WPFG

Natural Language, Graphics, Animation and Robotics: Formalizing the Semantic and Semiotic Nexus
Roger Marshall
Florida Institute of Technology

NATURAL LANGUAGE, GRAPHICS, ANIMATION AND ROBOTICS: A FUNCTIONAL GRAMMAR APPROACH TO FORMALIZING THE SEMANTIC AND SEMIOTIC NEXUS

R. Marshall
Computer Science Department
Florida Institute of Technology
Melbourne, FL 32901
internet: marshall@zach.fit.edu

"It has been said of him that he could smell an isosceles triangle." Nathaniel West

Abstract

We present a formal model of the semantic, pragmatic and semiotic connections and constraints inherent in mapping one-dimensional natural language constructs (which are ordered in space and time) to two-dimensional and three-dimensional spaces as typified by the two-dimensional graphics display and robotics environments respectively. The formalisms, based on extensions of Simon Dik's Functional Grammar, result in special linguistic knowledge representation structures and algorithms which can be integrated with robot vision and motion algorithms to facilitate robot activity in a real-world setting.

Keywords: Function (Representation)
             Knowledge(Language, objects, space, time)
             Domain(Physical world, robots)
             Foundation(Cognitive, semiotics)

1. Introduction

The open literature does not seem to contain very many articles on the topic of explicit and implicit connections between language structures and their static and dynamic representations as exemplified by graphics and robotics [Bertin 83, Loomis et al 83, Magnenat and Thalman 85, Sampath 91, Tufte 84]. Published work is disparate and does not capture the connections adequately. In this paper, we make a preliminary attempt at formalizing the semantic, pragmatic, and semiotic connections and constraints inherent in mapping linguistic constructs which are linear and temporal in nature to two-dimensional and three-dimensional spaces obtained in the conventional two-dimensional graphics display and robotics environments respectively. Our methodology is based on a semiotic, cognitive semantics model of language [Buchler 55, Eco 79, Eco et al 88, Kulas et al 88, Ehrlich and Johnson-Laird 82] for two reasons - a) such a model emphasizes the communicative function of natural language and b) each semantic representation within such a model is simply a transformation into another symbolic representation system.

The well-known symbol manipulation thesis of Newell and Simon [Newell and Simon 72] is based on the assumption that a physical symbol system has the necessary and sufficient means for intelligent action and that cognition is independent of learning; however, humans encounter new situations everday and try to adapt to the new circumstances on the basis of past experience and they possess the ability to decide which type of knowledge or which type of reasoning is relevant in a given situation without resorting to elaborate 'computations' in their mind. This may be viewed as akin to employing natural language without being too concerned about grammaticality to communicate in a manner appropriate to the communicational setting; obviously, this involves 'learning' in one form or another.

The formalisms discussed in the paper can be fruitfully employed, we believe, in the following research areas - a) natural language understanding (in particular the resolution of problems presented by ambiguity, anaphora and garden path sentences) [Abe et al 88, Nakagawa and Mori 88, Warner and Glass 87], b) tutoring systems, in particular, language tutoring systems, [Madrigal and Dulac 62, Weischedel et al 78, Wenger 87], c) studies of biological constraints on and neuro-physiological connections between verbal and visual semantic systems [Bellugi and Studdert-Kennedy 80, Poizner et al 87], and d) studies on whether word order in natural language constructs is influenced by the motion of objects in space when these objects are the denotata of language constructs. In obtaining our formalisms, we have borrowed ideas from the functional grammar papers of Dik [Dik 78, 80] (Dik's grammar is based on a communicative as opposed to a cognitive semantics model of language) and papers on the syntactic specifications of American Sign Language (ASL) [Fischer 75, Friedman 75, Liddell 80].

2

Our paper is organized as follows: in Section 2, we provide short discussions of the semiotics of representation of verbal and non-verbal languages. In Section 3, we focus on the semiotics of graphical representations of natural language. In Section 4, we develop formalisms for representing natural language constructs in a dynamic graphical setting. In Section 5, we present some examples based on the formalisms. In Section 6, we indicate how the same formalisms may be adapted to control the activity of a robot. A software system architecture based on the formalisms is presented in Section 7. In the concluding section of the paper, future research directions are indicated.

2. <u>Semiotics of Language Representations</u>
In this section, the semiotics of representation of the following types of languages is presented: - a) spoken or written language, b) American Sign Language (ASL), c) music, and d) other languages based on special symbols as in traffic signs, hieroglyphics and ideograms.

2.1 <u>Natural Language Representations</u>
Natural language representations are typically formulated at different levels such as phonological, morphological, syntactic, semantic, etc. and these representations can be generative, structural or transformational in nature. In any case, these representations are one-dimensional, meta-level specifications of mental concepts expressed by natural language constructs. Much of the real-world 'meaning' of a construct is missing from all of these representations.

Consider the simple sentence "The robot walks into the burning building". One may easily represent the syntactic structure of the sentence in graphical form by using a parse tree whose nodes have syntactic and lexical labels. If one wishes to represent the semantics of the sentence using graphics, then one would have to resort to depicting the underlying semantics of the sentence via graphical entities, assuming these exist. The issue at hand is one of determining, either a priori or otherwise, whether a given semantic structure can indeed be represented in graphical form. Even a partial resolution of this issue would require that an identification of the different possible semantic structures and their classification be undertaken first prior to their usage in a graphics setting. Obviously, any classification scheme would be based on the representability or non-representability of the semantic structures in graphical form. We will have more to say on this issue in a later section of the paper. Assume that in some applications environment electromechanical device (the robot) is programmed to walk into a building which is indeed on fire and that the activities of the robot are being monitored at a remote location. In such an environment, mapping the semantics of the sentence "The robot walks into the burning building" requires that one be cognizant of both the static and dynamic properties of the semantic constituents in the linguistic construct as dictated by

3

real-world constraints.

A natural language processing system which is designed to process normal discourse would have to, at the very least, include a real-world consensus model. This model would, in turn, have to contain aspects of ontology, physics and epistemology even if these are at a naive level. Such a model could account for a wide range of cognitive phenomena including the ability to understand and handle new situations. In order to represent this model and perform valid inferences about entities and their inter-relationships and associated constraints, suitable formalisms need to be devised in a broad linguistic framework. The current trend in semantic extraction from text seems to revolve around connectionist 'microfeature' approaches [McClelland and Kawamoto 88, Smolensky 88]. In Dik's Functional Grammar, the lexicon is used to record the semantic constraints on linguistic elements that surround verbs and nouns in each of their typical uses and each specific usage is recorded as a predicate in the lexicon. In this approach, the a priori specification of the basic lexicon entries is a distinct disadvantage. A better technique to creating the lexicon would be to extract constraints on word and phrase usage from a representative sample of linguistic expressions, build generalized structures (these would be the basic predicates), and use a set of inference rules to manipulate these structures to derive new ones. To formalize these constraint extraction and inference procedures, the real-world consensus model is essential since the model can account for both linguistic and extra-linguistic activity in a wider communicative setting where pragmatics is primary. The ability of certain neural network architectures [Kohonen 90, Ritter & Kohonen 89, Samarabandu and Jakubowicz 90] to generate semantic maps of linguistic data can be usefully employed to build the generalized structures mentioned above.

2.2 The Semiotics of American Sign Language

While there are sign languages based on languages other than English, we discuss the key features of American Sign Language (ASL) in this section only because there seem to be, at least in this author's opinion, more research articles devoted to ASL than any other sign language.

American Sign Language (ASL) is a manual-visual language produced by the hands and perceived by the eyes. ASL has a highly articulated grammar as complex and expressive as spoken language. A common process in ASL is the association of a noun with a location in space. Any subsequent references to the location is equivalent to pronominal reference in spoken language. Nouns are typically specified through sign classifiers where each classifier denotes a physical object with a specific shape. For example, the 4-CL and the bent-V classifiers represent 'some rectangular object with or without holes" and "some human or animal with bent legs", respectively. When the two classifiers are placed in spatial relationship to one another and hand motions are used, syntactic

4

and semantic information may be conveyed. For example, tracing an arc in space with the bent-V classifier over the 4-classifier and passing over it can be used to represent the construct "The cat jumped over the fence". As another example, the verb 'give' in its uninflected form ('The girl gives the book to the boy') and in its inflected forms ('The girl gives each boy a book' [exhaustive], 'The girl is giving the boy the book' [durational], etc.) and combinations thereof may be more or less accurately specified in ASL. The orientation, location and movement (or its absence) of individual classsifiers are used to determine the relations between objects [Liddell 80].

