

NATURAL LANGUAGE DRIVEN IMAGE GENERATION

Giovanni Adorni, Mauro Di Manzo and Fausto Giunchiglia

Department of Communication, Computer and System Sciences
University of Genoa
Via Opera Pia 11 A - 16145 Genoa - Italy

ABSTRACT

In this paper the experience made through the development of a NATURAL LANGUAGE driven Image Generation is discussed. This system is able to imagine a static scene described by means of a sequence of simple phrases. In particular, a theory for equilibrium and support will be outlined together with the problem of object positioning.

1. Introduction

A challenging application of the AI techniques is the generation of 2D projections of 3D scenes starting from a possibly unformalized input, as a natural language description. Apart from the practically unlimited simulation capabilities that a tool of this kind could give people working in the show business, a better modeling of the involved cognitive processes is important not only from the point of view of story understanding (Wa80a,Wa81a), but also for a more effective approach to a number of AI related problems, as, for instance, vision or robot planning (So76a). In this paper we discuss some of the ideas on which is based a NATURAL LANGUAGE driven Image Generation (NALIG from here on) which has been developed for experimental purposes at the University of Genoa. This system is currently able to reason about static scenes described by means of a set of simple phrases of the form: <subject> <preposition> <object> [<reference>] (*).

The understanding process in NALIG flows through several steps (distinguishable only from a logic point of view), which perform object instantiation, relation inheritance, translation of the surface expression into unambiguous primitives,

(*) NALIG has been developed for the Italian language; the prepositions it can presently analyze are: su, sopra, sotto, a destra, a sinistra, vicino, davanti, dietro, in. A second deeply revised release is currently under design.

This work has been supported by the Italian Department of Education under Grant M.P.I.-27430.

consistency checking, object positioning and so on, up to the drawing of the "imagined" scene on a screen. A general overview of NALIG is given in the paper, which however is mainly concerned with the role of common sense physical reasoning in consistency checking and object instantiation. Qualitative reasoning about physical processes is a promising tool which is exciting the interest of an increasing number of A.I. researches (Fo83a,Fo83b,Fo83c) ,(Ha78a,Ha79a) ,(Kl79a,Kl83a). It plays a central role in the scene description understanding process for several reasons:

- i. naive physics, following Hayes definition (Ha78a), is an attempt to represent the common sense knowledge that people have about the physical world. Sharing this knowledge between the speaker and the listener (the A.I. system, in our case) is the only feasible way to let the second make realistic hypotheses about the assumptions underlying the speaker utterances;
- ii. it allows to reach conclusions about problems for which very little information is available and which consequently are hard to formalize using quantitative models;
- iii. qualitative reasoning can be much more effective to reach approximate conclusions which are sufficient in everyday life. It allows to build a hierarchy of models in order to use every time the minimal requested amount of information, and avoid to compute unnecessary details.

Within the framework of naive physics, most of the current literature is devoted to dynamic processes. As far as we are concerned with the description of static scenes, other concepts are relevant as equilibrium, support, structural robustness, containment and so on. With few exceptions (Ha78a), qualitative theories to address these problems are not yet available even if some useful suggestions to approach statics can be found in (By80a). In this paper, a theory for equilibrium and support will be outlined. An important aspect of the scene description understanding process is that some amount of

qualitative analysis can never be avoided, since a well defined position must be completed for every object in order to draw the image of the scene on a screen. This computation must not result in an overspecification that masks the degree of fuzziness which is intrinsic in object positions (Wa79a), in order to avoid to unnecessarily constrain all the following reasoning activities. The last section of the paper will be devoted to the object positioning problem.

2. Object taxonomy and spatial primitives

Spatial prepositions in natural language are often ambiguous, and each one may convey several different meanings (Bo79a,He80a). Therefore, the first step is to disambiguate descriptions through the definition of a proper number of "primitive relationships".

The selection of the primitive relation representing the meaning of the input phrase is based mainly, but not only, on a taxonomy of the involved objects, where they are classified depending on attributes which, in turn, depend on the actual spatial preposition. An example may be given by the rules to select the relation $H_SUPPORT(A,B)$ (that is A is horizontally supported by B) from the phrase "A on B".

