Modelling Action in Verbal Command Context with Fuzzy Subsets and Semantic Networks

Christophe Brouard (Christophe.Brouard@lip6.fr)
Bernadette Bouchon-Meunier (Bernadette.Bouchon-Meunier@lip6.fr)
LIP6 Université Paris 6, Case 169, 4 place Jussieu,
75252 Paris cedex 05, FRANCE

Charles A. Tijus (tijus@univ-paris8.fr)
Laboratoire de Psychologie Cognitive, Université Paris 8, 2 rue de la liberté
93526 St Denis Cedex 02, FRANCE

Abstract

This study deals with the interpretation of verbal commands for action. After an experimental study of human interpretation of instructions for drawing geometrical figures, we have devised a model whose computerized version is called SIROCO. This model represents an attempt to simulate differents mechanisms implied in interpretation of verbal commands. These mechanisms exploiting contextual informations allow clarifying and completing propositions expressed in natural language. In the model, first, the constraints expressed in present command, in environnement (already present figures), and in background communication, are represented with fuzzy subsets and circumstantial semantic networks (well suited for flexible and dynamical representations). Subsequently, an optimization procedure integrating all this constraints allows finding a relevant response to the command. Finally a simulation which consists in translating instructor's verbal commands in a defined minimal language and making it interpreted by the system shows quite good results for the model.

Introduction

When you look at the content of verbal commands, they appear to be poor, ambiguous and elliptic. Nevertheless, they are in fact efficient as measured by the fit of operator (the person who executes commands) actions to the instructor's (the person who formulates the command) intended goal. In summary, a few words are enough to elicit complex and precise actions. How can the power of utterances be explained?

A partial explanation lies in the fact that the operator has mental models of situations, scenarios and procedures at his disposal. These comprise a general knowledge which allows him to complete the information received, to activate other knowledge in order to understand what is being asked of him and finally, to carry out the action. When, for example, someone is asked to post a letter, he knows that the letter needs a stamp, an address, that it should be dropped in a mail box or taken to the post office. Modelling the operator, (here, the person asked to mail the letter) calls for describing and representing the kind of general knowledge we have just described. This is what a number of recent systems have attempted to do, including CARAMEL (Sabah & Briffault, 1993) for understanding stories, CAMILLE (Hasting & Lytinen, 1994) for describing scenarios.

Pragmatic explanations might also be useful in explaining the power of utterances. Sperber and Wilson's communicational implications (1986) and Grice's maxims (1975) come to mind. Thus, in the above example, lacking any indication as to the cost of the stamps, the operator might rightly assume that the letter should be sent at a standard rate; because if it were to be sent express or recommended, this very relevant bit of information would surely have been provided. Modelling the operator thus calls for integrating pragmatic rules as well as general knowledge into the comprehension system. This is what has been done with DIABOLO, a system for analysing and generating dialogue (Vilnat, 1995).

The first explanation as to why utterances are so powerfull is about shared general knowledge, the second concerns the internal logic of communication. The situated action approach[1] provides a more circumstantial way of explaining the efficiency of speech. The proponents of situated action place less emphasis on the notion of internal representation and more on situational cues and action. For Olson (1970), who rejects the linguistic approach to studying the comprehension of verbal utterances, the meaning of an utterance should not be looked for in the proposition, but in the situation to which the utterance refers. This is the approach we are taking here: the power of language resides in its relation to a given situation. Important clues that allow completing vague and elliptical utterances are provided by (i) the environment, (ii) the information that has already been communicated (what we will call the "background") and (iii) the task (what must be done with the elements provided by the environment).

We thus propose such a system and called it SIROCO. Though it is currently outfitted to interpret verbal commands for drawing geometrical figures, it could be adapted to interpret other kinds of verbal commands. We have used it to study how operators interpret commands and make decisions. In the case of incompleteness, the system has to identify the instructor's intended categories. In the case of imprecision, it has to define the fuzzy boundaries of the categories. To this

[1] See Norman (1993), for an introduction to this situated action approach .

end, we used two tools for representing information that is incomplete or imprecise, namely: circumstantial semantic networks and fuzzy subsets.

