The Cognitive Style of Decision Making Narrations: INDIVIDUAL INVARIANCE AND COGNITIVE AUTOMATA

Alain Finkel and Isabelle Tellier

LIFAC (Laboratoire d'Informatique Fondamentale et Appliquée de Cachan)

ENS de Cachan

61 avenue du Président Wilson

94 235 Cachan Cedex

FRANCE

Tel : (1)-47-40-24-04

Fax : (1) 47-40-24-64

E-mail : Alain.FINKEL@lifac.ens-cachan.fr

Isabelle.TELLIER@lifac.ens-cachan.fr

A partial preliminary version of this paper has been presented at the Fourth International Colloquium on Cognitive Science (ICCS 95) which took place in Donostia-San Sebastian, (Spain), May 3-6, 1995 and will be published in *The Cognitive Level*, Endres-Niggemeyer (Ed.), Duisburg LAUD series of Cognitive Linguistics in 1996. Some of the main ideas of this paper have also been presented by the second author at the French Second Colloquium of Young Researchers in Cognitive Science, Giens (France), June 7-9, 1996.

The Cognitive Style of Decision Making Narrations: INDIVIDUAL INVARIANCE AND COGNITIVE AUTOMATA

ABSTRACT

This article proposes a new model for the analysis of texts, particularly adapted to the study of verbalizations. The model has been built to characterize the invariance between texts written by the same author. It includes two levels :

- at the enunciative level, lexical and syntactical regularities in the expressions of an identical intention are represented by « enunciative patterns » ;

- at the cognitive level, the global organization of discourses is taken into account and common points between various verbalizations of a same task are captured in « cognitive patterns ».

« Cognitive automata » are defined to sum-up all information collected. They provide a formal approach of what language tells us about the one that tells it.

1. INTRODUCTION

The usual viewpoint in computational linguistics and formal semantics considers natural language as a communication tool, as a code used for the exchange of information. Unfortunately, natural languages are defective, full of exceptions and ambiguities. So, some researchers in semantics try to define a universal and perfect code (for example a logical one) capable of translating the sense of every other language.

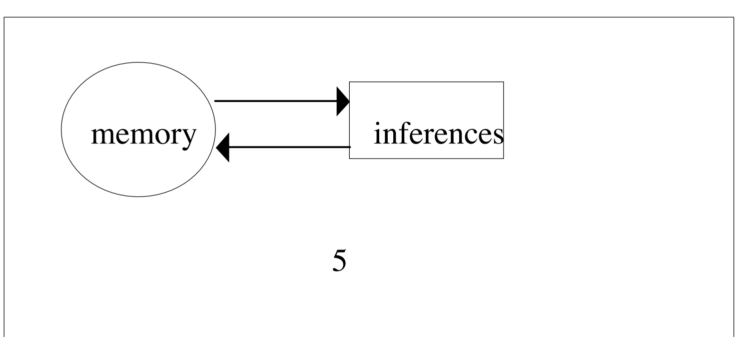
But, as cognitive linguists argue (Langacker 87, Lakoff 87), such an approach is a reducing one. Natural languages are the products of a long evolution. They are adapted to the special needs of human beings, not only for communication purposes but also for the expression of personal feelings or intimate deliberations. Thanks to the innumerable ways of saying the same meaning, each speaker is free to make *choices*. The systematic analysis of personal choices, leading to the study of personal invariance is the center of this paper.

In semantics, the objective meaning is preferred to the personal way to express it. Theories of sense refer meanings to *states of the world*. The approach advocated here focuses on the contrary on the structure of enunciation and refers it to *states of the mind* of the speaker. The paradoxical goal reached is the definition of a universal and formal symbolic language able to describe individual regularities in the production of natural language sentences and leading to a precise definition of the notion of style.

A particularly interesting application of the model proposed is the study of the verbalization of cognitive tasks. In these texts, the plan underlying the behavior is inferred from linguistic structures (Hamburger & Crain 87). So, from the analysis of texts, the representation built allows to reconstitute their conditions of productions. The task chosen for this purpose is the *decision making*.

In cognitive psychology, every task performed by the mind is analyzed in sub-tasks exchanging intermediary representations. For example, the process of understanding and producing propositions in natural language can be represented (in an oversimplified version) by the diagram of fig. 1. In this figure, dynamic processes are included in rectangles while static data appear in ovals. The arrows symbolize the flow of information. The purpose of computer scientists is generally to *simulate* the cognitive process studied by adapted software (Johnson-Laird 89). The flow of information circulating in this software follows the one postulated in the psychological model.

Our approach is radically different, as we go *against* the tide of information : from several productions by the same individual (this is why two « new sentences » ovals are represented in fig. 1), we want to infer (or abduct) general characteristics of the synthesis process of this individual. The corresponding « rectangle » in bold is the place where individual choices are performed. It is postulated that some of the properties of this process are independent of other internal representations : the individual invariance should be recognized whatever is the semantic content of the sentences produced. So, the *surface structure* of language will be preferred to its deep structure. This is the main difference between the work presented here and plan inferences (Cohen & Perrault 79, Allen & Perrault 80, Pollack 90) where part of the synthesis process is also reconstituted but only for understanding purposes.



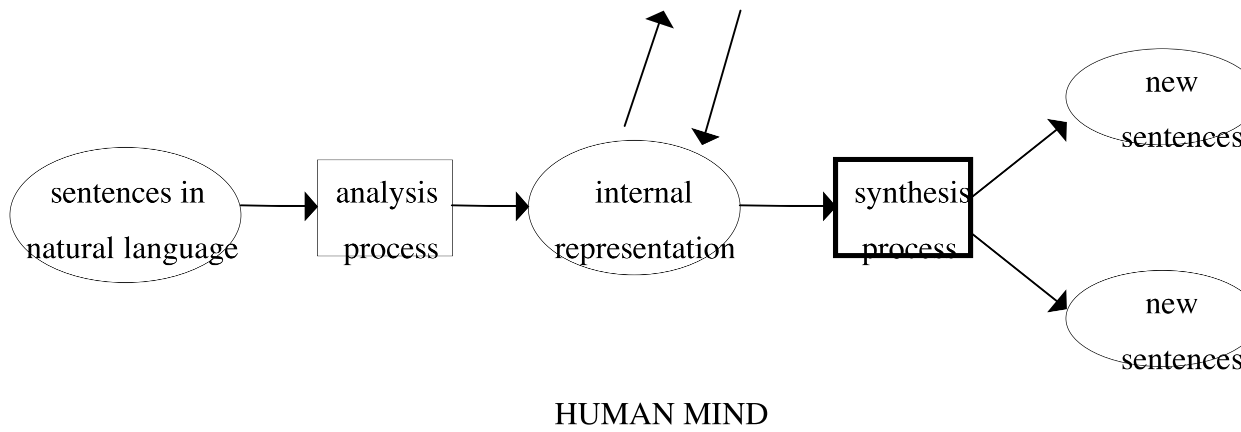


Fig. 1. Simplified schema of the process of understanding/producing natural language.

Like in (Grosz and Sidner 86) and (Daniel & al. 92), two levels of analysis, corresponding to two levels of regularities, of style, are to be distinguished. Linguistic patterns, first, are independent of the kind of text considered. They are defined at the propositional (or enunciative) level. At this level, the model describes how intentions are expressed. Cognitive patterns, on the other hand, are dependent on the cognitive task verbalized and are analyzed at the discourse level, by a sequence of attentional states represented by an automaton.