ASL incorporates both complex language structures and complex spatial relations. Space has a syntactic and topographic function in ASL. Superimposed changes in movement and spatial contouring of a sign stem convey inflectional and derivational processes in ASL. Spatial arrangements of nouns in a sentence need not correspond with spatial arrangements of signs in three-dimensional real world where sign language is being used to convey the same information. Function words (e.g. 'is', 'was') which commonly occur in written or spoken language are absent in ASL; however, pragmatic aspects such as topic and focus can be indicated via the order in which the signs are presented and the time delay between the presentation of one sign and the next. In the formulation of yes-no questions in ASL, non-manual activities such as facial expressions, body position and head position of the signer are important indicators of syntactic structure such as word order, relativization and subordination [Fischer 75, Friedman 75, Howard and Franklin 88].

The pervasive principle in ASL is the concurrent conveyance of information. In attempts to understand principles of neural organization underlying language, some researchers have tried to root language in movement control while others have tried to base it in the capacity to convey meaning through symbols. Questions which may be asked are - Is the underlying basis of left-hemisphere specialization for language tied to function rather than form ? Are left-hemisphere mechanisms in the brain, some of which are involved in sequential analysis called into play in a language which packages information in a spatial, concurrent manner [Corina et al 92]? Does right-hemisphere damage disrupt processing of this linguistic signal [Poizner et al 87, Shallice 88, Studdert-Kenndedy and Lane 80]. Neurophysiological evidence seems to suggest that there are separate verbal and nonverbal semantic systems [Coltheart et al 87, Howard 88, Warrington 75]. For example, 'optical aphasia' has been detected in hearing-impaired individuals [Beauvois 82].

2.3 Semiotics of Music
Schemes for representing music include the (now obsolete) tablature and the stave notation. In the stave notation, abstract symbols are used for representing music compositions. The shape of

the symbol and its placement on the 5-line staff signifies a musical sound of specified pitch and duration. Supplementary symbols such as sharps and flats are used to define the pitch more precisely whereas markings such as ties, dots and slurs direct the mode of performance. The notation is comprehensive enough to specify tempo or rhythm, repetition, looping back, etc. It is a two-dimensional representation which takes into account both time and space; however, in most cases, the medium on which the composition is to be performed is usually not explicitly specified. Incidentally, the parallelism inherent in the 5-line staff has been the basis of atleast one graphics animation language, DIAL [Foley at al 86].

This author is unaware of any published research which seek to explicate whether there are is any neurolinguistic basis for why the stave representation seems to capture the sound patterns of music better than other representations. One can only speculate on whether music sound patterns are processed differently, both auditorily and visually, in the brain compared to sound patterns of speech. A comparative assessment of tonal versus non-tonal languages from this perspective would be worth pursuing. Akin to dyslexia and optic aphasia, 'music aphasia' is a possibility ?

2.4 Semiotics of Other Language Systems
There are languages systems in which communication in effected through pictorial representations. In large part, the particular set of pictures employed to convey information are either culture bound or societally regulated. We shall have more to say on this in a forthcoming article.

3. Semiotics of Language Representations in Graphical Settings
If one wishes to capture the conceptual semantics of a linguistic construct in visual form to any degree of accuracy, then one must address a variety of semantic issues which arise from real-world constraints. We focus on three such issues - a) representability or non-representability of concepts, b) the semiotics of geometry and motion, and c) the choice of representation symbols.

3.1 Representability and Non-representability
We now address the issue of the representability, in graphical form, of the denotational and connotational aspects of linguistic entities. In comparing and contrasting two different natural languages, one encounters lexical, syntactic and semantic gaps. Likewise, in mapping natural language text into image form, one encouters 'visual' gaps. These are nothing more than those concepts which are expressible in natural language but which cannot be captured in visual form in an adequate manner. For example, the concepts underlying the phrases 'soft pillow' and 'tame lion' cannot be suitably represented in any simple fashion. Likewise, constructs involving aspect and mood ('She may go to the store') are quite difficult to represent.

6

3.1 Representability of Concepts

In the ensuing discussions, we use the phrase 'linguistic object' to mean any natural language construct at the lexical, syntactic, semantic or pragmatic level ('or' is used in an inclusive sense). A linguistic object falls under one of three classes of representation - a) directly representable through a primitive graphic symbol or compositions thereof, b) representable only indirectly using graphics primitives and their compositions, and 3) other. We surmise that phenomenological objects fall into the first two classes whereas noumenological objects are in the 'other' class. It is quite reasonable to assume that most objects of any syntactic or semantic category upon which a metric can be defined are directly representable. This is because such objects can be visually perceived and interpreted by the senses. Elements in the representable class of objects include animate and inanimate objects, their attributes (color, shape, size, etc.), motion, position, and direction indicators, and comparatives. Objects whose visual representations can only be achieved indirectly are those linguistic objects which cannot be perceived but can be accounted for by the other human faculties. Indirect representations require the use of literary devices such as metaphors (suggestive likeness or analogy between objects), metonyms (object description through object's attributes), similes (for comparing unlike objects) and the like. Even here a representation is possible only if a metric can be imposed on the elements defining the literary devices. For example, the concept which underlies "hot" and "hotter" can be visually represented using color gradations since color is an entity upon which a metric can be defined. However, in the metaphorical use of color, one has to be careful in the particular choice of color since many aspects of color are culture bound as has been established in several anthropological studies. For example, in Cherokee culture, 'red' symbolizes 'east', 'wounds', 'thunder', etc. while 'blue' symbolizes 'north', 'water', etc. Although descriptive attributes such as 'good' and 'bad' can be captured indirectly in visual representation, we have chosen to ignore these types of linguistic objects in this paper since highly subjective criteria are involved. This is in keeping with our philosophy of 'value neutrality' in representation schemes. Readers may seriously object to the notion of 'value neutrality' as being chimerical; indeed, they would be right.

Plain or unadorned text may be viewed as a homogeneous representation of ideas employing socially accepted signs (the code) in sequential fashion. The 'meaning' of each sign is known a priori but the 'meaning' of a composition of signs cannot be known a priori but can only follow. The intended meaning of a simple sentence such as 'I had to break my piggy bank for John's birthday' [Schank 86] cannot be established through traditional syntactic and semantic analyses unless one shares much of the utterer's world-view. If a natural language processing system is to have the capacity of 'understanding' sentences, a prime

7

requirement is that the system have incorporated in it a wealth of information on the very act of communication; such information ought to include, in addition to linguistic specifications, aspects of (naive) physics, logic, the semiotics of culture as dictated by custom, habit, consensus, beliefs, etc [Benedikt 91, Bosch et al 91, Hayes 85] . A cartesian account of space and time may be used to explicate the meaning of some, but not all, sentences involving spatial and temporal references [Habel 90, Hagiwara 91, Lang et al 91], since language allows liberties to be taken with these references; language allows speakers to refer to entities which do not exist ('The class was cancelled'); language allows self-referentiality ('This is a sentence'). It is important to consider utterances in their entirety since discourse strategy and tactics are largely dictated by the purpose of the discourse.

When text is transcribed into visual form (as exemplified by any number of elementary language learning books. For example, see [Madrigal and Dulac 62]), the symbols or signs that are used to represent the various ideas and concepts represented by individual words or word combinations are quite arbitrary and, to a large extent, reflect societal and cultural values; this is in contradistinction to language symbols which appear, to be value-neutral (exceptions include language systems based on hieroglyphics and ideograms). For example, what would be the appropriate choice of graphical symbols to be used in the visual representing the ideas conveyed by the simple sentence "The child plays with the doll" ? If the visual representation is that of a little girl in a pink dress playing with a blonde, blue-eyed baby doll, it is quite reasonable to assume that the social and cultural values of the individual creating the visual representation are being reflected in the representation. A similar remark would apply to any alternate representations of the sentence such as a male child playing with a GI Joe doll. A pertinent observation to be made here is that whereas syntactic representations of language appear to be value-neutral, visual representations do not. One of the questions we wish to pose is "Are there value-neutral symbols which currently exist or can such symbols be created so that they can be suitably employed in the visualization of the ideas conveyed by text ? ". If the answer is in the affirmative, then a prime requirement would be that such symbols (along with any nuances) be easily recognized by the user of the visual material. Stick figures of varying sizes appear to be value-neutral when they are employed as graphical elements in representing human beings of either gender and of different ages; however, a stick figure of an old man hardly captures the semantics of the phrase "old man". In mapping natural language to graphics, the linear and temporal nature of text is replaced by a planar, non-temporal representation. In many instances, the mapping contains 'visual gaps'; what was unambiguous in the original text is now rendered ambiguous. On the other hand, the visual representation may be such that it seeks to clarify what was previously ambiguous in the

text.

The concept underlying the sentence

"The girl is walking the dog"

cannot adequately be captured by a single image depicting a girl
either walking alongside a dog on a leash or walking alongside an
untethered dog since the image ignores tense aspects of the
sentence. In the case of an untethered dog, the relationship
between the two objects (the dog and the girl) can only be
inferred. In the visual representation, the spatial separation
between the girl and the dog is one of the deciding factors in
capturing the underlying semantics of the sentence. The semantics
of tense may be resolved by resorting to static multiple images or
through animation. In the case of the former, it would seem
necessary to indicate the passage of time by some means (e.g. a
clock). Note that in the visual mapping of a rather
straightforward sentence with unambiguous semantics such as the
sentence presented above, the choice of representation introduces
semantic aspects not present in the original sentence. We label
this the 'metaphrastic' problem. This problem cannot be adequately
resolved unless some basis for the equivalence of representations
in the two media, text and graphics, can be established. A similar
problem arises even if the same medium (e.g. text) is used; the
problems faced in translating constructs in one natural language to
constructs in another are quite well-known. Similar ambiguities
have also been shown to occur in mapping spoken language into ASL.