The study we present here was done in three phases: An experimental phase in which a human subject-operator was asked to interpret and carry out instructions for drawing geometrical figures given in natural language by a subject-instructor. The second phase consisted in designing a model of the subject-operator. Finally, a simulation allowed comparing SIROCO's responses to those of the subject-operator. This paper is organised following these three different phases.

Experiment

Objectives

The objective of this experiment was to provide empirical data on the degree of precision with which people interpret verbal commands for drawing geometrical figures. More importantly, it aimed at providing information on how missing information is completed and, more generally, on how concrete situations influence the precision with which a command is carried out. All data relative to instructor commands and operator actions was collected automatically to provide a precise record of input and output for the simulation.

Method

Participants Thirty five instructors were recruited from the undergraduate population of the University Paris 8, St.Denis-Vincennes. A single operator was recruited from the same population, his responses provided the data we analysed.

Materials A set of 35 drawings (8,2 cm large and 14,8 cm high), one for each instructor, were created with drawing software. Each drawing was composed of three simple geometrical figures. The set was designed to provide a wide range of property combinations for the geometrical figures. The different figure-properties were: rectangle, circle and square, for the shape; red, green and blue, for the color; small, medium and large, for the size; top, center and bottom for the vertical position; and, finally, left, middle and right for the horizontal position. The computer apparatus consisted of two large monitors placed back to back on a long table (Figure 1). The instructor could only communicate through verbal commands, the operator could not see the original drawing the instructor had in his hand.

General Procedure Each one of the 35 drawings was given to an instructor. The instructor was asked to make the operator reproduce this picture through verbal commands only. After the operator had finished carrying out a command, the instructor could correct the drawing with a new verbal command and so on, until the instructor was satisfied with the drawing the operator had produced.

Automatic Data Collection All action related to writing verbal commands (on the word processor) and drawing figures (on the graphic interface) was recorded with "spy" software.

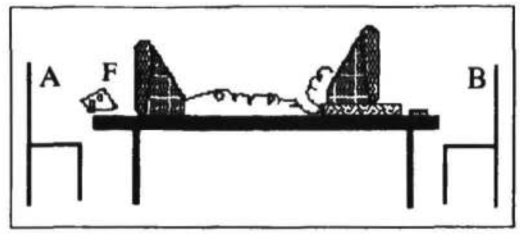

Figure 1: The instructor was placed in A and had in his hands a drawing (F). The operator was placed in B.

Results and Discussion

The Power of Utterances On average, 9 commands were necessary for a satisfactory reproduction of the original drawing. The minimum was 4, the maximun was 18 for a single drawing. On average, 3 commands were required for reproducing each figure. More precisely, 2 commands were sufficient to correct the first attempt to draw a figure.

The Precision of Commands for Discrete Properties There were just a few lateralisation errors ("not on the left, I said on the right"). Though information on size and color was not always given, these were correctly reproduced by the operator. The hypothesis that the semantic structure of the properties of the figures already in place (see Model section) allows completing the missing property information in a command was globally satisfied.

The Precision of Commands for Continuous Properties From a statistical point of view, there were no significant differences between the figure-values for the continuous properties of the operator's finished drawings and the corresponding values of the original drawings: for the X coordinates of the figure's top-left corner $p>.96$, for the Y coordinates (of the same point) $p>.17$, for width $p>.94$ and for height $p>.08$. The rate of correlation between the operator's drawings and the originals one was .86 for the X coordinate ($p<.0001$), .93 for the Y coordinate ($p<.0001$), .70 for W (width) and .86 for H (height) ($p<.0001$). The average correlation for each of the first seven commands is given in table 1.

Table 1: The average rate of correlation for each of the first seven commands

	C1	C2	C3	C4	C5	C6	C7
X:	.75	.95	.70	.80	.91	.91	.91
Y:	.88	.99	.76	.97	.98	.94	.97
W:	.96	.76	.27	.55	.53	.96	.50
H:	.94	.93	.80	.50	.93	.97	.88

It is clear that the operator faithfully reproduced the original values quite rapidly, because from the first try on, the commands were executed with an overall precision of 4 % for X, 1% for Y, 3 % for W and 2% for H. When the operator's figures did not fully correspond, a few more verbal commands were all that was needed to correct them.