The structure of automata has been defined by theoretical computer scientists. It has not often been used to represent texts, except in (Lehnert 81) and (Finkel 92). We think that it describes well a dynamic process, whereas usual discourse representations (logical languages, conceptual graphs, Kamp & Reyle 93's DRS...) are static. We propose then to characterize the cognitive style of an individual (relatively to a task) by a *cognitive automaton*. Our model is computable (automata will be built nearly automatically from texts) and can also be considered as a first formal approach to the *performance* of a speaker.

Decision making is an interesting task because it is both very personal and can be applied to a vast range of domains. It is a perfect example of « mental act ». Moreover, it is expected that the cognitive automaton built is closely related to the decision making process itself and not only to its verbalization.

This paper presents a formal model. It is illustrated by examples extracted from a real corpus.

2. FROM NATURAL LANGUAGE TO ENUNCIATIVE PATTERNS

Lexical and syntactical choices apply at the enunciative level. To analyze these choices, *a structured language of features* is defined. This language allows us to give account of each choice independently. The enunciative level includes two sub-levels :

- the elementary item of choice, corresponding to a unique feature, is the *substitutable morpheme*. Let us recall that the morpheme is the smallest signifying unit in linguistics. It will be said substitutable if it can be replaced by any other member of its lexical class without changing the grammaticality of the proposition it belongs to. It is obvious that every lexical morpheme (common noun, verb, adjective...) is substitutable. Some grammatical morphemes are also : it is the case, for example, for the grammatical tense and mood of verbs. Others (prepositions, conjunctions...) are definitely not. The use of a substitutable morpheme reveals, by definition, a *lexical choice*.
- the elementary structure of choice, corresponding to the basic structure of our language of features, is the *syntagm*, which mainly appears under two forms : the predicative

syntagm (or verb phrase) and the noun phrase. The organization of the syntagms appearing in a proposition characterizes the *syntactical choices*.

2.1 Analysis of the lexical choices

Let us start with verb phrases (in the restricted sense of « predicative syntagms » without complements). What are the relevant features (i.e. substitutable morphemes) it includes ?

First of all, morphological features are to be considered. As a matter of fact, the tense, mood and voice of conjugation of a verb reflect the way the speaker is related to what he (she) says. In the context of decision making, it is expected that some individuals will refer to past experiences while others will try to imagine the possible consequences of their decision in the future. This should be readable in the conjugation chosen for the narration.

The choice of a verb is of course also important. But, in our approach, semantic features are only taken into account as far as they reveal how the resources of the individual are used to take a particular decision. So, relative importance allowed to sensorial perception, abstract reflection, affective domain or action performance is taken up. States verbs (those eventually followed by attribute adjectives) are considered apart as well as those expressing an attitude (« believe », « expect »...) or a modality (« must », « can »...) and necessarily followed by a verbal or a propositional complement.

To describe the structure of the features appearing in a verb phrase, the formalism of (non commutative) « and »/ (exclusive) « or » trees will be exploited. Fig. 2 shows the tree corresponding to predicative syntagms.

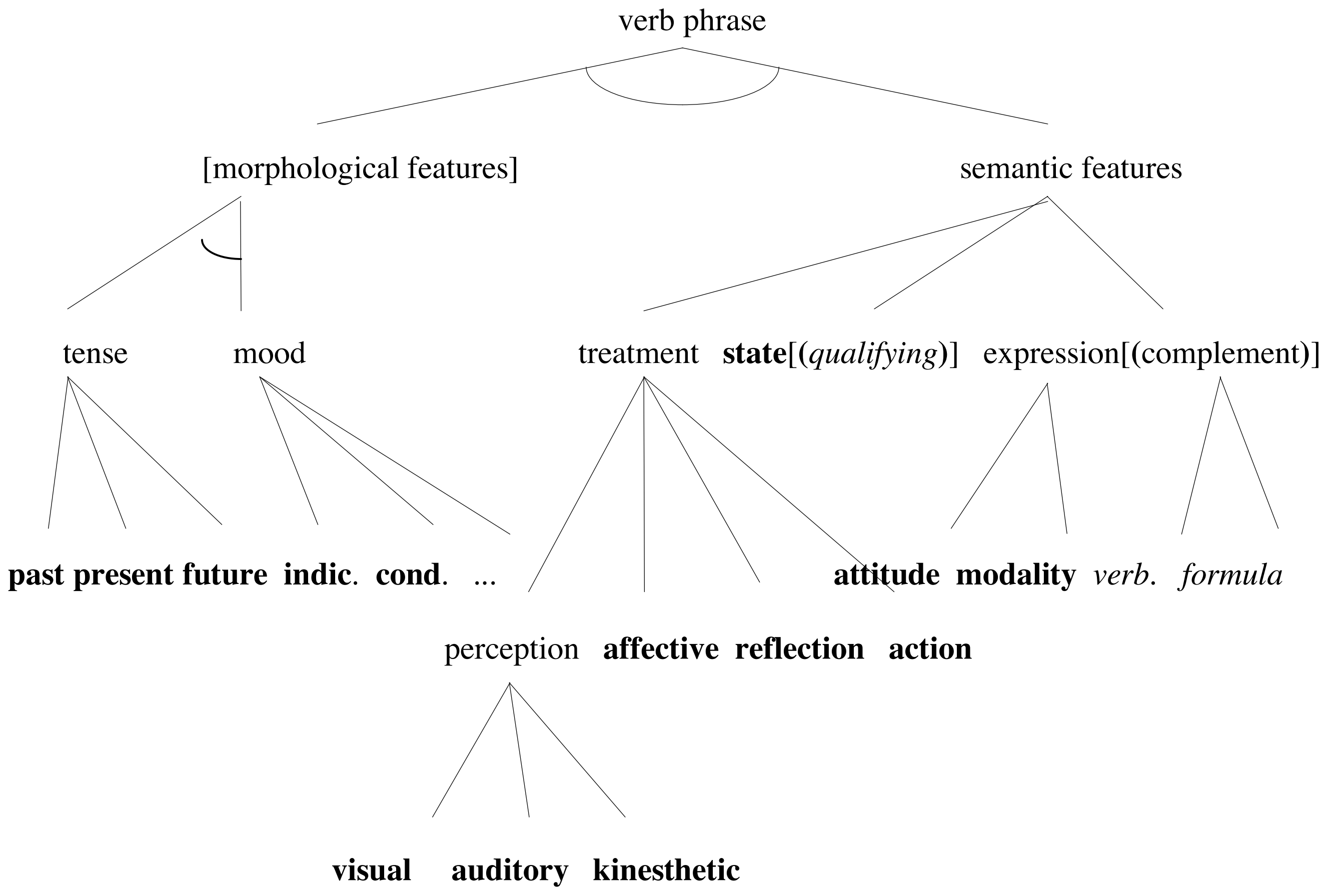


Fig. 2. And/or tree of the substituable morphemes taken into account in verb phrases.

