Two Approaches to Commonsense Inferencing for Discourse Analysis.

Marc Dymetman
Université Scientifique et Médicale de Grenoble
Groupe d'Etudes pour la Traduction Automatique B.P. 68
38042 Saint Martin d'Hères FRANCE

Abstract:

The dominant philosophy regarding the formalization of Commonsense Inferencing in the physical domain consists in the exploitation of the "tarskian" scheme axiomatization <-> interpretation borrowed from mathematical logic. The commonsense postulates constitute the axiomatization, and the real world provides the "model" for this axiomatization.

The observation of the effective activity of linguistic communication and of the commonsense inferencing processes which are involved in it show the unacceptability of this scheme.

An alternative is proposed, where the notion of "conceptual category" plays a principal role, and where the principle of logical adequation of an axiomatization to a model is replaced by a notion of "projection" of a conceptual structure onto the observed reality.

1. Commonsense Inferencing in context: Commonsense and the linguistic process.

1.1 Disambiguation and Commonsense.

Let's consider the following text fragment:

"Mary dropped the plate on the table. It broke." (T. Winograd)

It is quite clear that the disambiguation of the pronominal lexical reference of "it" (table or plate) involves, in addition to the syntactic and semantic constraints (where "semantics" is taken in a narrow sense of "formal semantics") a certain understanding of the natural structure of the described events. This understanding, shared by all members of the linguistic community, we can name it "Commonsense".

In the case at hand, commonsense informs us that:

i) a plate dropped on a table will in all likelihood collide with the table.

ii) when such a collision happens, it is not unlikely for the plate to break, but quite abnormal for the table to break.

This shared Commonsense knowledge is taken into account at the pragmatical discourse level both by the speaker and the hearer, and allow the speaker to reconstruct, with the help of Gricean-type *conventions of communication*, the meaning intended by the speaker and therefore the correct pronominal reference: "it" refers to the plate.

1.2 Linguistic Acceptability and Commonsense.

The disambiguation problem just considered is closely linked to the problem of the linguistic acceptability of a text fragment. In the following example, the acceptability/unacceptability of the 2 texts is due to the possibility/impossibility on the part of the listener to reconstitute a direct causal link between the first element and the second element of each text:

"John's parachute had teared to pieces. He crashed on the ground."
*"John's parachute had got damped by the rain. He crashed on the ground."

1.3 Reconstitution of the information implicit in discourse and commonsense.

The previous examples showed already a relatively limited form of implicit information reconstitution in a piece of discourse: the text fragment was acceptable if it was possible for the listener to restore an implicit causal link in the text at hand. Here is now a more impressive example of implicit information reconstitution by the listener of a piece of discourse (fig. 1).

Here, the reconstitution of the informations implicitely contained in the original text makes use of at least 3 different domains of commonsense:

- commonsense in the physico/biological domain: to fall, to collide, disruptive consequences of collision, to be hurt, to die.
- commonsense in the psychological domain: a murder consists in an *intentional* killing, i.e. with a plan the goal of which is the death of somebody.
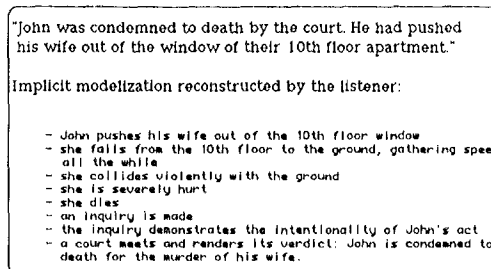- commonsense in the social domain: the way in which society treats crime.

Figure 1

This example may serve to remind us of the easiness and the swiftness with which we are able to reconstruct a complex set of informations implicit in a piece of discourse from our mastery of Commonsense in a variety of domains simultaneously. However, in the previously given examples, only the "physical" domain ("natural phenomena") was made use of, and in the remainder of this paper, we will essentially restrict ourselves to this domain of Commonsense.

2. The dominant model for Commonsense Inferencing in the physical domain.

A considerable interest has arosen recently in A.I. for Commonsense modelizations of natural phenomena. The articles of Pat Hayes <Hayes 1978> seem to have set the tone for the philosophy presently predominant in the domain, of which the collections of papers <Hobbs, Moore (eds) 1985> and <Hobbs et al. 1985> can provide recent examples.

