Radboud Repository

Radboud University Nijmegen

PDF hosted at the Radboud Repository of the Radboud University Nijmegen

The following full text is a preprint version which may differ from the publisher's version.

For additional information about this publication click this link.
http://hdl.handle.net/2066/60292

Towards a Theory of Meaning Extraction

Article · June 2004

Source: CiteSeer

2 authors, including:

Janos J. Sarbo
Radboud University
62 PUBLICATIONS **191** CITATIONS

Some of the authors of this publication are also working on these related projects:

Knowledge in Formation View project

Towards a Theory of Meaning Extraction

Janos Sarbo and József Farkas

University of Nijmegen, The Netherlands
`janos@cs.kun.nl`

Abstract. Using a Peircean model of sign recognition, we define meaning extraction as text summarization based on propositional signs. Such signs, as parts of sentences, can be conceived as transitions between successive states. Consecutive transitions can be combined to generate a single proposition (or sentence) comprising the summary of a text.

1 Introduction

Assume we are shown a picture, somewhat later another one, and so forth, repeatedly. What we see is a series of individual pictures. Then, we are presented the same pictures again, but now without interruption. If the difference between the consecutive pictures is not large, we may see motion. In this paper we argue that the summarized meaning of a text may arise analogously. A picture corresponds to an individual state and effect represented by a sentence, and the experience of the succession of pictures as motion to the semiosis of a single state and effect represented by the text as a whole. Such process of semiosis is called *meaning extraction*.

Recently we introduced a Peircean model for the recognition of signs [3]. This paper is an attempt to show that our framework could be suitable for meaning extraction as well. Earlier research on this subject has been presented in [10]. Other applications of our approach, for example, for modeling morphological, syntactic and semantic signs, and most recently, reasoning signs, have been reported in [8], [4], [9]. Meaning extraction is a complex process, potentially using *all* such types of signs. Due to lack of space, we cannot expand on their models, in this paper we will assume some familiarity with the underlying theory.

Traditionally, text summarization is typically based on statistical, or syntactical analysis [6], [2]. We hold that in virtue of their formal character such approaches may not be able to find the 'natural' meaning of a text (here, 'formal' means to be about the form). In this paper we will argue that our semiotic framework may provide such a representation of knowledge, by virtue of respecting its *natural* ontology.

1.1 Signs and aspects

Insofar signs are related to aspects, as Peirce maintains [7], and language consists of symbols, which are signs, also language phenomena must represent aspects.

Such aspects may arise via the semiosis of language phenomena, for example, a phrase or a sentence. In [3] it is advocated that such semiosis can be modeled by a process of sign interactions. The signs generated by such a process are not finished and can only appear via the mediation of a final proposition sign. Peirce's signs and their aspects are shown in fig. 1 (a horizontal line in the left-hand side diagram denotes a sign interaction). Let us emphasize that in our use of Peirce's semiotics the *interpreter* of a sign interaction is assumed to be the brain as a *computational system*. Although meaning in the most complete sense must be a sign of Peirce's decadic classification, in this paper we are concerned only with its approximations, which are signs of the 9-adic classification. We will use the notions 'sign' and 'meaning' in such a restricted sense.
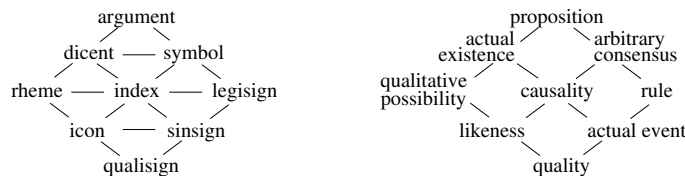

Fig. 1. Peirce's classification of signs and aspects

Meaning extraction is also strongly related to communication. We argue, it is due to the *universal* character of signs, and their *uniform* recognition in the different domains of knowledge, that communication is in principle possible between arbitrary speaker and hearer like child and adult, or layman and expert. In order to find out the relevant meaning of a message communicated by the speaker, the hearer has to re-present the received input, possibly by means of summarization. In the course of such a process, the symbols perceived may have to be interpreted in different ways (in general in any order), for example, as a logical, or a syntactic, or a semantic symbol, conform the holistic meaning of signification. For example, 'a Queen' can be interpreted as a syntactic subject, but also as a logical term of a premise.