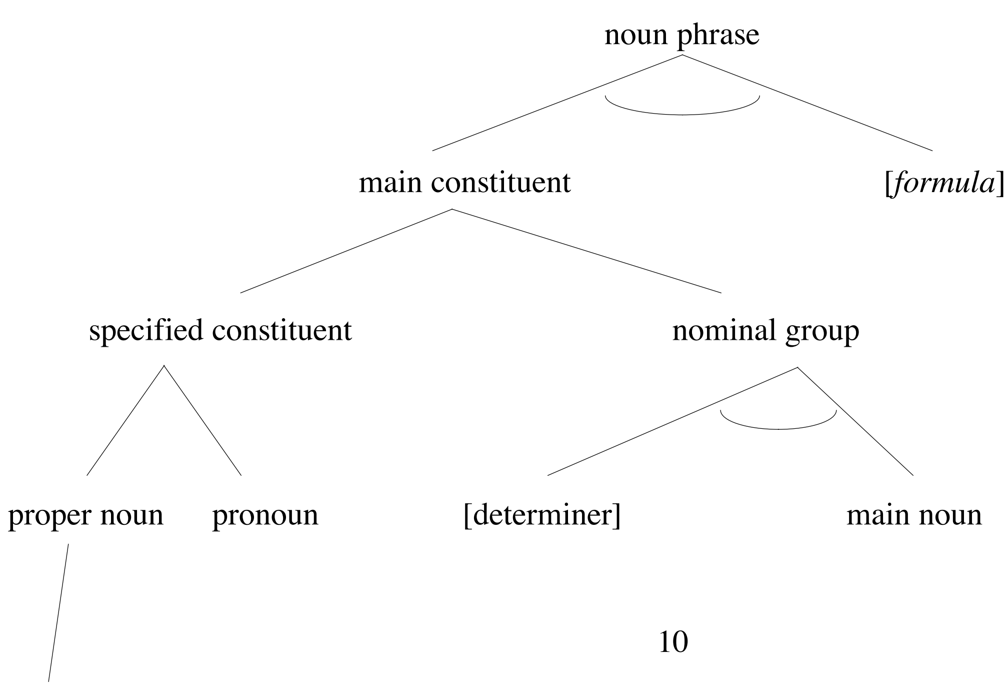
In this figure, the conventions of notation have the following meaning :

- « and » branches are linked together by an arc : they provide a decomposition of the father node into an ordered set of daughter features ; others are « or » branches : they enumerate in the daughter nodes the possible mutually exclusive values of the father node ;

- nodes written into square brackets are optional ones : for example, the morphological features of verbs in the infinitive form are omitted ;
- nodes in usual typography define *attribute features* ;
- nodes in bold are the *attribute value features* which are the possible final values of *attribute features* ([Le Ny 89]) : they only appear as leafs of the tree ;
- nodes in italics refer to other nodes in the same tree (for example, *verb.* stands for « verb phrase », so that the global structure is a recursive one) or in another tree to be defined later (this is the case for *qualifying* and *formula*, respectively corresponding to adjectives and to propositions).

It is also possible to write in a linear expression, a structured list, the set of features extracted from a verb phrase. It will be of the form : [[tense, mood], semantic feature]. So the syntagm « would like » is represented by : [[present,conditional],affective], « to see » by : [visual] and « will be able to see » by : [[future,indicative],modality([visual])].

The structure of relevant features extracted from nominal syntagms (or noun phrases) is given in the fig. 3.



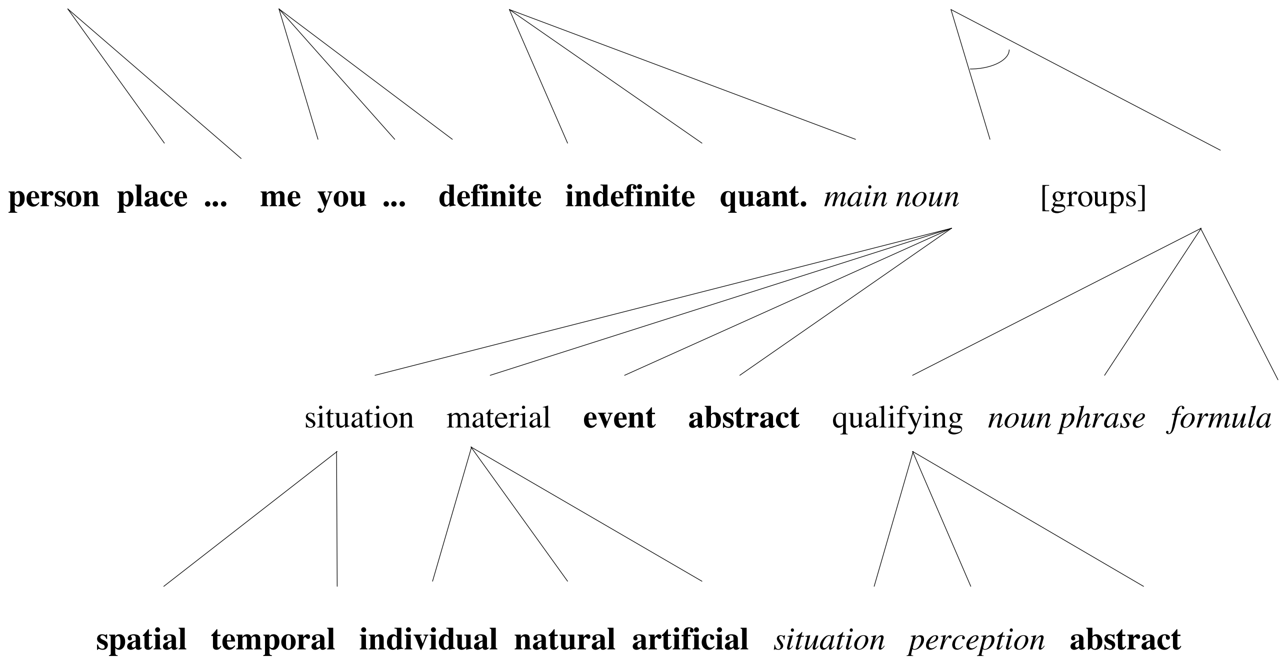


Fig. 3. And/or tree of the substituable morphemes taken into account in noun phrases.

This figure is written following the same conventional notations as the previous one. We will not develop here the choices made to built it.

It allows to represent « a black cat » by : [indefinite,[natural,visual]]. In this expression, the first feature comes from the « indefinite » determiner « a », « natural » is the final value associated with the main noun « cat », which is qualified by the « perceptive » qualifying adjective « black » (the feature « perception » is decomposed in fig 2.). The noun phrase « Paris that I like » is associated with [place,*formula*]. As a matter of fact, « Paris » is a « specified constituent » whose final value is « place » and « formula » (to be decomposed in the next section) comes from the relative proposition « that I like ».

Now that predicative and nominal syntagms are independently analyzed, the way they are combined in a proposition can be studied.

2.2. Analysis of the syntactical choices

The distinction between predicative and nominal syntagms is much more general than the case structure. It comes from a *functional* point of view that can be applied to every natural language (Shaumyan 87). In English, verb phrases are the functions and noun phrases are the arguments. The fundamental structure of propositions can thus be written :

[verb phrase]$_n$ (noun phrase)$_1$... (noun phrase)$_n$

where [verb phrase]$_n$ is a function of arity n (i.e. expecting n arguments) and (noun phrase)$_i$, for $1 \leq i \leq n$, is its i-th argument. This relation defines precisely what a *formula* is. Let us represent it in a new and/or tree (fig. 4).

