TOWARD A DETAILED MODEL OF PROCESSING FOR LANGUAGE DESCRIBING THE PHYSICAL WORLD*

Davii L. Waltz

Coordinated Science Laboratory and
Electrical Engineering Department
University of Illinois
Urbana IL 61801

ABSTRACT

This paper explores the problem of judging whether or not an English sentence could correspond to a real world situation or event which is is literally, physically plausible, and the related problem of representing the different possible physical situations. The judgement of plausibility can be made at a high level by checking semantic marker restrictions on verb case frame constituents. Often, however, plausibility judgement can only be based on the results of an attempt to construct (imagine) a scene that corresponds to the sentence, and which does not violate "common sense" (i.e. relevant physical laws and expected, stereotyped behavior). Methods are presented for constructing representations for different scenes which could correspond to a sentence. These methods incorporate (1) "subscripts" (sequences of scenes which comprise an event, with attached preconditions and postconditions) to express different verb senses, (2) object representations which express properties such as shape, size, weight, strength, and behavior under common conditions; (5) physical laws, encoded as constraints on behavior; (4) representation of context; and (5) robot problem solving-like methods to fit all this material together.

1. Introduction

It is clear that humans understand language in far greater detail than do any programs written to date; people can make fine distinctions based on apparently peripheral properties of objects or nuances of behavior, and can draw upon a vast number of possible inferences about any given sentence or situation. This paper examines the problem of adding greater detail to the internal representations and reasoning methods for a natural language understanding program which works in the domain of physical scenes and events. I show how we can begin to account for the fine-grained but important distinctions introduced by the use of alternative verbs (e.g. hit vs. graze vs. strike vs. smash) and adverbial modifiers (e.g. hit vs. hit hard vs. almost hit vs. hit squarely).

Understanding language involves (at least) the following processes:

(1) judging the plausibility of the language; various possible readings of a sentence or text may have to be compared for relative plausibility, both on the basis of the inherent content of the language in isolation, and within the context where the* language is encountered;

(2) representing meaning; a suitable representation of the meaning of the language, in context, must be constructed. The representation should be unambiguous, and should have anaphora and ellipsis resolved. Representations should be capable of distinguishing readings that people regard as different; they should allow one to make natural inferences about the the events or scenes described; and similar meanings should have similar representations;

("5) retrieving information from and/or making modifications and additions to relevant memory;

(4) taking action if appropriate; action can be physical, linguistic (e.g. performing a speech act), or mental (e.g. planning a strategy).

These actions should not necessarily be done in any particular sequence. For example, it may be that one has to find a representation from memory before one can judge its plausibility.

1.1 Representing Spatial Meaning in Ehglish

Ehglish is a particularly poor language on which to bise a study of language and perception. English spatial locative prepositions (on, in, at, etc.) are ambiguous and irregular, so Ehglish descriptions of the perceptual world which use them are in general highly ambiguous. We often need to have a great deal of a priori knowledge in order to understand the idiomatic meanings which result when Ehglish locative prepositions are used with particular objects.

For example, to understand "The chair is at the desk", one must know that the ordinary relationship between desks and chairs is one where the chair faces the desk, and is partially beneath it; if a chair is facing away from the desk, or is upside down, or is on top of the desk, it cannot be said to be "at the desk". Phrases such as "at the corner", "at the store", "at the door", and so on, also require idiosyncratic analyses, and thus their meanings must be stored as separate lexical items. The same is true for other Ehglish spatial locative prepositions (see 1.1,12]).

Ehglish verbs present similar problems. Take for instance, the verb "cut". "Cut" can take on a wide variety of spatial/temporal meanings, which depend upon the nature of its object and instrument. Compare for example, the images and inferences evoked by "cut the board with a hand saw", "cut my finger on broken glass", "cut through the mountain with bulldozers", "cut the paper with a scissors", "cut the steel with a torch", "cut the corn with a combine", and so on. A system must be able to judge that it is possible but peculiar to cut a grape with a jigsaw, that it is impossible to cut a diamond with a plastic knife. Although there seem to be shared aspects of meaning, each verb-object-instrument combination must have separate and extensive meaning representation scenarios.

