NL-20

ON MANAGING SENTENCE MEANINGS*

Robert F. Simmons
Professor of Computer Science
at
The University of Texas at Austin
Austin, Texas 78712

ABSTRACT

A deep case structure semantic analysis is shown for two sentences of encyclopedia text. A process modelling technique for representing lexical and discourse meanings is outlined and applied to the sentences. Implications for data management systems are discussed.

Some of the most recent and exciting research in natural language processing is characterized by the recognition that an English text translates into some form of procedural model of what that text is attempting to communicate. A clear and elegant example of this is Winegrad's (1972) model of a hand and table-top with blocks. In this model actions were mapped onto procedures that operated the hand, and objects referred to the blocks. The blocks were interrelated by such relations as "supporting", "on top of", "beside", "under", etc. A natural subset of English that a human uses to command the robot hand is well understood--i.e. formalized--by Winegrad's system which demonstrates its understanding by carrying out commands and by answering questions.

A system constructed by Heidorn (1972) translates English descriptions of queing problems into Simscript programs. An experimental system by Schank and his students (Schank et. al. 1973) generally represents the meanings of English words and sentences as conceptual dependency structures based on a language of 14 primitive actions--i.e. a model that provides a translation from English sentences into sequences of primitive actions and their arguments.

An unusual thesis by Ramani (1969) translates from English text that teaches elementary math into the procedures that are being taught. Charniak (1972) has attempted to model the world of children as expressed in a primer by using Planner procedures, assertions etc. Here, at the University of Texas, Hendrix, Slocum and Thompson (1973) and Hendrix (1973), have described language understanding systems and models that map English statements into procedures that store their meanings as relations among objects in a time ordered data base.

This brief outline of current research developments in computational linguistics is aimed toward suggesting that data management requirements for accessing the meaning of textual data will require a significant development step from the present procedures for accessing a largely static data base toward a system that will compile inference procedures corresponding to questions and apply them to dynamic models of what the text is about. In the remainder of this paper, I shall develop some of the structures required to make and understand a model of the following two sentences quoted from Compton's Encyclopedia (1962).

"Coconut Palm

> One of the first plants to appear on a newly formed tropical island is the stately and graceful coconut palm. The seed which is the coconut of commerce is securely protected from the action of seawater by its thick, fibrous husk and hard shell. ..."

*Supported by NSF Grant GJ 509X

This paper will first present a semantic representation of the meaning of these sentences. It will then develop process models for the concepts of the sentences and finally, integrate these into a model that purports to explicate the meanings of the two sentences. In previous papers I have dealt at greater length with semantic representations of English structures (Simmons 1972a), and with algorithms for transforming from English into these structures, answering questions, and generating English from them. (Simmons, 1972b, Simmons & Slocum, 1972, see also Hendrix, Slocum and Thompson 1973). The present purpose will be served by a statement of the modelling techniques which I believe will suggest the nature of advances that will eventually be required of data management systems.

To communicate the semantic structure, let us examine the sentences phrase-by-phrase, working out some of their logical meaning, then expressing it in the formalism of semantic network structures.

ONE of the first plants to appear on a newly formed tropical island-There is a member or a species of plants such that it appears early in the history of a new island located in the tropics--

One of the	(X1(SUP X2) (Q1) (NBR SING) (DET DEF) (T* R1))
first plants	(X2(TOK PLANT) (NBR PL) (DET DEF)) .

The form of the semantic net structure is an object--e.g. a node on a graph--that is the name of a set of relations with other objects. Relations in this example are SUP for superset, Q for quantifier, NBR for grammatical number, DET for determiner and T* to show that X1 is the Theme argument of a relation, R1. The expression shows that X1 is a subset of X3, having 1 member; it is grammatically singular and refers to a definite set. T* shows that it took the action--i.e. was the deep object of some verb shown in R1. The object, X2 names and specifies a particular set that are TOKens or occurences of PLANTS.