The outlines of this attitude - which we call the Direct Interpretation approach (DI approach) - towards Commonsense are the following:

- The relation between theory and predictions is conceived as similar to what it is in "scientific" physics, but the natural *laws* which are encoded in the theory are different: essentially they are of a more "qualitative" and less "numerical" kind.
- The inferential procedure is made in the deductive mode (in the sense of "deductive logic", opposed to "inductive").
- Like "scientific" physics, "naive" physics should be able to predict - in a deductive manner - the greatest possible number of phenomena from the smallest possible number of general "natural laws".
- "Naive" physics is a strictly objectivist physics: it describes reality "as it is": to be valuable, the laws and the consequences of the laws must be "true".
- The description of the physical world is done through an axiomatization, typically in first-order logic, which respects the "tarskian" schema of semantical adequation of a logical theory to its model. In this schema, borrowed from mathematical logic (fig. 2.a), the real world is supposed to provide a model for the axiomatization, which is a set of axioms representing physical laws which are verified in the model (fig. 2.b):
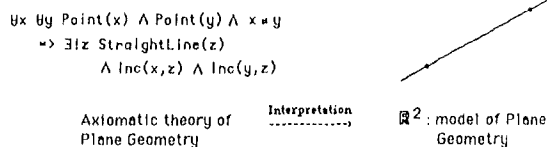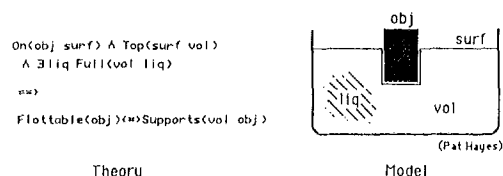

Figure 2a

Figure 2b

3. Problems with the dominant model.

3.1. Deduction or Induction?

3.1.1. The linguistic examples given in section 1 demonstrate how difficult it is to consider the commonsense inferences which appear there as a process of purely deductive type.

Thus, to take example 1.3, there is no logical necessity to the sequence of implicit events reconstructed by the listener: these informations are not *deducible* from the explicit informations "John was condemned by the court" and "John had pushed his wife out of the window": Maybe there was a safety net in place, and maybe John murdered a policeman 3 days later, and so on and so forth...

Actually, the analysis of the type of inferencing which are involved in the reconstitution of implicit informations in the examples of section 1 is complicated by the fact that Commonsense is not alone in the picture here: pragmatical *conventions* of communication come also into play and *exploit* the elements of pure commonsense knowledge which are the common beliefs of the speaker and the hearer; place is lacking here for full discussion of this problem, but let's be satisfied with the following brief comments:

> At the level of discourse pragmatics, several types of intersentential "syntactical" relations between two contiguous declarative statements are possible: *Cause, Parallel, Exemplification, Explanation,...* <Hobbs 1984>. In the case at hand, it is the syntactical relation of *Explanation* which is realized between the first statement "John was condemned by the court" and the second statement "John had pushed his wife out of the window".
> This relation of *Explanation* can only be validated because Commonsense Inferencing is capable to provide a "natural" explicative chain between the first statement and the second statement.
> Finally, the speaker and the hearer share conventions (in the sense of Grice's "conventional implicature" <Levinson 1983> which allow them to "agree" about which of the admissible reconstructions is the intended one, perhaps with the help of a common evaluation of the degree of "naturalness" of the different admissible evenemential chains.

At any rate, the commonsense inferencing level *per se* - which is exploited by the pragmatic level - seems difficult to reconcile with a purely deductive process. In fact, the type of commonsense inferencing used in the examples of 1.3 presents certain striking similarities with the two following inductive/recognitive processes (by order of increasing complexity):

> If I observe a bike leaning behind a tree trunk hiding its central parts from me, I will have no difficulty to induce the existence of a unique bike, and not of 2 bike-halves placed in this exact position by a malevolent aesthete.

> In a movie, a sequence of two shots, the first one showing a woman writing a letter, the second one showing a man reading a letter will be interpreted as having a certain structural unity: it is the same letter, it has been sent by the woman and received by the man, etc... (Here again, as cinema is a communication medium, one must take into account not only the commonsense level, but also pragmatical conventions).

3.1.2. Several attempts have been made recently to extend the standard deduction mechanisms of logic into "non-monotonic logics". These attempts have been relatively successfull in accounting for several types of "default reasoning", but I am not aware of any application of these techniques to the extended "gestalt-type" reasoning exemplified in 3.1.1.