The structure of this paper is the following. In the first part we briefly recapitulate our model of language. This is followed, in the second part, by the introduction of a theory of meaning extraction. In the final part, we illustrate our approach with an extended example, which is the summarization of the fairy tale "Snow White". Because the sample text contains complex language phenomena, preceding summarization we discuss some language modeling issues. The hidden agenda of this paper is an attempt to show the potential of our language framework for analyzing complex utterances in a simple way.

2 Language model revisited

In this section we recapitulate the primary concepts of our language model [8] that we will need later in our summarization example, in section 4.

Qualities as signs In [3] we argued that observations arise from 'real' world phenomena, which are interactions. We perceive such phenomena via the interaction between the input stimuli and the senses. The output of the senses, a bio-electric signal, is processed by the brain in percepts. By comparing the previous percept with the current one the brain is capable of distinguishing between two sorts of qualities, which we call, respectively *state* and *effect* qualities (or briefly, state and effect). The two types of qualities and, ambiguously so, their sets, are denoted by a and b. The set of state and effect qualities are collectively called the *input*.

a and b are *dual* re-presentations of a phenomenon, respectively, as a state, and an effect changing that state [4]. The interaction involved in the phenomenon is represented by the (sign) interaction between a and b, defining a *transition* generating a next state. The input signs (a, b) trigger the memory which in turn produces a response (a', b'). As such response can signify the input either in the sense of agreement ($*$), or possibility ('$+$'), the input can be finally represented by the observed and complementary state ($A=a*a'$, $\neg A=a+a'$), and effect ($B=b*b'$, $\neg B=b+b'$, respectively).

Language signs Language phenomena are symbol interactions. We argue that the subject and the predicate of a sentence, respectively, represent an observed state and effect, and the sentence itself a change, defined by the interaction between the subject and the predicate, generating a next state.

Language phenomena can be subject to nesting. A nested phenomenon, for example, a closed clause, is defined by a contiguous segment of input symbols. Such segment is analyzed recursively, its final proposition sign is represented in the nesting phenomenon, degenerately (in the semiotic sense), by a complementary quality, and eventually by an index sign.

A classification of morphological and syntactic signs on the basis of their lexical relational needs is depicted in fig. 2 (a, p, and n denote an active, passive and neutral type of relational need, respectively [8]). The interaction between language signs is called *binding*; the degenerate types of a binding are called *coercion* and *accumulation*.
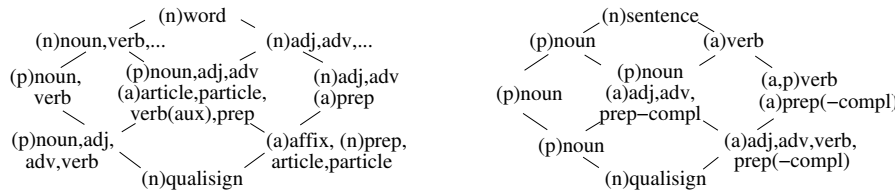

Fig. 2. The classification of morphological and syntactic symbols

Besides morphological and syntactic symbols language also contains *semantic* signs. In traditional sentence analysis it is usually assumed that semantics

is compositionally built on the syntactic structure. Contrary to this view, we hold that syntax and semantics are complementary representations, collectively defining the full meaning of a language sign, e.g. a sentence. Following our earlier results, we assume that syntactic and semantic signs can be isomorphically modeled. The difference between the two types of signs lies in the differences between their types of relational needs. Syntactic relational needs are abstract, and therefore formal; semantic ones are more specific, and are more closely related to the particular meaning of the 'real' world phenomenon signified. For example, in an adjectival modification, the syntactic meaning of the adjective only specifies the *type* of the complement, e.g. its type of speech (and the same holds for the complement, too). In the semantic interpretation, the relational needs of the adjective and the complement describe different, but at the same time compatible views of a single phenomenon. This may explain why "green ideas" can be interpreted as a syntactic, but not a semantic sign, inasmuch green can syntactically combine with any noun, but green and ideas are not known to collectively signify any 'real' world phenomenon. Another examples of semantic language phenomena are the case or thematic role relationships between predicate and argument(s), as well as the anaphoric relations between symbols.

