

**KNOWLEDGE in FORMATION**  
**A computational theory of interpretation**

**TUTORIAL NOTES**

Janos J. Sarbo

©2014 This work is subject to copyright



## Preface

These notes have been developed for the tutorials of the course “Cognition and Representation”, given at Radboud University, Nijmegen (NL), in 2009-2018. The theory of this course is described in the book, “Knowledge in Formation: A computational theory of interpretation”, Springer, 2011, ISBN 978-3-642-17088-1, e-ISBN 978-3-642-17089-8.

The relation between these notes and the tutorials can be illustrated with Aristotle’s correspondence with Alexander the Great (quoted by (Dumitriu, 1977), volume I, pg. 143).

(Alexander to Aristotle)

“*Salve.*

You were wrong to publish the acroamatic works.<sup>1</sup> How shall we be different from the rest, if the teaching which formed us becomes a common good for all? As for me, I should like better to be distinguished by my knowledge of the loftiest [things] in the world than by power. Health to you!”

(Aristotle to king Alexander)

“*Salve.*

You wrote to me about my acroamatic teaching, to tell me I ought to have them kept secret. Well, you must know they are published and not published at the same time, as they are intelligible only to those who attended my lectures. Health to you!”

Nijmegen, March 2014

---

<sup>1</sup>Acroamatic were called [those works] treating profounder and subtler problems, and belonging to the contemplation of nature [to physics] or to dialectical problems.



# Chapter 1

## Basic notions

### Exercise 1

A list of important terms used by this course is given below. Explain the meaning of at least five of them, briefly.

- meaning, meaningful, interpretation
- sign, object
- icon, index, symbol
- containment, involvement
- co-existence, co-occurrence
- background information, context
- concept, proto-type

### Answer

#### Meaning

In his pragmatic maxim, Peirce defined this concept as follows. “Consider what effects, which might conceivably have practical bearings, we might conceive the object of our conception to have. Then, our conception of these effects is the whole of our conception of the object.”

Alternatively, meaning can be said to be the consequence of testing a hypothesis by way of its conceivable results. Meaning is virtual, not actual (Brent, 1993, 1998), pg.12.

### **Meaningful**

Having a meaning or purpose. It refers to interpretation from some perspective. As interpretation involves a process, and processes are goal driven, interpretation must be purposeful, as well.

### **Sign, object**

A sign stands for something else than itself and signifies it for some interpreting agent. That what the sign stands for is called its object. For instance, the phrase, ‘an apple’ (sign), may stand for an apple on the table (object). An agent may interpret the above phrase as a request to take and an apple on the table and give it to somebody.

### **Containment, involvement**

Containment assumes familiarity with certain rules. For example, by saying that this box contains a pair of shoes, we suggest to have familiarity with the rules of opening the box and, this way, accessing the shoes contained. Involvement is a different concept. For instance, a piece of marble may involve –not contain– a statue.

### **Background information, context**

Interpretation always makes use of information by an agent. This information can be obvious or in the focus of the agent. An example is walking as falling. By perceiving the familiar feeling of falling, the agent may ‘know’ what (s)he has to do, that is, make a step ahead. If the feeling is not a familiar one, for instance, because there is a hole in the ground, background or context information can be used by the agent in order to decide if a jump can be necessary, instead of a simple step.

## **Exercise 2**

In your opinion, which of the following statement(s) can be true?

- a) The computer can generate meaning.
- b) The computer can generate meaning.
- c) The computer can obtain more complex meaning than humans can.

## Answer

- a) True. It is commonly accepted that human interpretation must be qualitatively more complex than computations generated by traditional, e.g., van Neuman computers. Meaningful interpretation can be obtained –not generated– by the brain/mind.
- b) False.
- c) False. If, as we assume, the brain/mind has ‘copyright’ for interpretation, then any concept of meaning which is more complex than meaning obtained by human interpretation, must be contradictory.





## Chapter 2

# Signs and sign interpretation

Phenomena can be classified in three categories, according to Peirce, which he called firstness, secondness, and thirdness. An example of a phenomenon is the following. Assume you are at home, sitting behind your desk or lying on the sofa. Suddenly you hear the noise of a car passing by in the street.

### Exercise 1

Explain the noise of the car passing by, as a firstness, a secondness, and a thirdness category phenomenon.

Hint: you are asked to specify three categorically different phenomena, commonly referred to as ‘the noise of a car passing by’.

### Answer

The noise, perceived as a feeling or a quality in itself, is a firstness. In this phenomenon, the noise does not interpreted as a meaningful quality, e.g., as the noise of passing car, which would be a thirdness, nor as a relation of a quality with something else, e.g., noise triggering your eardrum, which is a secondness. In order to appear, firstness needs secondness, and in turn, thirdness.

## Exercise 2

The noise of a car passing by can be interpreted as a sign. What is the object and interpretant of this sign?

### Answer

The object of the perceived noise (sign) can be “locomotion by a car”, the interpretant the thought sign “somebody must be in a hurry”, but there are many more answers possible. Note that the object of a sign can be determined through mediation by the interpretant, never directly by the sign.

## 2.1 Exercise 3

Offer an analysis to the noise of car passing by in itself (1), from the perspective of its relation to its object (2) and its interpretant (3).

### Answer

#### Relation of the sign to itself

The noise may appear as a quality in your eardrum (qualisign). It may appear as an event occurring now (sinsign), for instance, if you perceive temporal aspects of the noise triggering your sensory apparatus. And, it may appear as a habitual sign, if you are familiar with noise signals, and recognize its type, e.g., as “acceleration by a car” (legisign).

#### Relation of the sign to its object

The noise may appear in your perception as a signal of certain frequency and intensity, hence as form (iconic). It may refer to similar noise patterns experienced in the past (indexical). And, it may appear as a noise, conventionally known as “speeding up by a car” (symbolic).

#### Relation of the sign to its interpretant

The perceived signal may appear as a noise of any car (rhematic). It may be interpreted as a proposition of a fact (dicential), e.g., “this noise”. And, it

may be interpreted as a proposition of premise, or a hypothesis (argumentative), e.g., “this noise is by an accelerating car; the driver must be in a hurry”, or alternatively, “close the window, please”.



## Chapter 3

# From signs to interpretants

Interpretation can be based on strong or weak sameness. Explain the two kinds of an interpretation of our running example of the noise of a car passing by. Can be strong sameness practical in human processing?

### Answer

If interpretation is based on strong sameness, all interpretants of the input signal will be identical informationally. Interpretation is governed by the involved legisign, for instance, the habitual rule of “noise by a car”. As a result, a single argumentative sign (cf. DIR) may arise, e.g., the motor reaction “wait patiently until the car is away”. Strong sameness characterizing reactions such as reflexes can be most practical.

Interpretation based on weak sameness may obtain increasingly more developed interpretants, hence a grow of information. For example, noise of the car passing by can be explained this way as the thought sign “somebody must be in a hurry”, or the motor reaction, using complementary information about noise protection, such as “close the window, please” or “take earplugs”.



## Chapter 4

# A semiotic account of interpretation processes

Interpretation can be characterized by the perspectives or sign aspects involved. Offer an analysis to the noise signal in our running example, interpreted as an argumentative sign.

### Answer

Following our model of interpretation, the observer, occurring in some state, is in interaction with the input noise signal, appearing as an effect. Qualia representing this input as a state and effect involve the aspect of a qualisign. Input qualia sorted out in a form (such as a frequency and intensity value), and an actual event (e.g., a co-occurrence of qualia), involve the sign aspect iconic and sinsign, respectively. A representation of the input qualia as an abstract state, having the potential to be interpreted in any conceivable way, involves the sign aspect rhematic; an embedding the input qualia in habitual effect, the aspect of a legisign. An expression of the input, as a reference to complementary information or the context of the input interaction involves the sign aspect indexical. Complementary or background information can be used for a representation of the abstract input state as an actual existent or a fact, which involves the sign aspect dicential, and it can be used for an expression of the rule-like input effect as a characteristic property, involving the sign aspect symbolic. A combination of the input state and effect in context may obtain a representation of the input as a premise, which involves the sign aspect argumentative.





# Chapter 5

## A world of signs

### Exercise 1

Following Sect. 4.3, percept with the previous percept the brain can be able to distinguish between two sorts of qualia: one, which was there and remained there (cf. state); and another, which, though it was not there, is there now (cf. effect). Qualia that were there in the previous percept, but are not there in the current one, are disregarded by our model. Can you justify this decision?

### Answer

According to our theory, knowledge arises from observations that are event representations of an interaction or reactions on an input stimulus. Qualia that were there but are not there now may correspond to stimuli occurring in earlier interpretation processes. Qualia that ‘disappear’ may not trigger memory, neither enable an immediate interpretation. They can be derived through reasoning, however..