...to appear	(R1(TOK APPEAR) (T X1) (LOC X3) (MD INF, FIRST))
(first) on X3	

R1 is the name of a relation set that describes the verb and its arguments. We use a deep case structure analysis that relates each noun phrase argument to the verb by one of the following relations:

Actant--animate doer of the action
Theme--takes the main effect of the action
Source--place or state of the origination of the act
Goal--place or state of termination of the act
Instrument--some other process that contributes
Locus--general location of the action
Time--any time predicate, e.g. Saturday, on the action

In addition there is a Modality, MD which is used to relate Tense, Mood, and Adverbial modification to the verb. In the R1, above, the modality shows that the verb was in the infinitive, and modified by "first". We can see from R1 that X1 appeared at the location X3 by unspecified agency.

on...tropical	(X3(TOK ISLAND) (NBR SING) (DET INDEF) (MOD TROPICAL)
island	(T* R2) (LOC* R1))

X3 is a relation set that describes some island that is tropical, on which the plants appear, and which is theme of R2.

newly formed island	(R2(TOK FORM) (T X3) (MD NEW))

R2 specifies that a "forming" process was recently applied to X3, the "tropical island". Who or what did the forming, whether by volcanic action or the accretion of coral is not specified in the sentence so no actant or instrument arguments can be shown.

the stately and graceful coconut palm	(X4(TOK COCONUT PALM) (NBR SING) (DET DEF) (MOD STATELY, GRACEFUL))

X4 treats "coconut palm" as a single word signifying only that it is a unique lexical entry. The relation MOD shows that the particular attributes associated with the words, "stately" and "graceful" will characterize this object.

X1 is X4	(R3 (TOK SPECIES) (T1 X1) (T2 X4) (MD/INDIC PRES))

The use of "is" with a definitely specified complement shows the species relation. Thus the sentence is equivalent to "One...is the species coconut palm". The arguments T1, and T2 simply stand for two theme arguments. The modality "present indicative" signifies to later programs that this is a tre general statement.

The second sentence reads, "The seed which is the coconut of commerce is securely protected from the action of seawater by its thick fibrous husk and hard shell". The relation sets which represent its semantic analysis are shown below (in an abbreviated notation using commas instead of inner parenthesis).

The seed	(R3.5 TOK HASPRT, T1 X1, T2 X5) (X5 TOK SEED, DET DEF, NBR SING, T2* R3.5, T* R3, T* R4, T* R5, T* R6, T1* R7)
which is the coconut of commerce	(R4 TOK EXCHANGE, T X5, S 'person', G 'person' I 'money') i.e. 'people exchange coconuts with people for money'
is securely protected from R6 by X7	(R5 TOK PROTECT, T X5, S R6, I X7, MD PRES, SECURELY)
The action of X6	(R6 TOK ACT, S X6, T X5, MD POSSIBLE)
seawater	(X6 TOK SEAWATER, DET INDEF, NBR SING, S* R6)
its thick fibrous husk and	(R7 TOK HASPRT, T1 X5, T2 X7) (X7 TOK AND, T1 X8, T2 X9, I* R5, T2* R7) (X8 TOK HUSK, MOD THICK, FIBROUS, DET DEF, NBR SING, T1* X7)
hard shell	(X9 TOK SHELL, MOD HARD, DET DEF, NBR SING, T2* X7)

We need to notice that "the seed" is an elliptic reference to "the seed of the coconut palm" that requires a sentence analyzer to find the relation that holds between "seed" and anything so far mentioned in order to discover the relation that connects this new term with the previous discourse. In a similar fashion, the possessive pronoun in "its thick fibrous husk and shell" signifies that the coconut participates as a Theme argument in the EXCHANGE relation which we use to model the meaning of commerce. In general, it can be noticed that this semantic analysis considers nominalized verbs such as "action" and "distribution" as syntactic forms of the verb where prepositional phrases and/or a possessive pronoun signal one or more of its arguments. In the above examples adjectives are presented in an abbreviated notation. What X9 MOD HARD expands to is:

(X9...,T* R8)
(R8 TOK HARD, T X9)

The brief form is easier for a person to read.

The relation sets that have been presented are shown in graphic form as a semantic network in Figure 1. The back pointers that are signified in the relation sets by T*, S*, T2* etc. can be understood as reverse directions of arrows on the directed graph of that figure.

MODELLING CONCEPT MEANINGS

When a human hears a word in context--such as island in "newly formed Island" he reports a sequence of ideas that are aroused in his mind--frequently as a sequence of images. For "a newly formed island", I have an image of a volcanic island arising in great turmoil from the ocean, spewing lava into the sea, gradually growing in size, randomly changing the shape of its harbors, bays and beaches at first over decades, then hundreds or thousands of years. If the island is tropical, inevitably the mountains and hills become covered with a rich green vegetation, animal life is heard...