3.2 The Semiotics of Geometries and Motion

In animating text, the particular choices of graphical objects to
represent the various semantic constituents of the text is
necessarily highly constrained. The constraints are brought about
by an awareness of the euclidean geometry of the diverse physical
objects [Bellugi 80, Poizner 81] and the two-dimensional plane of
representation (viz., the visual display unit in a computer
environment), the kinetics and kinematics of motion, and the
limitations of human faculties, especially vision. Similar
constraints apply when one attempts to map text (speech) into
commands for manipulating electromechanical devices such as robots
in the three-dimensional physical world. In the blocks-world of
Winston [Winston 84], the epistemological issues surrounding the
geometries of the blocks play an important role in determining when
an arch is not an arch; however, motion is not an aspect of the
representation which has been discussed. It is quite well-known
that shape and motion interact in unexpected ways and that shape
can be inferred from motion. See, for example, works on ethology
[Tinbergen 59] and computer vision [Marr 82] for two very different
interpretations of this aspect.

Similar interactions occur in a linguistic construct; for example,
the particular choice of prepositions in a sentence involving

9

physical objects is automatically restricted by the geometries of the objects. Consider, for example, the two sentences

'He hid under the bed'
and   'He hid under the tree'.

In both of these sentences, the preposition 'under' implies that the spatial properties of the objects are such that they permit the concealment of the man. However, the everyday interpretation of the objects 'bed' and 'tree' would tell us that only the first sentence makes sense. Analogous remarks apply to the sentence pair

'He stood under the tree' / 'He stood under the bed'.

According to Webster's dictionary, the primary definition of the word 'under' is as follows: "under \prep\ - below and in such a position as to be overhung, surmounted, covered, protected, or concealed by". Additional information is necessary to establish which one of these word senses is applicable. The implication here is that the semantics of prepositions is dictated by the real-world semantics of other lexical elements which occur in linguistic constructs.

In a visual representation of the motion of an object in space, even under the simplest of assumptions, primary issues of concern are the physics and the semantics of the motion. For example, in the sentence

"Bill ran down the street to greet Tom",

(In Dik's FG theory [Dik 78], this is an 'action' sentence) the physics of the motion relate to the rate or speed at which the motion occurred whereas the semantics of the motion relate to the duration, direction and location of motion. Note that the duration is implicitly specified in the sentence. If the words 'ran', 'down' and 'greet' were to be replaced in the sample sentence by 'walked', 'up' and 'meet' respectively, the physics and semantics of motion would still be the same.

However, in the sentence

"Bill waited on the street corner to meet Tom",

the physics of the motion is the same as before but the semantics of motion is different. In Dik's FG terminology, the sentence is an example of a 'position' sentence. It should be pointed out that, in the above examples, one intransitive verb has been replaced by another (the word order remains the same) causing the semantics of motion and the sentence typology to change.

4. The Need for Special Formalisms and a Proposed Formalism

Most natural language constructs may be viewed as nothing more than one-dimensional representations of the concepts underlying the constructs. However, the concepts themselves need to be viewed as objects which involve the three dimensions of euclidean space and the dimension of time. When the construct is mapped into some syntactico-semantic representation, spatial and temporal aspects are ignored since the original construct does not contain these elements. By including this information in an appropriate representation scheme, the 'hidden' semantics of the construct can be effectively specified. The information in this new representation can now be effectively mapped into media other than natural language using a variety of transformations.

4.1 The Need for a Formalism

Take, for example, the sentence

"The teacher walks into the classroom and writes on the board" .

In a visual representation of this sentence, let us implicitly assume that a typical classroom scene is presented where the image includes a doorway, chalkboard, teacher's desk, and students sitting at their desks. In capturing the actions of the teacher, we would need to show an unobscured sideview of the teacher, the figure of the teacher partially hidden by the students and their desks as he/she makes his/her way into the classroom, the teacher with his/her back to the students as he/she starts to write on the board (the initial position of the teacher with respect to the board is pertinent in this context), the occlusion effects as the teacher continues writing on the board, etc. In order to achieve these effects, the geometry of each animate or inanimate object (both dynamic and static objects are present in the given example) and its static and dynamic relationship with all other objects in the scene would need to be taken into account. A detailed semantic description of the sentence would require that each semantic constituent (explicit or implicit) in the sentence be specified or at least be derivable; moreover, any existing inter-relationships among these would also need to be specified or derived. Semantic, case-based, syntactic and functional grammar specifications of language only account for elements explicitly occurring in linguistic constructs. As a result, none of these specifications is adequate to properly represent text in image form.

4.2 A Proposed Formalism

The representation scheme that we have devised is loosely patterned after Dik's Functional Grammar (FG). FG has been chosen as the basis for developing our formalisms for several reasons - a) pragmatics is an essential, if not the primary, component of FG, b) FG contains many language-independent features, c) FG is applicable to the analysis of both written and spoken text, and c) FG techniques may, in principle, be applied at the word, phrase,

11

sentence and discourse levels. While Joshi's tree-adjoining grammars [Joshi and Levy 82] and Schank's conceptual dependency structures [Schank 75, 86] are also good candidates for the intiial representation of a language construct, they would need to be substantially expanded and, possibly, revised for our purposes.

In FG, sentences are classified as belonging to one of four types - action, process, state and position. Action and position sentences involve animate objects whereas process and state sentences typically involve inanimate objects. In our reformulation of Dik's basic sentence typology, we allow a sentence to be transformed from one type to another by imposing selectional restrictions on the verbs in the sentence. The extended semantic markings of the new sentence will reflect the newly derived semantics. In our representation, we introduce compositions of the above four basic types; these compositions, in turn, are classified on the basis of real-world considerations such as physical geometry, motion dynamics, etc. Furthermore, whereas Dik's FG theory mainly revolves around nouns and verbs, our representation schemes rely primarily on adjectives, adverbs and prepositions since these are the entities which introduce the semantics of place, time, manner, shape, color, size, etc. In all fairness, it should be remarked that in a more recent work co-authored by Dik, his lexicon has been expanded; predicate frames are now associated with meaning postulates, definitions and paradigms which are non-derivable [Dik and Kahrel 92].

In our knowledge representation scheme, word categorization is based on the conceptual semantics of words, both in isolation (a priori semantics) and in context (a fortiori semantics). In our scheme, concrete nouns are identified with graphically realizable objects (e.g., 'sun', 'book', 'dog') while other types of nouns are identified with metaphorically realizable graphical objects (e.g., 'sunlight'). Verbs define the physical placement and positioning of these objects in a visual scene (e.g., 'The sun is setting'). Adjectives, adverbs and prepositions typically either introduce new semantics into, or clarify the existing semantics of, the objects which participate in the scene (e.g., 'The light brown dog sits down in front of the TV'). In a graphical representation of this sentence, the implied spatial and temporal semantics of the two nouns will need to be taken into account. The preposition introduces hidden surface aspects into the scene while the adverb introduces motion semantics. Even though only the color of the dog has been specified by the adjectives, in the visual scene, 'contrast' is implied. As an aside, the ISSC-NBS system (the Munsell scheme) of employing values for hue, lightness, and saturation to specify any one of 267 possible colors is quite adequate to depict color descriptives such as 'shiny', 'brilliant', etc. if dictionary definitions of acceptable usage of these terms are followed.

In figure 1 a schematic of a physical-world realization of a mental

concept is presented. A natural language construct is analyzed using Dik's Functional Grammar and a representation is obtained (representation A). Since this representation does not capture the implied time and space semantics of the underlying construct, these additional semantics are incorporated into it using an extension of functional grammar (Extended Functional Grammar - EFG). In FG, satellites are used in constructing nuclear predicates from basic predicates [Dik 78];consequently, they play a secondary role in his scheme. In EFG, satellites are primary in that they specify lexical items which have defining roles in the semantic resolution of a linguistic construct. The EFG representation, labelled representation B in the figure, is then transformed into representation G using graphical functional grammar (GFG). GFG is a set of transformation rules which permit the specification of three-dimensional spatial semantics in a two-dimensional framework. These rules allow one to capture of the spatial and temporal interrelationships among objects (which occur in the language construct) using relative position and relative time information. The Z-buffer algorithm, well known in graphics programming [Foley and Van Dam 84], is used to superimpose objects with lower z-coordinate values onto objects with higher z-coordinate values. This superposition of objects is what enables hidden surfaces to be identified. Representation G is then mapped into one or more static images using graphical primitives. These images, in turn, may be used in an animation graphics environment to convey concepts involving time such as physical motion, tense, etc. One may choose, however, to map the information contained in representation B directly onto a system which is capable of generating commands for manipulating robots. Pragmatic functions such as FOCUS or TOPIC which occur in Dik's theory are easily mapped into a graphics environment through the use of highlighting of the relevant object or objects. This is especially useful in an animation scene.