Because semantic symbols are more concrete than syntactic ones, such signs can more aptly signify the 'deep' meaning of language phenomena. Such relational meaning can be represented from the perspective of the sign's relation to its object. This boils down to a classification of semantic symbols as icon, index, and symbol signs. Such types of signs, which form a *semantic triad*, are related to Peirce's three categories [7] and define an ordering (the category aspect of a type of sign is given in parentheses): icon(1)<index(2)<symbol(3).

For example, verbs can be classified as existence<state<action, respectively referring to an act of existence (e.g. there is a book), a modification of a state (e.g. she was alone), and a state transition which is an action (e.g. John is running), denoting semantically increasingly more meaningful events.

Semantic triads can classify the relation, for example, between verb and complement, or verb and modifier (which we collectively call a *verb triad*), and between adjective and noun, and adverb and verb (in short, *adjective* and *adverb triads*). Notice that also the types of syntactic relational needs can be ordered in a triad: $n<p<a$.

Complex language phenomena For the presentation of the example of sect. 4 we need to discuss the modeling of two types of language phenomena. The first of them, which is related to the *morphological* meaning of article symbols, can be illustrated with the sample string: 'the beautiful Queen'. Following the classification of fig. 2 (on the left-hand side), this input can be analyzed as (beautiful)(the Queen) (here, and in later examples, a morphologically finished sign is given in parentheses). This example illustrates that in our model article symbols can become immediate neighbors of their references.

The second language phenomenon is coordination. The *kernel* of an algorithm for recognizing such phenomena can be specified as follows. Coordination

consists of three phases. In the first phase, the input symbols are analyzed up to the coordinator. In the second phase, first, all existing signs are saved (their relational needs are remembered as 'traces'). Then, an initial segment of the remaining input is analyzed recursively, as a nested sign. In the course of this, actually generated and saved signs of identical classes are coordinated (this may include the inheritance of syntactic properties). In the last phase, the remaining saved signs are restored. Save and restore, respectively, can be preceded or followed by the elaboration of a finite number of pending symbol interactions.

Memory signs Language recognition heavily relies on memory signs, which are prototypical representations of state and effect qualities arising in earlier observations. In [4] we argued that the memory response generated for an input state (a) and effect (b) can be represented, respectively, by an *averaged value* (a') and a *dense domain* (b') of qualities stored by the memory segment of a 'collection', and a 'chain' of neurons. Here, by denseness we mean that the chain representation of a domain is extended by a new neuron (and quality), if the input signifies a yet unknown measure of the type of effect.

In [9] we pointed out that by maintaining such representation of memory qualities, also the inferential meaning of a sign interaction can be derived. Accordingly, complementation has the aspect of deduction and induction, respectively, in the rheme–index and index–legisign symbol interactions. Sorting, abstraction and predication have the aspect of abduction. Sorting and abstraction amount to the postulation of a hypothesis concerning the potential role of an input symbol in a subsequent inference, and predication to the generation of a proposition, which is a hypothesis. The classification of reasoning signs (borrowed from [9]) is depicted in fig. 3.
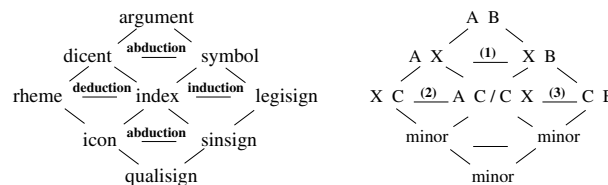

Fig. 3. The classification of inference modes and syllogistic schemes

3   Meaning extraction

Meaning extraction is text summarization based on *propositions*. A proposition represents a sentence or a clause (but which we will collectively refer to as *sentence*) and in turn a phenomenon, by the interaction of the subject (state) and predicate (effect) generating a next state (meaning). We will say that a pair

of sentences, *previous* and *current*, can be summarized, if there is a *transition* between their meanings. Such a transition may exist if the subjects and the predicates of the two sentences can be separately unified, and a common context shared by the two unifications can be defined (by *unification* we refer to the process deriving a synonymous interpretation of two symbols). If these conditions hold, a sentence containing the summarized meaning, or *summary*, of the two sentences can be generated; otherwise, the current sentence is considered to be the sign of a nested phenomenon (an "episode"), which may later intertwine with the meaning of the nesting phenomenon marked by the previous sentence. Notice that summarization is 'directed', from previous to current sentence.