### Exercise 2

The model of interpretation introduced in this chapter assumes an input consisting of a pair of qualia: state and effect. Is it possible to apply the model to phenomena involving any number of independent qualia? Illustrate your answer with an example.

## Answer

According to our model, state and effect represent collections of qualia that are independent. Establishing a relation between the two types of qualia is the goal of interpretation, as a process. We assume that, in human processing, referential information, including temporal information about a co-occurrence of perceived qualities, is represented by qualia as well. An example can be helpful.

Assume you are standing at a level crossing and suddenly you hear the idiosyncratic sound of the arriving train. At the same time you observe that a car is stopping in front of the barrier and, that a mother is grabbing the hand of her child. In this example, state qualia can be the perceived qualities of the car, barrier, mother and child; effect qualia the appearing sound of the train, the locomotion of the car slowing down and the mother grabbing her child's hand. We may assume the existence of referential information about the relation between the above state and effect qualia, as well as about qualia corresponding to different modalities. In language phenomena, it is traditionally assumed that referential information respects surface level contiguity. An example of a contiguous collection of qualia, in English, is the set of words comprising a closed clause. Following the rules of syntax, external references may refer to the entire closed clause only, not to the individual symbols contained.

Following cognitive theory, human processing is always subject to having a focus. In our example above we assumed that the focus is on the arriving train, all other input qualia, such as the idiosyncratic sound of the train, are complementary. Information processing is concerned with interactions between state and effect qualia that are in focus. Complementary qualia can be used as context information.

## Exercise 3

Explain the syntactic language phenomenon (phrase): **happiest girls**, as a result of a series of symbol (sign) interactions. Hint: input= happy, 'superlative', girl, 'plural'.

## Answer

Language symbols are conventional signs, traditionally distinguished in two major classes: morpho-syntactic and syntactic. Information processing in the two domains of signs can be quasi-simultaneous.

We assume that the morpho-syntactic symbol interaction (and relation) happy–‘superlative’ is represented by the syntactic symbol *happiest*; girl–‘plural’ by the syntactic symbol *girls*. The syntactic symbol interaction *happiest*–*girls* is represented by the phrase *happiest girls*. The suggestion of an alternative representation: (*happiest*–*girl*)–‘plural’, may not be correct, as a morphological plural symbol may not interact (or apply to) a noun phrase.

## Exercise 4

Assume you observe a vase visually, for some time. According to the theory of this book, you may perceive the vase only if it is appearing in your perception through an effect. Once qualia of the vase are present in the previous and current percept, both, you may not be able to process their input unless another effect may appear. Common experience suggests, however, that we are perfectly able to observe the vase for a longer time, even if there is no change, hence no novel effect in the input. Can you explain, why?

## Answer

This is a consequence of the cognitive phenomenon known as the saccadic movement of the eyes. The eyes are moving, even if a person tries to fixate them at a well-marked point as steadily as (s)he can. It is experimentally proven<sup>1</sup> that when a person views a stabilized image, the structure of the target fades after a short time and the visual field appears dark grey or black.

## Exercise 5

Consider the observation of a phenomenon illustrated with a pair of percepts, (a) previous and (b) current, in the diagram below. The symbols, ‘*e*’, ‘*f*’ and ‘*P*’ are added for referential purposes.

---

<sup>1</sup>See [www.answers.com/topic/stabilization-of-retinal-images](http://www.answers.com/topic/stabilization-of-retinal-images).

Define the input representations,  $[q_1 \ q_2 \ C]$  and  $(q_1, C) - (q_2, C)$ , for this phenomenon.



## Answer

A possible representation is the following.

$$[q_1 \ q_2 \ C] = [ \{e\} \ \{non\text{-parallel } f\} \ \{P \text{ edge}\} ];$$

$$(q_1, C) - (q_2, C) = \text{edge-}e \text{ IS non-parallel-to edge-}f$$

If for any reason the above proposition is found to be unsatisfactory, the interpreting system may try and find another context information, enabling a different conceptualization. A sample context information can be: ‘*intersecting-at*’.

$$[q_1 \ q_2 \ C] = [ \{e\} \ \{non\text{-parallel } f\} \ \{P \text{ edge intersecting-at}\} ];$$

$$(q_1, C) - (q_2, C) = \text{edge-}e \text{ IS shares-a-point-with edge-}f$$

If this proposition is rejected too and no additional context information (C) is available, the interpreting system may obtain a new input expression by shifting its focus. For example, by focussing on the input for a second time and processing some of the qualia that were complementary now as qualia that are in focus.

$$[q_1 \ q_2 \ C] = [ \{e\} \ \{non\text{-parallel } f \ P\} \ \{intersecting-at \ \text{edge angle}\} ];$$

$$(q_1, C) - (q_2, C) = \text{edge } e \text{ and edge } f \text{ IS intersecting-at } P$$

By making use of additional information from the context (memory), this can be paraphrased as: *edge e and edge f form an angle*.

## Exercise 6

In Sect. 4.1.3, it is suggested that the interaction between the input state and effect qualia re-represents a phenomenon in the ‘real’ world. Why is there a hyphen in *re-representation* and why are there primes in ‘*real*’?

## Answer

Sensory input qualia convey information about phenomena that are external to the brain as an interpreting system. As phenomena are assumed to be interactions, the interaction between the input state and effect qualia must be a re-presentation of a phenomenon, external to the interpreting system. In interpretation, as a process, the input interaction is represented by interactions between representations obtained by the interpreting system. This explains the use of a hyphen in ‘re-presentation’.

Although the real world is only there, we may talk about a specific type of phenomena that are internal to the brain, such as feeling and thoughts. Following our model, thought signs, that are internal, may arise via interpretation from input signs (stimuli), that are external. As a result of interpretation, increasingly more developed signs can be obtained, by means of memory information. As memory information may arise through generalization, interpretation may obtain representations that are only remotely related to the input stimuli. Such representations can be hallucinations, but may refer to phenomena that may not be subject to experience, hence may not be real. A famous example is the question raised in medieval angelology: “How many angels can dance on the head of a pin?”



# Chapter 6

## Perception and cognition

### Exercise 1

Explain the following quote, found in Sect. 5.1.

“... memory response signs ( $a'$ ,  $b'$ ) arise by means of input qualia ( $a$ ,  $b$ ) that trigger memory. Although we may distinguish memory response signs in two collections, which are independent, these signs have a shared meaning.”

### Answer

Memory information arises from input qualia via memorization. As the input qualia are in an interaction, they must be independent; as they are related, they must have a shared meaning. An example are the stone and the illuminating light, in Fig. 4.1, sharing the quality of reflectiveness. As memory information arises from input interactions, memory response qualia must have a shared meaning as well. As for a computational implementation of our model this means that a representation of ‘stone’ and ‘light’, must include information (qualia) representing reflectiveness, in addition to information concerning hardness (stone) and frequency (light).

Input qualia,  $a$  and  $b$ , can be subject to memorization if a corresponding memory value (quale) is as yet not available. If it is available, that is,  $a'$  and  $b'$  both exist, information about their values, as well as their combinatorial properties can be obtained via learning. This may apply to the used

complementary information (context), mediating  $a'$  and  $b'$  into their relation, as well. If, as a result of learning, the intensity of a relation between  $a'$  and  $b'$  grows above threshold, this information can become part of habitual knowledge by the interpreting system.

Computational implementation of the dependencies between  $a'$  and  $b'$  type values may require a complex organization. Those values are a set of qualia and the implementation must be able to represent a link between them, as well as any of their subsets, potentially. See Appendix C.

## Exercise 2

In our model, context information is represented by the expression:  $[\neg A, \neg B]$ ; see Fig. 5.3. Complementation of  $A$  ( $q_1$ ) and  $B$  ( $q_2$ ) is represented by  $(A, \neg B)$  and  $(B, \neg A)$ , respectively. Why are the combinations,  $(A, \neg A)$  and  $(B, \neg B)$ , omitted?

### Answer

Cognitive processing is concerned with a representation of state–effect interactions. Memory information designated by  $A$  and  $\neg A$  both are commonly triggered by  $a$  (cf.  $a*a'$ ,  $a+a'$ ), hence they are not independent. Information by  $a+a'$  involves information by  $a*a'$ , as a possibility (cf.  $a\cdot a'$ ).

## Exercise 3

Define  $a$ ,  $b$ ,  $a'$  and  $b'$ , for the running example of a car passing by. Also define  $A$ ,  $B$ ,  $\neg A$  and  $\neg B$  such that the relation,  $A$  is  $B$ , holds.

### Answer

A possible definition of the input qualia of the processes, perception and cognition, is given below.