Our formalism revolves around canonical predicates whose arguments have syntactic, semantic and pragmatic functions. The canonical predicates can be modified by satellites which are annotated by real-world semantics. These satellites dictate how and in what order functions may be assigned to the different arguments in the canonical predicate. The satellites, which are also canonical predicates, may be influenced by other satellites as well. The basic embedded predicate structure is as follows:

predicate (argument list) [satellite predicate list]
(predicate: predicate descriptor list)
(argument: argument descriptor list)
(satellite predicate: (predicate: predicate descriptor list))

Predicate descriptors are drawn from syntax lists (noun, verb, preposition, adjective, adverb, subject, object, etc.), semantic lists (agent, recipient, beneficiary, goal, location, time, etc.), pragmatics lists (topic, focus, theme, etc.), and real-world semantics list (reference position, reference time, action,

position, process, state, constraints, etc.).

Webster's dictionary defines a preposition to be a linguistic form which combines with a noun, pronoun, or noun equivalent to form a phrase that typically has an adverbial, adjectival, or substantival relation to some other word. Every preposition has an object and the typical relationships between a preposition's object and some other word in a sentence include actual space and time ('above', 'from', 'beyond', 'during', 'since'), metaphorical space or time ('beyond comprehension', 'from his experience'), exclusion ('but'), and cause ('due to'). While there are only about one hundred prepositions in English, since they are used very frequently in everyday speech and since their semantic content is very high, they cannot be viewed as mere fillers in sentence structures, irrespective of whether these structures are grammatical or not. This is especially so given our stated goal of obtaining formalisms suitable for the understanding of natural language by automated means such as robots in a real-world setting. For a treatment of prepositions as denotata of spatial cognition, see [Habel 90, Herskovits 86, Vandeloise 86, Wexelblat 91].

In this paper, we restrict our attention to prepositions which involve actual space and time. For example, in the sentence

   "After school, the student talked to the coach of the soccer
   team in front of the library"

the simple prepositions ('after', 'to', 'of') and the phrasal preposition ('in front of') which occur in the sentence show the spatio-temporal relationship that exists among the prepositions' objects ('school', 'coach', 'soccer team', 'library').

Adjectives are traditionally defined as words that modify nouns or pronouns by defining, describing, limiting, or qualifying those nouns or pronouns. Many adjectives can also be defined by their form (based on typical endings) or by the positions they occupy in a sentence. A majority of adjectives may be categorized as descriptive adjectives ('yellow house', 'large dog', 'angry man' etc.) which specify a quality or state of the noun or pronoun they modify. Other classes of adjectives include cardinal ('two birds') and ordinal numbers ('second bird'), definite ('the room') and indefinite (a picture') articles, and possessive ('his book'), demonstrative ('those boxes'), and indefinite (''any book') pronouns which are used to modify nouns. In obtaining the formalisms for predicate structures representing adjectives, we restrict our attention to those adjectives upon which metrics can be imposed so as to be able to capture the semantics involving comparisons, space, time, quantity, and color. For conceptual interpretations of dimensional adjectives, see [Bierwisch and Lang 89].

Adverbs are typically used to express some relation of manner or

quality ('hard', 'slowly', 'well'), place ('here', 'somewhere'), time ('now', 'later'), direction ('away', 'forward'), frequency ('often', 'sometimes'), affirmation or denial ('yes', 'no') and degree ('very'). They serve as modifiers for verbs, other adverbs, adjectives, prepositions, phrases, clauses and sentences. In this work, we focus on adverbs which specify relations of manner/quality, place and time and which act as modifiers for verbs, adjectives and prepositions. Once again, we are only interested in the spatio-temporal aspects of adverbial lexical elements in language constructs. Of particular import are prepositional adverbs ('out', 'on', 'off', etc.) since most of these express spatial relationships of one kind or another. The adverbs in a construct typically supply answers to such questions as 'How' (manner), 'When' (time or frequency) and 'Where' (place).

We have chosen to ignore, at least in this article, three other major parts of speech, namely, conjunctions, interjections and pronouns for the following reasons. Many conjunctions ('but', 'after', 'for') are identical in form with prepositions and only their function in the sentence identifies them as conjunctions. Consequently, remarks made earlier on the status of prepositions apply here as well. Interjections are grammatically independent words that express emotion.; emotive aspects of language and their representations will be dealt with in a subsequent article. Since pronouns traditionally substitute for nouns, a discussion of the latter would, for the most part, subsume any discussions of the former. A discussion of nouns has already been presented in earlier in this section.

As part of the real-world semantics list, a set of generalized constraints is also used. It is these constraints which permit the resolution of whether certain categories of adjectivals , adverbials, and prepositions can meaningfully occur in a linguistic construct once the objects (specified through nouns and their associated spatial semantics) and their motility aspects (specified through verbs and their associated spatio-temporal semantics) are available. In a visual graphics representation of a construct, spatial constraints are used to compute the relative placement of physical objects which occur, either explicitly or by (metaphorical) implication, in the construct. Likewise, in animated graphics representation, the spatio-temporal constraints play a major role. The generalized constraints should be viewed as mathematical operations which account for notions of inclusion, exclusion, precedence, hierarchy, greater than, equal to, less than, etc. An argument can be advanced that the set of constraints that occurs in Chomsky's Government and Binding Theory rears its head here too, albeit in different form.

Terms are associated with predicates and arguments and these terms are drawn from a lexicon which also contains the canonical predicate structures and satellite predicate structures.

15

In the rest of the paper, the following abbreviations are used in annotating the various predicate structures:
C canonical; V verb; N noun; ADJ adjective; ADV adverb; PREP preposition; CON conjunction; CN and CV represent canonical nouns and verbs, respectively; respectively; SUB subject, OBJ object; REC recipient; LOC location; and REFPOINT reference point.

A typical annotated predicate structure is shown below.

```
motion ( motile_object ( animate_object ) ( rate (t) ))
    CN             CV            CN              CN
                                 SUB
                ACTION         AGENT
```

The above structure may be used to represent any of several sentences including

         "The dog walks",
         "The boy runs fast",
         "The two women walked slowly"
and   "The cat ran".

It may not be used to represent the sentence

         "The car drove away quickly"

since an inanimate object is involved nor can it be used to represent the sentence

         "The cat walked across the street"

since no 'goal' or syntactic 'object' is specified in the predicate structure.

An appropriate structure to represent

         "The cat walked across the street"

would be the following:

```
motion(motile_object (animate_object, inanimate_object)
    CN             CV            CN              CN
                                 SUB             OBJ
                ACTION         AGENT             GOAL
                                                 LOC

                         (rate(t), position))
                            CN        CN
```

The above structure may also be used to represent

         "The dog climbed over the wall".

16

In the two structures presented above, the time argument, given by 't', for the rate predicate allows us to make the nature of the motion quite specific. Large values of t are representative of a whole class of motion verbs and adverbials whose semantics indicate heightened activity, and mutatis mutandis, for small values of t. For example, 'walk quickly/slowly, 'run', 'run fast', etc. Of course, default values for t will need to be specified. Note that when we are dealing with 'action' type sentences and 'position' type sentences, the particular choices for the values of t in the rate predicate can result in the sentence typology to change because a dynamic setting can be rendered static and vice versa.

The position argument has been introduced to account for a whole class of prepositions (across, on, over, etc.) which occur in a phenomenological context. It is also useful in constructions involving comparatives, as in 'taller than', and 'nearer'.

The predicate structure for noun classes has arguments which account for shape, color, size, etc. For example

```
inanimate_obj(state-verb(physical_object()(color)(size) (shape))
        CN              CV              N      ADJ    ADJ    ADJ
                                        SUB
                                        AGENT
                        STATE
```

may be used to represent

                "The big blue flower is round"

and similar types of constructs. Note that whole classes of adjectivals are specifiable in this manner in the satellites associated with a canonical object class.

5. Examples

To bring out the differences in the various representations for a given natural language construct, we use the following example "John gave the red book to Mary in front of the store last night". The phrases 'red book' and 'in front of' have been underlined to indicate that they are the linguistic objects of focus.

Functional Grammar Representation A:

                                                satellites

```
{give [(John(x1), book(x2), Mary(x3))(store(x4))(night(x5)]}
 VERB  NOUN       NOUN       NOUN       NOUN       NOUN
        SUB                  OBJ
        AGENT      GOAL       RECIPIENT LOCATION   TIME
                                                            ACTION
                  FOCUS
```

17

```
(x2: book(x2)) (red(y1))        satellite
     NOUN          ADJ
                             GOAL
```

In Extended Functional Grammar, physical objects which occur in a
construct, either explicitly or implicitly, are associated with
space and time coordinates. These coordinates are assigned values
based on the state of affairs of the sentence under analysis.