The above interpretation of transition is also supported by the semiotic meaning of language signs. Following our language model, the subject of a sentence can be represented by a dicent, and the predicate by a symbol sign, which arise, respectively, from the general meaning of a rheme ('possible subject') and a legisign ('law-like property') via complementation by an index sign ('selection'). Index signs, which are a representation of complementary phenomena, coordinate the relation between a rheme and dicent, and between a legisign and symbol sign.

Meaning extraction proceeds as follows. Initially, the input sentences are analyzed individually. In a single summarization step, first, the previous and current sentences are re-analyzed, respectively, 'bottom-up', starting from the input symbols, which are qualisigns, and 'top-down', from the proposition, which is an argument sign. Accordingly, the previous sentence is signified by a rheme, an index, and a legisign, and the current one by a dicent, and a symbol sign. The coordinated unification, respectively, of such a rheme and dicent, and legisign and symbol sign are used for the derivation of the subject and predicate, as well as the state and effect qualities of a summarizing observation, representing the event, linking the two sentences to each other.

We argue that meaning extraction, in the above sense, satisfies the conditions set for a process of sign recognition. Insofar such process can be isomorphically defined for the different domains of signs [3], meaning extraction can capitalize on the various interpretations of a symbol. For example, as a sentence from the logical point of view is a (logical) proposition, text summarization can alternatively be interpreted as a reasoning process. This is illustrated below with the analysis of the sample text 'John left the house. The door closed'. The subjects are John and the door (in short door), the predicates are left the house (in short left) and closed.

The inferential meaning of summarization can be elucidated by means of syllogisms [9]. A single syllogism corresponds to a unification, a coordinated pair of syllogisms to a transition. In the inferences below, the aspect of quantification is omitted; unification is denoted by a '∼' sign.

In our example, John∼door is possible, if we have memory knowledge about John's ability to open something. That ability, which characterizes door too, can unify the meaning of John and door, in a specific sense: as participants of an open event.

John *IS* open	;John opened something	
door *IS* open	;the door is opened	
⇒ door *IS* John	;the door (which is opened) is (opened) by John	

Analogously, left∼closed is possible if we know that left (and closed) may refer, in some sense, to door as well.

door *IS* left	;it is the door that somebody left	
door *IS* closed	;now the door is closed	
⇒ closed *IS* left	;(the door) which is closed, is (the door) that	
	;somebody left	

From the sample transition above we may have learned a new property (open) of the subjects (cf. deduction), and the existence of a new object (door) the predicates may refer to (cf. induction). These property and object, which function in the individual syllogisms as common terms, are interrelated by virtue of the co-ordination of the two inferences. Such common terms, that we call *counterparts* [10], are dual signs of a phenomenon representing a context, in which the two sentences can be merged to a summarizing single sentence: a summary, which is a hypothesis (cf. abduction). In our example, such a summarizing sentence is John left (if summarization is directed towards a "story" about John).

We may improve on the above syllogistic conclusions, by taking another domain and interpretation: the semantic one. Accordingly, the subjects, John and door, can be unified via accumulation, yielding John (opening the door), as John and door are, respectively, the agent and patient of the effect open, and in that context, semantically, door(2)<John(3) (the definition of the corresponding triad will be given later). Also the predicates, left and closed can be unified yielding left ('behind the closing door') denoting a value in the domain of 'left'-effects. In other words, closed is interpreted as a sort of leaving event. Due to the common terms, the two unifications have a shared aspect of interpretation: because John (agent) left, the door (patient) closed.

In sum, by meaning extraction we may find out how the effects of the individual sentences can be combined to a single effect changing the initial state of the first, to the final state of the last sentence. The process of summarization can make favorable use of the uniform character of sign recognition, by selecting suitable domains and interpretations of the input symbols. A hidden advantage of this approach lies in its potential for a *systematic* specification of knowledge in different domains, as relational needs.

4   An extended example

This section is an attempt to illustrate the potential of our theory for meaning extraction. We should emphasize that such a process may put a great burden on lexicon specification, but this aspect shall be left out of consideration in this paper.