- $a$  = **noise**
- $a'$  = car, any-engine, driver
- $b$  = **loud**
- $b'$  = ear-splitting, acceleration, hurrying



$A = a * a' = \textit{noise-of-a-car}$   
 $\neg A = a + a' = \textit{noise-of-any-engine, noise(-caused-by)-driver}$   
 $B = b * b' = \textit{ear-splitting(ly)-loud}$   
 $\neg B = b + b' = \textit{loud-acceleration, loud-hurrying}$

*A is B = Noise by the engine of an accelerating car, or  
A driver must be in a hurry*

## Exercise 4

Explain the following definition by Peirce (CP. 5.568):

[...] the percept's truth consists in the fact that it is impossible to correct it, and in the fact that it only professes one aspect of the percept.

## Answer

The input, which is forced upon, cannot be corrected. It signifies a phenomenon which is there, hence must be true (objectively). In a single run, the interpreting system makes use of a single information element from the context, hence may obtain an explanation of the input interaction from one aspect. For example, in visual information processing, qualities of the reflected light are perceived in qualia. This representation by the retina cannot be corrected. The reaction, e.g., “a stone (is there)”, may represent the observed phenomenon from the perspective of existence. Another representation, e.g., “a large stone (is there)”, may express the input phenomenon from the stance of quantity.

## Exercise 5

Offer an analysis to the utterance, *John likes Mary*, from the perspective of *rule* (cf. legisign aspect). Following our ‘naive’ logical model, this perspective or sign aspect can be represented by the logical relation ‘exclusive-or’. Give reasons for this logical expression of this sign aspect.

## Answer

In the syntactic phenomenon, *John likes Mary*, the subject is *John*, the predicate *likes Mary*. The predicate arises from the relation between the verb (*likes*) and the syntactic complement (*Mary*). The verb, *likes*, designates the habitual syntactic relation, the potential for a complementation with a syntactic complement: *likes*+<*sby*>. This relation can be interpreted in two ways: from the perspective of the verb (*B*), which is in relation with a possible complement ( $\neg A$ ), and from that of the nominal complement (*A*), which is affected by a verb ( $\neg B$ ), or in an active (cf.  $\neg A*B$ ) and passive sense (cf.  $A*\neg B$ ), respectively. The two perspectives cannot be observed simultaneously. This is expressed by the ‘exclusive-or’ relation, logically representing the legisign sign aspect.

# Chapter 7

## Language signs

### 7.1 Exercise

Analyze syntactically the following simple utterances. Utterances are presented by a string and a list of morpho-syntactically finished symbols, given in parentheses. Define the syntactic relational needs for each input symbol. Explain the relations generated by the process of syntactic interpretation (parsing). Also explain what the interpreting system ‘knows’ in each evaluation step (evaluation steps, depicted by an instance of the ‘diamond’ process, are defined by operations triggered by a single input symbol).

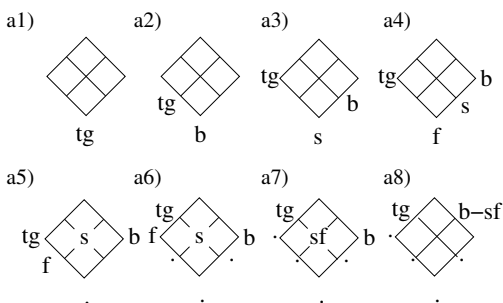
Following Sect. 6.4, lexical definition can be restricted to a specification of active relational needs of a symbol. In the diagrams below, a representation generated by ‘predication’ events can be omitted.

- a) The girl bought some flowers.  
*(The girl)(bought)(some)(flowers)*
- b) The lion in the cage is dangerous.  
*(The lion)(in the cage)(is dangerous)*
- c) Peter, the director, opened the school.  
*(Peter)(the director)(opened)(the school)*
- d) The joggers ran the pavement thin.  
*(The joggers)(ran)(the pavement)(thin)*

## Answer

a) Our model of syntactic parsing makes use of (i) rules dictated by sequential input processing schema, and (ii) those derived from the linguistic properties of the input symbols. An example of a rule of type (i) is the restriction that representations in the icon and sinsign positions cannot be simultaneously realized. An example of a rule of type (ii) is the combinatory need for establishing a relation of a noun, in the rheme position, with an adjective, in the index position.

From the assumption that interpretation may not be forgetful, it follows that appearing input symbols may force the interpreting system to re-present already existing representations through developing them into more meaningful ones. The appearance of a new input symbol, in the qualisign position, may require a representation of an existing symbol (in the same position), by a symbol in the icon or the sinsign position. If the existing symbol designates a nominal, hence a state, it will be represented by a symbol in the icon position; if it is a non-nominal, hence an effect, it will be represented by a symbol in the sinsign position.



**Figure 7.1:** The girl bought some flowers (tg b s f)

The evaluation steps depicted in Fig. 7.1 can be explained as follows:

a1) The interaction between the interpreting system (parser) and the input symbol *the girl* is represented<sup>1</sup> by an expression in the qualisign position (**tg**). At this stage, the system's knowledge is restricted to information about the existence of this symbol, as a quality.

<sup>1</sup>We use the terms 'represent' and 're-present', interchangeably.

a2) The next stage in input parsing is marked by the appearance of the next symbol, **bought** (b), in the qualisign position. As qualisigns are independent by definition (which property must be respected by their re-representation as a sign aspect, as well) and, because the process of interpretation is not-forgetful, the interpreting system is forced to represent the girl (qualisign position), this time from a more meaningful perspective. As **tg** is a nominal, hence a state, it can be represented from the perspective of constituency, in the icon position.

a3) The appearing symbol, **some**, forces the parser to represent **bought**, in an analogous fashion. As **bought** designates a non-nominal, hence an effect, it can be represented from an event perspective, by a symbol in the sinsign position. As in sequential processing, a representation of the input icon and sinsign aspects cannot simultaneously exist (cf. Sect. 6.1.1), the girl has to be represented, this time from the perspective of a qualitative possibility, in the rheme position. Following our theory that, in sequential processing all input symbols are equally considered to be in focus, the girl cannot be interpreted as complementary information (and represented in the index position). An exception of the above hypothesis are adjectives and adverbs, amongst others.

At this stage, the parser may 'know' that the girl is a candidate for the syntactic subject (possibly after a syntactic modification by a complementary symbol in the index position), that the input is an expression of an event (**bought**) and that there is also a quality (**some**), the import of which in the entire sentence is as yet not known. Note that information about the argument structure of **bought** is as yet not operational either. That information will be available later when **bought** is represented by a symbol in the legisign position.

a4) The appearance of **flowers**, in the qualisign position, forces the parser to re-represent **some**, in the sinsign position, conform the non-nominal type of this symbol. In turn, the event interpretation of **some** forces the parser to re-represent **bought**, in the legisign position. The involved law-like syntactic property follows from the syntactic need of **bought** for a complement, as a transitive verb.

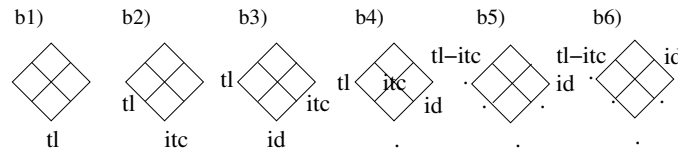
a5) The appearing dot symbol makes the parser represent **flowers**, as a constituent, in the icon position, and in turn **some**, as a syntactic modifier, in the index position. The second interpretation act, which is called a coercion, is a consequence of sequential processing, according to which, input representations in the icon and sinsign positions may not simultaneously exist. Note

that *some*, in the index position, cannot bind with *the girl*, in the rheme position, because of temporal information (cf. qualities) represented by *some*,<sup>2</sup> as well as, due to the already existent verb, *bought*, in the legisign position (which information can be mediated by the index position).

a6) The next dot symbol forces the parser to re-present *flowers*, in the rheme position, which in turn triggers a re-presentation of *the girls*, as the subject of the sentence, in the dicent position.

a7) The next dot symbol forces the establishing of a binding relation between the symbols *flowers* (rheme) and *some* (index). The import of the resulting phrase, *some flowers*, is interpreted as a complementary information. Accordingly, *some flowers* is represented, degenerately (in the process' sense), in the index position. Note that *some flowers* may not force the parser to represent *the girl*, in the argument position (this would make the process backtrack, because of the existence of pending symbols), neither to merge it with the *girl*, via accumulation (the two expressions are not anaphorically related).

a8) The active relational need for a complement, by *bought* (legisign), is satisfied by *some flowers* (index). Finally, the active relational need for the subject, by the verb, in *bought some flowers*, is satisfied by *the girl* (this final *predication* event is omitted in Fig. 7.1).



