

On Well-formedness in Requirement Elicitation Processes

Janos J. Sarbo

Radboud University Nijmegen, The Netherlands, janos@cs.ru.nl

Abstract: The cognitively based knowledge representation introduced in this paper enables specifications obtained in elicitation processes as well as the processes themselves to be verified for well-formedness from a semiotic point of view. By simultaneously elaborating object- and meta-level interpretations of a problem, the proposed theory may help speed up requirement elicitation thereby increase its efficiency.

Key words: Problem elicitation, Peirce, Cognitive, Information processing

1 Introduction

Computer scientists involved in requirement elicitation may reasonably be jealous of their colleagues, software developers, for the useful programs they have at hand, such as parsers, type checker and proof assistants. Problem elicitation^[3], this initial and potentially most influential phase of requirement engineering is as yet not facilitated by similarly powerful software tools. The premise of this paper is that the Knowledge in Formation theory (KiF) presented in [11], [9], amongst others, enables natural language (NL) specifications obtained in requirement elicitation processes as well as the processes themselves to be analyzed for semiotic well-formedness from a certain point of view. By simultaneously developing object- and meta-level concepts of an input problem, the proposed knowledge representation may help speed up requirement elicitation thereby increase its efficiency. Stake holders' time can be expensive after all.

Due to the complexity of language symbols, as opposed to the limitations of current linguistic tools, automatic well-formedness analysis of NL documents is beyond our current possibilities. By virtue of the simplicity of the proposed theory of knowledge representation, this task can be easily realized by the professional elicitor him/herself, however.

According to KiF theory (see also www.cs.ru.nl/kif), knowledge arises from the observation of phenomena, by means of signs. An observation can be modeled as a conceptualization *process*, generating representations expressive of the nine sign aspects introduced by Peirce^[6]. This also holds for NL phenomena such as specifications, in which, sign aspects are expressed by means of language clusters, e.g. phrases, clauses, sentences. In this paper I will suggest that language clusters of semiotically well-formed NL specifications can be arranged in a hierarchy, isomorphic to the Peircean classification of sign aspects (see fig. 1). Specifications satisfying this condition can be interpreted as representations of phenomena and have the potential to be experienced as meaningful. By linking language clusters, via sign aspects, with conceptualization events, the condition for well-formedness can be naturally 'lifted' from specifications to elicitation processes.

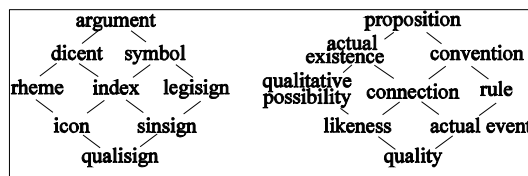


Figure 1 Peirce's sign aspects (left) and their mundane terms (right)

Past applications of KiF theory in different domains^[9] such as the domain of syntactic, reasoning, and mathematical symbols revealed the potential of Peirce's nine sign aspects for modelling information processing (at least in those domains). Results of an analysis of a professional elicitation process^[4] have

shown that contributions by stakeholders can be characterized as expressions of Peircean sign aspects of a common input problem. Only when all different sign aspects of the problem have been formulated, in agreement with all stakeholders, elicitation has been experienced to be complete and the process could be terminated (note that problems, which are phenomena, may include nested problems recursively).

Following KiF theory, meaningful interpretation arises from information processing of phenomena. In this paper I will suggest that the interpretation of an input problem can be related to the summarized meaning of the underlying phenomenon and, that an expression of *that* meaning may function as the governing ‘goal’ of problem elicitation as a process. From a methodological perspective, well-formedness analysis of elicitation processes requires (i) the generation of a summarized meaning of the input specification and, (ii) an interpretation of the language clusters comprised by the input text, as expressions of Peirce’s nine sign aspects. In this paper I will propose that KiF theory can be beneficially used for a realization of both sub-tasks. This will be illustrated with an extended example, elaborating the above two tasks in sequence (this can be interesting from a theoretical stance), as well as simultaneously (which can be more characteristic for human information processing).