Someone else might think of the formation of a coral island with its slow accretion of mass and there may be other modes of island formation known to specialists in geology. In every case, however, the image is a process; that is, it is a time-ordered series of connected events. It has a beginning, an ending, and at any given instant of observation, an age. Such processes can be modelled in various formal languages and particularly well in computer languages where time sequences are integral to the design of the language.

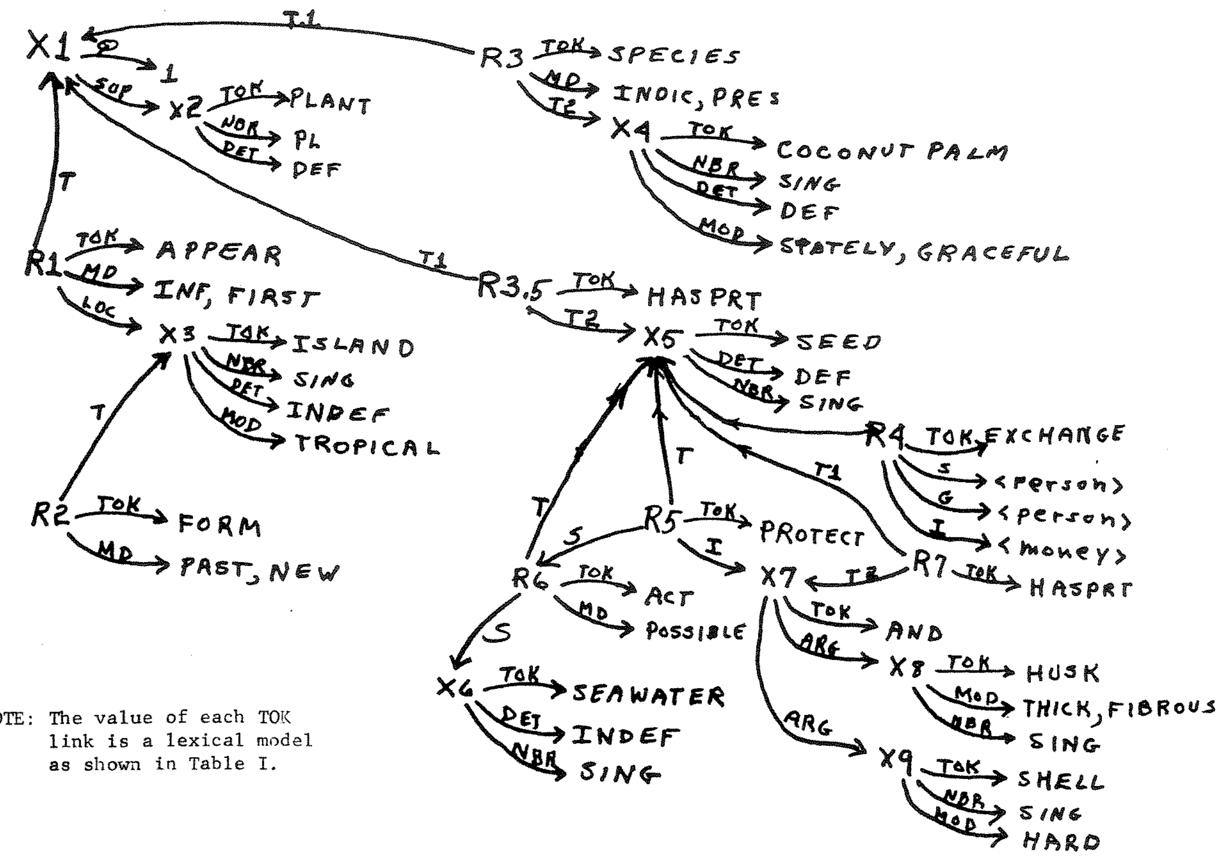
Let us see how we can formalize such an image as a computational structure.

```
ISLAND SUP LANDMASS
        HASPRT (EARTH, WATER, BEACH....)
        ENCLOSED (SEA, SKY)
        ISPART (OCEANBOTTOM
        DURATION EONS
        PROCESS ( (ti (FORM S SEABOTTOM G ISLAND I
                    (OR(CORAL VOLCANO))))
                  (t1 (FORM S ROCK G SOIL I (WEATHER,
                       ORGANISMS)))
                  (t2 (CARRY T SEED G ISLAND I (OR
                      (WIND, OCEAN, ORGANISM)))
                  (t3 (GROW T ORGANISM LOC ISLAND))
                   .
                   .
                   .
                  (tf-1 (OR(SINK T ISLAND LOC OCEAN)
                     (JOIN T1 ISLAND T2 CONTINENT)))
                  (tf (DISAPPEAR T ISLAND)))
```