Our sample text will be the fairy tale "Snow White" [1], displayed in fig. 4 (dialogs, not affecting the meaning of the tale, are removed). The presentation of summarization is simplified as follows. In an accumulation symbol interaction,

s1 In a far off land, there lived a very beautiful Queen who had a stepdaughter called Snow White. **s2** The wicked Queen ordered a servant to take Snow White to the forest and put her to death, but the poor man had not the heart to do it, and told her to hide in the forest.

s3 Snow White ran into the thickest part of the forest and walked for many long days, until she came upon a tiny little house. **s4** The Princess thought that in such a lovely place as that, there must live kind people who would give her shelter.

s5 The house, which at that moment was empty, was in a state of complete disorder. **s6** The furniture and everything inside was small.

s7 Snow White cleaned and tidied the little house until it shone. **s8** Then she lay down exhausted on one of the little beds and fell asleep. **s9** When she woke up, she found herself surrounded by seven little dwarfs who, on hearing what had happened to her, promised to protect her from her stepmother. **s10** They were all very happy and contented.

s11 One day, the Queen, who had heard that Snow White was still alive, disguised herself as an old woman and invited her to try an apple which she had poisoned. **s12** The Princess was tempted by the lovely apple the old woman was offering her. **s13** The old woman insisted, assuring her that she had nothing to fear, until at last, she accepted. **s14** But ... when she bit into the apple, Snow White fell senseless on the ground, and how those dwarfs cried and cried.

s15 A Prince who was passing by saw the beauty of Snow White and kissed her on the forehead, whereupon she wakened from that bad dream and the dwarfs were happy again. **s16** Snow White and the Prince married and were happy ever after.

Fig. 4. A sample text

a semantically less meaningful sign can be ignored. Index signs representing complementary meaning can be omitted, except those which refer to a lexically defined relation of a rheme or legisign symbol (such relations are not complementary, in the semantic sense). For example, an adjectival reference of a noun can be omitted, but an obligatory complement of a verb cannot. In a syntactic analysis, the two types of index signs [3] can be marked by a pointer indicating their reference, which is either a rheme ($\leftarrow$) or a legisign ($\rightarrow$).

Temporal aspects are left out of consideration. We assume that the order of appearance of the input sentences, and the temporal order of their effects, are isomorphic. Hence sentences which are neighboring can be summarized in any order (summarized sentences are conceptually replaced by their summary). In regard to thematic relations we refer to case theory [5]. Cases are interpreted as semantic signs referring to verb–complement relations. Semantic triads needed in the course of summarization are introduced 'on-the-fly'. Syllogisms underlying summarizations are not formally defined, due to lack of space.

The syntactic analysis of a sentence is given in a tabular form of the sign 'matrix'. A column corresponds to a sign class, a row to the recognition of an input symbol. Predication sign interactions are not represented; icons and sinsigns, which only represent the sorting of input symbols, are omitted. The following

abbreviations are used: input(i), accumulation(a), coercion(c), binding(b), degeneration (d). Accumulated signs are separated by a '/' sign; coordination is denoted by an '&', a syntactic symbol interaction by a '-' sign. Semantic sign interactions which coincide with syntactic ones are not displayed. Signs, which are used for summarization, are given in bold face.

s1 (In)(far off)(a land)(there)(lived)(very beautiful)(a Queen)(who)(had)(a stepdaughter)(called)(Snow White). **s2** (Wicked)(the Queen)(ordered)(a servant)(to take)(Snow White)(to the forest)&(put)(her)(to death)&(but)(poor)(the man)(had not)(the heart)(to do)(it)&(told)(her)(to hide)(in the forest).

Fig. 5. The morphologically finished signs of s1 and s2

The morphological and syntactic analyses of s1 are depicted in fig 5 and table 1, respectively. The nested clauses of this sentence, which are analyzed recursively and represented degenerately by index signs, are: in a far off land (inffl), a stepdaughter called Snow White (stcsw), who had a stepdaughter called Snow White (hadst). The signs of s1, which are used for summarization, are displayed in fig. 6.
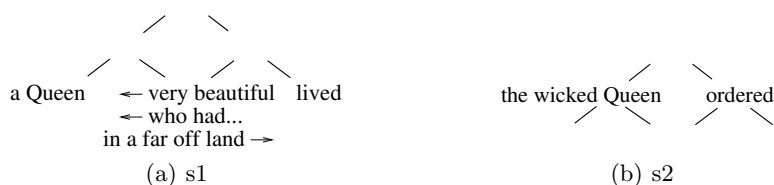

Fig. 6. The signs of s1 and s2

The syntactic analysis of an initial segment of s2 is depicted in table 2. This sentence consists of two clauses coordinated by 'but'; the signs of the first clause are displayed in fig. 6. Because the predicate complement is a clause itself, its analysis can be postponed. Temporarily, ordered will be the predicate (by using its meaning as an intransitive verb). The coordination, take Snow White to the forest and put her to death, is abbreviated by take&put.

