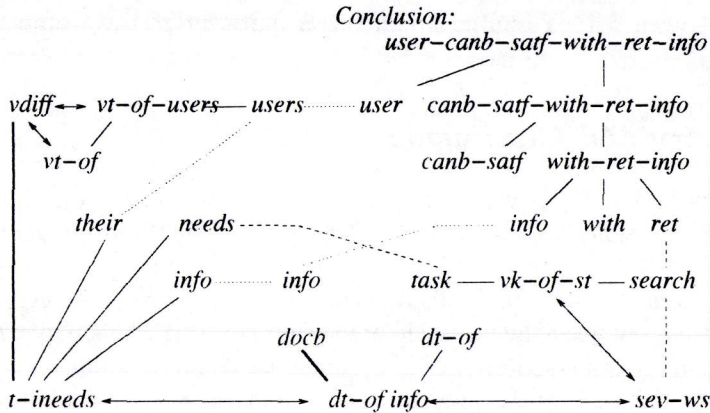
Conclusion:
user–canb–satf–with–ret–info

vdiff ←→ vt–of–users —— users .... user   canb–satf–with–ret–info

vt–of   canb–satf   with–ret–info

their   needs --------   info   with   ret

info .... info   task —— vk–of–st —— search

docb   dt–of

t–ineeds ←——————————→ dt–of info ←——————————→ sev–ws

Fig. 6: Sample text analyzed

Signs which are connected by a boldface or dotted line are equivalent and can be represented by any one of them. A constituent of a sentence must be represented by the sentence sign, but a sentence can alternatively be represented by all of its constituent signs. A possible summary generated from *user, t-ineeds, canb-satf, with, docb* is the following (notice that *docb* and *ret* are transitively equivalent, and *info* is represented by *t-ineeds; their* is replaced by a syntactic alternative):

Users' information needs can be satisfied with document-bases.

Notice that the phrasing of an actual, syntactically correct sentence expressing a summary is a challenge that is beyond the focus of this paper.

4.2   Analysis Revisited

We may conclude that the above summary correctly describes the meaning of the basic functionality of an IR system, precisely as the thesis suggests. Notice that the subject and the predicate of the entire text are necessary for the representation of the summary as a sentence, but it is the set of context signs which help us to understand why such a summary is indeed a *meaningful* characterization of the given text, and in which sense.

Another possible summary, *There are several ways*, requires knowledge about the specific meaning of *several ways*, the qualities of which are basically due to the context signs. We can acknowledge this fact by re-phrasing the above summary as: *There are several ways which amount to various kinds of search tasks*, where *amount* is used as a semantic equivalent of *be*. But we should not overestimate the potential of such syllogistic summarization. Optimally, our approach should model the entire text as a single *phenomenon* and derive the corresponding qualisigns. Such an effort, which would allow for sophisticated re-phrasings, might need, besides syllogism (basically deductive), inductive or abductive reasoning. The above result

149

indicates that even with a limited apparatus a quite interesting summary can be produced.

5 Summary and Conclusion

We have presented a framework for meaning extraction based on cognitive and semiotic principles strictly using lexical information about the words involved. The ultimate goal of this framework is that we want to boil down a text (consisting of cohering sentences) to a single phenomenon and represent it by signs. Such an effort may ultimately need the full scale of logical reasoning. This paper shows that interesting results can be achieved even by applying deduction alone.

In our view, the basic question concerning summarization of a given text relates to the correspondence between the entire text and its summary: how can we explain that there is a meaningful relation between subject and predicate of such a summarizing sentence? Also, how does (previously acquired) lexical knowledge contribute to the anticipation of possible meanings?

Our answer could be illustrated well by the classical syllogistic conclusion 'Socrates is mortal', taken as an instance of a summary. The concept of 'Socrates' and the concept of 'mortal' are related, because the other premises of the syllogistic scheme (considered here to be the entire text) imply via the *common term* 'human' that such a relation exists. Notice that the common term is not part of the summarizing sentence. However, it is part of our knowledge about the *generation* of the syllogistic conclusion.