Model

The results of the experiment show that the situation is indeed an aid in interpreting commands. The present model replicates the way in which the situation provides information by taking advantage of the dynamicity of circumstantial semantic networks and the flexibility of fuzzy subsets.

Incompleteness Processing with Semantic Networks

The propositional meaning of an instruction is first analysed as to the objects and their associated properties. Subsequently, objects and properties are used in order to construct a semantic network which reflects an understanding of the proposition (Poitrenaud, 1995). In this network, properties are grouped together when they are shared by several objects in order to constitute categories (Figure 2).

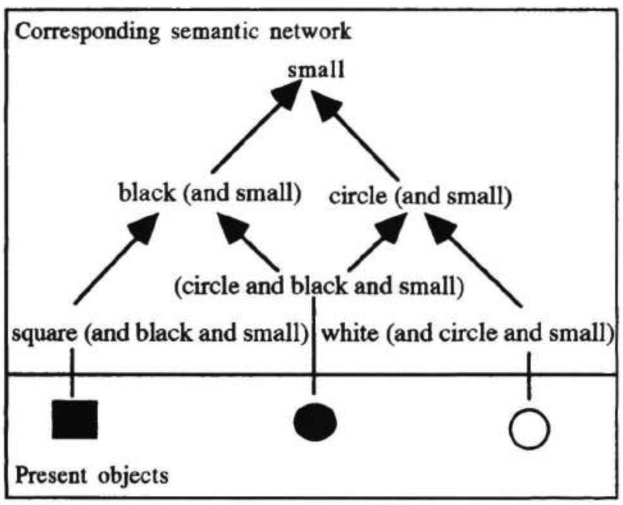

Figure 2: Example of a semantic network constructed from the object properties of the situation.

The underlying mathematical structure of this property network is the Galois lattice (Barbut & Monjardet, 1970). This network allows different logical operations. For example, if among different geometrical coloured figures, all the squares are black, it is possible to predict the black property from the square property because of the inherited properties of the square category in the semantic network. For example, it is always a problem to categorise an incompletely described new object. A good solution (from the point of view of modelisation) consists in choosing or constructing a category that does not alter the structure of the network (as far as possible). For example, if a white square has to be drawn, without any specification as to its size, in the situation described in Figure 2, it will be small. More generally, this circumstancial semantic network can be very useful for modelling action taking the situation into account (Richard, J.F., Poitrenaud, S., & Tijus, C.A., 1993).

Representing a Command with Fuzzy Subsets

A drawing command specifies size, shape, colour and position categories. Except for colour categories which are precisely defined (there is just one kind of blue, green and red available), the other kinds of categories (for example, large, rectangle) have imprecise boundaries. Thus an element (like a value corresponding to a surface in square centimetres) can have an intermediate degree of membership between 0 and 1 in a category. So we have chosen to represent these categories with a fuzzy subset. The concept of fuzzy subset (Zadeh, 1965) is a generalisation of the concept of set, it is characterised by its membership function (Figure 3).

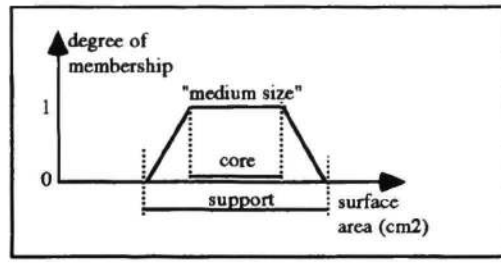

Figure 3: Membership function of a fuzzy subset representing "medium size" category. Note that the "core" is comprised of elements which belong to the fuzzy subset with a membership degree equal to 1, and that the "support" is comprised of elements which belong to the fuzzy subset with a non-zero degree.