This meaning is chosen by default when some conditions are satisfied. First of all, A must not belong to that special category of objects which, when properly used, are flying, as aircrafts, unless B is an object expressly devoted to support them in some special case: so, "the airplane on the runway" is likely to be imagined touching the ground, while for the "airplane on the desert" a flying state is probably inferred (of course, the authors cannot exclude that NALIG default reasoning is biased by their personal preferences). $FLYING(A)$ and $REPOSITORY(A,B)$ predicates are used to formalize these facts. To be able to give horizontal support, B must have a free upper surface ($(FREETOP(B))$), walls or ceilings or closed doors in an indoor view do not belong to this category. Geographic objects ($GEO(X)$) impose a special care: "the mountains on the lake" cannot be interpreted as the lake supporting the mountains and even if only B is a geographic object, but A can fly, physical contact seems not to be the most common inference ("the birds on the garden"). Hence, a first tentative rule is the following (the actual rule is much more complex):

not $GEO(A)$ and not($FLYING(A)$) and
not $REPOSITORY(A,B)$ and

$$\begin{aligned} & ((FREETOP(B) \text{ and not } GEO(B)) \text{ or} \\ & (\text{GEO}(B) \text{ and not } CANFLY(A))) \\ & \implies H_SUPPORT(A,B) \end{aligned}$$

A complete discussion of NALIG's taxonomy of objects is in (Bo83a). Both the set of primitives and the set of attributes have been defined on the basis of empirical evidence, through the analysis of some thousands of sample phrases. Besides the fact that NALIG works, there are specific reasons to accept the current taxonomy, and it is likely that further experience will suggest modifications; however, most of knowledge in NALIG is descriptive, and the intrinsic flexibility of an expert system approach an easy stepwise refinement.

The values of some predicates are simply attempts to summarize large amounts of specified knowledge. For example, $CANFLY(X)$ is true for birds, but $FLYING(X)$ is not; the last predicate is reserved for airplanes and similar objects. This is a simple trick to say that, in common experience, airplanes can be supported by a very limited set of objects, as runways, aircraft carrier ships and so on, while birds can stay almost everywhere and to list all possible places is too space wasting.

However, most of them are directly related to geometrical or physical properties of objects, to their common uses in a given environment and so on, and should be always referred to underlying specific theories. For instance, a number of features are clearly related to a description of space which is largely based on the Hayes' model to develop a theory for the containment of liquids (Ha78a). Within this model some predicates, as $INSIDE(O)$, can be evaluated by means of a deeper geometric modeling module, which uses a generalized cone approach to maintain a more detailed description of the structures of objects (Ad82a,Ad83a,Ad83b). Some of these theories are currently under development (a naive approach to statics will be outlined in the following), some others are still beyond the horizon; nevertheless, for experimental purposes, unavailable sophisticated theories can be substituted by rough approximations or even by fixed valued predicates with only a graceful degradation of reasoning capabilities.

Taxonomical rules generate hypotheses about the most likely spatial primitive, but these hypotheses must be checked for consistency, using knowledge about physical processes (section 4) or about constraints imposed by the previous allocation of other objects (section 5). Moreover there are other sources of primitive relations besides the input phrase. One of the most important sources is given by a set of rules which allow to infer unmentioned objects; they are briefly

outlined in the next section. Other relations may be inferred as side-effects of consistency checking and positioning activities.

3. Object instantiation

Often a natural language description gives only some details about the scene, but many other objects and relations must be inferred to satisfy the consistency requirements. An example is the phrase "a branch on the roof" which is probably interpreted as "a tree near the house having a branch on the roof". Therefore a set of rules has been defined in NALIG to instantiate unmentioned objects and infer the relations holding between them.