**Figure 7.2:** The lion in the cage is dangerous (tl itc id)

b) An explanation of the events depicted in Fig. 7.2:

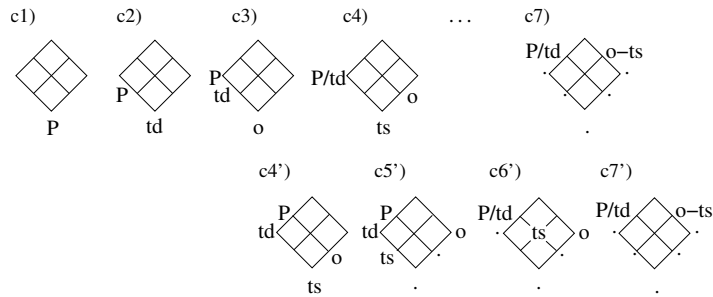
The subject is defined by the syntactic modification event of *the lion*, by the prepositional phrase (PP), *in the cage*.

c) An explanation of the evaluation steps illustrated by Fig. 7.3:

In this analysis, the interesting moment is depicted in diagram c4, in which the symbol, *Peter*, is accumulated with the symbol, *the director*, in the rheme position. This combination of the two symbols is only possible if they are

<sup>2</sup>*Some* follows *the girls* in the input and we assume that post-modification of a noun by an adjective is not enabled by the interpreting system

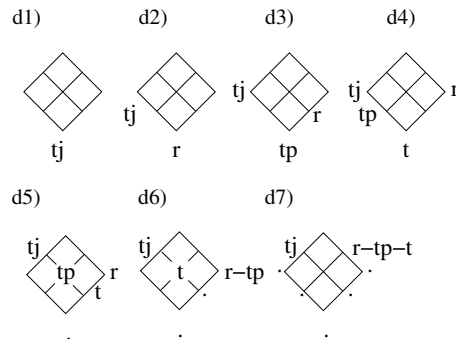
lexically known to be synonymous. Alternatively, the two symbols can be merged via accumulation, in the dicent position, as synonymous expressions of the subject of the sentence. The events of this analysis are depicted in c4'–c7'.



**Figure 7.3:** Peter, the director, opened the school (P td o ts)

d) An analysis of the evaluation steps displayed in Fig. 7.4:

This utterance is known as a syntactic post-modification phenomenon. The binding between the transitive verb, run, and its complement, the pavement, is represented by the symbol, ran the pavement, in the legisign position. The syntactic modification relation of this verb phras with the symbol, thin (index position) eventually obtains the predicate of the sentence, in the symbol position.



**Figure 7.4:** The joggers run the pavement thin (tj r tp t)





# Chapter 8

## Language signs (cont.)

### 8.1 Exercise

Offer a syntactic analysis to the following complex utterances.

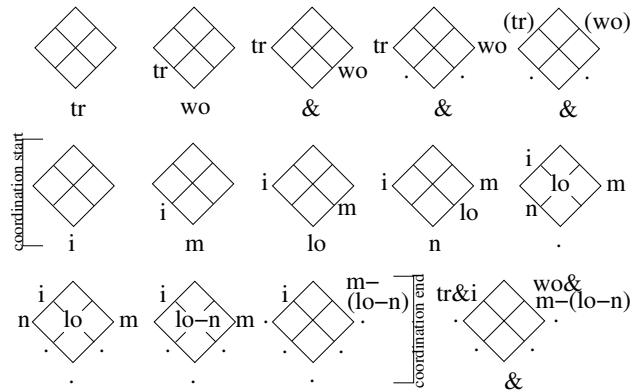
- a) The radio was on and it made a lot of noise.  
*(The radio)(was on)(and)(it)(made)(lot of')(a noise)*
- b) Finally John said, Mary does not like gardening.  
*(Finally)(John)(said)(Mary)(does not like)(gardening)*
- c) The job finished, he went to the pub.  
*(The job)(finished)(he)(went)(to the pub)*

### Answer

As in sequential processing all input symbols are equally considered to be in focus, which hypothesis may have to be withdrawn, Syntactic analysis may ask for a re-analysis, via backtracking. In order to keep our presentation concise, we focus on successful input analyses only.

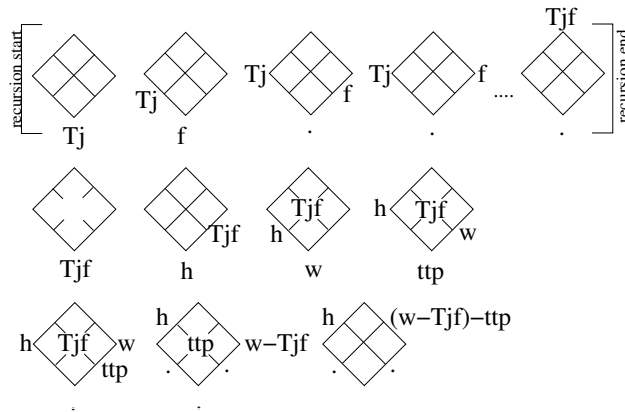
a) In this example we assume that in the clause, preceding the coordinator (*and*), the subject is defined by *the radio*, the predicate by the term *was on*. As, in the clause following the coordinator, the symbol, *it*, is anaphorically related to *the radio* (enabling *made a lot of noise* to coordinate with the predicate), we may conclude that, after parsing the coordinator symbol, *and*, the symbol, *the radio* (rheme position) can be re-presented in the dicent position, as the subject of the sentence. Note that *the radio* (rheme position) involves the sign aspect of a qualitative possibility, conform its potential to be subject to

syntactic modification, e.g., by an adjective (this potential is not actualized in the current example). See Fig. 8.1.



**Figure 8.1:** The radio was on and it made a lot of noise (tr wo & m lo n)

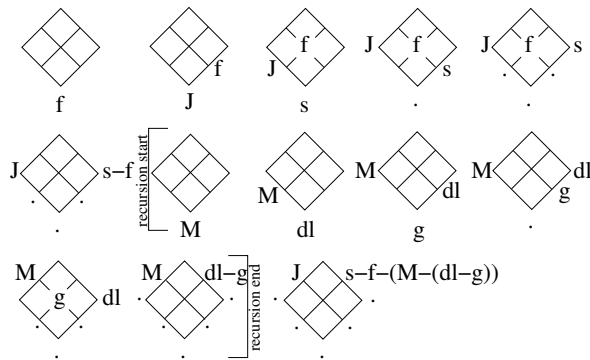
b) Syntactic analysis can be enhanced through substituting complex phrases by a simple term, conceptually. For example, by replacing the closed clause, ‘the job finished’, by the adverb ‘finally’. See Fig. 8.2



**Figure 8.2:** The job finished, he went to the pub (Tj f h w ttp)

c) Another example is the replacement of the complex complement of the verb (said), by a simple nominal: Finally, John said <something>. The complement, which is a closed clause, can be parsed by a nested, recursive process. Note that a recursive analysis may not be necessary if the positions of the

processing schema required by such an analysis are free (i.e. do not represent an input symbol). This does not apply to the current example. The symbol, *said*, in the legisign position, needs a complement, represented in the index position. As an analysis the complement asks for a representation of the phrase, *does not like*, in the legisign position (which is not free), recursive analysis is a must. See Fig. 8.3.



**Figure 8.3:** Finally John said, Mary does not like gardening (f J s M dl g)

## 8.2 Exercise

Offer a morpho-syntactical analysis to the following strings. Specify the morpho-syntactic relational needs of all input symbols. An analysis of layout symbols can be omitted.

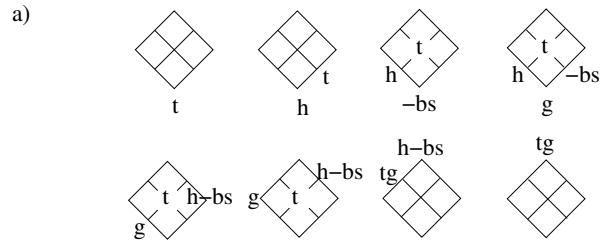
- a) the happy -bs girl
- b) in a far -bs off land

### Answer

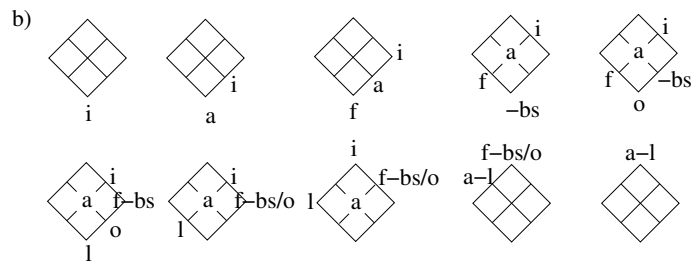
An analysis of (a) is depicted in Fig. 8.4, that of (b) in Fig. 8.5.

## 8.3 Trichotomic specification

Give a trichotomic specification of the symbols:



**Figure 8.4:** the happy -base girl (t h -bs g)



**Figure 8.5:** in a far -base off land (i a f -bs o l)

- a) reason
- b) read

## Answer

There may be numerous definitions possible, that differ in the perspective taken. The answers below correspond to a syllogistic perspective (a), and a depth of interpretation (b).