As we are focusing upon general principles, we chose simple variables (like surface area for size categories, and abscissa for horizontal position) and trapezoidal fuzzy subsets. We represent a command by associating a fuzzy set to each dimension of the description. Zadeh (1975) introduced the concept of linguistic variables which consist of a variable, a universe in which the variable is defined (real numbers for example), and a set of fuzzy subsets which represent different characterisations of the variable (for example, small medium and large for a size variable). Here we use four linguistic variables to represent a command: (i) the size linguistic variable which is the surface area of the figure and which is characterised by "small", "medium" and "large", (ii) the shape variable which is the width/height quotient and which is characterised by "upright" "equal" and "reclining", (iii) the horizontal position on the abscissa which is characterised by "left", "middle" and "right" and, finally, (iv) the vertical position which is on the ordinate and which is characterised by "top", "centre" and "bottom". Two discrete variables complete this representation: the colour which can be blue, green or red and the software's shape tools, one for drawing rectangles and the second for drawing ellipses.

Applying a Linguistic Modifier We aim here to represent utterances like "very large" or "toward the left" (as opposed to "on the left"). Linguistic are mathematical transformations which allow constructing new fuzzy subsets from initial ones. The initial fuzzy subset represents an initial category ("large"). The new fuzzy subset represents a modified version ("very large") of an initial category. Numerous modifiers exist. Here, we use those introduced by Bouchon and Yao (1992) which exploit the distribution of defined categories in a single universe (size, for example). The mathematical transformation of these linguistic variables corresponds to a shift (see Figure 4).

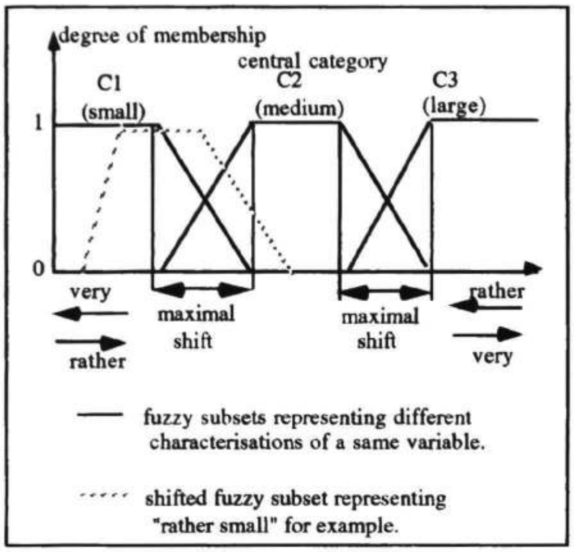

Figure 4: Illustration of linguistic modifier mechanisms.

From a given characterisation and a given modifier, these mechanisms allow deducing the shift (defined by a direction and an amplitude) to be applied. Thus, for modifiers like "very" the direction of the shift is toward an extreme and for modifiers like "rather" the direction is toward the centre (Figure 4). The amplitude of the shift is defined as a proportion of the maximal shift which corresponds to the distance between initial category cores. Thus a modified category will never overlap a neighbouring category. Moreover, the maximal shift automatically defines a scale regardless of the type of variable. Finally, it is possible to use modifiers with different strengths. Thus "very very" is a modifier of the same kind as "very" but the amplitude of the shift associated to it is larger. To be more precise, the coefficient associated to "very very" is larger than the one associated to "very"

Applying a Fuzzy Relation We aim here to represent utterances like "larger" or "a little bit less to the left". The concept of fuzzy relations is a generalisation of the concept of relation as it allows intermediate degrees (between 0 and 1) of relation between elements. It corresponds again to a fuzzy subset. Unlike the modifiers, this fuzzy set will not be constructed from a fuzzy set but from a value (the surface area of the figure, if the command is "larger").

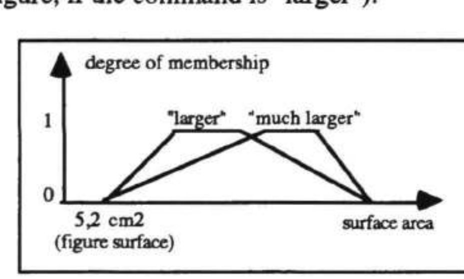

Figure 5: Illustration of fuzzy subset construction for utterances containing "larger" and "much larger".