The structure of this paper is the following. KiF theory is briefly recapitulated in sect. 2. This is followed by a definition of well-formedness (sect. 3), a sample analysis (sect. 4) and remarks concerning the efficiency of the proposed approach (sect. 5). The paper is closed by a summary in sect. 6.

2 A process model of cognitive activity

According to the used theory of knowledge representation, in an observation the brain is interpreting the input phenomenon as a sign. A phenomenon is considered to be an interaction between independent, dual entities. In KiF theory, human information processing is modelled on the basis of an assumption of cognitive theory^[10] that external stimuli appearing as qualities are processed by the brain in percepts. In a single operation, the brain compares the current percept with the previous one, and this enables it to distinguish between two sorts of input qualities (in short, *input*): one, which was there and remained there, which can be called a ‘state’; and another, which, though it was not there, is there now, which can be called an ‘effect’. The interaction between that ‘state’ and ‘effect’ can be used for a representation of the interaction between the brain (occurring in some state) and the external stimulus (appearing as an effect). Following Harnad^[2], percepts can be defined as collections of qualia (the terms ‘quality’ and ‘quale’ designate, so to speak, the outside and inside perspective of an entity; ‘effect’ and ‘state’ are used for the qualia of the external stimulus and all other input qualia, respectively).

As there are two types of input qualia, state and effect, the brain has to identify them individually (‘what is known about them in earlier observations’), as well as in relation to each other, in order to (re)cognize their meaning (‘why this effect is occurring to this state’). The above two phases of interpretation are acknowledged in KiF, by introducing two sub-processes of cognitive activity: *perception* and *cognition*.

In the process of perception, the brain compares the input qualia with the memory response triggered by them (memory response signs are assumed to be represented by dual collections of state and effect qualia obtained in earlier observations). Depending on the activation of the memory, there may be qualia in the memory response having an intensity: (i) above or (ii) below threshold, referring to an interpretation of the input which is in the brain’s ‘*focus*’, and which is only ‘*complementary*’, respectively. A memory response of type (i) signifies the recognition of the input in the sense of agreement: the input is recognized or ‘known’ as such an entity. A type (ii) response refers to input recognition in the sense of possibility: the input is not recognized or ‘not known’, indicating that the memory response is only representing a secondary or even less important aspect of the input qualia.

The above relation between input and memory signs enables the brain to distinguish between four different interpretations of the input stimulus: as a focused state (**A**) and effect (**B**), and a complementary state (\neg **A**) and effect (\neg **B**). Note the use of ‘ \neg ’ symbols, as an indication of the *denial* of a positive identification, not as a Boolean operator. Due to its minor importance for this paper, a definition of perception is omitted. A full account of that process may be found in [9], in which we also show that the

two phases of information processing can be modelled by isomorphic instances of a single type of process (this may also explain why the focus of this paper could be restricted to a definition of cognition as a process).

The input qualia of the process of cognition (see fig.2) are defined by the final representations generated by perception as a process. The ‘goal’ of the cognition process is the establishing of a relation between the input qualia that are in focus (**A**, **B**), in the light of the complementary qualia comprising the context ($\neg\mathbf{A}$, $\neg\mathbf{B}$). By assuming that the input qualia appear in a ‘primordial soup’ ($[\mathbf{A} \ \mathbf{B} \ \neg\mathbf{A} \ \neg\mathbf{B}]$), the events of the cognition process may be defined as follows. Below the convention is used that square brackets indicate that an entity is not yet interpreted as a sign, and no bracketing or the usual bracket symbols, if *some* interpretation is already available.