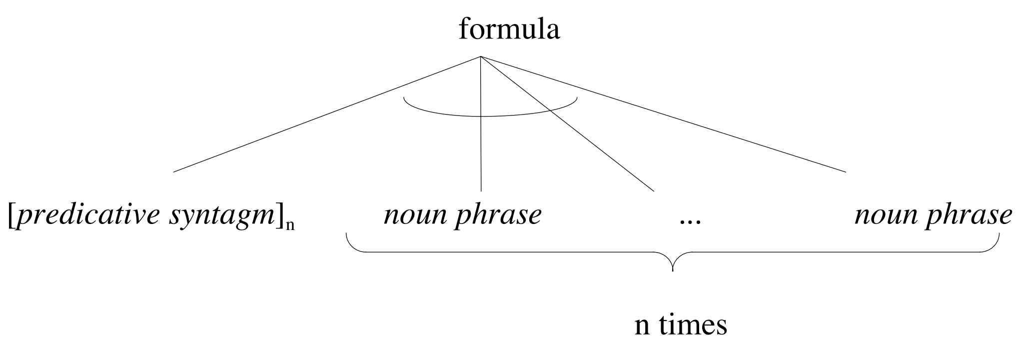


Fig. 4. And/or tree giving the structure of propositions.

The arity of a verb phrase depends of its transitive or intransitive nature. This tree is a condensed version of the « real » tree containing as many « or » branches as possible values admitted for n (in the following we can restrict ourselves to $n \leq 3$).

By a convention inherited from logical formalisms, the last argument of a verb phrase of arity n (i.e. the n-th noun phrase) always represents its grammatical subject, whereas the others (index from 1 to n-1) are its direct or indirect object complements, in their order of appearance in the proposition.

For example, the proposition « The boy looks at a car » is associated with the formula : $[[present,indicative],visual]_2([indefinite,artificial])_1([definite,individual])_2$. The complex noun phrase « Paris that I like », left partially undefined in the previous section, will finally be represented by : $[place(x),[[present,indicative],affective]_2(x)_1(me)_2]$. Here the « marker » x is introduced to identify the place « Paris », main constituent of the noun phrase, with the direct object of the predicate « like » (and its first argument) of the relative proposition which qualifies it.

Several extensions are still necessary to represent every proposition in natural language. We give here how to treat adverbial complements.

The number and the nature of every possible such complements are controversial, but it is possible, in first approximation, to consider an elementary list including time, location, mean, manner, purpose... An adverbial complement can be expressed by an adverb (« quickly », « tomorrow »), by a noun phrase eventually preceded by a preposition (« with a knife »), by an infinitive predicate (« to test ») or by a complete subordinated proposition (« when he entered the room »). The global structure of a formula should then include all these possibilities. A and/or tree can again express this situation (fig. 5).

Exceptionally, the highest « and » branch between « adverbial complement » and « formula » is a commutative one (the complement can appear either *before* of *after* the main proposition). Notice that the recursivity of this structure allows to combine several adverbial complements in a same proposition.

The trees of fig. 2 to 5 define a complete formal language that will be called the *enunciative language* L. L can also be characterized as the language generated by the context-free grammar whose non terminal vocabulary is the finite set of attribute features, terminal vocabulary is the finite set of attribute value features, axiom is the feature « formula » and set of rules include :

- all the rules of the form A ——> [B_1 ... B_k] where A is a father node of a « and » branch and B_1, ..., B_k are its daughter ;
- all the rules of the form A ——> B_1, ..., A ——> B_k for corresponding « or » branches.

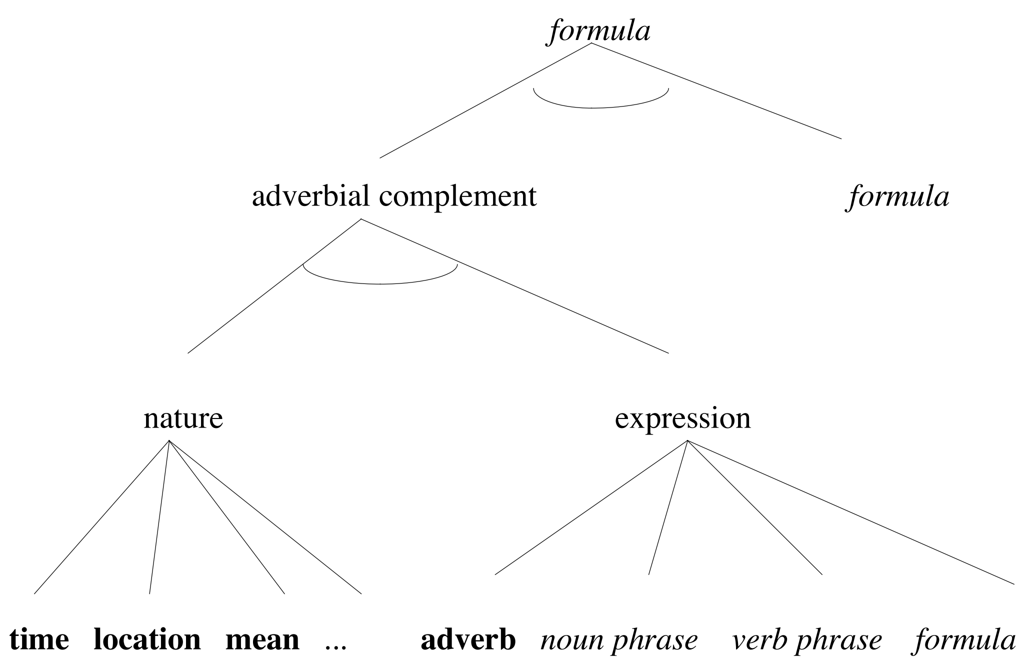


Fig. 5. And/or tree giving the structure of propositions with adverbial complements.

Some more rules are also needed to introduce the « markers », but won't be developed here. Every proposition in natural language can now be translated into a word of L.

Examples :

« I want to steel the car that I like. » is associated with the following word of L :

$[[\text{pres.,ind.}],\text{attitude}([\text{action}])]_2([\text{definite,artificial}(x),[\text{pres.,ind.}],\text{affective}]_2(x)_1(\text{me})_2])_1(\text{me})_2$

The proposition « When I saw it, I wanted it. » is translated by :

$[[\text{time},[[\text{past,indicative}],\text{visual}]_2(\text{it})_1(\text{me})_2], [[\text{past,indicative}],\text{attitude}]_2(\text{it})_1(\text{me})_2]$.

2.3. The building of enunciative patterns

What are the properties of L and what does it *represent* ? How can it be compared to semantic languages ?

- L is a *formal symbolic* language. It can be automatically generated from an analysis of propositions in natural language. Such a translation has been implemented. The analysis is performed by a Categorial Grammar (Oehrle, Bach & Wheeler 88) based on Lambek calculus (Lambek 58, Moortgat 88) and the translation process is inspired by Montague's works (Dowty, Peters & Walls 81). In this process, substitutable morphemes are associated with single features and the others are associated with λ-expressions and include « markers » (they provide the rules of combination between features thanks to which the structure of the formula is built).
- It is a *positional* language : the place where a feature appears in a list associated with a syntagm, and the place where this list appears in the global formula associated with the proposition expresses the functional and syntactical role played by the morpheme from which the feature is extracted. The lexical choices are symbolized by the value of features and the syntactical choices are coded by their position in a formula.

- It is a *recursive* language : it allows to combine a finite set of primitives (the attribute value features) into infinite possible ways. The expressivity power of L is the one of context-free languages.