We can divide this kind of command into two parts: the relation part which is, for example "much more", "less" or "same" and a category part which is, for example "on the left" or "large". It is possible to define mechanisms such that from a given relation and category, the fuzzy set representing the utterance can be constructed. Like modifiers , different coefficients are associated with each relation expressing a different strength. "Much more" indicates a stronger variation than "more" (Figure 5).

Background communication

At any point in a verbal exchange involving commands, what has already been said and done, the exchange's background, is decisive for interpretation. For example, what is meant by "larger" can vary according to whether it is an initial correction whose aim is to get the operator to draw a figure of roughly the right size or whether it is a final correction aimed at precision.

Background Construction During communication, various indications and corrections are given. This can be represented by a list of slopes of different constructed trapezia in the prior commands. For each variable, there is one background. Fixing a maximal length for this list allows taking the operator's limited memory into account.

Taking Background into Account Only two slopes are useful for each variable. They correspond to the more restrictive constraints (right, left constraints, for instance, respectively, much less big than 10.3 centimetres and larger than 5.4 centimetres) and allow constructing a fuzzy set. So, background is taken into account by intersecting this last fuzzy subset with the current command associated to the fuzzy subset. When this intersection is small (under a given threshold), we can decide to forget the background in order to produce an appropriate response despite contradictory commands.

Choosing an Appropriate Solution

Choosing a Relevant Point According to the specified position in the drawing command, the relevant point varies. For example, if the command calls for drawing a figure at the top left corner, the top left corner of the figure is the relevant point. From the 3 vertical position characterisations and the 3 horizontal position characterisations, we defined 9 relevant points.

Defining the Degree of Acceptability for all Points of the Drawing Area For a point p of the drawing area, the acceptability degree is computed by aggregation of two intermediate degrees $d_1(p)$ and $d_2(p)$. We chose the min operator for expressing conjunctions:

$$d(p) = \min(d_1(p), d_2(p)),$$

where $d_1(p)$ indicates the degree to which point p (the relevant point) is a good point from which to begin drawing the figure specified in the command ("in the top-left corner" or "near the circle") and where $d_2(p)$ indicates the degree to which it is possible to place at p a figure of the size and shape corresponding respectively to the size and the shape of the characterisations of the command. It is computed as:

$$d_2(p) = \sup\{(\min(\mu_{size}(l), \mu_{shape}(l,h))\}$$
$$\text{with } 0 < l < lmax, 0 < h < hmax(l)$$

where lmax and hmax are respectively the largest width and height possible taking into account the figures already present and μ_{size} and μ_{shape} are respectively the membership functions of the size and shape fuzzy subsets constructed from the command. Computing d(p) for all drawing area points allows defining favourable areas (Figure 7).

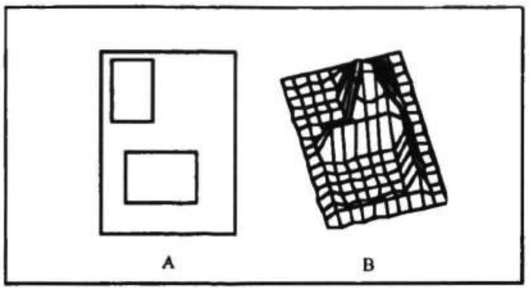

Figure 7: A) Figures already present. B) Visualisation of favourable areas for drawing "a large circle at the centre of the drawing area".

A Solution Suited to the Situation The general optimization procedure allows choosing a solution suited to the situation without explicitly describing the situation beforehand (Figure 8).

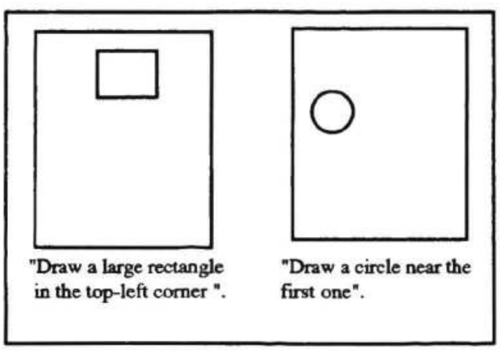

Figure 8: In these two situations, the optimisation procedures decide respectively to draw an upright rectangle and a circle to the right of the first one.