If we apply an English sentence generator to this structure, (see, for example, Hendrix et.al.), the following ordered set of sentences would be generated (providing a translation for the reader).

AN ISLAND IS A LANDMASS

IT HAS AS PARTS EARTH WATER AND BEACH
SEA AND SKY ENCLOSE AN ISLAND
IT ENDURES FOR EONS
INITIALLY CORAL OR VOLCANOES FORM AN
ISLAND FROM THE SEABOTTOM
THEN WEATHER AND ORGANISMS FORM SOIL



NOTE: The value of each TOK link is a lexical model as shown in Table I.

FIG. 1 SEMANTIC GRAPH OF EXAMPLE SENTENCES

FROM ROCK
THE WIND, OCEAN OR ORGANISMS CARRY SEEDS
TO THE ISLAND
THEN ORGANISMS GROW ON THE LAND
⋮
FINALLY THE ISLAND SINKS INTO THE OCEAN
OR JOINS A CONTINENT AND DISAPPEARS

Certainly, a very coarse outline of our idea of an island but it is a formalization that can be used by other programs--such as, for example, a picture generator.

Let us look at another process, this time the process of "forming". If we look in a very good dictionary such as Webster's 3rd International--the big one--we will discover that most of the sense meanings listed for "form" fit the paradigm, "Someone forms something out of or into something with some instrumental process". This fact gives us a paradigmatic deep case structure for this verb as follows: (FORM A S G I). That is, "form" takes Agent, Source, Goal and Instrument arguments and the G argument--that which resulted from the forming process--is required.

It is not easy for a person to form an image of an abstract notion such as a forming process, but let us look again at the paradigm sentence, "Someone forms something1out of something2with some instrumental process". We start with SOMEONE acting on SOMETHING2, with INSTRUMENT. The INSTRUMENT changes SOMETHING2 toward SOMETHING1 and finally, SOMETHING1 appears to SOMEONE. We can formalize this as follows:

```
FORM PDIGM (A S G I)
     PROC ((ti(ACT A A', T S', I I') )
           (t2(CHANGE S S', G G', I I') )
           (tf(APPEAR T G', G A')))
```

If in the PDIGM we let A take the value John, S clay, G pot, and I his hands, then a sentence generator would produce the following:

JOHN FORMS CLAY INTO POTS WITH HIS HANDS

INITIALLY, JOHN ACTS ON THE CLAY WITH HIS HANDS
THEN THE CLAY CHANGES TOWARD A POT
FINALLY A POT APPEARS TO JOHN

What we are doing in this process description is to list a sequence of events that characterize a forming in terms of what happens to the various arguments of the paradigm. The form, (ACT A A', T S', I I') means the Agent of ACT will take the value of A' the Agent of the paradigm. The Theme of ACT will take the value of S' the Source of the paradigm, and the Instrument of ACT will be the Instrument of the paradigm. In other words, this formalism is essentially a program that will bind the arguments in the paradigm to a sequence of actions that define a "forming" process--in a very limited manner.

An adjective in context--such as "tropical island" conjures certain ideas. In this case, the location, the temperature level, the rainfall, the type of vegetation etc. These notions are not particularly ordered in time, but they can be expressed in a model similar to that used for verbs. The simple adjective, in contrast to the comparative and superlative inflections takes one argument, its theme ; it is thus treated as a verbal with one argument:

```
TROPICAL PDIGM T
          PROCESS
          ((LOC T1 T' T2 TROPIC)
           ( TEMP T1 (LOC T') T2 HOT)
           (RAINFALL T1 (LOC T') T2 HEAVY)
           (VEGETATION T1 (LOC T') T2 LUSH))
```

A generator given "tropical island" and this defining model will produce:

> THE ISLAND IS LOCATED IN THE TROPICS
> AT THIS LOCATION THE TEMPERATURE IS HOT,
> THE RAINFALL IS HEAVY AND THE VEGETATION IS LUSH.