3.1.3. One of the other problems with the deductive approach of figure 2 is due to the difficulty there is to maintain the coherence of the thousands of axioms likely to be involved in any serious attempt to axiomatize any commonsense domain, at least if one supposes as direct a relation of interpretation as the one in figure 2. The coherence of such an axiomatization is easy to guarantee in planar geometry - for instance -, thanks to the fact that the structure of the $\mathbb{R} \times \mathbb{R}$ model is mathematically limpid, and that the correctness of the axiomatization is assured by its adequation to its model. Quite other is the situation when we endeavour to use "reality" as our model, for we have no clear idea of the structure of this model prior to our attempt at an axiomatization.

3.2. Commonsense observed.

Another important difficulty with the approach of fig. 2 is due to its stricly "objectivist" character, i.e. its insistance that Commonsense formalizations should strictly respect the "reality" of natural phenomena.

However, numerous psychological studies have unambiguously shown that our intuitive comprehension of physical phenomena differs considerably from the objective (scientific) physical state of affairs. Thus, the following experiment has been reported <Mc. Closkey 1983>: a sample of "naive" subjects were asked what they believed is the trajectory followed by a ball thrown horizontally. The results were:

> 5 % of the subjects believed that the ball travels first in a straight horizontal line, then falls straight down.
> 35 % of the subjects believed that it travels first horizontally in a straight line, then curves down for a while, then falls straight down.
> 30% of the subjects believed that it begins to curve down immediately but eventually falls straight down.
> (The scientific answer is that the ball follows a downward-oriented curve (close to a parabola) all the while).

The many observations of this kind make it difficult, if not downright impossible, to establish a system of physics which is at the same time "naive" and "objective", unless one considers that naive physics does not necessarily have to copy the commonsense structure which is observed in human individuals. This attitude might be admissible - at least theoretically -, if one were attempting to give *robots* some understanding of the external world, but not if one seeks to modelize the processes which are at the root of linguistic communication, for these constantly make use of commonsense abilities actually present in *Man*.

3.3. The diversity of phenomena accounted for by Commonsense.

Perhaps the most important source of troubles with the deductive approach of section 2 has to do with the fact that, contrarily to "scientific physics", Commonsense is not free to delimit its experimental domain in an "ideal" way. It has to give an account, at its own level, of our crude experience of the world, as it presents itself "naturally", whether or not this experience conforms to universal physical *laws*. But this experience is extremely varied and undisciplined: for instance, although most physical bodies show a "natural tendency" to "fall" if unsupported, this is neither the case with spinning-tops, nor of moving bicycles, nor the case of mercury in an upside-down thermometer, nor clouds, nor planes, nor the case of the Moon, etc...

It is difficult to imagine that there exist universal qualitative "naive" postulates from which one might deduce at the same time both on one hand the unstability of a pencil set up on its point, and on the other hand the stability of a spinning-top or of a bicycle in movement.

Also, many phenomena which Commonsense has no trouble giving an *a posteriori* account for are impossible to deduce from general laws, not only Commonsense laws, but often scientific laws too (in this instance, they are impossible or very difficult to deduce in practice, but more rarely in principle): for instance, the fact that a stone dropped into water will produce such-and-such a splash (a complex physico-mathematical fact); or, more dramatically, the fact that a cherry kernel planted into the ground will produce a cherry-tree, or even the fact that a moving bicycle is highly stable (even the experts do not seem to agree on the exact reason why it is so), etc...

However, although all these phenomena are impossible to deduce from general physical laws, Commonsense succeeds in partially motivating them and structuring them *a posteriori* with the help of general principles and schemata. (An enlightening discussion of the opposition between deduction and motivation in the domain of lexical semantics can be found in <Lakoff 1985>.)

Let's give an illustration of what is to be understood by this *a posteriori* "structuration" of experience: our understanding of the behaviour of a spinning-top.