•i'nis work is supported by the Office of Naval Research under Contract NOOO14-75-C-O612.

The critical problem we face is this: what is an appropriate way to represent various possible meaning of language describing spatial relationships and physical events? There seem to be four main possibilities:

(1) case frame representation.

(2) sentences of spatial primitives using either predicate calculus (as in [18 and 11)) or a conceptual dependency- or script-lika form.

(3) schematic modeling, using data structures to directly represent qualitative spatial knowledge in 2-D diagrams [7,9,19,20,22,26];

(4) direct geometric modeling, using 3-D coordinate systems and specification of locations, velocities and directions [1,17,24,29J;

It seems clear that all flour types of representation are important, and that they stand in a hierarchical relationship with each other: alternative (4) is the "deepest," i.e. closest to the representations of high-level perception, alternatives (3) and (2) are progressively more abstract, and (1) is closest to a surface linguistic representation. Alternatives (I) and (2) are attractive options, since their representations are bounded, linear forms, whereas (3) and (4) require the ievelopment of novel representation schemes.

2. Biting Dachshunds

Let us look in some detail at the way sentence
(31) My dachshund bit our mailman on the ear.
would be processed by a program which generates a model of the physical world correlates of the sentence*. As a general strategy, I assume that the program should start its processing at the most abstract level, and only go to deeper levels of simulation if necessary.

(Step 1) The first things to check are the case frame for the verb "bite" and any possible semantic interpretation patterns which match the words of the sentence. The case frame for bite:
Bite (animate-entity physical-object)
is satisfied with the arguments "dachshund" and "mailman's ear", respectively. General world knowledge** e.g.
(goal dog (bite dog physical-object) often)
(goal dog (bite dog man) sometimes)
(avoid man (bite dog man) usually)
would also match the given sentence, so the general plausibility of the sentence would be immediately established, and initial goals for the man and dog would be hypothesized. In a quick reading or shallow understanding mode, processing might terminate here. Similarly, if general arguments were the only ones given in the sentence, as in:
(S2) The dog bit the man.
then there would be no reason for further processing, since the case frame for this sentence already matches exactly one of the semantic patterns known to the system.

(Step 2) The mention of "dachshund" and "mailman" cause the 'instantiation' of internal data structures, consisting of default postures and settings for both the dog and mailman. The mailman is instantiated on 1 residential sidewalk, since this is a strongly coupled default, and the dog is put in the same scene with the mailman, since dog is weakly coupled to a number of default locations (yard, indoors, pet store, kennel, etc.). The dog and mailman are both assumed to be standing. The scene is thus as shown in figure 1. The structures for each scene entity actually represent only the coordinates of its enclosing parallelipiped, along with a stick figure skeleton to which named parts can be attached. The mailman's enclosing volume is the standard one for adult males, and the dachshund's is found by scaling the dimensions of the dog enclosing volume appropriately. "Midget" or "puppy" would select different enclosing volumes, as would explicit mention of height (a3 in "the six-foot mailman").

'Programs have not yet been written to handle these examples. I believe, though, that there will be no insurmountable obstacles to writing such programs.

**This knowledge may be listed explicitly, deduced from general facts, or deduced from the analysis of scripts for typical actions.

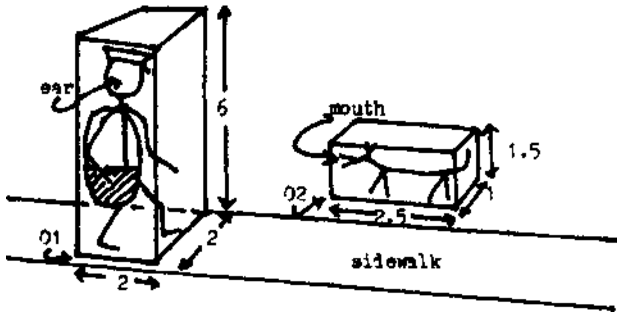

Figure 1. Default instantiation for (51).