Validation

The above model has been computerised and called SIROCO. This system allowed simulating the operator-subject in order to validate the model by comparing system responses to the operator ones.

Model Parametrization

The experiment provided thirty five communication records. Ten records were kept in order to test the model. The others were used for teaching the fuzzy subsets of the different characterisations, the modifiers and relation parameters to the system. More precisely, the first drawings for each communication (which correspond to a minimal context) allowed defining the cores for all characterisations. Supports were then defined in order to construct a fuzzy partition for each variable (see above). Analysing experimental results allowed defining modifier and relation coefficients. Relation coefficients express similarity, these similarity relations are not necessarily linear. For example, for an equal difference of surface, the smaller the two compared surfaces are, the more they are perceptively different. However, we considered these relations to be linear, and chose average coefficients because the experimental material did not allow inferring their exact shape.

Simulation

A Description of SIROCO The system we developed in C++ includes a graphic interface that allows visualising system and subject drawing responses. It also allows running a commands file, typing commands interactively and readjusting the system's responses to the subject's responses at will. Finally it allows visualising favourable drawing areas (by creating a matlab file).

Definition of a Minimal Language The commands that were kept in order to test the model were translated into a minimal language consisting of a limited number of words and with a strict structure[2] . Most of these words indicate the linguistic variable characterisations, and also, the modifiers and relations often used in commands. This language aims at representing commands without interpreting them. For example, "nearer the edge" is not translated as "more toward the left" (if the figure is near the left edge) but by "more extreme" ("extreme" is automatically replaced with the object category, in this case, the "left"). Likewise, "make the rectangle longer" is translated by "more extreme". Thus, if the shape category is upright, the height will be increased. If the category is reclining, the width will be increased. We use also the OR connector in order to express utterances like "rectangle" (an upright rectangle OR a reclining rectangle) or "next to" (to the left OR to the right of another figure).

Simulation with Readjustment For this simulation, the operator-subject comparison was made command by command. Each of the system's responses was automatically readjusted to the subject's, just before the next command was interpreted. The communication background was also readjusted. Thus, for each new command, the system was placed in precisely the same interpreting situation as the subject.

Results and Discussion

In order to evaluate model validity, we compared the figures the human operator drew with the ones the system drew. More precisely, we compared the X and Y coordinates of the figure's top-left corner, the width W and the height H. The average error margin for the position and the size of the different figures for the first seven commands is 0.3 centimetres.

Compared to the figure variance in the initial drawings, there is no significant difference between subject and system drawings for the X coordinate ($p>.2$) and for the width W ($p>.32$). On the other hand, we found differences for the Y coordinate ($p=.02$) and for the height H ($p<.01$). Positions

[2] The commands which could not be translated correctly with the minimal language, were not taken into account for the results. For more expressiveness, we should add variables and objects (like edges). Indeed, it seems more appropriate to translate "in the top corner" by "very near the top edge" than by "very very at the top". However the defined translation tables allow expressing correctly most of the commands.

and sizes have a very important correlation rate: .93, .84, .93 et .92, respectively for X, Y, W, H.

Table 2: For the first to the third figure, the degree of correlation

X:	.94	.93	.91
Y:	.93	.98	.30
W:	.91	.90	.97
H:	.83	.90	.97

Overall, the results showed a good simulation of a human operator. The error of margin with which the system operated might be, but is not necessarily, due to the model. The system chose one solution from a set of equally possible solutions and, under the same conditions, a human operator might also give different responses.

General Discussion

This modelisation is based on the integration of a set of flexible constraints into an optimization procedure. It allows explaining the adaptability of the cognitive system. A small number of combined cues are enough to allow defining a precise solution. The method we follow, first, determination of fuzzy meaning for a set of variables, and second, definition of a solution maximizing the satisfaction degree of all variable constraints and integrating all environnement constraints, seems well adapted to model action. Compared with a rule system where the rules have to cover all situations and have to be explicited, this method appears more adaptative and more simple to implement.