Some of these rules are based on knowledge about the structure of objects, so that, under proper conditions, the whole can be inferred when a part is mentioned. Other rules take into account state conditions, as the fact that a living fish need water all around, or containment constraints, as the fact that water is spread on a plane surface unless it is put into a suitable container. The inferred objects may inherit spatial relations from those explicitly mentioned; in such a case relation replacement rules are needed. A simple example is the following. Geographic objects containing water, as a lake, can be said to support something (the boat on the lake), but the true relation holds between the supported object and the water; this fact must be pointed out because it is relevant for consistency conditions. Therefore a replacement rule is:

$$\begin{aligned} &\text{ON}(A,B) \text{ and } \text{GEO}(B) \text{ and } \text{OPENCONTAINER}(B) \text{ and} \\ &\text{not } \text{GEO}(A) \text{ and not } (\text{FLYING}(A) \text{ and} \\ &\text{not } \text{REPOSITORY}(A,B)) \text{ and not } \text{CANFLY}(A) \\ \Rightarrow &\text{ON}(A,\text{water}) \text{ and } \text{CONTAINED}(\text{water},B) \end{aligned}$$

where $\text{ON}(X,Y)$ represents the phrase to be analyzed; $\text{OPENCONTAINER}(X)$ has the same formal meaning defined by Hayes (Ha78a) and describes a container with an open top.

When relation inheritance does not apply, relative positions between known and inferred objects must be deduced from knowledge about their structures and typical positions. For instance the PART_OF instantiation rule, triggered by the phrase "the branch on the roof" to infer a tree and a house, does not use the relation inheritance (the tree is not on the house), but knowledge about their typical positions (both objects are usually on the ground with assumed standard axis orientations) or structural constraints, as the house cannot be too high and the tree too far from the house, otherwise the stated relation between

the branch and the roof becomes unlikely. A deeper discussion of these inference rules is presented in (Ad83c).

4. Consistency checking and qualitative reasoning

Objects which do not fly must be supported by other objects. This seemingly trivial interpretation of the law of gravity plays a basic role when we check the consistency of a set of given or assumed spatial relationships; no object is properly placed in the imagined scene if it is not possible to relate it, possibly through a chain of other supporting objects, to one which has the role of "ground" in the assumed environment (for instance floor, ceiling and interior surfaces of walls in an indoor view). The need of justifying this way all object positions may have effects on object instantiation, as in the phrase "the book on the pencil". Since the pencil cannot give full support to the book another object must be assumed, which supports the pencil and, at least partially, the book; both objects could be placed directly on the floor, but default knowledge about the typical positions that books and pencils may have in common will probably lead to the instantiation of the table as the most likely supporting object, in turn supported by the floor.

The supporting laws may also give guidance to the positioning steps, as in the phrase "the car on the shelf" where, if there are reasons to reject the hypothesis that the car is a toy, then it is unlikely to have the shelf in its default position, that is "on the wall".

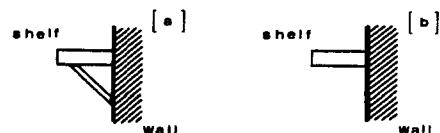


fig.1:assumed and default shelf structures

Another example of reasoning based on supporting rules is given by assumptions about the structure of objects, in those cases in which a number of alternatives is known. For instance, if we know that "a shelf on the wall" must support a heavy load of books, we probably assume the structure of fig.1a, even if fig.1b represents the default choice.

To reason about these facts we need a strategy to find the equilibrium positions of an object or a pattern of supports, if such positions exist, taking into account specific characteristics of the involved objects. This strategy must be based, as

far as possible, on qualitative rules, to avoid unnecessary calculations in simple and common cases and to handle ill-defined situations; for instance, rules to grasp objects, as birds, are different from those holding for not grasping ones, as bottles, and nearly all situations in which birds are involved can be solved without any exact knowledge about their weight distributions, grasping strength and so on.

An example of these rules, which we call "naive statics" is given in the following. Let us consider a simple case in which an object A is supported by another object B; the supported object has one or more plane faces that can be used as bases. If a face f is a base face for A ($\text{BASE}(f,A)$), it is possible to find the point e, which is the projection of the baricenter of A on the plane containing f along its normal. It is rather intuitive that a plane horizontal surface is a stable support for A if the area of physical contact includes e and if this area is long and wide enough, in comparison to the dimensions of A, and its height in particular. Hence a minimum equilibrium area ($\text{M_E_AREA}(a,f)$) can be defined for each $\text{BASE } f$ of A (this in turn imposes some constraints on the minimal dimensions of f).