The use of (LOC T') is required to refer to the Location of the Theme of the adjective so that in "tropical fish" it is the location, not the fish, that has the attributes of rainfall, temperature and vegetation.

Adverbs such as "first" and "newly" are apparently to be modelled as procedures to be applied to other models in a manner analogous to the way in which an adjective adds assertions to a noun model. Applied to the result of a forming process, "newly" signifies that the assertion (AGE T1 G' T2 YOUNG) will be added to the Goal argument of the process. "First" must add an assertion (TIMEORDER T1 X1 T2 FIRST) to whatever X1 it modifies. In general, we have very little knowledge at the moment of how to express comparative adjectives or adverbs in the models we devise.

Table 1 shows further examples of lexical entries for modelling other concepts in the two sentences. In general, it can be noticed that a conceptual definition makes available a set of assertions that correspond to the meaning of a word.

```
PLANT SUP ORGANISM
       HASPRT (STALK, ROOTS, LEAVES, BUDS, FLOWERS,
                      FRUIT, SEEDS)
       T* (GROW, LIVE, PLANT, FERTILIZE, HARVEST, EAT,
                      DIE, OXIDIZE)

       DURATION YEARS
       PROCESS  ( (ti (COVER T SEED, I SOIL) )
                  (t2 (GROW S SEED, G PLANT I (SOIL,
                            SUN, WATER)))
                  (t3 (APPEAR T PLANT S SOIL))
                  (t4 (GROW S PLANT G (GET "PLANT
                         HASPRT) I"))
                  (t5 (DIE T PLANT))
                  (tf (OXIDIZE T PLANT G SOIL I
                         MICROORGANISMS)))

SEED SUP ORGANISM
     ISPRT PLANT
     PROCESS (ti (GROW S PLANT T BLOSSOM))
             (t2 (FERTILIZE T BLOSSOM S (PLANT DET
                                          OTHER)
                I (OR WIND, INSECT, OTHER)) ))
             (T3 (MOVE T SEED S PLANT G ELSEWHERE I" ))
             (tf (GET "PLANT PROCESS )
     HASPRT (KERNEL SHELL)

APPEAR PDIGM  (T S G <animal>)
       PROC  (ti (SEE A G', T T', LOC S' MD NOT ))
             (t2 (OR (MOVE T T') (MOVE T G') (MOVE
                                          T S')))
             (tf (SEE A G', T T', MD POSSIBLE)))
```

TABLE 1 SOME LEXICAL STRUCTURES

MODELING SENTENCE MEANINGS

From the previous section it is apparent that lexical entries for underatanding English concepts are fairly complicated in structure and must be interpreted by a function which will bind values into the variables of the definition. Combining these definitional models into a structure that models the meaning of sentences is clearly a more complicated process and one for which only the beginnings of an understanding exist. Let us look again at the first example sentence:

> "One of the first plants to appear on the newly formed tropical island is the stately and graceful coconut palm".

If we imagine a movie that is consistent with our understanding of "appear", "form", "plant", etc., we might suppose the following sequence:

> UNINTERRUPTED OCEAN AND SKY
> A TURBULENT OCEAN FROM WHICH THRUSTS A FLAMING VOLCANO SPEWING LAVA
> LAVA FLOWS DOWN THE MOUNTAIN AND CONTACTS THE SEA AMID NOISE AND CLOUDS OF STEAM
> OVER DECADES THE VOLCANO QUIETS, THE ISLAND ENLARGES,
> OVER CENTURIES THE ROCK CRUMBLES INTO SOIL
> SOIL, SEEDS, SPORES ETC. DRIFT IN FROM WIND, WATER, AND BIRDS
> GREEN SPROUTS APPEAR IN THE SOIL
> SOME OF THESE SPROUTS GROW INTO TREES WITH FRONDS AND BEAR COCONUTS
> THE TREES ARE LARGE-GIRTHED, TALL AND GRACEFUL

Let us see how closely we can come to this sequence with an algorithm that uses the semantic structure and lexical entries such as those outlined. The abbreviated form of writing semantic structures will be used as much as possible in this exposition.