In usual semantics, where the purpose is to characterize the *sense* of propositions, equivalence classes of representation regroup the translations of paraphrastic sentences. These sentences correspond to *different ways of saying the same thing*. As we are interested in the individual invariance, our equivalence classes should regroup the *same way of saying different things*. For example : « Mary is looking at a playing cat. » and « A cat that Mary is looking at is playing. » are paraphrastic because they can both be translated into the logical formula : $\exists x[cat(x) \wedge look(x)(M) \wedge play(x)]$. But the words of L associated with each of them are different. They are respectively :

$[[present,indicative],visual]_2([indefinite,natural(x),[[present,indicative],action]_1(x)_1])_1(person)_2$

$[[present,indicative],action]_1([indefinite,natural(x),[[present,indicative],visual]_2(x)_1(person)_2])_1$

The structural differences between these formulas reflect the enunciative differences between the two original sentences. Another example would be given by propositions in the passive voice. They are usually paraphrastic with the corresponding proposition in the active voice but they would be translated by different words of L because the same morphemes have different syntactical functions in the two propositions (to represent passive propositions in L, a « voice » feature has to be added in « verb phrase»).

Conversely, propositions like « I can see a cat. » and « I am looking at a dog. », non paraphrastic (!), are with our definitions translated by the same word of L :

$[[present,indicative],visual]_2([indefinite,natural])_1(me)_2$

as they express very similar relationships between a narrator (« I ») and how he (she) perceives (« can see », « am looking at ») the outside world represented by a natural entity (« a cat » or « a dog »).

L is a propositional language but, in our approach, the notion of proposition is not the same as in logic. It is closer to the use of cognitive psychologists, for whom truth values and quantifiers are not relevant. Psycholinguists studying language production have come to suppose that, between the semantic level and the phonological level of the production process, there is a « functional level » where the lexical choices are represented in an abstract form (which includes semantical and grammatical characteristics) and where the functions of these words are decided (Garrett 80, Levelt 89, Caron 89). L is a formal representation of this level.

In a simplified way, logic-like formalisms represent the *sense* of propositions and words of L represent their structure, their *form*. The intuitive notion of style is precisely a link between what is said and how it is said. So, the couple (semantic representation, enunciative representation) would be a first interesting formal definition of this notion. But, in this case, there would be as many couples as there are propositions considered. The style seems to be more general.

Let us define first an extension of L called L'. In the new language L', some features can be only partially specified : they can take several possible values enumerated with an OR operator, or even all possible values acceptable in their position : in this case, the feature is replaced by the symbol « _ », like in the Prolog language. For example, the expression :

w=[[present OR past, indicative],action]$_1$(_)$_1$ is a word of L'.

L' is a *language of specification.* Each word of L' describes the lexical and syntactical common points between several formulas of L, it defines a sub-set of L. For a given word w of

L', and for any word v of L, we will say that v *satisfies the specification* w if it is compatible or unifiable (in the Prolog sense) with w.

For example the formula $[[\text{present,indicative}],\text{action}]_1(\text{me})_1$ satisfies the specification $w=[[\text{present OR past, indicative}],\text{action}]_1(\_)_1$ but $[[\text{past,indicative}],\text{affective}]_1(\text{you})_1$ does not, as the feature « action » is not compatible with the feature « affective ».

Now, let us have a more general view on meaning. The level of *intention* has been shown to be relevant for linguistic studies (Searle 83, Grosz & Sidner 86). It seems to be very difficult to identify an intention on linguistic parameters in a general way. Nevertheless, our hypothesis is that every individual tends to privilege lexical and syntactical structures in the expression of his (her) own intentions.

Definition 1

We call an *enunciative pattern* every couple (I, w) where I is an intention an w is a word of the language L'.

Hypothesis 1

For every individual P and every intention I there exists a (non empty) word w of L' so that every expression of the intention I by P is performed through a proposition associated with a formula of L that satisfies the specification w. The corresponding enunciative pattern is then said to be relevant for P.

2.4. An example

Let us give here the enunciative representation of the verbalization (*a posteriori*) of a real decision making (translated from French). The narrator of text 1 is on a tourist trip in London and is looking for souvenirs.

text 1

> (*I am looking at a board with pin's*)$_1$. *Although* (*I don't intend to buy any*)$_2$*,* (*the « bobby » and the « telephone cabin » attract my look*)$_3$. (*The cabin is definitely preferred*)$_4$. (*It evokes well my idea of London*)$_5$*, Soon after,* (*the image of my cousin's board of pin's appears in my mind*)$_6$. (*I think that* (*this pin's would be a good present for him*)$_8$)$_7$. *So,* (*I buy it*)$_9$.

The propositions are put in brackets and are given a number. The translation of this text in our L language gives the following sequence of formulas :

1. [[present,indicative],visual]([indef,[artificial,artificial]])(me).

   [[opposition,

2. NOT[[present,indicative],attitude([action](it)(me))](me)],

3. [[present,indicative],action]([def,visual])([def,artificial] & [def,artificial])].

4. [[mean,adverb],[[present,indicative],state([affective])([def,artificial]).

5. [[mean,adverb],

   [[present,indicative],visual]([def,abstract(x),[[present,indicative],reflect.](x)(place)(me)])(it).

   [[time,adverb],

6. [[present,indicative],action]([def,abstract])([def,[visual,[art.,[art.,[def,individual]]]]])].

7. [[[present indicative],reflection(*proposition 8*)](me)

8. [[[present,conditional],state]([indef,[art,[abstr,individual]]])([def,art]]).

9. [consequence,[[present,indicative],action](it)(me)]].

How can we use such a representation ? First, we can look for the *enunciative constants*, i.e. the words of L' that are satisfied by every formula (this search is obviously computable and can be done automatically). Let us define P=present. The most general constant of our example is obviously : [[P,_],_](_)...(_).

Then, we can try to define words of L' corresponding to global classes of propositions. In text 1, it is possible to distinguish two main classes. The first class regroups propositions whose grammatical subject is « I » and translated by formulas satisfying the specification : [_](_)...(M), with M=me. The second class includes the propositions whose grammatical subject is an artificial thing (« the telephone cabin ») or the image or the idea of this thing (« the image of my cousin's board »). These subjects are translated into noun phrases of the form : A=([def,artificial OR [visual OR abstract,_]] OR it). Adverbial complements are not taken into account here.

Now, let us define some enunciative patterns corresponding to the propositions of this text :

- 1 : (*observation*,[[P,indicative],visual](_)(M))
- 2 : (*intention*,[[P,indicative],attitude(*choice*)](M) OR [[P,indicative],attitude](_)(M))
- 3 : (*attraction*,[[P,indicative],action]([def,visual])(A))
- 4 : (*preference*,[[P,indicative],state([affective])](A))
- 5 : (*evocation*,[[P,indicative],visual]([_,abstract OR visual(x),_](A))
- 6 : (*solicitation*,[[P,indicative],action]([def,abstract])(A))
- 7 : (*judgment*,[[P,indicative],reflection(*possible-fact*)](M))

- 8 : (*possible-fact*,[[P,_],state]([indef,[art,_]])(A) OR [[P,_],state(_)](impersonal))

- 9 : (*choice*,[[P,indicative],action](it)(M))

The first element of each couple, in italics, is an intention expressed in a proposition. The second one is a word of L' satisfied by the formula translated the corresponding proposition. These patterns are built so as to be considered as relevant after the study of other texts (see below). Note that the name of an intention can be used as an abbreviation inside the specification formula (like in items 2 and 7). This allows a new level of recursivity.

3. FROM DISCOURSES TO COGNITIVE PATTERNS

At the enunciative level, a text is translated, proposition after proposition, in formulas of L. But this translation does not give account of the global organization of this text. If this text is the verbalization of a task, the sequence of the processes performed during the execution will only appear at the discourse level. The notion of enunciative pattern just proposed precisely allows us to define subsets among the set of propositions constituting the text. It is the first step towards a reconstruction of the « cognitive structure » of this text on formal criteria.