(Step 3) The next step is to attempt to "run" (interpretively evaluate) the definition of "bite". This definition is shown pictorially in figure 2a, and part of the definition in assertion form is shown in figure 2b.

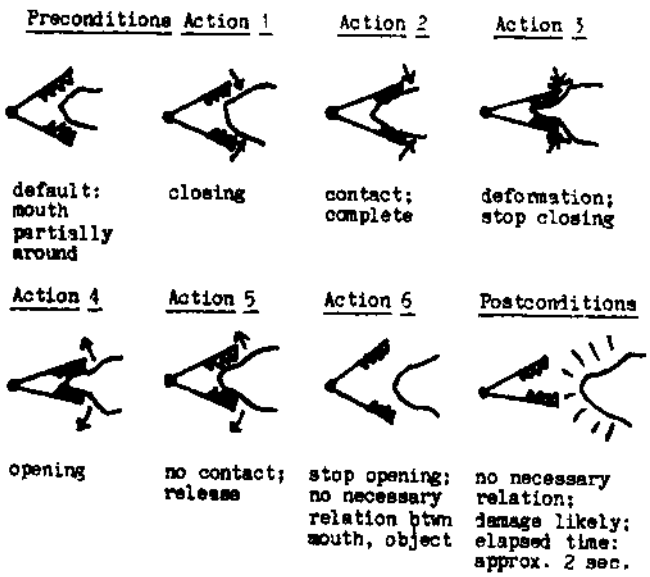

Figure 2a. Pictorial definition of "bite".

BITE (Animate-entity Physical-object)

Preconditions	Action 1	Action 2
1: (deformable P)	1: (closing M)	1: (contact M P)
2: (and (mouth M) (part-of A M))		2: (complete BITE)
3: (partially-around M P)		
4: (open M)		

Action 3	Postconditions
(deformed P)	1: (may-not-be (partially-around M P)
(not (closing M))	2: (open M)
(connected M P)	3: (may-be (damaged P))
(apply-force-on M	4: (time BITE (aprx 2 sec))
(may-not-be (open	

Figure 2b. Assertions for definition of "bite".

The definition includes preconditions, postconditions, and a body. The only postcondition we need for this example is Bite-PostC-1 , which states that there is (probable) damage to the thing bitten. The preconditions are of two kinds: "static preconditions" on the nature of the individual entities involved (analogous to semantic marker restrictions) and "dynamic preconditions" on the relations between the individual entities. In this case the static preconditions are Bite-PreC-1, which states that the object bitten must be delbmable (or else we would say "bite down on" instead of "bite"i, and Bite-PreC-2, which states that the thing that bites must be something that can close in the manner of a hinge, preferably a mouth of an animal or person*. The single dynamic precondition is Bite-PreC-3 which states that the thing that bites must partially surround the object bitten (if it totally surrounded the object, we would probably use "chew on rather than "bite").

Checking preconditions, we see that Bite-PreC-1 and Bite-PreC-2 are satisfied, since ears are deformable, and the dachshund has a mouth**. Precondition Bite-PC-3 cannot be directly derived from general knowledge, however, so "bite" tries to make it come true.

(Step 3.1) To this end, the program would attempt to plan a sequence of steps which could simulate reaching the goal:

(Be (Partially-around dachshund's-mouth mailman's-ear))).

To accomplish this goal, new preconditions must be satisfied. Basically they are Partially-Around-PreC-1 , that the mouth be open, and Partially-Around-PreC^T; that the location of the object be located between the ends of the "jaws" of the mouth. Partially-Around-PreC-1 is trivially satisfied, since we assume that opening a mouth is something that agents can always do unless explicit conditions have been mentioned which preclude this action. However problems arise in trying to satisfy precondition Partially-Around-PreC-2, since the dachshund's mouth is not around the mailman's ear in the default situation instantiation (see figure 1).