<u>Extended FG Representation B</u>:

```
                                              satellites

{give [(John(x1),  book(x2),  Mary(x3))(store(x4))(night(x5)]}
 VERB    NOUN        NOUN        NOUN        NOUN        NOUN
         AGENT       GOAL      RECIPIENT  LOCATION     TIME
                                                                ACTION
                    FOCUS

[(x1: John(x1))(x, y, z, t)]
      NOUN        POS
                             AGENT

{[(x2: book(x2))(red(x2) (x, y, z))]}       satellites
       NOUN          COLOR      POS
                                        GOAL

                                FOCUS
[(x3: Mary(x3))(x, y, z, t)]               satellite
      NOUN        POS
                        REC

[(x4: store(x3))(x, y, z)(y1: front)]      satellites
      NOUN        POS            PREP
                                       LOC
                                FOCUS

[(x5: night(x5))(x, y, z, t)]              satellite
      NOUN        POS
                        TIME
```

Note that in a visual depiction of 'night', if a moon is the symbol
used, its position needs to be specified. This is why the word
'night' has been associated with space and time coordinates.

<u>Graphical Functional Grammar Representation G:</u>

```
                                                        satellites

{give  [(John(x1),  book(x2),  Mary(x3))(store(x4))(night(x5)]}
 VERB   NOUN        NOUN        NOUN        NOUN        NOUN
        AGENT       GOAL        REC         LOC         TIME
                    FOCUS

                                                             ACTION

[(x1: John(x1))(x, y, z, t)]   (x0, y0, t0)        satellites
    NOUN        POS
                        AGENT       REFPOINT


{[(x2: book(x2))(red(x2) (x, y))]} (x0, y0, t0)    satellites
    NOUN        COLOR      POS        REFPOINT
                        GOAL
                        FOCUS

[(x3: Mary(x3))(x, y, t)]   (x0, y0, t0)          satellites
    NOUN        POS              REFPOINT
                        REC

[(x4: store(x3))(x, y)(y1: front)] (x0, y0)       satellites
    NOUN        POS        PREP        REFPOINT
                                LOC
                        FOCUS

[(x5: night(x5))(x, y, t)]   (x0, y0, t0)         satellite
    NOUN        POS              REFPOINT
                        TIME
```

Note that, in the GFG representation, the specification of a reference point (with reference to space and time) has a two-fold purpose - a) the positioning of each object relative to the other objects is uniquely specified, and b) whatever object has been specified as the FOCUS of the construct can be made the reference point. Simple comparisons of the numerical values of the x, y, and z coordinates allows us to capture the semantics of words such as 'left', 'right, 'wider' [x-dimension], 'up', 'down', 'over', 'under', 'taller' [y-dimension], 'front', 'behind', [z-dimension], etc. The set of generalized constraint rules (see section 4.2) is used to ensure the validity of a given visual representation.

As a second example, consider the sentence, "The blue car runs faster than the red car". For this 'process' type sentence, the GFG representation is shown below.

```
run [ car(x1) blue(x1) ] (y1: fast(y1) (z1: car(z1) red(z1))
VERB NOUN   ADJ      AGENT    ADV              NOUN     ADJ
                     SUB                                        OBJ
                  PROCESS                                     MANNER

[x1: car(x1)(x, y, t)(rate(r1))]
                              SUB
                              AGENT

[z1: car(z1)(x, y, t)(rate(r2))]
                              OBJ
```

When the GFG contents are mapped into a static graphic scene, the
objects in the scene will be the animate and inanimate physical
objects which occur in the GFG representation. In a dynamic scene,
those objects which have motion aspects specified in their
satellites will be impacted.

One may easily represent the syntactic structure of the sentence
"The blue robot lifts the hammer slowly" in graphical form by using
a parse tree whose nodes have syntactic and lexical labels. If one
wishes to represent the semantics of the sentence using graphics,
then one would have to resort to depicting the underlying semantics
of the sentence via graphical entities. In such an environment,
mapping the semantics of the sentence requires that one be
cognizant of both the static and dynamic properties of the semantic
constituents in the linguistic construct as dictated by real-world
constraints. Since this information in our representation scheme,
the 'hidden' semantics of the construct can be effectively
translated into commands for manipulating the robots. This is
discussed in detail in the next section.

6. Semiotics of Language Representations in Robotic Settings
Current research in robotics and machine intelligence is mostly
focused on vision algorithms for object identification, motion
facilitation and motion planning. A diversity of approaches and
techniques are employed in this context. Some of the approaches
include a) visual cue extraction for robot navigation and obstacle
avoidance, b) design of complex sensor systems for exploratory
perception- and motion-based visual determination of navigational
obstacles, c) machine learning and geometric reasoning about
spatial relationships between parts of an object in the presence of
uncertainty, d) providing qualitative descriptions and models of
relationships between motion and induced change in three-
dimensional appearance employing aspect space (cross product of
image plane and viewpoint space), e) simultaneous tracking of
objects in position and depth, and f) employing thermal and visual
imagery to extract internal object properties to serve as
physically meaningful features for region labeling.

Very little research has been done on how to devise and link

knowledge representation (KR) structures of natural language with real-world activity (as in robot limb manipulation and machine vision). In our research, we tend to view KR structures as constituting models in their own right and not as entities which are devised on an ad hoc basis. Consequently, a knowledge representation structure is a model which meets several requirements: a) the structure exhibits the characteristics of a system composed of several elements, none of which can undergo a change without effecting changes in all the other elements, b) for a given model, the possibility of ordering a series of transformations should result in a group of models of the same type, c) the possibility of predicting how the model will react if any of its elements is subject to modifications, d) the constituents of the model are such that all observed facts are made immediately intelligible.

Convincing arguments may be advanced that visual semantic representations of linguistic constructs such as "Locate the red switch in the electrical panel and turn it off" are possible only if the linguistic structures used to encode the information contained in the descriptions are capable of representing both the semantics embedded in the text/speech and the extra-textual real-world semantics.

In the above example, natural language comprehension is involved. The robot operator would need to address a variety of issues including what an object looks like, in isolation and as part of some other object or objects, what an object's components are, what is the purpose of each object, individually and as part of a larger ensemble of objects, what are the temporal and spatial aspects of objects, etc. Commands to the robot using natural language (assume the robot has vision, tactile and other sensory devices attached to it) would require that the robot be capable of a) identifying objects in three-dimensional space, their absolute positions (for a given reference point) and their positions relative to other objects in the same visual scene, and b) performing causal, spatial and temporal inferencing. Many robot motion algorithms and vision algorithms have been developed for the above purposes; however, these do not have any linguistic components associated with them.

To represent these types of information on the screen and in the robot's memory, the real-world semantics of a diverse set of linguistic objects would need to be properly represented. The formalisms we have proposed serve this purpose. We present two simple examples to try to convince the reader that this is the case.

i)   "The blue robot lifts the hammer slowly"

```
lift((robot, blue), (hammer)) [position(Agent), position(Goal)]
                                                    [rate(t)]
  V     N    Adj      N
            Agent    Goal

  Action
```

<u>Robot Control Script</u>

```
     robot  (Agent)    simple physical object
            color      attribute
     hammer (Goal)     simple physical object
     lift   (Action)   verb of motion
                       Move Agent.arm to position(Goal)
                       Move Agent.arm at rate r1
```

ii)   "The robot removes the bomb from the middle of the street"

```
{remove [((robot(x1)), (bomb(x2))) (street(x3))]}
   V          N            N            N
          AGENT          GOAL          LOC
                                               ACTION

[x1: robot(x1) (x, y, z, t)]
      N               POS
                      AGENT

[x2: bomb(x2) (x, y, z)]
      N            POS
                   GOAL

[x3: street(x3) (x,y,z) (y1: middle)]
      N            POS      PREP
                            LOC
```

<u>Robot Control Script</u>

```
     robot   Agent      complex physical object
     bomb    Goal       physical object
     street  Location   physical object
     remove  Action     verb of motion
                        Move Agent to street (Location of
                                                  Goal)
                        Move Agent.arm to position(Goal)
                        Move Agent to street (Location -
                                                  other)
```

22

7. Software System Architecture

A software system architecture for a robot-world realization of voice commands is shown in figure 2. The voice command (natural language construct) is analyzed using Dik's Functional Grammar and an FG representation is obtained. Since this representation does not capture the implied time and space semantics of the underlying construct, these additional semantics are incorporated into it using an extension of functional grammar (Extended Functional Grammar - EFG). The EFG representation is then transformed into a partially filled representation RFG using robotic functional grammar (RFG). RFG is a set of transformation rules which permit the specification of three-dimensional spatial and temporal semantics. The partially filled representation RFG is completely filled by input obtained from sensory data and data provided by vision and motion algorithms after this input is checked against the lexicon and the set of generalized constraint rules. It is then mapped into one or more commands using robotic action primitives. We hope to be able to create a significant portion of the lexicon automatically by processing speech and text samples using an associative neural network [Ritter and Kohonen 89] or similar learning nets [Oosthuizen 89, Powers and Turk 89]. Such nets are capable of classifying lexical items and semantic structures (words and phrases) using contextual information. An example is presented in figures 3a, 3b and 3c. These figures have been taken from [Ritter and Kohonen 89].