4. Structuration of Experience: the Spinning-Top.

Informally, and without claim to psychological accuracy (no experimental verification was made), Fig. 3 gives an idea of what an intuitive structuration of a spinning-top could look like:

- pointed object, symmetrical around an axis
- set upright on its point
- having a fast rotational movement around its axis
- the axis oscillates around the vertical direction
- the "rotational impetus" is slowly wearing out
- two "antagonist tendencies" are fighting against each other :
  a) "natural tendency" of the object to fall down
  b) tendency of the axis to stay close to the vertical direction
  this last tendency is in a relation of monotonic-increasing with its
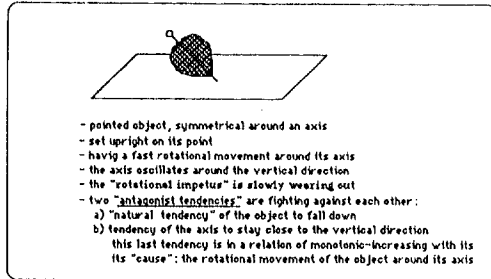 its "cause": the rotational movement of the object around its axis

Figure 3

Once structured in this manner, the spinning-top phenomenon is not any more a raw experience, but has become an informed experience: what we can call, following <Lakoff 1985> a "natural *category*" .

Several points should be noted:

The structure of Fig.3 has only a remote connection to the actual mechanical structure of the phenomenon: there is no physical sense in speaking about a "tendency to stay vertical", and it is very problematical from the physics point of view to divide the phenomenon into causes and consequences.

The spinning-top category has been structured with the help of a process of *metaphorical importation* <Lakoff, Johnson 1980> of pre-existing conceptual structures (which are categories too, but of a more abstract type, see section 5), most notably the structure of "two antagonist tendencies fighting against each other". This type of structure seems to be of wide use in language and thought and has been extensively studied by Len Talmy <Talmy 1985>. Another imported conceptual structure which has served to build the spinning-top category is the category of "natural impetus loss" which specializes here to a structure of "rotational impetus loss". (For a related notion, see the discussion of "Ohm's p-prim" in <DiSessa 1980>).

The structure which has been given to the spinning-top category is "explicative" (and "predictive") for a whole class of behaviours of the spinning-top: thus, the fact that it is impossible for the top to stay vertical if it is not rotating, the top's stability loss with time (because of the loss of rotational impetus), etc...

The structure of the spinning-top category is, in an essential way, post-experimental: for an observer without experience of gyroscopic-type behaviour, the top should have "fallen" and it would not have constituted *a priori* a special category. From the standpoint of such an observer, a cartoon showing a "spinning-top" losing its balance in the same way as a pen set on its point would have been perceived as the normal thing and would have constituted a *dynamic category* (see section 5) built with the *parallel-event* construct (which is an abstract category) out of 2 subcategories: "fall of an unbalanced body in contact with a surface" and "rotative movement of a symetrical body around its axis". What happens when the first actual spinning-tops experiments are observed is the fact that the observer becomes aware that this possible structuring of the experiments has to be rejected - it obviously does not fit the observed facts - and that a new "reading" of the experiments has to be built: a new category must be created, which integrates in a novel manner, on the one hand the rotational movement and on the other hand the natural tendency of the top to fall, by rendering these two factors causally dependent in the way given in fig. 3.

5. An alternative to the Direct Interpretation Approach.

In the remaining space, we would like to point out to another direction that research on Commonsense Inferencing could pursue. This will be done with the help of a "speech analysis metaphor". This metaphor should really be justified by a careful analysis of the inferencing processes that are at work in the modern

generative approach to linguistics (as exemplified e.g. in <Kay 1979>), which space does not permit to be given here. Such an analysis would show that generative syntax - although it can indeed be reinterpreted in FOL, in about the same way as a high-level-language program can be reinterpreted in machine code - really relies on a quite different approach to interpretation than the tarskian theory/model approach. Namely, the same linguistic input can be interpreted by different competing structural descriptions. Adding new predefined structural descriptions to the linguistic theory (i.e. the grammar) will *increase* the range of the accepted linguistic-imputs, whereas adding axioms to a logical "theory of the world" *decreases* the range of accepted situation-inputs.