(Step 3.1.1) To achieve Partially-Around-PreC-2, one can either bring the mouth to the ear, or the ear to the mouth. Case grammar analysis has already identified the dachshund as the agent, so the preferred method is to bring the mouth to the ear.

(Step 3.1.1.1) The first goal tried is thus:
(Reach dachshund's-mouth mailman's-ear)
Through match of general patterns for accomplishing Reach, at least four possible subgoals can be suggested: (reach-by-body-posture), (jump-up-to), (climb-up-using physical-object), and (make someone (move someone physical-object)). The library of dog body positions should be checked first***.

None of these postures allows the mouth to be around the ear, so jump-up-to is made the subgoal, and dynamic behavior of the dog checked. Using high level knowledge of animals' abilities attached to jump-up-to, for instance that animals can typically jump no higher than twice their longest dimension, it can be deduced that the dachshund could not j\mp high enough either.

The next thing to try is (climb-up-using physical-object). Methods attached to this goal attempt to use any objects that are known to be around in order to let the dachshund climb up to the mailman's ear. However, since no objects have been mentioned explicitly, and no default objects (e.g. mailbag, uniform, dog collar) are good for climbing on, this possibility fails also.

The last possibility is to induce someone to lower the ear so that the dachshund can reach it. The only other possible agent is the mailman, and dogs can send only a very limited repertoire of messages. A dog might make sounds to induce the mailman to pet it, but other options, e.g. convince, bargain, etc. are only available to people. Other messages a dog can send, such as "threaten", would probably lead the mailman to flee rather than to come closer.

(Step 3.1.1.2) Having failed at getting the dachshund reach the mailman's ear, the program would then check to see if the mailman's ear might be gotten to the dog. Here too there are two main possibilities: the mailman may have put his ear within range of the dog's mouth intentionally or unintentionally. Having earlier noted the general fact:
(goal man (avoid (bite dog man)))
we can conclude that the goal
(goal mailman (bite dachshund mailman's-ear))
has a low probability of being true, and we can thus preclude the possibility that the mailman might intentionally put his ear within the dachshund's range for the purpose of having it bitten.

Another possibility is that the mailman got bitten unintentionally; this could happen if the mailman typically assumed postures that would allow the dachshund to reach him. To check out this possibility, the script for "mailman" must be consulted. No mention of postures other than standing, walking, getting mail out of a mailbag, and putting mail in boxes are to be found in the script, so there is no reason to believe that a mailman would get to where a dachshund could reach him unintentionally either.

Note that we could also find range of body positions possible for a dog by simulating the movements of its joints. However, I do not believe that people behave this way, and for a program, it is a great deal easier to simply store common postures (and posture sequences) directly.

"The meaning of "bite" corresponding to the bite of a bee or mosquito would have a different representation.

**These facts could be stored with the concepts of "ear" and "dachshund" or, more likely, with their ancestor concepts in a "structured inheritance network" (e.g. [2|) — i.e. these facts can be derived automatically from more general facts, in this case, that the parts of the body of a creature, except bones and teeth, are deformable; likewise all animals have mouths.

The overall conclusion of the program would thus be that while it is plausible that a dog could bite a man, the program cannot imagine how a dachshund could hive bitten a mailman's ear. How can we avoid pssible explanations based on the observations that people do occasionally squat or lie down, making them accessible to bites by small dogs? During informal experiments, people have suggested only possible explanations that tie in somehow with the particulars of mailmen and dachshunds; they do not suggest explanations such as "The mailman was a friend of the owner of the dog and was lying down on the owner's sofa when bitten". This seems consistent with a policy of omitting concepts that do not occur in the sentence explicitly (e.g. sofa, owner of dog, friendship) unless they can be assumed as part of a relevant script (mailbag, letter, dog collar, mailbox, and so on could thus be used as part of an explanation). Knowledge about apparently peripheral items (e.g. the likelihood that the mailman is a friend of the dog's owner) is not necessarily irrelevant, and may in fact be used explicitly by people in reasoning about event plausibility. I see no reason to believe that a system could understand natural language in any epistemologically non-trivial sense without having available a great deal of information (see also [15]).