Is a transition possible from s1 to s2? In order to find an answer for this question we need the rheme, index and legisign from the bottom-up analysis of s1, and the dicent and symbol signs from the top-down analysis of s2. Summarization may then proceed as follows.

Let us begin with the unification of the subjects. The rheme, a Queen, and the dicent, the wicked Queen, can be unified by the Queen. Indeed, by assuming that the triad of referential signs can be defined as general<indefinite<definite

input	rheme	index	legisign	dicent	symbol	action
recursion						
in						
far off(foff)			in			c
a land(alnd)		foff	in			c
	alnd	foff	in			c,b
		foff-alnd	in			b
					in-foff-alnd	
return						
there(t)		inffl→				d
lived(lvd)	t	inffl→				c
very beautiful(vb)		inffl→	lvd	t		c,c
a Queen(aQ)		←vb,inffl→	lvd	t		c,a
	aQ	←vb,inffl→	lvd	t		c
recursion(1)						
who(wh)						
had	wh					c
	wh		had			c
recursion(2)						
a stepdaughter(st)						
called(ca)	st					c
Snow White(SW)	st		ca			c
		SW	ca	st		c
				st	ca-SW	b
return(2)						
	wh	stcsw	had			d,c
		stcsw	had	wh		b
				wh	had-stcsw	b
return(1)						
	aQ	**←vb/hadst, inffl→**	**lvd**	t		d,b
. . .						
				t/aQ-hadst…	lvd-inffl	b

Table 1. The syntactic analysis of s1

(a 'general' reference is, for example, 'one'), we have that a(2)<the(3), hence the indefinite article can be removed. Following the triad of adjectives introduced in [8], wicked(1) and very beautiful(1) are semantically less meaningful than who had...(2), hence they can be ignored. We tacitly assume that the indexical meaning of the dicent sign (cf. wicked) can be abstracted to an index sign.

The unification of the predicates yields ordered, because lived(1) refers to an act of existence, but ordered(2) to a modification of a state. Summarization is possible, due to the referential meaning of the context (who had...) coordinating the individual unifications of the subjects and the predicates (see fig. 7).

The complement of ordered contains embedded clauses, allowing further summarization. In what follows, we will need the triad of action(3) verbs, defined as neutral<modulate<change, respectively denoting a neutral (e.g. Snow White

input	rhme	indx	legi	dcnt	symb	action
wicked(wk)						
the Queen(tQ)		wk				c
ordered(ord)	tQ	wk				c,b
			ord	**wk-tQ**		
recursion						
a servant(se)						
to take(ta)	se					c
Snow White(SW)	se		ta			c
to the forest(tf)		SW	ta	se		c,c
		tf	ta-SW	se		b
				se	ta-SW-tf	b
coordination						
put(pu)						
her(he)			pu			c
to death(td)		he	pu			c,b
		td	pu-he			b
					pu-he-td	
restore						
				se	**take&put**	
return						
		se-take&put	ord	wk-tQ		d,b
				wk-tQ	ord-se-take&put	b

Table 2. The syntactic analysis of an initial segment of s2

ran), a reversible, and an irreversible action (the last two types are illustrated below). We will tacitly assume that also prepositional complements can be specified analogously. Accordingly the complement, to take Snow White to the forest and put her to death can be summarized as follows. To take(2) and to put(2) are reversible actions, but to the forest(2)<to death(3), therefore the first conjunct can be removed.

The resulting sentence can be summarized with the second conjunct of 'but'. This requires the unification of the subjects, servant and the man, and of the predicates, to put her to death... and had not the heart.... The result of the first yields the servant, inasmuch 'but' can be interpreted as an 'and', coordinating the second conjunct with the first one, in the sense of complementation. The indexical poor(1) does not refer to a lexically defined relation, therefore removed.