5.1. A speech-analysis Metaphor for Commonsense Inferencing.

The main features of the metaphor are illustrated in fig. 4:

structure "read" by the hearer   structure "read" by the observer

interpretation

untreated acoustical-speech chain   situation in the real world

words ●   simple conceptual categories

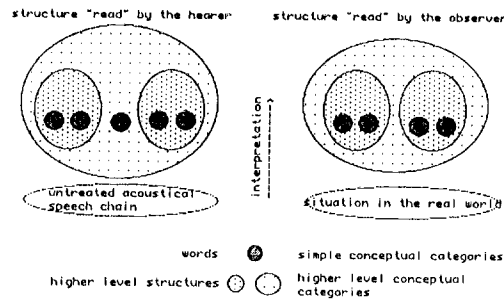higher level structures ○ ○   higher level conceptual categories

Figure 4

In the metaphor, a parallel is made between:

The unanalysed real world, which Commonsense "observes" and has to "make sense" of. / The unanalysed acoustical string, which the linguistic agent (here the hearer) "observes" and has to "make sense" of.

The FD (Functional Descriptions <Kay 1979>) of words are the elementary "blocks" out of which the linguistic structures are built / The simple conceptual categories (e.g. *cat, dance*) are the elementary blocks out of which higher conceptual structures are built.

The higher level FDs (e.g. those corresponding to VP) are used to build more complex linguistic structures. (There is really only a difference of level between this case and the preceding one.) / The higher level conceptual categories (e.g. the *parallel-events* category, the *time-sequence-of-events* category (see section 5.3)) are used to build more complex conceptual structures. (Same remark as before.)

The linguistic structure "*projects* " onto the sound-chain. There must be some "fit", but it need not be perfect: the sound-chain may be noisy and lacunary / The conceptual structure "*projects* " onto the observed situation in the real world. There must be some "fit", but it need not be perfect: in fact the interpretative freedom is much more important here then in the speech-analysis case (see section 3.2).

5.2 The different dimensions along which categories can be ordered.

One can distinguish 4 axes along wich conceptual categories can be ordered informally:

Along the complexity dimension, one can place, from simple to complex, categories representing:

"simple objects": a chair, an apple
"simple events": to eat
"complex objects": a pair of shoes, a road network
"complex events": a spinning top, a plane flying

Along the abstraction dimension, one can place, from concrete to abstract, categories representing:

"concrete" categories: an apple, a cat
more "abstract" categories: free fall of physical bodies, quantity-conserving pouring
"abstract" categories: an oriented graph, causal sequencing of events

513

Along the productivity dimension, one can place, from fixed to productive, categories representing:

fixed categories: John's bicycle
more productive categories: an apple
productive categories: causal sequencing of events

The fourth axis, that of *staticity/dynamicity* is discussed in next section.

5.3 How new categories are formed.

The metaphor of section 5.1 between commonsense inferencing and the process of linguistic interpretation permits to transpose the generative approach of syntax to the plane of commonsense inferencing (fig. 5):
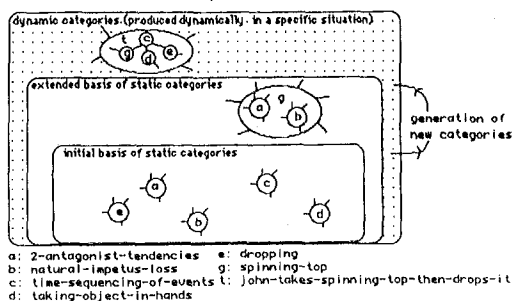

Figure 5

In this figure, there appear several types of categories:

- i) The categories belonging to the "initial basis of static categories".

They constitute the initial structural "stock" from which new categories are formed by the process of generation. In fig. 5 a, c, etc... are the categories *two-antagonist-tendencies, time-sequencing of events*, which form the basis upon which higher categories are built.

- ii) The categories belonging to the "extended basis of static categories".

They are obtained through integration and coordination of several already existing static categories. An example that we have already seen is the example of the *spinning-top* category, which integrates (among others) the structures of *two-antagonist-tendencies , natural-impetus-loss*, and other structures which we have not spoken of, as for instance that of *abstract-containment* .

Static categories, contrarily to dynamic categories (see iii next), correspond to classes of reproducible phenomena, which, once structured into a conceptual category, become of general utility for commonsense processes and therefore need to be memorized permanently. For instance, once the *spinning-top* category has been formed to account for the experiences one has had of effectively observed spinning-tops, this category becomes of a general utility to:

1) identify a spinning-top phenomenon as an instance of the *spinning-top* category.

2) infer properties of this phenomenon thanks to the previous identification of the category. These properties are not directly perceivable in the phenomenon, but are described in the structure of the category which has just been identified.

In the same way, to take a simpler example, once the static category of *bicycle* has been identified in a real situation, the perception of a fuzzy spot on the bicycle's front is sufficient to infer immediately the presence of a headlamp which is not directly perceivable (gestalt-type processing).