3.1. Representation of the example

Now that intentions have been associated to words of L', text 1 can be described as a path between such intentions. As we have already noticed, the corresponding words belong to two main classes, depending on their grammatical subject. So, intentions are also included in two main « boxes ».

Text 1 from 2.4. is then represented by fig. 6.

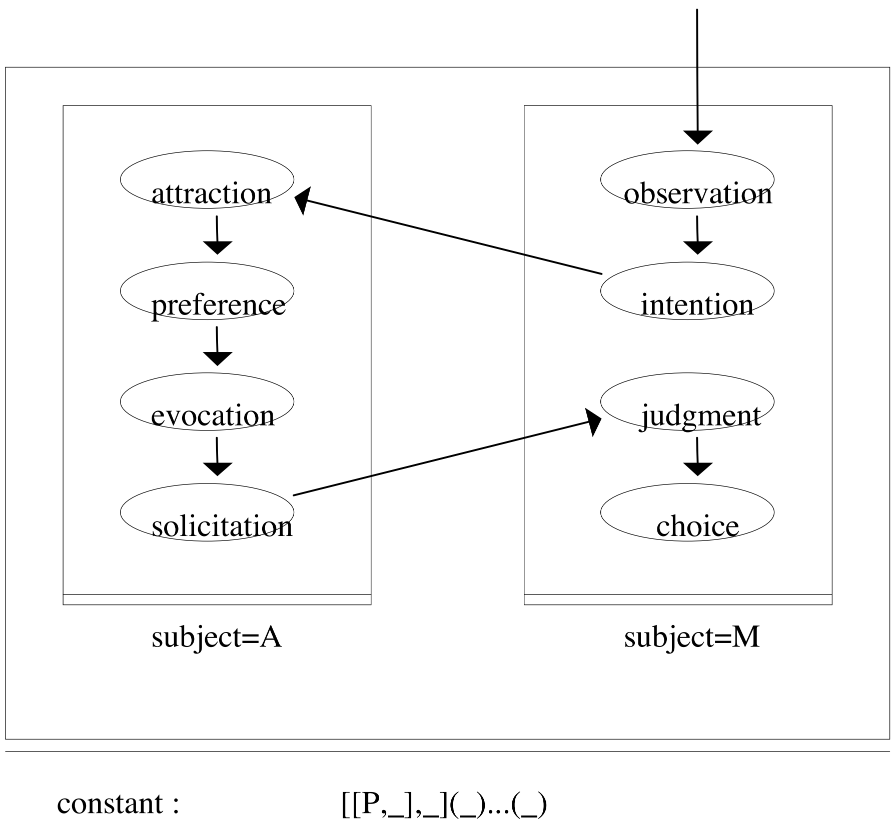


Fig. 6. Representation of text 1.

In this figure, the boxes represent the « scope » of the constants displayed at their bottom. The ovals include the name of the intentions expressed. The arrows follow the sequence of the propositions in the text. But this first analysis still needs to be confirmed by others.

3.2. Common patterns between two texts

Let us study now another verbalization by the same narrator of another decision making, whose subject is completely different.

text 2

> *In the morning,* (*I hesitate between two pullovers*)$_1$. (*The idea of the second one evokes an image in which I am wearing it*)$_2$*, In this image,* (*I look at the bottom of the pullover*)$_3$ *and* (*I find it inelegant and too long*)$_4$. (*The first one sets me the problem of taking also a sash*)$_5$*, Finally,* (*I put it on*)$_6$ *and* (*open the window*)$_7$. (*I think that* (*it is cool enough*)$_9$)$_8$ *and* (*I keep it*)$_{10}$.

The translation of text 2 in our L language provides the following sequence of formulas :

1. [[time,[def,[temporal]],[[present,indicative],attitude]([def,artificial])(me)].

2. [[present,indicative],visual]

 ([indef,visual(x),[[location,x],[[present,indicative],action](it)(me)])([def,[abstract,[def,art]]]).

 [[location,[def,visual]],

3. [[present,indicative],visual]([def,[spatial,[def,artificial]]]])(me)]

4. & [[present,indicative],reflection(abstract & visual)](it)(me)].

5. [[present,indicative],reflection]([def,abstract([action]([def,art]))([def,art]).

 [[time,adverb],

6. [[present,indicative],action](it)(me)

7. & [[present,indicative],action]([def,art])(me)]

8. [[present,indicative],reflection(*proposition 9*)](me)

9. [[mean,adverb],[[present,indicative],state(kinesthetic)](impersonal)]

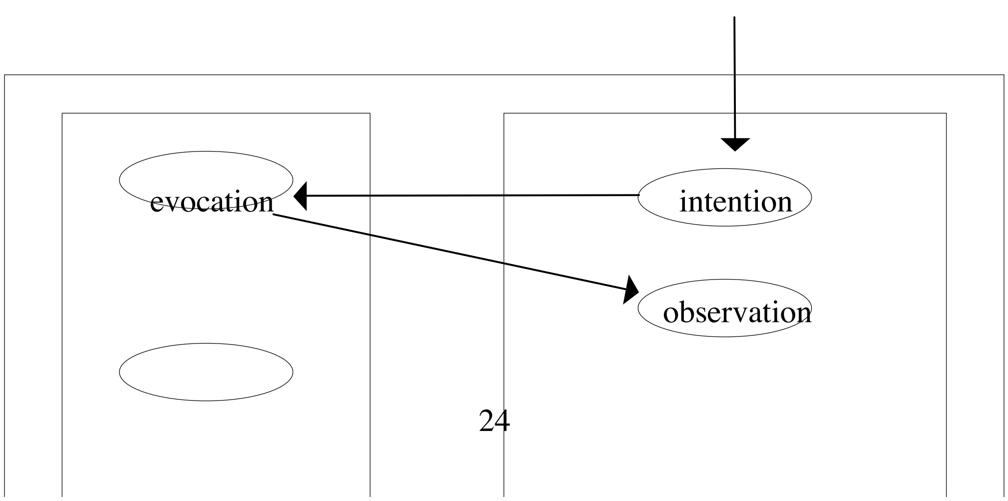
10. & [[present,indicative],action](it)(me).

Here, the constant P'=[present,indicative] applies and it is striking to notice that, like in text 1, propositions are equally devised into the class of those whose grammatical subject is M=me and those for which it is of the form A=([def,artificial OR [visual OR abstract,_]] OR it). Most of the enunciative patterns previously defined are relevant, as they also apply here. Only two new items need to be introduced :

- 5 : (*problem*,[P',reflection]([def,abstract(_)])(A))
- 6 and 7 : (*test*,[P',action](it OR [def,artificial])(M) AND *test*)

These definitions allows us to build for text 2 the representation of fig. 7.

It can be noticed that, in the second text, the narrator hesitates between two options (two « pullovers »). His decision is thus composed of two sub-decisions : the first option is considered and abandoned, then the second option is considered and admitted. Each of these sub-decisions has the same structure displayed in fig. 7 by a « come and go » path between the two main boxes. As this structure is also the one that emerges from fig. 6, it is probably characteristic of the decision process of the narrator.