- a) reason: (1) proposition < (2) premise < (3) conclusion (act)
- b) read: (1) articulate < (2) skim < (3) explain

# Chapter 9

## Reasoning signs

### 9.1 Exercise: Sample abduction

Assume you are travelling in the train and you make acquaintance with a young girl, Mary, and an elderly man, John, sitting next to her. The man is telling you that he has five daughters. In the noise you hear four names only: Nana, Nene, Nini, Nono. Surprised, as you are, you say: “So you have four daughters only!”. Upon which, the man is answering: “No, five!”. From this you conclude that the fifth daughter must be Mary.

Model your reasoning, as a process. You may assume the potential by the interpreting system for counting one-by-one and representing the result by a single symbol. E.g., the result of counting three similar input items may obtain the symbol ‘111’.

#### Answer

A solution can be given analogously to the example of sect. 7.3, in the book (cf. roe dear). The observation of an injured tree is comparable to the observation of four names only.

We assume that the names are interpreted from the perspective of number and counted in an accumulative fashion. As a result, each one of the names, Nana, Nene, Nini, Nono, is represented by a unary number ‘1’; their collection by the symbol ‘1111’. Eventually, this enables John’s current information (“Nana, Nene, Nini, Nono”) and his earlier one (“five daughters”), which is memoried, to be represented as follows.

$$a_f = \{\text{Mary, John, 1111}\},$$

$$a'_c = \{11111\}.$$

In addition, we may assume the existence of complementary information about 'Mary' and 'John', such as, 'daughter', 'father-of', 'John-has-a-daughter'.

John's negative response makes you reconsider your earlier interpretation. To this end, the process model offers the following possibilities:

- a) try and use another complementary memory information;
- b) shift your focus to other input qualities;
- c) introduce *new* input qualities, abductively:

$$new-a := a' * a; \quad new-b := a' \setminus a.$$

We assume that alternative (c) above is only applicable. In addition we assume that all earlier information is still available (interpretation is not forgetful).

$$a_f = \{\text{Mary, John, 1111}\}$$

$$a_c = \{\}$$

$$a'_f = \{\text{has-a-daughter}\}$$

$$a'_c = \{11111, \text{father-of}\}$$

This enables new information to be generated:

$$new-b := a'_c \setminus a_f$$

$$= 11111 \setminus 1111$$

$$= 5th_{new}$$

$$new-a := a'_f * a_f =$$

$$= \text{has-a-daughter} * \{\text{John, Mary}\}$$

$$= \text{John-has-a-daughter-Mary}_{new}$$

As a result of the new effect, Mary may appear as John's 5th daughter.

$$a_f = \{\text{Mary, John, 1111, John-has-a-daughter-Mary}_{new}\}$$

$$b_f = \{5th_{new}\}$$

Analogously to the roe deer example, the current interpretation process may enable the development of a new habit.

$$a_c = \{\text{noise-qualia}\}$$

$$a'_c = \{\text{concept-of-noise}\}$$

By shifting your focus to the qualia above and interpreting them in the context of information about similar phenomena observed in the past, you may conclude: "In the case of noise, the input can be corrupted. So it is reasonable to try and find an interpretation in an abductive fashion".

# Chapter 10

## ‘Naive’ mathematics

### 10.1 Exercise 1

Assume you are familiar with rational numbers, but have not yet studied irrationals. One day your teacher gives the task: “*Show that  $\sqrt{2}$  is rational*”. Because you think the task must be reasonable, you try to ‘solve’ it in a usual fashion, as is illustrated with definitions below. We use a ‘\*’, and a ‘+’ (or alternatively, a ‘,’) symbol, for a relation between qualia in the sense of agreement, and possibility, respectively. E.g., **sqrt\*2** stands for an agreement relation between memory qualia of **sqrt** and 2. The symbols, *a*, *a'*, *b*, *b'*, may designate a single quale, as well as a collection of qualia, ambiguously. We use abbreviations: ‘sqrt’= square root, ‘rat’= rational, RP= relative prime, (f)= focus, (c)= complementary.

```
a=  sqrt(2)
a'= (f) sqrt*2                % sqrt(2) is related to sqrt and 2
    (c) p/q                    % sqrt(2) has the form p/q (cf. state)
b=  rat
b'= (f) type                  % rat is related to a mathematical type
    (c) p/q                    % rat has an idiosyncratic shape,
    % as a property (cf. effect)
```

This enables *A* and *B* to be defined as follows:

```
A=  sqrt(2)*(sqrt*2)
B=  rat*type
```

As you are familiar with this mathematical type only, you conclude:

$$\begin{aligned}
A \text{ is } B &= \mathbf{sqrt(2)} * (\mathbf{sqrt*2}) \text{ IS } \mathbf{rat*type} \\
&= \textit{'sqrt(2) IS rational'}.
\end{aligned}$$

Your teacher is telling you that you are mistaken. Explain how an answer, proving that  $\sqrt{2}$  is not rational, can be generated in an abductive fashion.

Hint:  $\sqrt{2}$  is irrational means that it cannot be expressed by a division of a pair of integers. Assume, nevertheless, that  $\sqrt{2}=p/q$ , for some relative primes,  $p$  and  $q$  (this property means that  $p/q$  cannot be simplified). By taking the square of both sides of the above expression you may get:  $2=(p/q)*(p/q)=(p*p)/(q*q)$ , revealing 2 to be a common denominator of  $p$  and  $q$ , contradicting the assumption that they are relative primes.<sup>1</sup>

## Answer

In this example we restrict ourselves to the interpretation process involved in the proof of  $\sqrt{2}$  is irrational. A derivation of a proof itself is beyond our horizon. Your teacher's 'no' may refer to the use of wrong (background) information (wrong reasons), as well as the correctness of your conclusion (proposition), ambiguously. Below we focus on the second possibility, immediately above, and assume that the teacher's response make you obtain a revised conclusion, abductively.

## Abductive reasoning

From the existing input qualia (cf. 'primordial-soup') new qualia can be derived, abductively. In this section we make use of the convention that rational numbers can be expressed by a division of a pair of integers, that are relative primes ( $\mathbf{p/q}$ ). This symbol can be used as a property (or effect), as well as a form (cf. state). Analogously, '(gcd=1)' may designate the algorithm, 'greatest common divisor', 'yielding 1' (or '=1'), as a property or an effect, as well as, the result obtained by the algorithm ('1'), as a value or state. Finally, ' $\uparrow^2$ ' may stand for 'the second power of (-)', as an operation or effect, and a function application, hence a form or state, ambiguously. The subscript, '*new*', indicates that a symbol is defined via abduction.

---

<sup>1</sup>For a pair of relative primes  $p$  and  $q$ , their square,  $p*p$  and  $q*q$ , are relative primes too. See also [www.homeschoolmath.net/teaching/proof\\_square\\_root\\_2\\_irrational.php](http://www.homeschoolmath.net/teaching/proof_square_root_2_irrational.php) (2012).



$$\begin{aligned}
new-b &= p/q_{new} \\
&\quad \% a' \setminus a: \text{ an effect that may turn } \mathbf{sqrt(2)} \text{ into } p/q \\
new-b' &= (f) RP_{new} \\
&\quad (c) ((gcd=1)+\uparrow^2)_{new} \\
&\quad \% (gcd=1), \text{ as a property or value 1;} \\
&\quad \% RP \text{ is invariant under } '\uparrow^2' \\
new-a &= (\mathbf{sqrt*2})_{new} \\
&\quad \% a' * a: \text{ what is common in } (\mathbf{sqrt*2}) \text{ and } \mathbf{sqrt(2)} \\
&\quad \% \text{ e.g., } \mathbf{sqrt(2)}, \text{ as a number} \\
new-a' &= (f) p/q_{new} \\
&\quad \% 2 \text{ is a rational, hence involves the form } p/q \\
&\quad (c) (\uparrow^2+(gcd=1))_{new} \\
&\quad \% \uparrow^2 \text{ can be applied to } \mathbf{sqrt(2)}
\end{aligned}$$

Existing and abductively generated input qualia can be combined (interpretation is not forgetful).

$$\begin{aligned}
a &= \mathbf{sqrt(2)}, (\mathbf{sqrt*2})_{new} \\
a' &= (f) (\mathbf{sqrt*2}), p/q_{new} \\
&\quad (c) p/q, (\uparrow^2+(gcd=1))_{new} \\
b &= \mathbf{rat}, p/q_{new} \\
b' &= (f) \text{ type, } RP_{new} \\
&\quad (c) p/q, ((gcd=1)+\uparrow^2)_{new}
\end{aligned}$$

## Input processing re-visited

The arising new input qualia may trigger further interpretation, as a process. We assume that information processing has the potential to consider any subsets of  $a$ ,  $a'$ ,  $b$ , and  $b'$ , to be qualia that are in focus, or complementary. Below, we restrict ourselves to an illustration of successful input representations, as usual (for the sake of simplicity, the subscript 'new' can be omitted).

$$\begin{aligned}
A &= a*a' = (\mathbf{sqrt*2})*p/q && \% \text{ new qualia that are in focus} \\
B &= b*b' = p/q*RP \\
\neg A &= a+a' = \uparrow^2+(gcd=1) && \% a'_c = p/q \text{ is omitted} \\
\neg B &= b+b' = p/q+(gcd=1)+\uparrow^2
\end{aligned}$$

Here,  $A=(\mathbf{sqrt*2})*p/q$  stands for the function 'sqrt' and the value '2', in relation to 'p/q', as an expression of division, as a form;  $B=p/q*RP$  designates the agreement sense relation between  $p/q$  and  $RP$  as a property.