8. Conclusions

In semantic representations of natural language constructs, the semantics of the real-world entities denoted (and connoted) by the lexical units occurring in a language construct are usually ignored. We have presented a representation scheme which allows these important extended semantics aspects to be incorporated into the representation using a variant of Dik's FG. The information contained in these representations may be mapped into a graphics environment quite directly. They may also be directly mapped into commands for manipulating robots in a real-world setting. It would be interesting to study whether there is any relationship between the movement rules used in transformational grammars (Chomsky) or tree adjoining grammars (Joshi) and the rules governing the motion of objects in a physical world setting.

Bibliography

Abe, Y., K. Hatasa and J.R. Cowan 88, "Investigating universals of sentence complexity" in M. Hammond, E. Moravcsik and J. Wirth (eds.) Studies in Linguistic Typology, Philadelphia:John Benjamins.

Beauvois, M-F. 1982, "Optic aphasia: a process of interaction between vision and language", Philosophical Transactions of the Royal Society of London, vol. B298: 35-48.

Bellugi, U. and M. Studdert-Kennedy (eds.) 1980, Signed and Spoken Language: Biological Constraints on Linguistic Form, Deerfield Beach, Florida: Verlag Chemie, 115-140.

Benedikt, M. 1991, "Cyberspace: Some Proposals" in M.Benedikt (ed.): Cyberspace, Cambridge, Massachusetts: MIT Press.

Bertin, J. 1983, Semiology of Graphics, Madison, Wisconsin: University of Wisconsin Press.

Bierwisch, M. and E. Lang (eds.) 1989, Dimensional Adjectives: Grammatical Structure and Conceptual Interpretation, New York: Springer Verlag.

Bosch, P., C.R. Rollinger, and R. Studer (eds.) 1991, Text Understanding in LILOG: Integrating Computational Linguistics and Artificial Intelligence, New York: Springer Verlag.

Buchler, J. (ed.) 1955, Philosophical Writings of Peirce, New York: Dover Publications.

Corina, D.P., J. Vaid and U. Bellugi 92, "The Linguistic Basis for Left Hemisphere Specialization", Science, 255: 1258-1260.

Coltheart, M., K.E. Patterson and J.C. Marshall 1987, Deep Dyslexia, London: Routledge Kegan Paul.

Dik, S. 1978, Functional Grammar, Amsterdam: North Holland.

Dik, S. 1980, Studies in Functional Grammar, New York: Academic.

Dik, S.C. and P. Kahrel 1992, "ProfGlot: a multi-lingual natural language processor", Working Papers in Functional Grammar, wpfg no. 45, University of Amsterdam Press.

Eco, U. 1979, A Theory of Semiotics, Bloomingon, Indiana: Indiana University Press.

Eco, U., M. Santambrogio and P. Violi 1988, Meaning and Mental Representation, Bloomington, Indiana: Indiana University Press.

Ehrlich, K. and P.N. Johnson-Laird 1982, "Spatial descriptions and referential continuity", Journal of Verbal Learning and Verbal Behavior, 21: 296-306.

Ekman, P. and W.V. Friesen 1975, Unmasking the Face, Englewood Cliffs, New Jersey: Prentice Hall.

Farah, M.J. 1988, "Is Visual Imagery really Visual ? Overlooked evidence from Neuropsychology", Psychological Review vol. 95, no. 3: 307-317.

Fischer, S. 1975, "Inferences on word-order change in ASL", in C.N. Li (ed.) Word Order and Word-order Change, Austin, Texas: University of Texas Press.

Friedman, L. A. 1975, "Space, time and person reference in ASL", Language, 51: 940-961.

Foley, J.D and A. Van Dam 1984, Fundamentals of Interactive Computer Graphics, Reading, Massachusetts: Addison Wesley.

Habel, C. 1990, "Propositional and Depictorial Representations of Spatial Knowledge: The case of Path Concepts", in R. Studer (ed.),

24

Natural Language and Logic. Lecure Notes in Artificial Intelligence, 459: 94-117, Berlin-Heidelberg-New York: Springer Verlag.

Hagiwara, M. 1991, "Self-organizing neural network for spatio-temporal patterns", International Joint Conference on Neural Networks, Seattle, July, IJCNN-91, 2: 521-524.

Hayes, P. 1985, "The second naive Physics Manifesto" in J.R. Hobbs and R.C. Moore (eds.) Formal Theories of the Commonsense World, 1-36, Norwood, New Jersey, USA: Ablex.

Herskovits, A. 1986, Language and Spatial Cognition: An Interdisciplinary Study of Prepositions in English, Cambridge, England: Cambridge University Press.

Howard, D. and S. Franklin 1988, Missing the Meaning ?, Cambridge, Massachusetts: MIT Press.

Joshi, A. and L.S. Levy 1982, "Phrase Structure Trees Bear More More Fruit Than You Would Have Thought", American Journal of Computational Linguistics, 8: 1-11.

Kohonen, T. 1990, "The self-organizing map", Proceedings of the IEEE, vol. 78, 9: 1464-1480.

J. Kulas, J.H. Fetzer and T.L. Rankin 1988, Philosophy, Language and AI, Kluwer Press.

Lang, E., K-u Carstenson and G. Simmons 1991, Modelling Spatial Knowledge on a Linguistic Basis - Theory, Prototype and Integration", Lecture Notes in Artificial Intelligence, no. 481, Subseries of Lecture Notes in Computer Science, New York: Springer Verlag.

Liddell, S.C. 1980, American Sign Language Syntax, New York: Mouton.

Loomis, J.H., H. Poizner, U. Bellugi, A. Blakemore and J. Hollerbach 1983, "Computer graphic modeling of American Sign Language", Computer Graphics, 17: 105-114.

Madrigal, M. and C. Dulac 1962, See it and say it in French, New York: New American Library.

Magnenat-Thalman, N. and D. Thalman 1985, Image Synthesis, New York: Springer Verlag.

Marr, D. 1982, Vision, San Francisco: Freeman.

McClelland, J.L. et al 1988, Parallel Distributed Processing: Explorations in the Microstructure of Cognition, volumes 1 and 2, Cambridge, Massachusetts: MIT Press.

McClelland, J.L. and A.H. Kawamoto 1988, "Mechanisms of sentence processing: assigning roles to constituents of sentences" in McClelland et al (eds.).

Nakagawa, H. and T. Mori 1988, "A parser based on the connectionist model", Proceedings of the 12th International Conference on Computational Linguistics, Budapest, August, 454-458.

Newell, A. and H. Simon 1972, Human Problem Solving, Englewood Cliffs, New Jersey: Prentice Hall.

Oosthuizen, G.D. 1989, "Machine learning: A mathematical framework for neural network symbolic and genetics-based algorithm", Proceedings of the 3rd International Conference on Genetic Algorithms, San Diego, 385-390.

Poizner, H. 1981, "Visual and phonetic encoding of movement:

Evidence from American Sign Language", Science, 212: 691-693.
Poizner, H., E.S. Klima and U. Bellugi 1987, What the Hands Reveal about the Brain, Cambridge, Massachusetts: MIT Press.
Powers, D.M.W. and C.C.R. Turk 1989, Machine Learning of Natural Language, London: Springer Verlag.
Ritter, R. and T. Kohonen 1989, "Self-organizing semantic maps", Biological Cybernetics, 61: 241-254.
Samarabandu, J.K. and O.E. Jakubowicz 1990, "Principles of sequential feature maps in multi-level problems", Proceedings of the International Joint Conference on Neural Networks, IJCNN-90, Washington DC, 683-686.
Sampath, G. 1991, "An animated graphics environment for language tutoring using a natural-language-graphics extension", Proceedings of the 1991 Long Island Conference on Artificial Intelligence and Graphics, March, 74-80.
Schank, R. 1975, Conceptual Information Processing, Amsterdam: North Holland.
Schank, R.C. 1986, "Language and Memory" in B.J. Grosz, K.S. Jones and B.L. Webber (eds.) Readings in natural language processing (eds.), Los Altos, California: Morgan Kaufman.
Shallice, T. 1988, From Neuropsychology to Mental Structures, Cambridge, England: Cambridge University Press.
Smolensky, P. 1988, "Information Processing in dynamical systems: foundations of harmony theory" in McClelland et al (eds.).
Studdert-Kennedy, M. and H. Lane 1980, "Clues from the differences between signed and spoken language", in U.Bellugi and M. Studdert-Kennedy (eds.) Signed and Spoken Language: Biological Constraints on Linguistic Form, Deerfield Beach, Florida: Verlag Chemie, 29-39.
Tinbergen, N. 1959, Animals in the World, Cambridge, England: Cambridge University Press.
Tufte, E. 1984, Envisioning Information, Cheshire Connecticut: Graphics Press.
Vandeloise, C. 1986, "L'espace en Francais. Semantique des Prepositions Spatiales", Paris: Editions du Seuil.
Warren, J. and A.L. Glass 87, "Context and Distance-to-Disambiguation Effects in Ambiguity Resolution: Evidence from Grammaticality Judgments of Garden Path Sentences", Journal of Memory and Learning, 26:714-736.
Warrington, E.K. 1975, "Selective impairment of semantic memory", Quarterly Journal of Experimental Psychology, 27: 635-657.
Weischedel, R., W.M. Voge and M. James 1978, "An artificial intelligence approach to language instruction", Artificial Intelligence, 10: 225-240.
Wenger, E. 1987, Artificial Systems and Tutoring Systems: Computational Approaches to the Communication of Knowledge, Los Altos, California: Morgan Kaufman.
Wexelblat, A. 1991, "Giving meaning to places: semantic spaces" in M. Benedikt (ed.).
Winston, P.H. 1984, Artificial Intelligence, Reading, Massachusetts: Addison Wesley.
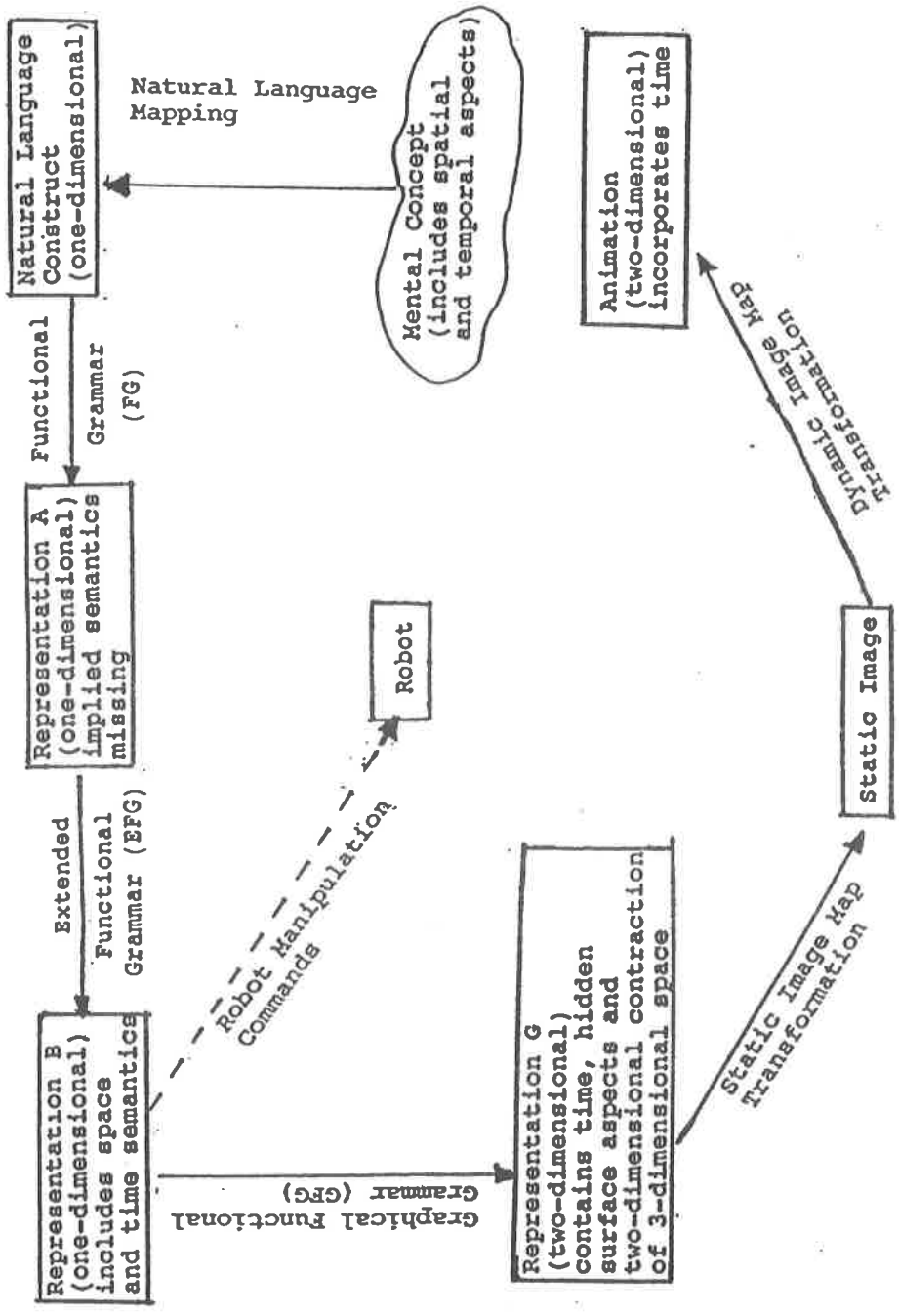
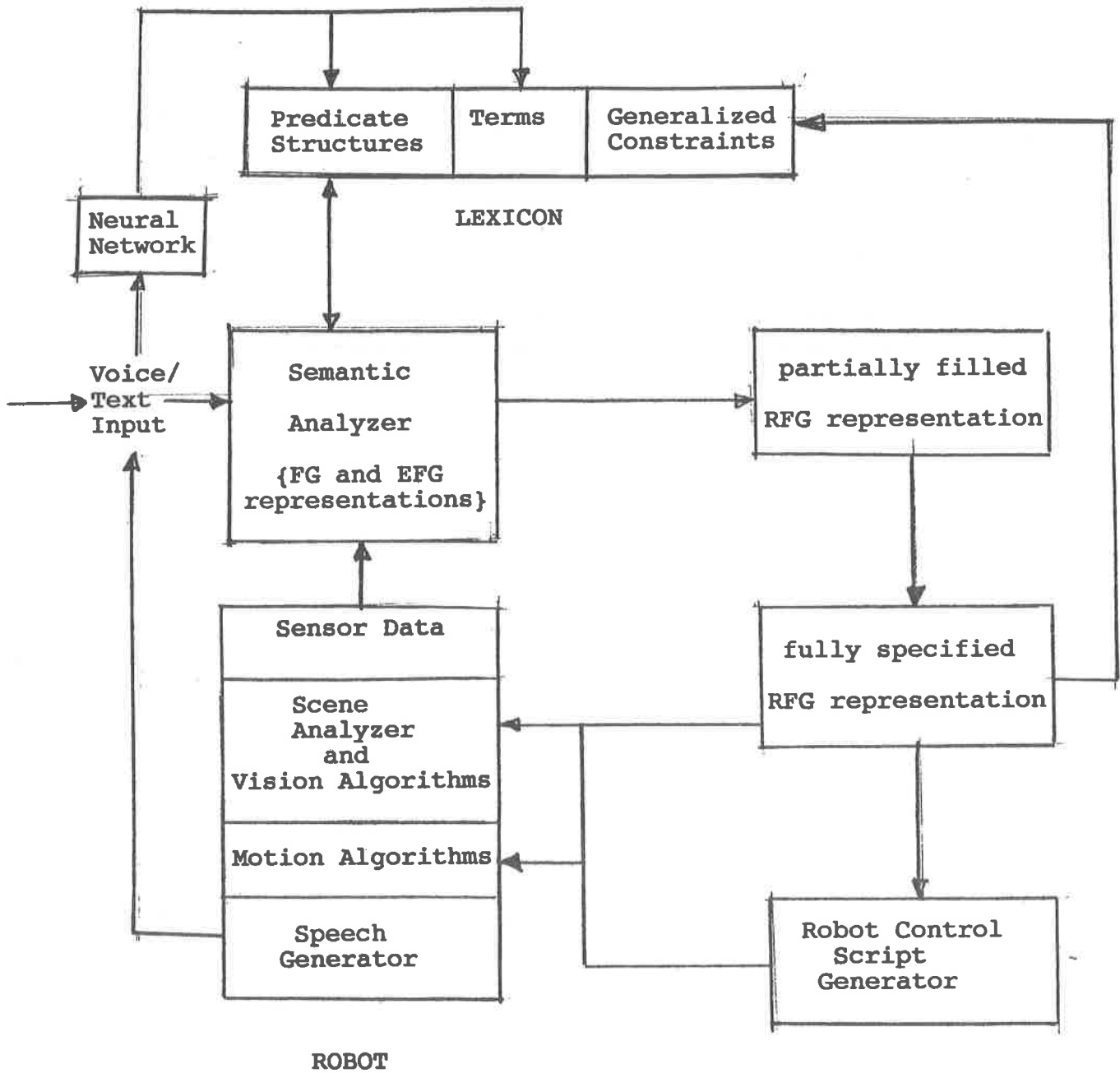Figure 1. Physical-world realization of mental concepts