The upper surface of B may be of any shape. A support is a convex region of the upper surface of B; it may coincide with the whole upper surface of B, as it happens with a table top, or with a limited subset of it, as a piece of the upper edge of the back of a chair. In this example we will consider only supports with a plane horizontal top, possibly shrinking to a line or a point; if s is such a part of B, it will be described by the predicate $\text{P_SUPP}(s,B)$.

Let's consider now an object A, with a regular base f, lying on one or more supports whose upper surfaces belong to the same plane. For each position of A there is a pattern of possibly disconnected areas obtained from the intersection of f with the top surfaces of the supports. Let be a the minimal convex plane figure which include all these areas; a will be referred to as a supporting area ($\text{S_AREA}(a)$). A rather intuitive definition of equilibrium area is that A is stable in that position if its $\text{M_E_AREA}(a,f)$ is contained in the supporting area. A further condition is that a free space V around the supports must exist, large enough to contain A; this space can be defined by the smallest convex volume Va enveloping A which is part of the description of A itself. Therefore conditions of stable lying can be formulated as follows:

$$\begin{aligned} &\text{BASE}(f,A) \text{ and } \text{LAY}(A,B) \text{ and} \\ &\text{FREE}(V) \text{ and } \text{ENVELOP}(Va,A) \text{ and } \text{CONTAINED}(Va,V) \\ &\Rightarrow \\ &\text{STABLE_H_SUPPORT}(A,B) \end{aligned}$$

where:

$$\begin{aligned} \text{LAY}(A,B) \equiv & \text{P_SUPP}(s_1,B) \text{ and... and } \text{P_SUPP}(s_n,B) \\ & \text{and } \text{S_AREA}(a) \text{ and } \text{M_E_AREA}(e,f) \text{ and} \\ & \text{CONTAINED}(e,a) \end{aligned}$$

The evaluation of the supporting area (i.e. to find an area a for which its predicate $\text{S_AREA}(a)$ is true) may be trivial in some cases and may require sophisticated positioning strategies in other cases. The most trivial case is given by a single support S, in this case we have $\text{S_AREA}(\text{TOP}(S))$, which means that the supporting area a coincides with the top surface of S.

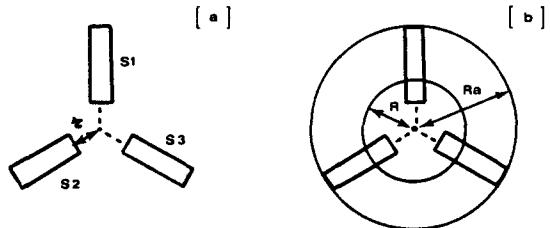


fig.2: radial simmetry

Another simple but interesting case is given by regular patterns of supports, where it is possible to take advantage of existing symmetries. Let's consider, for instance, a pattern of supports with radial simmetry, as shown in fig.2a, which may resemble a gas_stove. If the base f of a has the same kind of approximately radial simmetry (a regular polygon could be a good approximation) and if the projection c of the baricenter of A coincides with the center of f, then the supporting a is the circle with radius Ra under the condition $r \leq R$, where r is the radius of the "central hole" in the pattern of supports and R is the (minimal) radius of f. This simply means that the most obvious positioning strategy is to center A with respect to the pattern of supports; their actual shape is not important provided that they can be touched by A. In case of failure of equilibrium rules a lower number of supports must be considered and the radial simmetry is lost (for instance, the case of a single support may be analyzed).