If forced to produce a plausible explanation, the program should be able to select ways of getting the dog's mouth to the mailman's ear which were not precluded, but for which there was no support (e.g. the mailman may have squatted for some reason, there might have been an object that the dachshund could climb up on, the mailman might have been very short, etc.).

3. Slightly different sentences, radically different processing

Let us now look at the ways in which the processing of similar sentences would differ from the processing of (31).

Let us start with

(53) My dachshund bit our mailman.

As in the case of (S2) *"The* dog bit the man" (mentioned earlier in section 2), this sentence would be processed only to the case frame level. If pressed, of course, a system could imagine a scene to simulate this sentence, e.g. the dachshund biting the mailman on the leg.

Consider next (S4), "My doberman bit our mailman on the ear". The first step in processing would be exactly the same as that for (31), i.e. the general plausibility of dogs biting people would be noted along with the fact tht people usually avoid being bitten. The second step would likewise be similar, but in this case the size of the dog would be much larger. In step 3, however, the processing would follow a very different line. In step 3.1.1.1, (Reach doberman's-mouth mailman's-ear) would succeed via reach-through-body-posture, since the pre-stored postures of a dog would allow the mouth to be at the height of the ear when scaled appropriately for a doberman.

The program would then continue the running of the event simulation fbr bite (see figure 2), simulating in turn the closing of the mouth on the ear, the contact between the mouth and ear, the deformation of the ear with probable damage, and the opening of the doberman's mouth again. Finally, through knowledge that only a few body postures are stable (i.e. can be held for a long time), the program would simulate the return of the doberman from its standing-on-two-legs posture to a standing-on-four-legs posture. In this case, then, nothing remains to be explained, the program could judge that the event described in the sentence was plausible, and the simulated subevents could become part of its memory.

Consider now *(55),* "My dachshund bit our gardener on the ear". The processing of (35) would be different from any of the cases above; it would parallel the processing of (31) until the point where the mailman script is checked to see what body postures are ordinarily assumed by mailmen in the course of their work. The script for a gardener involves getting into a Kneeling or sitting position frequently, so there would be a ready explanation for how the person's ear could have gotten in range of the dachshund's mouth. The processing could then continue as in the case of (34) to completion of the biting action and return of the dog to a standing position on the ground.

The processing of (56), "My dachshund almost bit our mailman", is especially interesting. Here we have to decide what "almost" could mean. Referring to the definition fbr "bite" shown in figure 2, we can see that the snapshot frame in which the teeth make contact with the object is marked "complete", meaning that if this snapshot portion ever occurs, then the action of biting will have been completed, regardless of exactly what course the action takes afterwards. On the other hand, until the action in this snapshot actually occurs, no biting can be said to have taken place. With this knowledge available, we are now in a position to understand (36).

Basically, we know that at least some of the preconditions for bite must have been satisfied, or else "almost bite" would not be an appropriate phrase to use to describe the event (this idea is related to Grice's maxims for communication [10]). To process this sentence the way a person would, we would have to be able to generate the range of types of events that could qualify as "almost bite"; some would be easy, e.g. cases where the dog closed his mouth but failed to contact the mailman because either he or the dog moved, or where the dog opened his mouth near the mailman s body but never closed it on him. Generating other cases would require special knowledge of how dogs bite, as opposed to the knowledge of biting in general, which we have been using so far. Special knowledge of dogs could allow us to judge that a threat (such as baring the fangs and/or growling) might qualify as "almost biting". If we knew a great deal about the particular dog referred to in the sentence, we might envision arbitrary behavior that had in the past preceded biting incidents (e.g. the dog might always jerk its head rapidly to the left, or bark twice just before biting someone).