- (1) *sorting*: $[\mathbf{A}]$, $[\mathbf{B}]$, $[\neg\mathbf{A}]$, $[\neg\mathbf{B}]$
the identification of the qualia in the ‘primordial soup’, which are in focus ($[\mathbf{A}]$, $[\mathbf{B}]$), and which are complementary ($[\neg\mathbf{A}]$, $[\neg\mathbf{B}]$);
- (2) *abstraction*: \mathbf{A} , \mathbf{B}
the separation of the collections of the two types of focused qualia;
- (3) *complementation*: $(\mathbf{A}, \neg\mathbf{B})$, $(\mathbf{B}, \neg\mathbf{A})$
the linking of focused qualia with information by the context;
- (4) *predication*: $(\mathbf{A}, \neg\mathbf{B}) - (\mathbf{B}, \neg\mathbf{A})$
the establishment of a relation between the completed focused qualia.

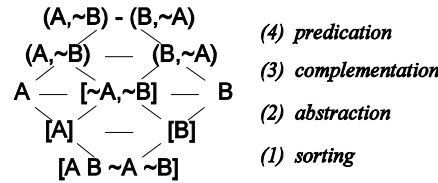


Figure 2 A schematic diagram of the cognition process. A horizontal line stands for an interaction between neighbouring entities.

In conformity with the duality required for an interaction, in step (3), there may be an interaction between **A** and $\neg\mathbf{B}$, and between $\neg\mathbf{A}$ and **B**, but no interaction may occur between **A** and $\neg\mathbf{A}$ or **B** and $\neg\mathbf{B}$. This is because **A** and $\neg\mathbf{A}$ (but also **B** and $\neg\mathbf{B}$) arise due to the same input trigger, meaning that the two signs are *not* independent. Let me emphasize the mediation function of context signs, in step (4). Through the common meaning shared by $\neg\mathbf{A}$ and $\neg\mathbf{B}$, the context implicitly determines the meaning of the relation between **A** and **B**. In the rest of this paper I will alternatively refer to the process depicted in fig. 2 as the *process model*.

The ‘goal’ of the process model is an interpretation of the entire input as a (single) sign. As this goal can only be achieved upon termination (if ever), representations generated by the process can only be interpreted as signs that are on way towards becoming meaningful. In [8], I called such representations a ‘proto-sign’. The difference between signs and proto-signs may be illustrated with the metaphor of apparent motion perception. In that phenomenon, a series of steady pictures are presented and, although each picture may be meaningful in itself, combined they are interpreted as parameters in the experience of the series of pictures as motion. According to a conjecture of KiF theory, an analogous relation may exist between the individual interpretation moments of the process model on the one hand, and an experience of their process as meaningful on the other.

2. 1 ‘Naive’ logical and semiotic analysis

The completeness of the process model depicted in fig.2 can be demonstrated by offering a ‘naive’ logical analysis to the events of this process. Such an analysis can also be useful as an intermediate step towards an interpretation of the process model as a meaningful representation.

An essential element of a ‘naive’ logical interpretation^[9] is the abstraction of a common meaning for the two types of input qualia (state and effect), which is the concept of a *logical variable*. By virtue of the duality characterizing input qualia, we need two variables, that we denote A and B . The difference between qualia that are in focus, and those that are complementary can be represented by the difference in their expression. Perceived state and effect qualia which are in focus are indicated by A and B , respectively; those which are complementary by $\neg A$ and $\neg B$. Notice the use of ‘ \neg ’ as a complementation operation on collections. For example, the complementary sub-collections of A -type qualia are denoted by A and $\neg A$ (the label A is used ambiguously). In the logical expressions depicted in fig. 3, ‘or’ (‘+’) and ‘and’ (‘*’) are used as representations of a relation in the sense of possibility and agreement, respectively, which aspects are included in the meaning of these logical operators. For instance, **[B]** can be assigned the logical expression $A*B$, by virtue of an assumption of KiF theory, that the existence of an effect (cf. B) always implies the existence of a state (cf. A) and, that the simultaneity included in their relation can be expressed by the simultaneity included in the meaning of ‘*’. The logical expressions assigned to the interpretation moments are depicted in fig. 3; a full account of ‘naive’ logic may be found in [9].