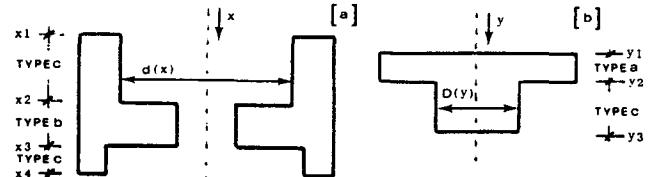


fig.3: axial simmetry

As a third example let us consider a couple of supports with an axis symmetry as shown in fig.3a (straight contours are used only to simplify the discussion of this example, but there are not constraints on the actual shapes (besides simmetry). If the face f for A exhibits the same kinds of simmetry (fig.3b) the simplest placement strategy is to align the object axis to the support one. In this case the interior contours of each support can be divided into a number of intervals, so that for each interval $[X_i, X_{i+1}]$ we have:

- a. $\min_{[X_i, X_{i+1}]} d(x) \geq \max_y D(y)$ or
- b. $\max_{[X_i, X_{i+1}]} d(x) < \min_y D(y)$ or
- c. $\left\{ \begin{array}{l} \min_{[X_i, X_{i+1}]} d(x) \geq \min_y D(y) \\ \max_{[X_i, X_{i+1}]} d(x) < \max_y D(y) \end{array} \right\}$ and

Analogously the object contour can be divided in intervals, so that for each interval $[Y_j, Y_{j+1}]$ we have:

- A. $\min_{[Y_j, Y_{j+1}]} D(y) > \max_x d(x)$ or
- B. $\max_{[Y_j, Y_{j+1}]} D(y) \leq \min_x d(x)$ or
- C. $\min_{[Y_j, Y_{j+1}]} D(y) > \min_x d(x)$ and
 $\max_{[Y_j, Y_{j+1}]} D(y) \leq \max_x d(x)$

Of course, some situations are mutually exclusive (type a with type A or type b with type B intervals).

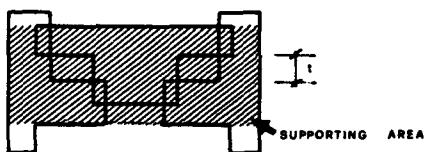


fig.4:supporting area

Equilibrium positions may be found superimposing object intervals to support one by means of rules which are specific for each combination of types. For example, one type A and one type b intervals can be used to search for an equilibrium position by means of a rule that can be roughly expressed as:

"put type A on type c and type C on type b so that the distance t (see fig.4) is maximized".

The supporting area a obtained this way is shown (the dashed one) in fig.4. This kind of rules can be easily generalized to handle situations as a pencil on a grill. Some problems arise when the supports do not lie on the other plane, as for a book supported partially by the table top and partially by another book; in this case the concept of friction becomes relevant. A more detailed and better formalized description of naive statics can be found in (Di84a).

5. Positioning objects in the scene

A special positioning module must be invoked to compute the actual coordinates of objects in order to show the scene on the screen. This module, which we mention only for lack of space, has a basic role, since it coordinates the knowledge about the whole scene, and can therefore activate specific reasoning activities. For instance, there are rules to handle the transparency of some objects with respect to particular relations and possibly to generate new relations to be checked on the basis of the previously discussed criteria. An example is the phrase "the book on the table", which is accepted by the logic module as H_SUPPORT(book,table) but can be rejected at this level if there is no enough free space on the table top, and therefore modified into a new relation H_SUPPORT(book,B), where B is a suitable object which is known to be supported by the table and is transparent to respect the On relationship (another book, for instance). A more detailed description can be found in (Ad84a).

6. Conclusions

NALIG is currently able to accept a description as a set of simple spatial relations between objects and draw the imagined scene on a screen. A number of problems are still open, mainly in the area of knowledge models to describe physical phenomena and in the area of a suitable use of fuzzy logic to handle uncertain object positions. Apart from these enhancements of the current release of NALIG, future work will be also focused on the interconnection of NALIG with an animation system which is under development at the University of Genoa (Mo84a), in order to explore also those reasoning problems that are related to the description of actions performed by human actors.