Some of the important interpretation moments of the above input are explained below. According to our model, the rheme position can be represented by  $A$ , the legisign position by  $B$ , the index position by  $\neg A$  and  $\neg B$ . We assume that  $\text{sqrt}$  and  $\uparrow^2$ , as references to algorithms (cf. memory information), are related by a relation of conversion (see Sect. 6.7.5). As we look at our input through the glasses of computation, the symbol interactions (‘-’) between the positions, rheme – index and index – legisign can be explained as follows.

$$\begin{aligned}
& \text{rheme}(A) - \text{index}(\neg B) \\
&= (\text{sqrt} * 2) * p/q - p/q + (\text{gcd}=1) + \uparrow^2 \\
&= (\text{sqrt} * 2) * p/q * p/q \\
&\quad + (\text{sqrt} * 2) * p/q * (\text{gcd}=1) \\
&\quad + (\text{sqrt} * 2) * p/q * \uparrow^2 \\
&= (\text{sqrt} * 2) * p/q * \uparrow^2 \\
&= 2
\end{aligned}$$

$$\begin{aligned}
& \text{index}(\neg A) - \text{legisign}(B) \\
&= \uparrow^2 + (\text{gcd}=1) - p/q * RP \\
&= \uparrow^2 * p/q * RP \\
&\quad + (\text{gcd}=1) * p/q * RP \\
&= 1
\end{aligned}$$

Note that, from a computational stance,  $(\text{gcd}=1)$ , hence also 1,  $p/q$ , and  $RP$ , all stand for the same property; and that the properties,  $p/q$  and  $RP$ , are invariant for the operation  $\uparrow^2$ .

The goal of ‘naive’ interpretation is the proposition of a relation between the subject and predicate of a phenomenon. As, ‘2’ (subject) cannot be predicated by ‘=1’ (predicate), the proposition “False” can only be obtained. By assuming that all possibilities of interpretation have been exhausted, we may conclude that the original input must have been false, hence  $a'(c)=p/q$  must be incorrect, and so the proposition “**sqrt(2) IS rat**”, must be incorrect as well. Eventually we may conclude: *sqrt(2) is not rational*.

Note that a conceptualization of ‘*sqrt(4) is rational?*’ may proceed along the same lines. Indeed, if somebody is telling you that ‘*sqrt(4) is rational*’ is false, you may conclude that this proposition is wrong, only if you are familiar with the property that  $\text{sqrt}(4)=2$  and know that 2 is rational.

## 10.2 Exercise 2

What is the difference between the mathematical and the ‘naive’ concept of infinite?

### Answer

The concept of countably infinite is a mathematical abstraction. An example is the numerosity of the set of natural numbers,  $1, 2, 3, \dots$ . ‘Naive’ infinity, which refers to a property that we may perceive, is different from the formal mathematical concept. Indeed, for a sufficiently large integer value, say  $k$ , we may not be able to experience a difference between  $k$  and  $k+1$ , and between  $k+1$  and  $k+2$ , and so on. Because of a lack of a difference (each one of these numbers arises from the previous number through incrementation), these values must be synonymous, from a number perspective. An experience of their synonymous character is our perception of ‘naive’ infinity.



# Chapter 11

## Text summarization

### 11.1 Exercise 1

Consider the text below. Generate a summary in a single sentence, from the perspective of Prince.

- (s<sub>1</sub>) There lived a Princess.  
(There)(lived)(a Princess)
- (s<sub>2</sub>) A Prince saw her beauty and fell in love with her.  
(A Prince)(saw)(her)(beauty)(and)(fell)(in love)(with her)
- (s<sub>3</sub>) They married.  
(They)(married)

### Answer

Following our theory, summarization must be preceded by a syntactic analysis of the input sentences and clauses, commonly refer to as sentences. Summarization of a pair of sentences asks for a unification of the subjects, as well as the predicates of the individual sentences. A succesful unification enables a pair of sentences to be replaced by their summary. In the course of this process, less meaningful symbols can be omitted.

- s<sub>1</sub>  
'There', which is a placeholder for the syntactic subject, can be removed. The subject is 'a Princess', the predicate is 'lived'.

A Princess lived.

- $s_2$

The coordination structure, ‘saw her beauty and fell in love with her’, can be summarized by making use of the following trichotomies:

verb:                   existence < modification < transformation  
 transformation:   neutral < modulation < change

Indeed, saw(1)<fell in love(3), as ‘saw’ is a neutral effect, as opposed to ‘fell in love’ which may designate an irreversible change (cf. transformation). As the complement, ‘with her’, cannot be removed, because of the anaphoric reference to ‘Princess’ (in  $s_1$ ), we may simplify  $s_2$  as follows:

A Prince fell in love with her.

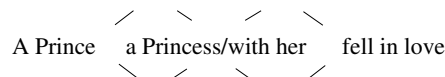
- $s_3$

An analysis of this most simple sentence does not provide new information.

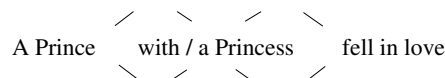
They married.

A summarization of  $s_1$  and  $s_2$  may *not* be possible. This is a consequence of the fact that the subjects in two sentences are different. As  $s_2$  is anaphorically related to  $s_1$  (through ‘her’), we may assume that  $s_1$  signifies an episode, functioning as a context, in the analysis of  $s_2$ . To this end, we represent the sentence symbol, ‘A Princess lived’, by the phrase, ‘A Princess’. Although the representation is degenerate (in a syntactic sense), the anaphoric relation with ‘her’ is kept invariantly.

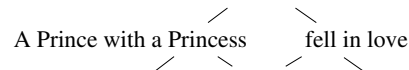
Following the above line of thinking, we assume that  $s_2$  is syntactically analyzed in the context of ‘A Princess’, used as complementary information. The verb phrase, ‘fell in love with her’, can be interpreted as a representation of an interaction between the symbols, ‘fell in love’ (legisign) and ‘with her’ (index). Luckily, the second symbol can be interpreted as a converse of ‘a Princess’, which enables the following instantiation of the interpretation process.



The above representation can be simplified by means of the anaphoric relation between ‘her’ and ‘Princess’.

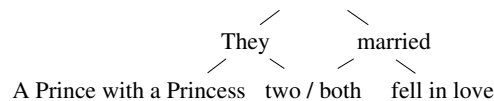


The symbol interaction between ‘A Prince’ (rheme) and ‘with a Princess’ (index) can be represented by ‘A Prince with a Princess’ (dicent), designating the subject of the summarizing sentence of  $s_1$  and  $s_2$ . The predicate can be defined by the verb phrase, ‘fell in love with a Princess’, or simply, ‘fell in love’. This simplification is enabled by the idempotence of information for addition (cf. repetition), which property is frequently used in natural language.



**Remark** The anaphoric relation of ‘her’ with ‘Princess’ enables an alternative summarization of  $s_1$  and  $s_2$ . As ‘fell in love’ entails the existence of two actors (cf. a number interpretation), by introducing a common expression for the actors, e.g., the symbol ‘a couple’ (index position), ‘the Prince’ and ‘with a Princess’ can be combined into a single expression. This alternative, which is semantically more demanding, is not considered in our analysis. (**End**)

Summarization of  $s_2$  and  $s_3$  ( $s_2 \Rightarrow s_3$ ) may proceed as follows. Through a ‘naive’ mathematical (number) interpretation, from the subject of  $s_3$ , ‘they’, we may deduce that it must stand for two or more persons. As ‘A Prince with a Princess’ refers to two persons, this symbol can be unified with the symbol, ‘they’. This way we may obtain the subject of a summarizing sentence, ‘A Prince with a Princess’ (the semantically less meaningful symbol, ‘they’, can be omitted). Information about the involved numerosity can be represented by the symbol ‘two’ and its converse, the symbol ‘both’ (index position). The second symbol facilitates a unification of the predicates, ‘fell in love’ and ‘married’, in the symbol ‘married’. This can be justified by the trichotomical specification of ‘partnership relation’, as bachelor < engaged < married. According to this, ‘(both) fell in love (synonymously expressed by engaged) can be semantically less meaningful than ‘married’.