If the sentence had been (S7) "My dachshund tried to bite our mailman", then the envisioned possibilities would be similar to those generated for (S6), although we would strongly prefer that all the preconditions of bite be satisfied, along with all actions up to the snapshot frame marked "complete"; idiosyncratic or ordinary threat signals would not really be appropriately described as "try to bite".

Finally for (38), "My dachshund bit our mailman hard", we would have to find a step where there was an application of force, in this case step Action} of "bite" (see figure 2). *The* representation of this portion of the subscript would then be simply modified, and the additional observation made that biting hard ==> more deformation (probably) ==> more damage (probably).

4. How much knowledge is necessary to do these examples?

A lot, though not an impossibly large amount. Let us here take stock of the kinds of different knowledge necessary for processing the sentences above, and also look at how a system could know that it should retrieve this particular taiowledge at the right times.

(l) Case frames. Garden variety case frames would suffice here, retrieved as dictionary entries; the only even mildly nonstandard feature I envision here would be to use structured inheritance for case frames (as well as most other types of information) so that, for example, verbs that describe events that have duration would have ancestors that elaborated on the use of time expressions for verbs with duration (as distinct from verbs describing events which occur at a point in time such as leave or hit, or verbs which describe states rather than events).

(2) Semantic patterns. As I envision these, they would be similar in many ways to the patterns which form the heart of PARRY [5] and also similar to the patterns used by Wilks in his preference semantics [28]. Many of these patterns would be stored directly (e.g. <avoid animate-entity pain usually>) and others would have to be inferred (e.g. <avoid man (bite dog man) usually>). Generally these patterns would describe goals, beliefs, and combinations of concepts that would usually co-occur with a verb. I believe that it will prove worthwhile to tag these patterns with probability information of a rather broad type (always, usually, often, sometimes, rarely, never). The system would use this information in the manner of expert systems [6]. Generally the system would begin with patterns involving as many of the entities in the sentence as possible, and attempt to find patterns that would express the goals and/or beliefs of each of the participants.

(3) Spatial descriptions. Each physical object known to the system should have a spatial description, which would allow the system to Judge for each object its size, weight, stable postures or orientations, positions of parts within an enclosing volume for the object, dynamic behavior (how fast the object can move, what sequences of positions are possible), etc. These descriptions should be hierarchical, so that, for example, "jaw" would have its own spatial description which is merely pointed to by the descriptions of entities which have jaws, along with appropriate scaling information. Again here, structured inheritance would be useful to store information about the spatial world efficiently.

(4) "Subscripts" for verbs. In addition to case frames, verbs would also have much more spacific information about the time sequence of subparts of the action described by the verb, including a set of preconditions, a body of actions, and a set of postconditions. The effects of the action could be found either in the general semantic patterns (item (2) above) or in the subscript corresponding to a given verb sense, or both places.

(5) Scripts. These serve to connect goals with methods for achieving them. In addition, scripts may be used to describe typical or stereotyped sequences of events (e.g. earthquakes cause structural damage to buildings, may lead to tidal waves, aftershocks, rescue of victims, repair of damage, etc.). The scripts may in turn include subscripts. In general each goal may refer to a number of scripts (see [21]).