Thus, the stock of already existing static categories serves to give structure to new categories of phenomena. We have seen in section 4. that these structures are, in many cases, necessarily post-experimental and cannot be deduced *a priori*.

The existence of categories pre-existing to the experiencing of new phenomena permits to the active observer to recognize in the experience of a phenomenon the presence of simple or complex categories which are already known by him.

Thus, if the observed phenomenon *had been* different (for instance if the top behaved differently from what is the actual case), other pre-existing structures would have come into play: there would be in any case adaptation of the modelization to the observed experience. The difference with the philosophy of section 2 appears clearly here.

- iii) The "dynamically produced categories".

Dynamic categories (which we have also called "theorizations") differ mainly from static categories in that dynamic categories only "serve once" to structure an actually present concrete situation which is irreproducible (by definition, for it is "this" situation here and now). They are therefore created on the spot and are only temporary structures. For instance, the dynamic category t of fig. 5 corresponds to the concrete situation "John takes the spinning-top in his hand then drops it on the ground": The structuration of this category is partially shown on fig. 5: it is a category formed by integration of the categories *spinning-top, taking-object-in-hands, dropping* with (and "under") the productive category *time-sequencing-of-events* (we do not elaborate here on the question of what happens to a spinning top when it is taken in hands!).

Space is lacking here for us to give a detailed discussion of dynamic categories and of the inferencing processes which manipulate them, which discussion must be postoned to a later study, but let us just give another illustration, related to the example of section 1.3: "John was condemned to death...". In this case, the speaker, who is supposed to have been the direct "observer" of the situation, has built a dynamic category which represents his understanding of the situation he has observed, and which is an integration of the dynamic categories "John pushes his wife out of the window", "she falls from the 10th floor to the ground", etc... themselves constituted from static categories: "pushing", "falling", "dying", etc... Furthermore, it is this dynamic category (or a part of it) that he tries to transmit to his listener through an efficient linguistic coding which is highly elliptical, and it is this dynamic category (or a part of it) that the listener succeeds in reconstructing.

The principal feature of the process of formation of new categories, whether static or dynamic, is that it is not basically a process of deduction, but a process of recognition of structures. This recognition is done on the basis of a set of partial informations about the phenomenon, which is what gives it its status of inferencing. This is a crucial fact that cannot be detailed here.
Let's only give one example: A sufficiently old child will recognize the abstract category *quantity-conserving-interpouring* in an actual situation where Peter pours some wine from a bottle into a glass, although the invariance described by the category is not directly observable by her - indeed younger children, who do not yet possess this category, or a more abstract one over which this one is built, are not aware of this conservation, see <Piaget 1979>.

Conclusion.

We have tried to demonstrate the difficulties there are to apply a DI (direct interpretation) approach to the formalization of our commonsense understanding of the natural world. From a confrontation with linguistic phenomena and linguistic methodology, we have advocated a more indirect approach to interpretation. Work on a formal framework for this approach is currently under way.

REFERENCES
Disessa, Andrea. 1983. *Phenomenology and the evolution of Intuition* in (Gentner, D. and Stevens, A.L.) *Mental Models*, Lawrence Erlbaum Associates.
Hayes, Patrick J. 1978. *The naive physics Manifesto*, ISSCO, Geneva.
Hobbs, Jerry. 1984. *On the Coherence and Structure of Discourse*, SRI
Hobbs, Jerry. and Moore, Bob. 1985. *Formal Theories of the Commonsense World*, Ablex, Norwood.
Kay, Martin. 1979. *Functional Grammar*, Berkeley Linguistics Society.
Lakoff, George. 1985. *Women, Fire and Dangerous Things. What Categories Tell Us About The Nature of Thought.*, University of Chicago Press.
Lakoff, George. and Johnson, Mark. 1980. *Metaphors we live by*, University of Chicago Press.
Levinson, Stephen. 1983. *Pragmatics*, Cambridge University Press.
McCloskey, Michael. 1983. *Intuitive Physics*, Scientific American, April 1983 .
Piaget, Jean. 1979. *L'épistémologie Génétique*, Presses Universitaires de France, Paris.
Talmy, Len. 1985. *Force Dynamics in Thought and Language*, University of Chicago Linguistics Society.