REFERENCES

- Ad82a. Adorni,G., Boccalatte,A., and DiManzo,M., "Cognitive Models for Computer Vision", Proc. 9th. COLING, pp. 7-12 (Prague, Czechoslovakia, July 1982).
- Ad83a. Adorni,G. and DiManzo,M., "Top-Down Approach to Scene Interpretation", Proc. CIL-83, pp. 591-606 (Barcelona, Spain, June 1983).
- Ad83b. Adorni,G., DiManzo,M., and Ferrari,G., "Natural Language Input for Scene Generation", Proc. 1st. Conf. of the European Chapter of the ACL, pp. 175-182 (Pisa, Italy, September 1983).
- Ad83c. Adorni,G., DiManzo,M., and Giunchiglia,F., "Some Basic Mechanisms for Common Sense Reasoning about Stories Environments", Proc. 8th. IJCAI, pp. 72-74 (Karlsruhe, West Germany, August 1983).
- Ad84a. Adorni,G., Di Manzo,M., and Giunchiglia,F., "From Descriptions to Images: what Reasoning in between?", to appear in Proc. 6th. ECAI, (Pisa, Italy, September 1984).
- Bo79a. Boggess,L.C., "Computational Interpretation of English Spatial Prepositions", TR-75, Coordinated Sci. Lab., Univ. of Illinois, Urbana, ILL (February 1979).
- Bo83a. Bona,R. and Giunchiglia,F., "The semantics of some spatial prepositions: the Italian case as an example", DIST, Technical Report, Genoa, Italy (January 1983).
- By80a. Byrd,L. and Borning,A., "Extending MECHO to Solve Static Problems", Proc. AISB-80 Conference on Artificial Intelligence, (Amsterdam, The Netherlands, July 1980).
- Di84a. DiManzo,M., "A qualitative approach to statics", DIST, Technical Report, Genoa, Italy (June 1984).
- Fo83a. Forbus,K., "Qualitative Reasoning about Space and Motion", in Mental Models, ed. Gentner,D., and Stevens,A., LEA Publishers, Hillsdale, N.J. (1983).
- Fo83b. Forbus,K., "Measurement Interpretation in Qualitative Process Theory", Proc. 8th. IJCAI, pp. 315-320 (Karlsruhe, West Germany, August 1983).
- Fo83c. Forbus,K., "Qualitative Process Theory", AIM-664A, Massachusetts Institute of Technology, A.I. Lab., Cambridge, MA (May 1983).
- Ha78a. Hayes,P.J., "Naive Physics I : Ontology for liquids", Working Paper N.35, ISSCO, Univ. of Geneve, Geneve, Switzerland (August 1978).
- Ha79a. Hayes,P.J., "The Naive Physics Manifesto", in Expert Systems in the Micro Electronic Age, ed. Michie,D., Edinburgh University Press, Edinburgh, England (1979).
- He80a. Herskovitz,A., "On the Spatial Uses of the Prepositions", Proc. 18th. ACL, pp. 1-6 (Philadelphia, PEN, June 1980).
- Kl79a. de Kleer,J., "Qualitative and Quantitative Reasoning in classical Mechanics", in Artificial Intelligence: an MIT Perspective, Volume I, ed. Winston,P.H. and Brown,R.H., The MIT Press, Cambridge, MA (1979).
- Kl83a. de Kleer,J. and Brown,J., "Assumptions and Ambiguities in Mechanistic Mental Models", in Mental Models, ed. Gentner,D., and Stevens,A., LEA Publishers, Hillsdale, N.J. (1983).
- Mo84a. Morasso,P. and Zaccaria,R., "FAN (Frame Algebra for Nem): an algebra for the description of tree-structured figures in motion", DIST, Technical Report, Genoa, Italy (January 1984).
- So76a. Sondheimer,N.K., "Spatial Reference and Natural Language Machine Control", Int. J. Man-Machine Studies Vol. 8 pp. 329-336 (1976).
- Wa79a. Waltz,D.L. and Boggess, L., "Visual Analog Representations for Natural language Understanding", Proc. 6th. IJCAI, pp. 926-934 (Tokyo, Japan, August 1979).
- Wa80a. Waltz,D.L., "Understanding Scene Descriptions as Event Simulations", Proc. 18th. ACL, pp. 7-12 (Philadelphia, PEN, June 1980).
- Wa81a. Waltz,D.L., "Toward a Detailed Model of Processing for Language Describing the Physical World", Proc. 7th. IJCAI, pp. 1-6 (Vancouver, B.C., Canada, August 1981).