A summary of the entire input text can be ‘A Prince with a Princess married’, paraphrase by the syntactically correct sentence, ‘A Prince married with a Princess’.

## 11.2 Exercise 2

Summarize the following short poem:

“Roses Are Red”  
Roses are red violets are blue  
But nothing is as pretty as you!!

We may begin with an analysis of the clause, preceding the coordinator ‘but’. The phrases, ‘violets’ and ‘roses’, can be summarized in the symbol ‘flowers’. We may assume that the unknown poet used ‘is red’ and ‘is blue’ as an expressions of attractiveness or a measure of ‘beauty’ (such as nice, lovely, etc.). This must be so, as the alternative interpretation, ‘Flowers are colorful, but nothing is as pretty as you’, for instance, does not make much sense.

Roses    are    red  
Violets    are    blue    Following step 3, in Sect. 9.3, the sentence headed  
⇒Flowers    are    nice

by ‘but’ can be interpreted as a syntactic complementation, in a negative sense, logically. This is designated below by the symbol ‘neg’.

We may distinguish between three cases of an effect of negation. We make use of the convention that (neg) nothing may equivalently be represented by the symbol everything and, that (neg) as pretty can be interpreted as less or more pretty, ambiguously (in a semantic sense). An analysis of the three cases of an effect of negation is the following.

a) Negation applies to the subject and predicate, both, distributively.

(neg) [nothing    is    as pretty as you]  
=(neg)nothing    is    (neg) as pretty as you  
=everything        is    less or more pretty than you

By assuming that everything<flowers, e.g., from the stance of referential content, then we may get the conventional summary: ‘Flowers are less pretty than you’. By taking an alternative interpretation, using the order relation nice<less or more pretty, as an expression of growing referential information, we may derive the summary: (Even) flowers are more pretty than you.

b) Negation applies to the subject only

(neg) [nothing    is    as pretty as you]  
=(neg)nothing    is    as pretty as you



If everything < flowers and nice < as pretty as, we may obtain the summary:  
'Flowers are as pretty as you' or 'You are as pretty as flowers'.

c) Negation applies to the predicate only

(neg) [nothing is as pretty as you]

=nothing is (neg)as pretty as you

=nothing is less or more pretty as you

By assuming the ordering, nothing < flowers, we may obtain the summary  
derived in (b), again.



# Chapter 12

## Sample text analysis

In the below text fragment, in Table 12.1, the author expounds his concept of ‘Information Retrieval Systems’ (Huibers, 1996). Arguably his definition can be found plausible. We ask: what can this be attributed to? To this end we consider the question how the words, phrases, clauses, etc., and their order of presentation may contribute to an efficient communication of the intended meaning by the author?

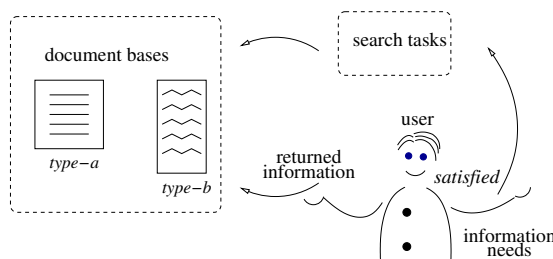
<i>label</i>	<b>Information Retrieval Systems</b>	<i>abbreviation</i>
s <sub>1</sub>	There are several document-bases.	sev-docb
s <sub>2</sub>	Each document-base contains different types of information.	each-docb dt-of-info
s <sub>3</sub>	There are various types of users and there are vast differences between their information needs.	vt-of-users vdiff-inneeds
s <sub>4</sub>	There are various kinds of search-tasks, or stated differently, there are several ways in which	vk-of-st sev-ways
s <sub>5</sub>	a user can be satisfied with the returned information.	canb-satf-w- -ret-info

**Table 12.1:** Sample text

## Answer

Below we assume that the text, in Table 12.1, is an account of a phenomenon in the author's perception. In order to get hold of the meaning of this phenomenon, the reader is invited to draw a picture of his/her understanding of the text. A possible solution is depicted in Fig. 12.1.

We also assume that the author's interpretation of his phenomenon must be the ground for his formulation of the text. We suggest that, through communicating his words, phrases and clauses, in the given order, the author provides a 'guided tour' of his concept of 'Information Retrieval Systems', as a phenomenon. The author's formulation of the text is such that it enables the reader to 'generate' a similar interpretation through language processing.



**Figure 12.1:** Sample phenomenon

## Text analysis

An interpretation of our text fragment assumes a syntactic and semantic analysis of some sort, and therefore we need to provide at least a reasonably detailed analysis at that level. Note, however, that there are certainly other possible analyzes than the one presented here, quite probably rendering different interpretations. The primary goal of this analysis is merely to describe the clusters of language symbols we take to be the input of a 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 (Quirk, Greenbaum, Leech, & Svartvik, 1985). As a full syntactic analysis using the language model introduced in this book is more tedious, it is omitted.

- In  $s_1$  we have so-called existential *there* as the 'grammatical subject', followed by main verb *be*, and *document-bases* as the 'notional subject' with the

quantifier *several*. There is a syntactic place-holder, in the regular subjective position, for the ‘notional subject’.

- In  $s_2$  we observe: *Each document base* (subject); *Each*, which relates to the quantifier *several* in  $s_1$ ; *contains*; a 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  $s_3$  is similar to that of  $s_1$ ; coordinating and simply joins the two clauses of this sentence. We again have existential *there* and *various types of users* (subject). The adverbial prepositional phrase makes explicit that the *vast differences* hold between the information needs of the users. Through the anaphoric reference of the possessive pronoun *their*, *vast differences between their information needs* becomes the predicate of *various types of users*.

- $s_4$  is analyzed similar to  $s_1$ . *Or stated differently* is an explicit statement at text grammatical level, indicating that the first clause is paraphrased by the second clause of the coordination structure in  $s_4$ . 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  $s_4$ , not unlike in  $s_2$ , 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 *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 *different types of information* in  $s_2$ ; *returned* is never explicitly mentioned, only hinted at in  $s_4$  (assuming that tasks have something to do with the returning of information). In fact, it could be argued that, given that  $s_4$  already introduces *search tasks*, the explicit in-

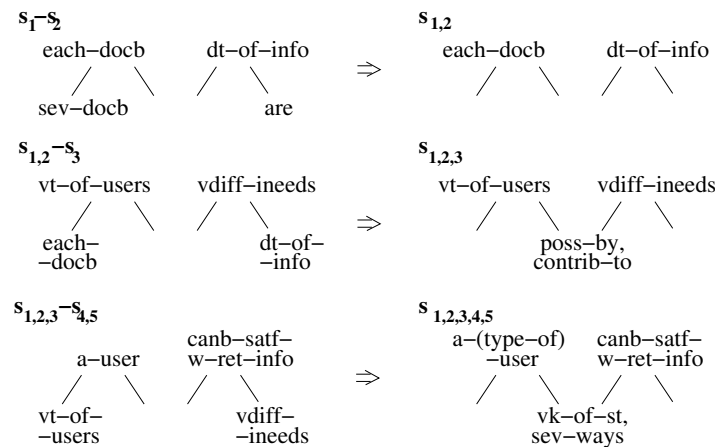
roduction of returned information in the text is the most important semantic contribution of  $s_5$ .

Having identified clusters of symbols, we now return to the question why these clusters, in the particular order in which they occur, do indeed enable an interpretation intended by the author.

### Series of phenomena analysis

Following the theory of this book we may give an answer to our question in two ways. The first one, text summarization, may provide an answer through considering pairs of sentences and generating a summarizing single sentence, in an iterative fashion. This approach may require the use of ‘naive’ reasoning. The second way, single phenomenon conceptualization, may obtain an answer through interpreting sentences as expression of sign aspects of a phenomenon. This approach is analogous to apparent motion perception, as explained earlier in the book.

Below we refer to language clusters of the input text by means of labels and abbreviations. See Table 12.1. A summarization event is designated by a ‘ $\Rightarrow$ ’ symbol, the obtained summary by the labels of the used language clusters, separated by a comma. For instance,  $s_1-s_2 \Rightarrow s_{1,2}$  designates a summarization of  $s_1-s_2$ , in the summary  $s_{1,2}$ . Trichotomies used by our analysis are defined ‘on the fly’. An overview of the summarization events is depicted in Fig. 12.2.



**Figure 12.2:** Summarization events

- $S_1-S_2 \Rightarrow S_{1,2}$

By making use of a semantic classification of referential information, *ref-to-whole*<*ref-to-element*, we may deduce that *sev*<*each*, enabling *sev* to be omitted. An analogous classification of verbs, *existence*<*modification* (e.g. a ‘modification’ of an entity by a property), enables *are* to be removed. These operations are depicted in Fig. 12.2, in the first row (context signs imputed by the used semantic orderings are omitted in this diagram).