4.1 Subscripts

Subscripts are at the heart of the work I am describing here. They serve as a bridge between the case frames, selection restrictions, etc. on one hand, and spatial descriptions on the other. They are important for several reasons:

(1) Adverbial modifier representation. Subscripts provide a framework in which adverbial modification has a natural representation. To take some specific examples, consider "almost"; as suggested in the example of handling of (38) verbs capable of being meaningfully modified by almost (e.g. hit, break, reach, finish, buy, etc.) each have a step in the body of their associated subscript which, if not completed, allows the action up until that step to be described as "almost having happened", whereas if that step does occur, "almost" can no longer be used to refer to the action. Adverbs such as hard, gentle, soft, violently, etc. refer specifically to subscript steps that involve application of force, transfer of momentum, or related concepts. Adverbs such as fast, slow, smoothly, abruptly, suddenly, etc. refer to subscripts involving motion, and require that the system have a notion of "normal" duration or speed of motion in order to interpret them properly. Other modifiers such as painfully, successfully, happily, etc. emphacize effects; still others, such as gently, quietly, rudely, cleverly, etc. require special knowledge of behavior that is beyond the scope of ray present work.

(2) Verb sense representation. Subscripts permit easy distinctions between verb senses, even where the case frame constituents do not allow for any distinctions. Thus, for example, fly (meaning "go by plane") can be easily distinguished from fly (meaning "pilot a plane"), because the subscripts attached to each sense represent very different sequences of actions. The detail in subscripts, when compared with the current context, adds a great eal of information useful for selecting appropriate word senses.

(5) Common sense. Subscripts provide ways of organizing pointers to relevant real world physical forces, behaviors, constraints and properties, which make it possible to bring general knowledge to bear on language understanding.

5. Related AI work

The two questions of plausibility and representation have not usually been linked. Case frames and semantic markers have been used to judge plausibility, as have Wilks* "preference semantics" structures [28] but neither is a serious knowledge representation candidate. Knowledge representation schemes (KL-ONE, NETL, SNePS, KRL, etc.) have not been much used to judge plausibility — rather, they have been used for storing information which is fed to them by other programs. A partial exception is the work of Schank et al (on scripts as well as primitives and structures of conceptual dependency) which try to combine both plausibility and representation, although at a much more abstract level than the spatial models being proposed here. The representation of adverbial modifiers is also central here. Relatively little work has been done on this topic. Overall, however, HcDermott's work, both on TOPLB [16] and the more recent work on spatial inferences [17], seems to me to be most similar to what is suggested here. Other clear intellectual predecessors are [14], [29]. [22], [19]. and [13].

6. Promising Directions

A key problem is the development of an adequate s*t of spatial primitives. English prepositions are inadequate, so where can we look for guidance? we have been studying non-Indo-European languages, at least some of which (e.g. Jinghpaw, a Tibeto-Burmese language, and Tarascan, a native American language) are far more regular and precise than English in their expression of spatial and temporal ;meaning. Jinghpaw, for example, has some fifty different verbs for describing cutting actions (compare with English examples above), and a like number for describing different finger positions and motions. More will be reported on this study in the near future.

Suggestive and very interesting material on grammatical clues for spatial and temporal information is found in Talmy [23] and Fillmore [3], and on building and switching contexts in Chafe [4]. I also intent to integrate my work on object shape description [25].

References

In the references below. TINLAP-1 refers to R. Schank and B. Nash-Webber (eds.) Tneoretical Issues in Natural Language Processing, Arlington, VA: ACL, 1975; and TIINUP-2 refers to D. Waltz (ed.) Theoretical Issues in Natural language Processing—2, New York: ACM, 1978

[I] Boggess, L. C. 1979. Computational Interpretation of English Spatial Prepositions. Tech Rpt T-75. CSL, Univ. of Illinois.

[2] Brachman, R. J. 1979. On the Epistemological Status of Semantic Networks. In N. Findler (Ed.), Associative Networks. New York: Academic.

[3]Bundy, A. 1978. Will It Reach the Top? Prediction in the Mechanics World. Artificial Intelligence 10, 2.

[4] Chafe W. L. 1979. The flow of thought and the flow of language. In T. Givon (Ed.) Discourse and Syntax., New York: Academic.