Other more elaborate experiments could reveal other important cues. Even in the particular case of this experiment, we do not pretend to have tackled all the facets of command interpretation. Category learning, that is to say the adjustment of interlocutor categories, is not taken into account here. This is not very important here because communication between the operator and the instructor took place very quickly (the instructor was replaced for each new drawing).

As we mentioned in the introduction, our study is about a particular contextual explanation of the power of language. Thus, some implicits of communication were not taken into account, whereas their effects were not negligible from the point of view of the results. For example, when the command was to draw a figure on the left and there already was a figure on the left, the system chose to place the new figure very near the first one (it placed it as far to the left as it could). The implicit information in this command is that the two figures can not be stuck together, because if they were, this information would be given. To explain this kind of implicit principle of relevance introduced by Sperber and Wilson (1986) seems well suited.

Conclusion and Perspectives

The aim of this interdisciplinary study was double. On the one hand, our goal was to model the processes of command interpretation (through cognitive psychology) and on the other, it was to create a system capable of responding consistently to verbal commands, of detecting implicit information and of adapting itself to a given situation (through artificial intelligence). These two aspects of the study are by no means opposed because devising a system that models a human subject has every chance of being a system whose behaviour is adequate. This is all the more true given that verbal communication is a specifically human activity.

This study must be considered as a first step. Our futur work will consist in expanding the scope and the analysis of the human experiment. In our first experiment, we chose to keep the same operator (in order to have enough data for modelling him). We could now compare different operators. We could also study the effect of stronger regularities (with drawings composed of more than 3 figures). Finally, it would be interesting to make the system interact directly with a human instructor.

Acknowledgement

This research was funded by DGA/DRET Contract N° 93-166. The authors express their thanks to E. Hamilton for her assistance in research and translating.

References

Barbut, M., & Monjardet, B. (1970). Ordre et classification: algèbre et combinatoire. Paris: Hachette.

Bouchon-Meunier, B. & Yao, J. (1992). Linguistic modifiers and gradual membership to a category. *International Journal on Intelligent Systems,* 7, 25-36.

Grice, H.P. (1975). Logic and conversation. In P.Cole and J.L. Morgan (Eds.). *Syntax and semantics*, Vol 3: Speech acts, New York: Academic Press.

Hasting, P.M., & Lytinen, S.L. (1994). Objects, Actions, Nouns, and Verbs. In Ashwin Ram & Kurt Eiselt (Eds.). *Proceedings of the Sixteenth Annual Conference of the Cognitive Science Society.*(pp. 397-402). Hillsdale, NJ: LEA

Norman, D. A. (1993). Cognition in the head and in the world: an introduction to the special issue on situated action. *Cognitive Science*, 17, 1, 1-6.

Olson, D. R. (1970). Language and Thought: Aspects Of a Cognitive Theory Of Semantics. *Psychological Review*, 4, 257-273.

Poitrenaud, S. (1995). The Procope Semantic Network : an alternative to action grammars. *International Journal of Human-Computer Studies*, 42, 31-69.

Sabah, G., & Briffault (1993). CARAMEL: step towards reflexion in natural language understanding systems. *Actes IEEE International Conference on Tools with Artificial Intelligence,* Boston, (pp. 258-265).

Sperber, D. & Wilson, D. (1986). Relevance : Communication and cognition. Cambridge, MA: Harvard University Press.

Richard, J.F., Poitrenaud, S., & Tijus C.A. (1993). Problem-solving restructuration: elimination of implicit constraints. *Cognitive Science*, 17, 497-529.

Vilnat, A. (1995). STANDIA: A pragmatic-driven Man-Machine System, the structure of multi-modal dialogue. Amsterdam: North-Holland.

Zadeh, L.A. (1965). Fuzzy sets. *Information and Control* , 8, 338-353.

Zadeh, L. A. (1975). The concept of a linguistic variable and its application to approximate reasoning. *Information Sciences*, 8, 199-249, 301-357.