The unification of the predicates proceeds as follows. As told her...(1) is a neutral action, it can be removed. In the first conjunct, had not the heart to do it, negation can be moved (syntactically) to the front, its logical meaning emphasized ('neg'). This leaves the summarization of to put her to death, but (neg) had the heart to do it. Here, had the heart to do it(2)<put to death(3), since the first conjunct refers to an intension (cf. 'modulate'), but the second to an irreversible change. Because (neg) signifies a 'negative' effect which has to be represented

(conform the anaphoric reference of it), and put to death is a lexical synonym of murder (her refers to Snow White), we finally get: (neg) to murder her.

The summarized complement of ordered is a proposition, represented by a non-finite clause with an explicit subject. By summarizing this clause, with the summary of s1 and s2 depicted in the right-hand side diagram in fig. 7, we get the signs displayed in fig. 8 (the index sign who had… is omitted, as it does not refer to a lexically defined relation of the Queen).
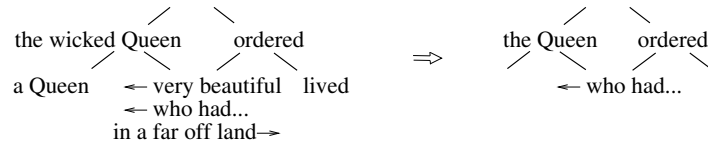

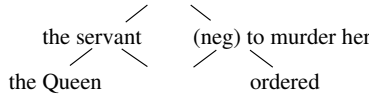**Fig. 7.** The summarization of s1 and s2

Fig. 8. The summarization of s1 and s2 (cont.)

Again, summarization is possible if Queen and servant, and ordered and murder her can be unified, as well as a common context exists. Queen∼servant is possible, because, semantically, the Queen can be the 'master of' the servant (we assume this to be a lexically defined semantic relation; 'master of' is a context sign). We introduce a semantic triad of thematic roles as follows: experiencer<patient<agent. Because Queen is the agent, and servant the patient of ordered, and ordered 'transitively links' its agent with its patient (which is a context sign, representing the meaning of ordered as a transitive verb), we may finally conclude that the servant (patient) can be removed. The existence of a summary additionally requires that the same contextual interpretation applies to the unification of the predicates as well. That this condition holds indeed, is explained below.

The unification of the predicates can be justified by using the induction meaning of complementation. The legisign, ordered, denotes a law-like property represented by a domain. A specific value of this domain, which can be selected by 'to-make-somebody-do-something' (a context sign), is denoted by the predicate: ordered to murder her.

Summarization is possible, for 'the master of' and 'to-make-somebody-do-something' can be lexically defined counterparts, and are synonymous to 'transitive linking'. This defines the final subject, the Queen, and the final predicate, (neg) ordered to murder her. Because ordered is an action, and (neg) refers to the

failure of its effect, the meaning of the predicate can be expressed by using failure's counterpart (failed) added to the verb (ordered) via coordination in a complementary sense: (A) The Queen ordered to murder her, but it failed.

We may further summarize this sentence by interpreting the title of the text as an incomplete sentence, the predicate of which being defined by the tale itself. Unification of the Queen and Snow White is possible, by assuming that the personages of a story can be semantically classified as unnamed<episodic<title. Because the predicate of (A) denotes a value in the domain of 'order' representing a kind of 'to murder', which is lexically related to its counterpart symbol ('murder'), we may get the final summary:

Snow White: A murder failed

Interestingly, this sentence may very well function as the summary of the whole story. The analysis of the rest of the tale shows that the remaining sentences do not change the above summary (assuming summarization is directed towards a story about Snow White). Indeed, when the Queen discovers that the servant has not been loyal to her, she begins her 'apple project', but which eventually fails as well.

The summarization of the remaining sentences goes as follows. Due to lack of space, we restrict ourselves to the specification of a simplified meaning of the sentences and the result of the individual summarizations. A transition is denoted by a '⇒' sign.

s3: Snow white ran into the forest indicates the beginning of a new episode. This sentence can be summarized with s4, assuming Snow White and the Princess are the same person. s3⇒s4: (B) Snow White ran into the forest and thought.... Also s5 is marking the beginning of a new episode. As in the summarization of s3 'house' has been removed, s5 and s6 (which can be summarized with s5) can be ignored.