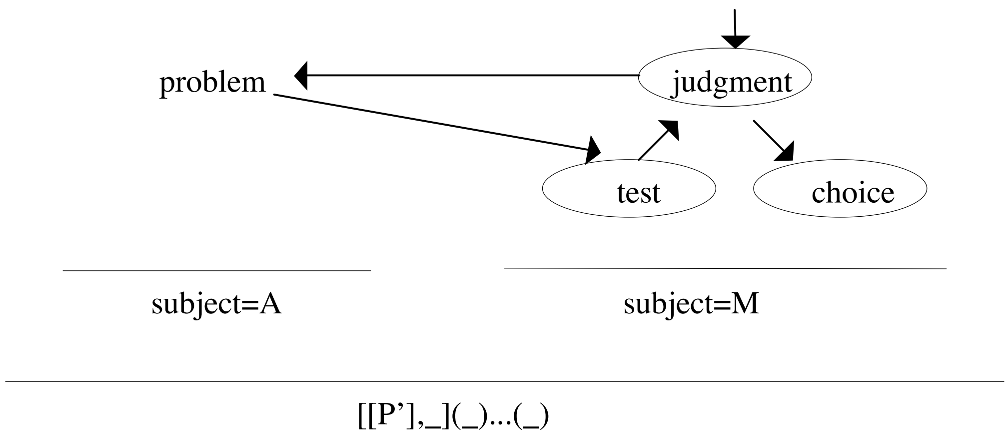


Fig. 7. Representation of text 2.

To give a unified representation of this process, let us define global enunciative patterns associated with global intentions defined as followed :

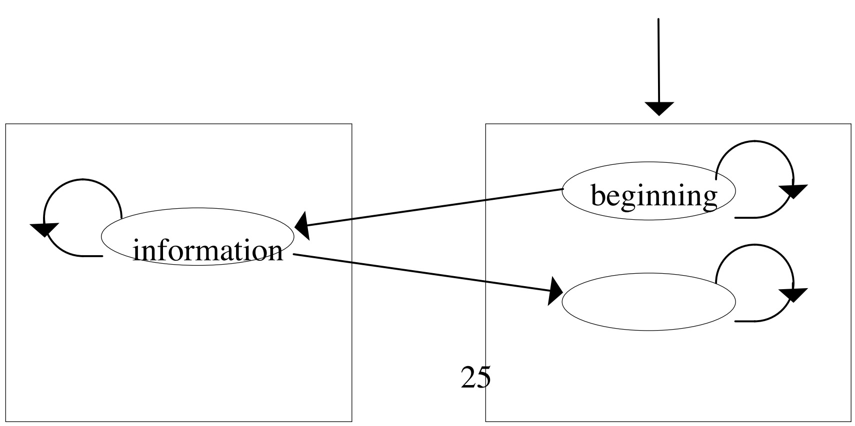
information = attraction OR preference OR evocation OR solicitation OR problem

beginning = observation OR intention OR judgment

reflection = judgment OR observation OR test

The intention « choice » can be left alone for the moment. The corresponding words of L' are of course defined as the disjunction (with the OR operator) of the words associated with each elementary intention. The « information » set (the intention meant is in fact « taking of information ») is also more simply characterized by the constant : subject=A.

The decision process of the author of both texts can now be expressed by the structure represented in fig. 8, called a « domain specific graph ».



reflection

subject=A subject=M

Fig. 8. Representation of the common points between text 1 and text 2.

The loop appearing on each global intention state means that this intention can be expressed by a sequence of several propositions. This structure is instanciated once in the first text and twice in the second text. We expect to find it in every verbalization of decision making by the same author.

Definition 2

We call a *cognitive pattern* every couple (T, G) where T is a task and G is a domain-specific graph (in our example, T=« decision making » and G is given in fig. 8).

Hypothesis 2

For every individual P and every task T there exists a (non empty) structure S so that every instance of verbalization of the task T by P is a discourse whose representation in our formalism includes an instance of S. The corresponding cognitive pattern is then said to be relevant for P.

3.3. The building of cognitive automata

The notion of cognitive pattern allows us to make precise expectations on future verbalizations by the same narrator. But it describes a *local* invariant and does not really give account of the global common structure of both texts. For example, it can be argued that, as the « beginning » and the « reflection » states both include the « judgment » and « observation » intentions, they can be included into a still more general intention that can be called « decision ». Furthermore, both texts end with an instance of the « choice » intention that does not appear in the cognitive pattern defined.

To give account of these remarks, fig 6 and fig. 7 can be considered as *refinements* of the more general cognitive automaton of fig. 9.

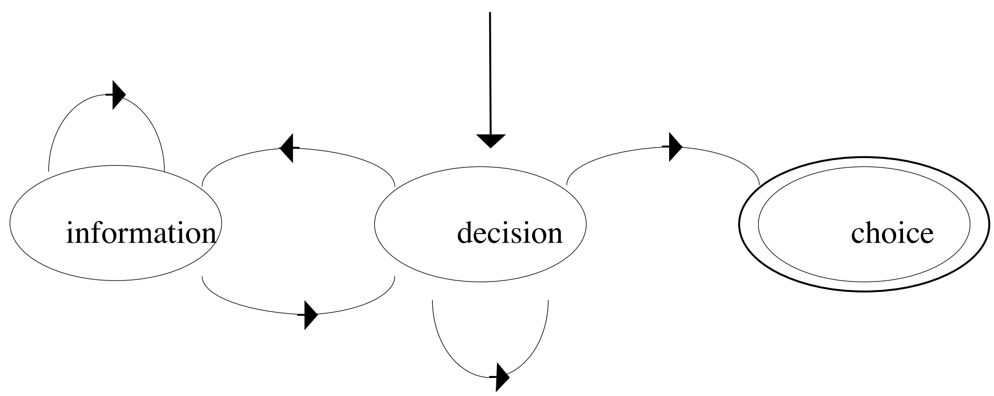


Fig. 9. Cognitive Automaton representing the decision process of the narrator.

where :

information = attraction OR preference OR evocation OR solicitation OR problem

decision = observation OR intention OR judgment OR test

choice = choice

Let us recall that the model of automata comes from theoretical computer science. An automaton is constituted of a finite set of states (among which there are one or several initial states and one or several final states) and a finite set of labeled transitions between states.

In our representation, the states are defined by enunciative patterns and the transitions are simple arcs (the alphabet of possible transition labels is reduced to one element). In fig. 8.,

« decision » is the initial state (it is pointed by he initial arrow) and « choice » is the final one (this is why it is included in a double oval).

A cognitive automaton, as it is previously defined, includes a lot of information. It is much more constrained as it seems, because each of its states is associated with a word of L' which is a specification constraint. A new text will be recognized (or accepted) by this automaton if its translation into the L language produces formulas that satisfies the specifications of the states and whose transitions are also admitted.

Automata are dynamic representations, as opposed to usual semantic representations that are mostly static. Although the one of fig.8 has been built after an analysis of texts, it is also a good model for the language synthesis process. Associated with a context, it can be used to write texts « the way the narrator would do ». But it maybe even more than this. Text 1 and text 2 are both characterized by a particular path in it. A decision making is also necessarily a path between an « indecisive » mental state and a « decisive » one. If the corresponding verbalization is faithful and sincere, it is to be expected that the states of our cognitive automaton are plausible representation of these mental states. As a matter of fact, words of L' represent filters among the linguistic features taken into account by the narrator. They may also represent filters among the features of the world that attracted the *attention* of the narrator.

4. EVALUATION AND CONCLUSION

Our purpose was to give a formal way of representing individual invariance. Two levels of regularities have been introduced and each of them is characterized by symbolic patterns. The cognitive automaton finally built provides a sum-up of all extracted information and can be considered as a task-oriented representation of the cognitive style.