Figure 2. Robot-world realization of natural language commands

28

Bob/Jim/Mary	1
horse/dog/cat	2
beer/water	3
meat/bread	4
runs/walks	5
works/speaks	6
visits/phones	7
buys/sells	8
likes/hates	9
drinks/eats	10
much/little	11
fast/slowly	12
often/seldom	13
well/poorly	14

Sentence Patterns:

1-5-12	1-9-2	2-5-14
1-5-13	1-9-3	2-9-1
1-5-14	1-9-4	2-9-2
1-6-12	1-10-3	2-9-3
1-6-13	1-11-4	2-9-4
1-6-14	1-10-12	2-10-3
1-6-15	1-10-13	2-10-12
1-7-14	1-10-14	2-10-13
1-8-12	1-11-12	2-10-14
1-8-2	1-11-13	1-11-4
1-8-3	1-11-14	1-11-12
1-8-4	2-5-12	2-11-13
1-9-1	2-5-13	2-11-14

Mary likes meat
Jim speaks well
Mary likes Jim
Jim eats often
Mary buys meat
dog drinks fast
horse hates meat
Jim eats seldom
Bob buys meat
cat walks slowly
Jim eats bread
cat hates Jim
Bob sells beer
(etc.)

Fig. 3. a List of used words (nouns, verbs and adverbs), sentence patterns, and some examples of generated three-word-sentences
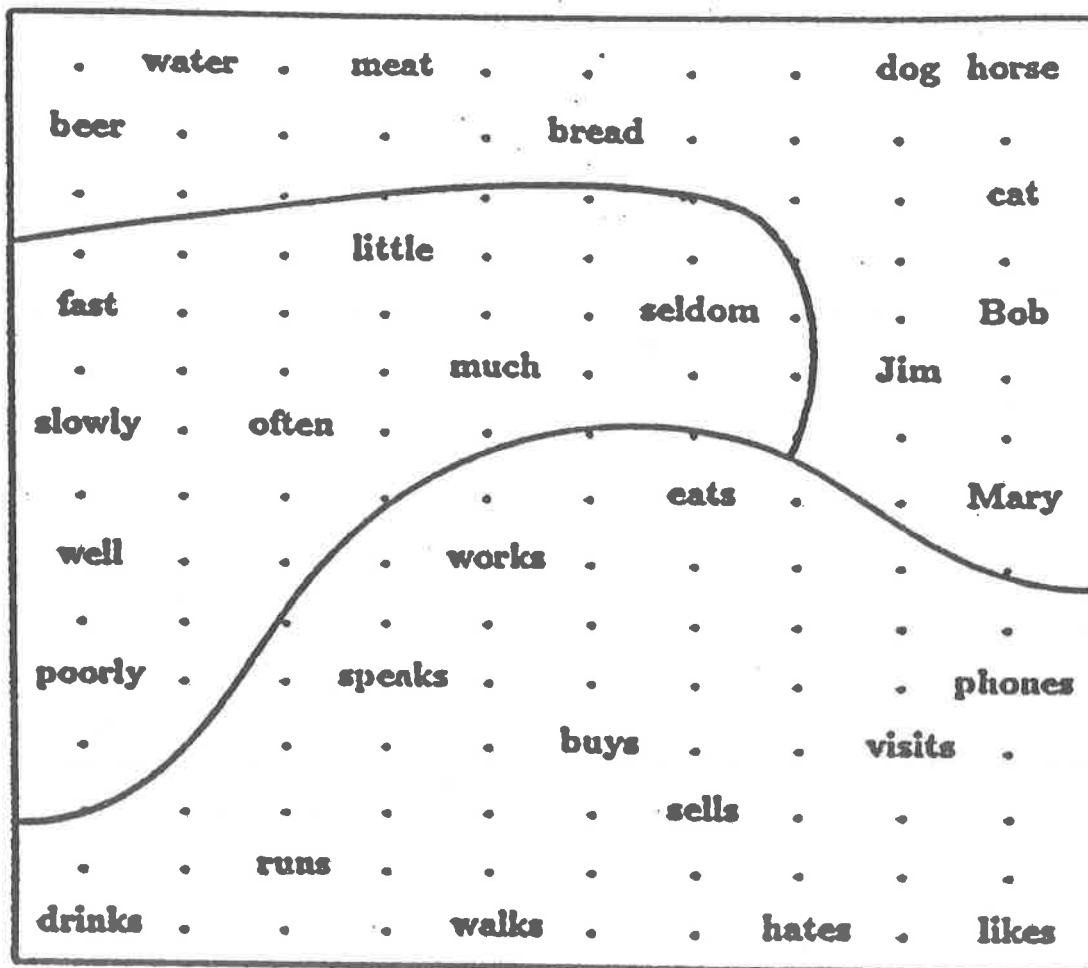
Fig. 3b. "Semantic map" obtained on a network of 10 × 15 cells after 2000 presentations of word-context-pairs derived from 10,000 random sentences of the kind shown in Fig. 3a. Nouns, verbs and adverbs are segregated into different domains. Within each domain a further grouping according to aspects of meaning is discernible
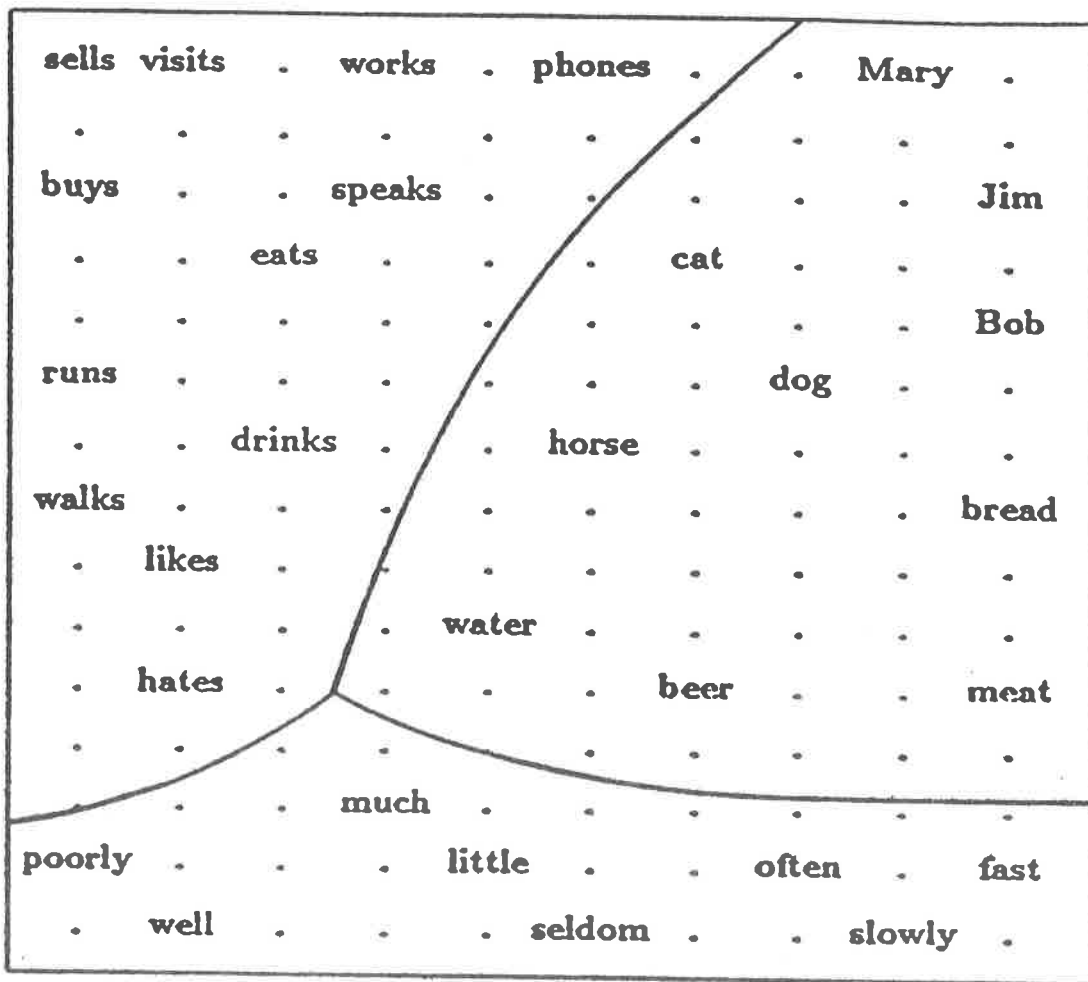
Fig. 3c. This map has been obtained by the same procedure as the map in Fig. 3b but with a more restricted context that included only the immediate predecessor of each word