(B)⇒s7: (B1) Snow White ran into the forest and cleaned and tidied (a) house. s8: She lay day and fell asleep. (B1)⇒s8: (B2) Snow White ran...and lay down and fell asleep. s9: She found herself.... (B2)⇒s9: (B3) Snow White ran...and lay down...and fell asleep and found herself.... s10: They were.... Because the dwarfs (referred to by they) have been removed, this sentence as a whole can be ignored.

s11: The Queen disguised herself and invited her to try an apple. (A)⇒s11: (A1) The Queen ordered...but it failed, and disguised herself.... Inasmuch as her is a reference to Snow White, the stories of the Queen and Snow White may become joined again. What links the two stories is the event invited. Therefore, the other events which only refer to one of the characters like ran, found herself, etc. can be removed. (A1)⇒(B3): (AB) The Queen invited Snow White to try an apple (assuming her is a reference to Snow White).

s12: The Princess was tempted.... This sentence can be summarized with (AB), but it will not generate any new meaning. s13: The woman insisted. Again, summarization with (AB) will make no change. s14: Snow White bit...and fell. As the dwarfs have been ignored already in s9, they can be removed now too. (AB)⇒s14: (AB1) The Queen invited Snow White to try an apple and Snow White bit into the apple and fell.

s15: A Prince saw Snow White and she wakened from (a) dream is again the beginning of a new episode. s16: Snow White and Prince married. s15⇒s16: (C) A Prince saw Snow White and they married.

Finally we have two sentences, (AB1) and (C). What might link them to form a single story? By taking Snow White bit into the apple and fell as the first premise, and our earlier knowledge about the apple (which she had poisoned) as the second, we may deduce that Snow White must have been poisoned. Inasmuch as a poisoned person (who fell) must be dead, we may abductively infer that Snow White is dead. But according to (C), which follows (AB1) in time, Snow White gets married, from which it deductively follows that she must be alive. Because poisoning is a counterpart of murder, we get again: a murder failed.

Summary

Meaning extraction is a most complex problem which, as we argue in this paper, might have a systematic solution in the cognitively based Peircean model of signs [8]. The advantage of such an approach lies in its potential for a uniform specification of symbols as relational needs, and a uniform interpretation of the combinations of such symbols as sign interactions.

References

1. *Festival Fairy Tales*. Peter Haddock Limited, England, 1988.
2. B. Endres-Niggemeyer. *Summarizing Information*. Springer Verlag, Berlin, 1998.
3. J.I. Farkas and J.J. Sarbo. A Logical Ontology. In G. Stumme, editor, *Working with Conceptual Structures: Contributions to ICCS'2000*, pages 138–151, Darmstadt (Germany), 2000. Shaker Verlag.
4. J.I. Farkas and J.J. Sarbo. A Peircean Ontology of Semantics. In U.Priss and D. Corbett, editors, *Conceptual Structures: Integration and Interfaces (ICCS'2002)*, volume 2193 of *LNAI*, pages 177–190, Borovets (Bulgaria), 2002. Springer-Verlag.
5. C.J. Fillmore. The case for case. In R.T. Harms E. Bach, editor, *Universals in linguistic theory*, pages 1–90, New York, 1968. Holt, Rinehart and Winston.
6. K. Sparck Jones. What Might Be a Summary. In *Information Retrieval '93: von der Modellierung zur Anwendung*. Universitatsverlag Konstanz, 1993.
7. C.S. Peirce. *Collected Papers of Charles Sanders Peirce*. Harvard University Press, Cambridge, 1931.
8. J.J. Sarbo and J.I. Farkas. A Peircean Ontology of Language. In H. Delugach and G. Stumme, editors, *ICCS'2001*, volume 2120 of *LNAI*, pages 1–14, Stanford (CA), 2001. Springer-Verlag.
9. J.J. Sarbo and J.I. Farkas. Logica utens. In A. de Moor and B. Ganter, editors, *Using conceptual structures*, pages 43–56, Dresden (Germany), 2003. Shaker-Verlag.
10. J.J. Sarbo, S. Hoppenbrouwers, and J.I. Farkas. Towards thought as a logical picture of signs. *International Journal of Computing Anticipatory Systems*, 12:137–152, 2002.