[5] Colby, K., B. Paught, and R. Parkinson. 1974. Pattern Matching Rules of the Recognition of Natural language Dialogue Expressions. Memo AIM-234, Stanford AI Lab.

[6] Davis, R. 1978. Knowledge Aquisition in Rule-Based Systems. In D. Waterman and F. Hayes-Roth (eds.) Pattern-directed Inference Systems. New York: Academic.

[7] de Kleer, J. 1977. Multiple Representations of Knowledge in a Mechanics Problem-solver. IJCAI-77, 197-203.

[3] Fillmore, C. J. 1977. The Case for Case Reopened. In J. Kimball (Ed.) Syntax and Semantics. New York: Academic.

[9] Forbus, K. D. 1981. A Study of Qualitative and Geometric Knowledge in Reasoning about Motion. Tech Rpt AI-TR-615, MIT AI Lab.

[10] Grice, H. P. 1975. Logic and Conversation. In P. Cole and J. Morgan (Eds.) Syntax and Semantics: Speech Acts. New York: Academic.

[ITr"Hayes, P. J. 1978. The Naive Physics Manifesto; and Naive Physics 1: Ontology for Liquids. Tech. Rpts., Institute for Semantic and Cognitive Studies, Geneva.

[12) Herskovitz, A. 1980. On the spatial uses of prepositions. Proc. of 18th Annual Meeting of the ACL, Univ. of Perm., Fnila.

[13] Johnson-Laird, P. N. 1980. Mental models in cognitive science. Cognitive Science 4, 1, 71 -115.

[14] McCarthy, J. 1968. 'Programs with common sense. In Semantic Information Processing, M. Minsky (ed.), MIT Press Cambridge, MA, 403-418.

[15] McCarthy, J. 1977. Epistemological problems of artificial intelligence. IJCAI-77.

[16] McDermott, D. 1974. Assimilation of New Information by 1 Natural l/inguage-Understabling System. Tech Hpt.TR-241, MIT AI lab.

[17] McDermott, D. 1980. Spatial Inferences with Ground, Metric Formulas on Simple Objects. Rpt. 173, Yale Univ. Computer Science Dept.

[18] Miller, G. A. and P. Johnson-Laird. 1976. language and Perception, Cambridge, MA: Harvard Univ. Press

[191 Novak, G. S. 1976. Computer understanding of physics problems stated in natural language. Tech. Rept. NL-30, Dept. of Computer Science, Univ. of Texas, Austin*

[20] Rieger, C. 1975- The comraonsense algorithm as a basis for computer models of human memory, inference, belief and contextual language comprehension. TINLAP-1, 180-195.

[21] Schank, R. C. and R. Abelson. 1977. Scripts, Plans, Goals, and Understanding. Hillsdale, NJ: Erlbium.

[22] Simmons, R. F. 1975. The Clowns Microworld. TINUP-1. 17-9.

[25] Talmy, L. 1979. The Relation of Grammar to Cognition - A SSjmopsis. TINLAP-2, 14-24.

[24| Waltz, D. L. and L. Boggess. 1979a. Visual analog representations fbr natural language understanding. LJCAV79. 926-34.

—H75J Waltz, D. L. 1979b. Relating images, concepts, and words. In N. Badler and R. Bajcsy (eds.) Representation of 3-D Objects. Verlag, to appear.

[26] Waltz, D .L. 1980. Understanding scene descriptions as event simulations. Proc. 18th Annual Meeting of the ACL, June, Univ. of Perm., Phila., 7-12.

[27] Waltz, D L 1981. Generating and understanding scene descriptions. In Joshi, Webber, and Sag (eds.), Elements of Discourse Understanding, Cambridge Univ. Press.

[28] Wilks, Y. A. 1975. A preferential, pattern-seeking, semantics for natural language inference. Artificial Intelligence 6, 1, 53-74.

[29] Winograd, T. 1972. Understanding Natural Language. New York: Academic. Winograd T. 1972. Understanding Natural Language. New York: Academic.