In our semiotic approach, such a common term as a sign amounts to a *context*. In the text (premises), the common term is semantically used in two ways: as a *subject* and as a *predicate*, requiring the possibility for conversion. Our framework shows that such a conversion may emerge from concepts (for example, words) which are *counterparts*. Such counterparts are based on qualities which belong to the same phenomenon, and define it. Counterparts represent those qualities, alternatively as parts and as a whole (typically by using different denotations for the two). For example, in this context the noun 'human' represents the *parts* that constitute the concept and '(to be) human' the *properties* that characterize those parts as a whole. 'Human' as a common term then represents the *coherency* of these symbols.

The above understanding of the notion of context, which relates subject and predicate of the sign of a particular phenomenon, can be seen as an attempt to represent contextualized meaning. It is often emphasized (for example in [WF86]) that contextualized meaning of words is much richer than their lexical meaning, and therefore is so complex that its representation presents a serious implementation problem. Practical approaches try to circumvent this by restricting meaning to (finite) sets of features. Despite such restriction, the complexity of such approaches is very high, which may be due to their formal nature. Indeed, features and the combination of such features are typically formally defined. Because such features

refer to qualities of *real-world* phenomena, a cognitively based approach would more appropriately reflect the real nature of (lexical) properties, and a semiotic framework may offer a more simple treatment of their combination.

In our approach, a representation of contextualized meaning emerges dynamically as a "side effect" of sign generation. For example, *information* (as it appears in our summary) receives its contextualized meaning via the other signs (forming the context of the summarizing sentence). In context, *information* refers to that particular instance of the general (lexical) notion which is of *their*, where *their* refers to *users* which are of *various types*, and that this information is the one that is *returned* via *various kinds of search tasks*. We maintain that such a web of relations can be represented as a coherent set of logical consequences, as suggested by the title of this paper.

References

[Boc61]   I.M. Bochenski. *A history of formal logic.* University of Notre Dame Press, Notre Dame (Indiana), 1961.

[DFS99]   G.Y. Debrock, J.I. Farkas, and J.J. Sarbo. Syntax from a Peircean perspective. In P. Sandrini, editor, *5th International Congress on Terminology and Knowledge Engineering,* pages 180–189, Innsbruck (Austria), 1999.

[EN98]   B. Endres-Niggemeyer. *Summarizing Information.* Springer Verlag, Berlin, 1998.

[FS99]   J.I. Farkas and J.J. Sarbo. A Peircean framework of syntactic structure. In W. Tepfenhart and W. Cyre, editors, *ICCS'99,* volume 1640 of *LNAI,* pages 112–126, Blacksburg (VA), 1999. Springer-Verlag.

[FS00]   J.I. Farkas and J.J. Sarbo. A Logical Ontology. In G. Stumme, editor, *Working with Conceptual Structures: Contributions to ICCS2000,* pages 138–151, Darmstadt (Germany), 2000. Shaker Verlag.

[Har87]   S. Harnad. *Categorical perception: the groundwork of cognition.* Cambridge University Press, Cambridge, 1987.

[Hui96]   T. Huibers. *An axiomatic theory for information retrieval.* PhD thesis, University of Nijmegen, 1996.

[Jon93]   K. Sparck Jones. What Might Be a Summary. In *Information Retrieval '93: von der Modellierung zur Anwendung,* pages 1–19, Universitatsverlag Konstanz, 1993.

[Pei31]   C.S. Peirce. *Collected Papers of Charles Sanders Peirce.* Harvard University Press, Cambridge, 1931.

[QGLS85]   R. Quirk, S. Greenbaum, G. Leech, and J. Svartvik. *A Comprehensive Grammar of the English Language.* Longman, London and New York, 1985.

[SF01]   J.J. Sarbo and J.I. Farkas. A Peircean Ontology of Language. In H. Del-ugasch and G. Stumme, editors, *ICCS'2001*, volume 2120 of *LNAI*, pages 1–14, Stanford (CA), 2001. Springer-Verlag.

[Tej88]   V. Tejera. *Semiotics from Peirce to Barthes*. E.J. Brill, Leiden, 1988.

[WF86]   T. Winograd and F. Flores. *Understanding computers and cognition*. Ablex Publishing, Norwood (NJ), 1986.