Note, in fig. 3, the presence of all Boolean relations on two variables, indicating the completeness of the cognitive process: We look at our input from *all* possible angles!

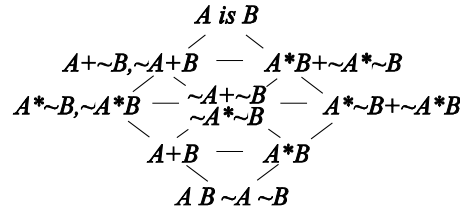


Figure 3 A ‘naive’ logical interpretation of the events of the process model

An important contribution of the above ‘naive’ logical interpretation is the introduction of meaningful concepts, opening the way towards a Peircean semiotic analysis of the process model (see fig. 1). For instance, **[B]** can be interpreted as a representation of the input as an *actual event* (cf. *sinsign*), by virtue of the aspect of simultaneity included in the logical meaning of $A*B$. In [9] we have shown that the nine interpretation moments of the process model, depicted in fig. 2, can be associated with the nine sign or meaning aspects introduced by Peirce (cf. fig. 1). What makes this link with Peircean theory especially important, is its potential for interpreting the process model as a ‘meaningful’ process (as the Peircean classification of sign aspects may be interpreted as a ‘process’, but this possibility shall not be considered in this paper). See fig. 4. By virtue of this relation, representations generated by the process model can be interpreted as meaningful representations or ‘solutions’ of an input problem. The above link also enables the Peircean sign aspects to be used as pointers to the status of a sign in the interpretation process. For example, an expression in the index position can be alternatively referred to as an index(ical) expression or, simply, an index.

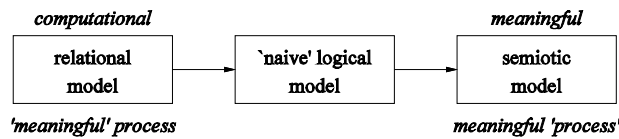


Figure 4 A model of information processing

3 Processing complex phenomena

The above definition of the process model capitalizes on the assumption that the qualia of an input phenomenon are present in a single interaction. This property may not hold for natural language

phenomena, which, due to the sequential character of the input symbols, may consist of a series of interactions and (nested) phenomena. For example, a sentence may include a series of words.

Following [9], sequential input processing can be modeled by assuming that the interpreting system occurring in some state is affected by the next symbol appearing as an effect. By processing a series of such interactions, a relation between individual input symbols can be established by making use of their relational properties. The ‘goal’ of syntactic analysis as a process is an interpretation of the entire input as a syntactically well-formed sentence (individual symbols are interpreted as ‘proto’-signs with respect to the entire input as a sentence). The sign aspect exhibited by a language cluster can be determined on the basis of the syntactic relational properties of the symbols included by the cluster.

The above view of sequential symbol processing can be applied to texts as well, except the assumption for the relational properties of symbols, which has to be adjusted (an example for a dependency between sentences are anaphoric relations). According to [9], a series of sentences (cf. pictures) can be processed in a summarizing single transition (cf. motion), transforming the initial state represented by the first, to the final state represented by the last sentence. Following this line of thinking a model of text summarization can be defined as follows.

In a pair of previous and current (or next) sentences, the current sentence appearing as an effect is affecting the previous sentence, representing some state. Their interaction, appearing as a change, can be interpreted as an event, expressed by a (single) sentence. The generation of such a sentence can be modelled, by interpreting the change due to the differences between the subjects as well as the predicates of the two sentences, as a transition of the state represented by the previous sentence. If a transition is possible, the two sentences can be merged in a summarizing single sentence, called a *summary*; otherwise, the current sentence is interpreted as a nested phenomenon or ‘episode’, that will later intertwine with the nesting phenomenon, marked by the previous sentence (here I make use of the tacit assumption that the entire text can be summarized in a single sentence). A remotely similar approach to text summarization may be found in [5].