- $S_{1,2}-S_3 \Rightarrow S_{1,2,3}$

In  $s_3$ , the ‘focus’ is shifted from documents (*docb*), to their owners (*users*). A classification of ownership information, *owned*<*owner*, enables *each-docb*, interpreted as *owned*, to be removed. Similarly, we may deduce that *dt-of-info*<*vdiff-inneeds*, as *dt-of-info* (cf. *owned*) form the basis for and contribute to *inneeds* (cf. *owner*). The adjective *vdiff* is an expression of a 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 (if an act of possession can take place then there must be something contributing to hence enabling that event). See Fig. 12.2 (second row).

- $S_{1,2,3}-S_{4,5} \Rightarrow S_{1,2,3,4,5}$

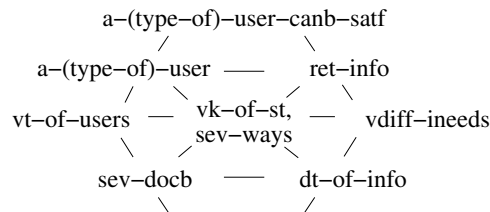
Coordination, defined by *vk-of-st* and *sev-ways*, in  $s_4$ , is degenerately represented in the index position, by the expressions ‘*vk-of-st, sev-ways*’. That a synonymous interpretation of these symbols is possible, is by virtue of the coordinator, or stated differently, presenting them as converse formulations of ‘search’, as an action (cf. *effect*) and an act (cf. *state*). Put differently, if *vk-of-st* may occur, then there must be *sev-ways*, in which ‘search tasks’ and ‘returning of information’ (or ‘information returning’) can be realized. A degenerate representation of the above expressions, as complementary symbols, is motivated by the fact that there is no reference to them 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-inneeds* 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 between *a-(type-of)-user* and *canb-satf*. A linguistic expression of the summary is omitted.

## Single phenomenon analysis

Following this approach, the input sentences can be interpreted as expressions of sign aspects of a phenomenon. To this end we define a collection of qualia potentially involved. By showing that those qualia can be represented by a proposition (cf. argument position), we may say that their collection can be well-formed (from some perspective).

Well-formedness of our text, in Table 12.1, can be shown by taking the summary `a-(type-of)-user-canb-satf`, generated earlier, and deriving interpretation moments of a process, on the basis of the semantic order relations used. See Fig. 12.3. The only interesting event is the interpretation of modality, involved in the predicate (`can be`), as a contribution of the *predication* symbol interaction itself. This enables the symbol position to be represented by the event ‘information returning’ (`ret-info`).



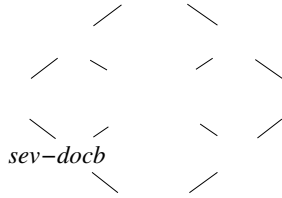
**Figure 12.3:** Single phenomenon analysis (a representation of the qualisign sign aspect is omitted)

We assume that, in human conceptualization, summarization and an interpretation of (summarized) text clusters as sign aspects of a hypothetical phenomenon may proceed simultaneously. An analysis of the second way of processing is the subject of this section.

Below, the assignment of a sign aspect to a language expression is indicated by an “=” symbol; positions of the process model are referred to by the Peircean sign aspects.

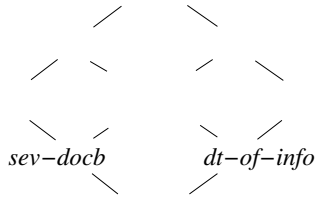
- ( $s_1$ ) `sev-docb=icon`  
A postulation (are) of `sev-docb` as constituent entities, not as qualitatively possible ones (rheme), nor as such entities in context (dicent). The hypothesis immediately above can be justified by the fact that, besides  $s_2$ , `sev-docb` has no later references in the text.





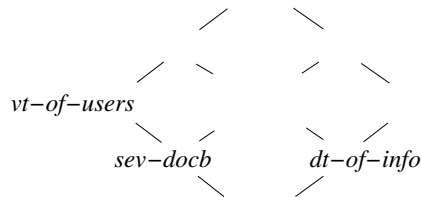
- (s<sub>2</sub>) dt-of-info=sinsign

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



- (s<sub>3</sub>) vt-of-users=rheme

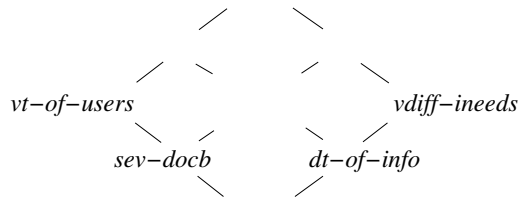
An expression of existent entities (*are*). The later anaphoric reference to *users*, by *their*, in *s<sub>4</sub>*, enables *vt-of-users* to be interpreted as an expression of a range of possibilities ('what can types of users be in general'). Following the dependencies between the Peircean sign aspects, the interpretation of *vt-of-users* in the *rheme* position implies a representation of *users* in the *icon* position (in the diagram below this is omitted), enabling *users* and *docb* to be synonymously interpreted as constituents (this is represented by *sev-docb*).



- (s<sub>3</sub>) vdiff-inneeds=legisign

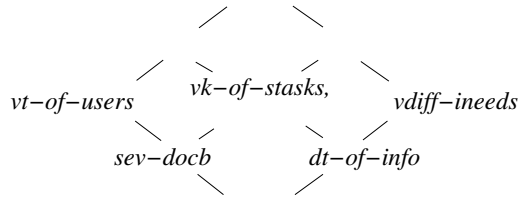
A generalization of the event of an appearing new property, *dt-of-info*, in the type of an event, represented by *information-needs*. A generation of this symbol is possible through interpreting the relative difference between the

expressions in the icon and sinsign positions, each and several, and different-types-of, respectively. The hypothesis immediately above is confirmed by the rule-like compatibility of vast differences (effect) and information needs (state), which is expressed by *vdiff-ineeds*. Due to the dependencies between the Peircean sign aspects, the interpretation of *vdiff-ineeds*, in the legisign position, implies an interpretation of vast differences and between in the sinsign position (in the below diagram this is omitted). Note that the meaning of these expressions is already included in the meaning of *dt-of-info*.



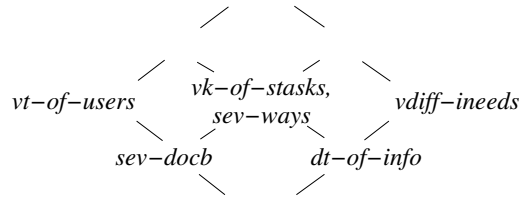
- (s<sub>4</sub>) *vk-of-st=index*

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 an expression of the input in the rheme or dicent position either. According to the preferred interpretation, *vk-of-st* is representing an event, not by explaining it in any way, but by pointing in its direction. Due to the dependencies between the Peircean sign aspects, the interpretation of *vk-of-st*, in the index position, implies the existence of complementary qualia (in the qualisign position), that must be included in the meaning of the rheme and legisign expressions as well.



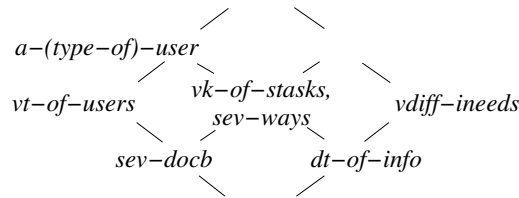
- (s<sub>4</sub>) *sev-ways=index*

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



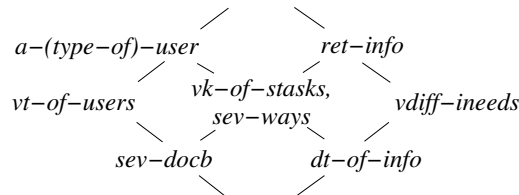
- (s<sub>5</sub>) a-user=dicent

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



- (s<sub>5</sub>) 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 ‘information returning’. Following this line of thinking, canb-satf can be interpreted as a contribution by the predication symbol interaction between ret-info and a-(type-of)-user, generating a representation 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 called ‘information returning’ indeed.



By summarizing the results of the two ways of analysis above we may conclude that, by formulating language expressions, in the given order, the author is likely to have attempted to enhance the reader’s conceptualization through respecting the types of events of his/her process of interpretation. We believe that an analogous analysis of well-formedness of documents in organizations could be practical as well.

## References

- Brent, J. (1993, 1998). *Charles sanders peirce: A Life*. Bloomington and Indianapolis: Indiana University Press.
- Dumitriu, A. (1977). *History of logic* (Vols. 1–4). Tunbridge Wells: Abacus.
- Huibers, T. (1996). *An axiomatic theory for information retrieval*. Unpublished doctoral dissertation, University of Nijmegen.
- Quirk, R., Greenbaum, S., Leech, G., & Svartvik, J. (1985). *A Comprehensive Grammar of the English Language*. London: Longman.