Our model makes use of various tools from linguistics (notion of morpheme, of syntagms, feature analysis, functional structure of propositions) and computer science (and/or trees, context-free grammars, unification mechanism, automata). It is mostly computable and has already been partially implemented. Human expertise is nevertheless still necessary for the identification of the intentions expressed in a text and for the definition of the words of L' to be associated with.

Our psychological hypotheses 1 and 2 are formally expressed. We think that the execution of some daily simple tasks such as decision making is partially automatically performed. Behavioral redundancy indicate that individual attitudes are underlain by mental structures. The structures we propose can be compared to schema (Bartlett 32) or scripts (Schank & Abelson 77) which would not be related to a situation (a diner in a restaurant) but to a mental attitude. This special kind of schema is at the same time universal and specific to everyone (Diguier 93). The data structure of automata seems to be a valuable representation for it.

As far as we know, the stylistic regularities had till now mainly been approached thanks to statistical calculation. But the figures obtained are usually limited to lexical regularities and explain nothing. Our model is much closer to cognitive studies. Our « enunciative level » is closely related to the « functional level » postulated by psycholinguists while the « cognitive level » can be compared to the planification step. We have already noticed that the automaton obtained is a plausible model for the production process. Whereas logical formalisms are *prescriptive*, it is a *descriptive* model : it describes the execution of a task *as it has really been* *performed* (in fact *as it is said to have been performed*). It is now admitted that in human decision making, affective and emotional parameters cannot be ignored (Damasio 95).

Of course, our hypothesis still needs to be confirmed by experiments. We have already collected and carefully analyzed a corpus (from which texts 1 and 2 were extracted) containing 10 texts written by 4 different individuals. Individual regularities and inter-individual variations have been proved to be very relevant for this corpus, both at the enunciative and cognitive levels (Finkel & Tellier 96, Tellier 96). More data are still needed. We want for example to test our model for the analysis of particular decision making situations like diagnosis, where both the choice of parameters (i.e. features) taken into account (Patel & Groen 1986) and the general reasoning process (Lemieux & Bordage 1992) seem important.

But whatever would be the modalities of these tests and even their results, we provide a systematic analysis process for verbalizations that can be applied to a vast range of domains. The status of our cognitive automata relatively to the texts they have been extracted from is original : from a surface analysis, they describe a form, a dynamic structure. The links between language and thought may be more complex than the one postulated here but we want to promote a forsaken approach to language. We hope to open a possible way towards formal cognitive linguistics. The first goal aimed here is the characterization of an individual by his (her) language. Further research in this direction seems promising.

BIBLIOGRAPHY

Allen, J., Perrault, C.R. (1980). "Analyzing Intention in Utterances", *Artificial Intelligence* 15, p143-178.

Bartlett, F.C. (1932). *Remembering*, Cambridge University Press.

Benveniste, E. (1964). *Problèmes de linguistique générale*, (2 tomes), Paris, Gallimard.

Caron, J. (1989). *Précis de psycholinguistique*, Paris, PUF, coll : "Le psychologue".

Chomsky, N. (1965). *Aspects of the theory of syntax*, MIT Press, Cambridge.

Cohen, P.R., Perrault, C.R. (1979). "Elements of a Plan-Based theory of Speech Acts", *Cognitive Science* 3, p177-212.

Damasio, A. (1995). *L'erreur de Descartes*, Paris, Odile Jacob.

Daniel, M.P., Nicaud, L., Prince, V., Pery-Woodley, M.P. (1992). "Apport du Style Linguistique à la modélisation Cognitive d'un élève", p 252-259 in the proceedings of Intelligent Tutoring Systems 92 (Montréal), LNCS 608, Goos & Hartmanis (Eds.), Springer-Verlag.

Diguier, L. (1993). *Shéma narratif et individualité*, Paris, PUF, coll : "Le psychologue".

Dowty, D., Wall, R.E., Peters, S. (1981). *Introduction to Montague semantics*, Reidel, Dordrecht.

Finkel, A. (1992). "A formalization of subjective experience", LIFAC internal report n°92-4.

Finkel, A., Tellier, I. (1996). From natural language to cognitive style, in *The Cognitive Level*, Endres-Niggemeyer (Ed.), Duisburg (to be published).

Fodor, J. (1975). *The Language of Thought*, Thomas Y. Crowell, New-York.

Gardner, H. (1985). *The Mind's New Science; A History of the Cognitive Revolution*, Basic Book.

Garret, M. F. (1980). "Levels of processing in sentence production", in Butterworth, B. (Ed.), *Language production : speech and talk*, vol. 1, New York Academic Press.

Ghiglione, R., Bonnet, C., Richard, J.F. (1990). *Traité de psychologie cognitive; Cognition, Représentation, Communication* (3 tomes), Paris, Dunod.

Grosz, B., & Sidner, C. (1986). "Attention, Intention and the Structure of Discourses", Computational Linguistics 12, n°3, p175-204.

Hamburger, H., Crain, S. (1987). "Plans and semantics in human processing of language", *Cognitive Science* 11, p101-136.

Jakobson, R. (1963). *Essais de linguistique générale,* Paris, Minuit.

Johnson-Laird, P.N. (1983). *Mental Models*, Cambridge, Massachusetts, Harvard University Press.

Johnson-Laird, P.N. (1989). *The Computer and the Mind; An introduction to Cognitive Science,* Harvard.

Kamp, H., Reyle, U. (1993). *From discourse to logic*, Kluwer Academic Press, Dordrecht.

Langacker, R.W. (1986). "An introduction to Cognitive Grammar", *Cognitive science* 10, p1-40.

Langacker, R.W. (1987). *Foundation of Cognitive Grammar* (2 volumes), Stanford University Press.

Lehnert, W.G. (1981). "Plot Units and Narrative Summarization", *Cognitive Science* 4, p293-331.

Le Ny, J-F. (1989). *Science cognitive et compréhension du langage*, Paris, PUF, coll : "Le psychologue".

Lemieux, M., Bordage, G. (1992). "Propositional versus structural semantic analyses of medical diagnosis thinking", *Cognitive Science* 16.2, p185-204

Levelt, W.J.M. (1989). *Speaking : From Intention to Articulation*, Cambridge, MIT Press.

Montague, R. (1974). *Formal Philosophy; Selected Papers of Richard Montague*, Yale University, New Haven.

Patel, V.L., Groen, G.J. (1986). "Knowledge Based Solution Strategies in Medical Reasoning", *Cognitive Science* 10.1, p 91-116.

Pollack, M.E. (1990). "Plans as complex mental attitudes", in : *Intention and Communication*, p77-103, MIT Press, Cambridge.

Sabah, G. (1988). *L'intelligence artificielle et le langage* (2 volumes), Paris, Hermès.

Shaumyan, S. K. (1987). "The goals of linguistic theory and applicative grammar", *Mathématiques et Sciences Humaines* 77, p7-43.

Schank, R.C., Abelson, R.P. (1977). *Scripts, Plans, Goals and Understanding*, Lawrence Erlbaum Associates Publisher, Hillsdale, New Jersey.

Searle, J. (1983). *Intentionally*, Cambridge University Press.

Tellier, I. (1996). *Définition et implémentation par les grammaires catégorielles d'un modèle cognitif formel de l'énonciation*, PhD thesis in Computer Science, ENS de Cachan.

Weil-Barais, A. (directed by, 1993). *L'homme cognitif*, Paris, PUF.

Fig.