In order to generate a summarizing single sentence, the process may need semantic order relations on symbols (‘<’)¹, enabling less meaningful symbols to be removed in favour of more meaningful ones. For instance, if the subjects of the previous and current sentences are *several document-bases* and *each document-base*, respectively, and, according to a semantic ordering of referential symbols, *several*<*each*, then the subjects can be summarized by the symbol *each document-base*. The use of semantic orderings requires a perspective of interpretation, for instance, that *several* may refer to a collection as a whole, as opposed to *each*, which may refer to an element. Semantic orderings can be defined on a categorical basis, as shown in [12].

3.1 Well-formedness

Summarization is a necessary first step towards establishing well-formedness of a text obtained by problem elicitation. In this paper, a text, and so, an elicitation process is called (*strongly*) *well-formed* iff from the summarized meaning of the text generated by the processing schema (argument position), expressions for all less developed interpretation moments can be derived, satisfying the condition that an (*order-preserving*) *bijective* mapping exists between them and a set of language clusters induced by the text. An (ordered) set of language clusters is defined as a partitioning of the entire text in (a sequence of) disjoint fragments.

Well-formedness analysis involves a validation of the dependencies between expressions corresponding to different interpretation moments, by means of a quasi-parallel (e.g. tabular) or backtracking algorithm. In practical requirement elicitation processes by humans, such an analysis can be simultaneously elaborated with the elicitation process itself. Although now a final summary may not be available, the analysis can make favourable use of the hypothesis that, eventually, the set of language clusters defined by the input text can be expressed by a single sentence. Note that the title of a text (if available) may sometimes be used as a potential summary, as well.

¹ ‘<’ is used as a polymorphic operator.

4 An extended example

The theory of this paper is illustrated in this section, with an analysis of a text, found in a PhD thesis. In this fragment (see fig. 5), the author defines his concept of ‘Information Retrieval Systems’. Why these text elements did he communicate with the reader? Why in this order?

Below I attempt to answer these questions by showing that the sample text satisfies the conditions set for well-formedness. I will assume that the text displayed in fig. 5 is an account of a hypothetical phenomenon and, as such, can be represented by means of the processing schema. By articulating language clusters in a certain order, the author provides a ‘guided tour’ of his thought. By interpreting those language clusters as signs, the listener is recapitulating their intended meaning, as sign aspects of a phenomenon. A definition of an underlying collection of qualia is beyond the goal of this paper.

no.	String	Abbreviation
1	There are several document-bases	sev-docb
2	Each document-base contains different types of information.	each-docb dt-of-info
3	There are various types of users and there are vast differences between their information needs.	vt-of-users vdiff-ineeds
4	There are various kinds of search-tasks, or stated differently, there are several ways in which	vk-of-st sev-ways
5	user can be satisfied with the returned information.	canb-satf ret-info

Figure 5 Sample text defining the concept ‘Information Retrieval Systems’

4.1 Well-formedness analysis

Well-formedness analysis requires that, first, the input text is summarized, in order to derive a consistent collection of interpretation moments, second. An illustration of this theoretical approach will be the subject of the first part of this section. Such a laborious development of well-formedness analysis may not be necessary in practical elicitation, in which the above two phases can be simultaneously realized by virtue of human dexterity for finding solutions in a complex web of dependencies. This may hold for elicitation processes involving different stakeholders, as well as for processes realized by a single person. Either way, this approach requires that the elicitor makes consistent use of context information corresponding to different interpretation moments. This, more practical approach is illustrated in the second part of this section.

In the analysis below, language clusters can be referred to by their labels, as well as by the abbreviations introduced in fig. 5. A summarization event is indicated by a ‘→’ symbol, a summary by the labels of constituent language elements, separated by a comma. For example, (1)-(2)→(1,2). Semantic orderings used in the example are defined ‘on the fly’.

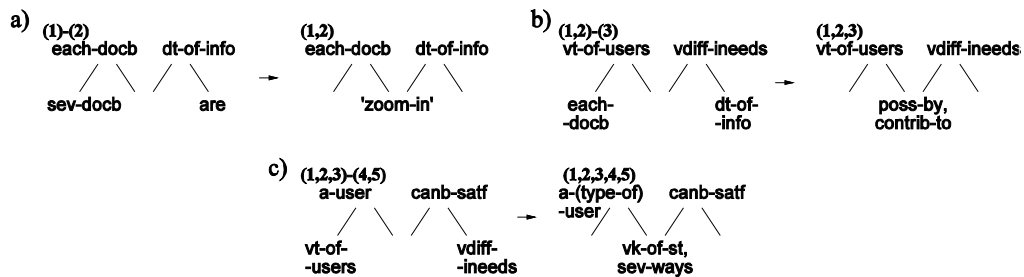


Figure 6 Summarization events

(1)-(2)→(1,2) By making use of a semantic ordering of referential content *ref-to-whole*<*ref-to-element*, the adjectives can be ordered: *sev*<*each*. As a result, *sev* can be omitted. A semantic ordering of verbs, *existence*<*modification* (e.g. a ‘modification’ of an entity by a property), enables *are* to be removed. The above operations are depicted in fig. 6(a). The context sign imputed by the used semantic orderings is called ‘zoom-in’ (that is, from collections of *sev-docb* to individual documents designated by *each-docb*).

(1,2)-(3)→(1,2,3) In (3), the ‘focus’ is shifted from documents (*docb*), to their owners (*users*). The semantic ordering *owned*<*owner*, enables *each-docb* (interpreted as *owned*) to be removed. Similarly, *dt-of-info*<*vdiff-ineeds*, as *dt-of-info* (cf. *owned*) form the basis for and contribute to *ineeds* (cf. *owner*). The adjective *vdiff* is an expression of the rule-like meaning induced by the differences included in *dt-of-info*. The used semantic orderings are represented in the index position, by the expressions *poss-by* (short for ‘possessed by’) and *contrib-to* (short for ‘contributes to’), which are each other’s converses (in an act of possession (effect) there must be something (state) that contributes to and enables that event). See also fig. 6(b). Here and in later examples, a “,” symbol is used for an expression of a synonymous representation from a certain perspective (converse symbols are alternative expressions (state/effect) of a phenomenon and are synonymous in this sense).

(1,2,3,4)-(5)→(1,2,3,4,5) The coordination, defined *vk-of-st* and *sev-ways*, in (4), is degenerately represented in the index position, by the expressions ‘*vk-of-st, sev-ways*’. A synonymous interpretation of these symbols is possible due to the coordinator, or stated differently, assigning the two expressions the converse meaning of a ‘search’ phenomenon (in a *vk-of-st* act, there must be *sev-ways* in which ‘search tasks’ (‘returning of information’) are realized). A degenerate representation of these expressions as a complementary symbol is motivated by the fact that the corresponding language clusters have no reference elsewhere in the text. In context, *vt-of-users* is interpreted as *a-user* (a more precise definition would be *a-(type-of)-user*), and *vdiff-ineeds* as a result of ‘search tasks returning information’. Some of the ‘returned information’ may satisfy *a-(type-of)-user*. This is expressed by the *predication* symbol interaction (cf. fig 2) between *a-(type-of)-user* and *canb-satf*. A linguistic expression of the summary is omitted.

Well-formedness of the input text can be shown, by taking the above summary, *a-(type-of)-user-canb-satf*, and deriving a concise collection of interpretation moments, on the basis of the used the semantic orderings. The result of such a derivation is presented in fig. 7. The only interesting decision is the interpretation of the modality exhibited by the predicate, *can be satisfied with ...*, as a contribution of *predication* itself. As a consequence, the symbol position can be represented by the event of ‘information returning’ (*ret-info*). As the order of expressions as shown in fig. 7 respects the order of language clusters induced by the text, the text as well the corresponding elicitation process (which is assumed) must be strongly well-formed. This closes the illustration of the theoretical approach.

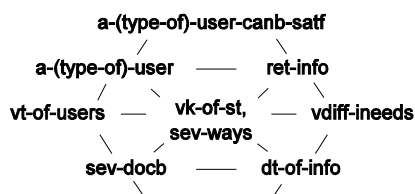


Figure 7 A strongly well-formed text

Following an assumption of this paper, in human problem elicitation, the two phases of well-formedness analysis can be simultaneously elaborated. In that process, which is the subject of the second part of this section, the input is processed sequentially and language clusters are hypothetically associated with sign aspects, ‘on the fly’. In the presentation below, a hypothesis is indicated by an “=” symbol. Non-

deterministic analysis is restricted to the presentation of a single, correct solution; positions of the process model are referred to by Peircean sign aspects. See also fig. 1 and fig. 7.

(1) **sev-docb=icon**. A postulation (are) of **sev-docb** as constituent entities, not as qualitatively possible ones (rheme), nor as such entities in context (dicent). The hypothesis immediately above can be justified by the fact that, besides (2), **sev-docb** has no later references in the text.

(2) **dt-of-info=sinsign**. An expression of **dt-of-info** as an appearing new property of **docb** (cf. actual event). By means of the adjective **different**, this **sinsign** lays the ground for an interpretation of its relative difference with each (icon), as a **legisign**.

(3) **vt-of-users=rheme**. An expression of existent entities (are). The later anaphoric reference to **users**, by their, in (4), enables **vt-of-users** to be interpreted as an expression of a range of possibilities ('what can types of users be in general'). Following the dependencies between the Peircean sign aspects, the interpretation of **vt-of-users** in the rheme position implies an induced representation of **users** in the icon position hence a synonymous interpretation of **users** and **docb**, as constituents of a phenomenon (this is represented by **sev-docb** in the icon position).

(3) **vdiff-ineeds=legisign**. A generalization of the single event (the appearance of a new property) **dt-of-info**, in the type of events represented by **information-needs**, on the basis of the relative difference between the icon and **sinsign** positions, marked by **each** and **several** and, by **different-types-of**, respectively. The hypothesis immediately above is confirmed by the rule-like compatibility of **vast differences** (effect) and **information needs** (state), expressed by **vdiff-ineeds**. Due to the dependencies between the Peircean sign aspects, the interpretation of **vdiff-ineeds** as a **legisign** implies an induced representation of **vast differences** and **between** in the **sinsign** position (also these qualia are represented by **dt-of-info**), and a representation of **ineeds** in the icon position (also **ineeds** is included in the meaning of **sev-docb**).

(4) **vk-of-st=index**. As there is no reference to **vk-of-st** later in text, this symbol may not be a representation of a meaning which is in focus. For this reason, **vk-of-st** may not be interpreted as a rheme or dicent expression of the input either. According to the preferred interpretation, **vk-of-st** is representing an event, not by explaining it in any way, but only by pointing in its direction. Due to the dependencies between the Peircean sign aspects, the interpretation of **vk-of-st** as an index implies the existence of complementary qualia (**qualisign** position), that must be included in the meaning of the rheme and **legisign** expressions of the input as well.

(4) **sev-ways=index**. By virtue of the coordinator, or stated differently, and the complementation by **in which ...**, also this symbol can be interpreted as an indexical expression of complementary qualia. Note the converse meaning of **vk-of-st** and **sev-ways**, in the index position.

(5) **a-user=dicent**. An expression of **vt-of-users** in context (more precisely **a-(type-of)users**), representing users demanding various kinds of search tasks.

(5) **ret-info=symbol**. From a syntactic point of view, the predicate (symbol position) is defined by the phrase can be satisfied with the returned information. The complement (**ret-info**) can be interpreted as a representation of a nested phenomenon, 'information returning'. Following this line of thinking, **canb-satf** can be interpreted as a representation of an interaction between **ret-info** and **a-(type-of)-user**, hypothetically expressing the meaning of the entire text (argument position). Note that **ret-info** can be a representation of the conventional meaning of **vdiff-ineeds** in context, expressed by a combination of different information needs and search tasks, that may be correctly called 'information returning'.

5 Efficiency of elicitation

Following a conjecture of this paper, by simultaneously developing object- and meta-level concepts of a problem, the efficiency of elicitation can be improved. Meta-level, semiotic concepts or sign aspects can be used in combination with object-level symbol interactions, in order to introduce goals during problem elicitation. For example, if the stakeholders are stuck at an expression of their problem as a range of

possibilities for an action (rheme position), then, by raising questions motivating the introduction and/or use of context information (index position) and the generation of an expression of the actual subject of the input problem (dicent position), the elicitor may help the elicitation process to be further developed.

An application of the proposed theory requires some familiarity with Peirce's sign aspects. Experience obtained in a professional elicitation process at the Dutch software enterprise Sogeti^[4] has shown that the necessary knowledge can be shared with the stakeholders in a short training of 15...30 minutes. Knowledge about the process model itself is not required (except by the elicitor, of course). The results of that research revealed that also practical elicitation processes respect the events dictated by the process model. An experimental validation of the effects of a combined object- and meta-level analysis during elicitation is the subject of current research.

6 Summary

Well-formedness of natural language specifications developed in elicitation processes can be verified by analyzing semiotic conditions set for language clusters. Such an analysis is possible by making use of the cognitively based knowledge representation presented in this paper. Theoretically, this requires (i) the representation of a summarized meaning of the text and, (ii) a subsequent derivation of expressions exhibiting sign aspects of the input. A specification (text) as well as the corresponding elicitation process is called well-formed, if a bijective mapping exists between those expressions and the set of language clusters induced by the text. Due to human dexterity for finding solutions in a web of conditions, the above two sub-processes can be simultaneously realized in practice. The two approaches, theoretical and practical, are illustrated with an extended example.

References

- [1] P. Hagoort, L. Hald, M. Bastiaansen, K.-M. Petersson. Integration of word meaning and world knowledge in language comprehension. *Science* (304): 438-441, 2004.
- [2] S. Harnad. *Categorical perception: The groundwork of cognition*. Cambridge University Press, 1987.
- [3] J.Krogstie, G. Sindre, H. Jorgensen. Process models representing knowledge for action: a revised quality framework. *European Journal of Information System* (15): 91-102, 2006.
- [4] E. Klomp. Conceptualisatie in een requirements development proces. Master thesis (in Dutch), J.J. Sarbo, A.J.J. van Breemen (supervisors), Radboud Universiteit, Nijmegen (NL), 2008.
- [5] W. Mann and S. Thompson. *Rhetorical Structure Theory: Toward a functional theory of text organization*. *Text*, 8(3): 243-281, 1988.
- [6] Ch.S. Peirce. *Collected Papers of Charles Sanders Peirce*. Harvard University Press, Cambridge, 1931.
- [7] R. Quirk, S. Greenbaum, G. Leech, J. Svartvik. *A Comprehensive Grammar of the English Language*. Longman, London and New York, 1985.
- [8] J. Sarbo. Peircean proto-signs. D.M. Dubois, editor, *AIP Conference Proceedings* (839): 474-479, 2006.
- [9] J. Sarbo, J. Farkas, A. Breemen. *Natural Grammar*. In R. Gudwin, J. Queiroz, editors, *Semiotics and Intelligent System Development*, 152-175, Hersey (PA), Idea Group Publishing, 2006.
- [10] R. Solso. *Cognitive Psychology*. Harcourt Brace Jovanovich, New York, 1988.
- [11] A. van Breemen, J. Sarbo. Sign processes and the sheets of semeiosis (S_s). In K. Liu, editor, *ICOS'2007*, 89-98, Sheffield (UK), 2007.

Author in Brief

Janos J. Sarbo – PhD (1985), is mainly engaged in research on knowledge representation, including applications in requirement engineering and language modelling.