(X1 Q 1, SUP (X2 TOK PLANT, Q SOME, TIMEORDER FIRST, T* R1, T1* R3))

This says there is one X1 which is a member of X2 which is some subset of plants whose time of existence is earlier than that of the remainder. If we look at the process model for PLANT, we discover that t2 is the first subprocess that mentions the existence of a plant.

(t2(GROW S SEED G X1 I (SOIL SUN WATER)))

We have substituted X1 for the reference to PLANT, and T2 says that soil, sun and water help a seed to grow into X1 a plant. GROW is a verb that takes Source and Goal arguments and for change verbs of this type, the final subprocess states, (tf(APPEAR T G')). Thus, expanding GROW we can add the subprocess.

(t3(APPEAR T X1 S SEED))

A result of growing a plant is that a plant appears.

Now let us look at some more of the sentence.

(R1 TOK APPEAR, T X1, LOC X3, MD POSSIBLE)

The fact that the verb was an infinitive shows that the author is talking about a possible action rather than one that actually happened. We can bind the arguments of APPEAR into its subprocesses, but in the sentence we are given only Theme and Location arguments. First we get the cononical case structure for

APPEAR:

(APPEAR T S G)

A Theme can appear from some Source to some observer called the Goal of the appearance. All verbs can take Location and Time arguments, so they are not specified in the canonical forms. If we now consult the predications developed so far for X1 we discover (APPEAR T X1 S SEED) so we can fill in the Theme and Source arguments above and add the LOC given in the sequence to give:

(APPEAR T X1, S SEED G UNKNOWN LOC X3)

We can now bind these arguments into the subprocesses as below:

(ti(SEE A UNKNOWN S SEED T X1) LOC X3, MD NOT)

(t2 (OR (MOVE T T') (MOVE T A') (MOVE T G'))

(tf(SEE A UNKNOWN, T X1, LOC X3))

This says that initially X1 is not seen at X3 and finally it is. We might at this point ask why, go to the lexical entry for seed and discover that:

(ti BURY T SEED, I SOIL))

By expanding BURY we would find a final condition such as the following:

(tf(PREVENT S SOIL, T (SEE A UNKNOWN, T SEED))

showing that the soil the seed is buried in prevents us from seeing it as it first produces a sprout.

Let us now examine X3.

(X3 TOK ISLAND, Q 1, T* (FORM G X3, S SEABOTTOM, I(OR VOLCANO, CORAL) (TIM PAST, NEW))

By looking at the first subprocess of ISLAND in which an ISLAND is mentioned:

(ti(FORM S SEABOTTOM, G ISLAND, I(OR(VOLCANO, CORAL))))

We can add arguments to "newly formed island" to show from what and by what possible processes it came into existence. We can expand the meaning of "tropical" and add,

(X3 TOK ISLAND,...LOC TROPIC, VEGETATION LUSH, RAINFALL HEAVY, TEMPERATURE HOT)

Finally, we can add to X1 as follows:

(X1 SUP X2 ... SPECIES (X4 TOK COCONUT PALM, HEIGHT TALL, APPEARANCE GRACEFUL)).

Thus, we have used the lexical models to form a model of the meaning of the sentence which makes explicit many things not said which are important in understanding the statement.

Table 2 shows a representation for both this and the second sentence. On the right hand side of the table are shown the sentences which a generator can produce from these deep case structures. Not too surprisingly, the resulting text is not very close to the image description presented earlier; but it clearly demonstrates how appropriate lexical models can be expected to explicate some of the implicit meanings of sentences.

This section was introduced with the disclaimer that we have very little understanding of how to model sentence meanings. It is appropriate to reemphasize in closing that the procedures for using lexical models to form models of sentence meaning are still highly exploratory, as is structure and content of the lexical models that were shown.

(X1 SUP X2, Q1)	<u>One of</u>
(X2 TOK Plant, Q Some Tim FIRST)	<u>the first plants</u>
(ti (BURY T SEED, I SOIL))	Initially, a seed was buried in soil
(ti(MAKE T HOLE, LOC GROUND))	by making a hole in the ground
(t2(PUT T SEED, LOC HOLE))	putting it in the hole
(t3(PUT T SOIL, LOC SEED))	and putting dirt on it.
(tf (SEE T SEED, MD NOT))	finally the seed can't be seen)
(t2(GROW S SEED G X1, I (SOIL, SUN, WATER))	then the seed grows into a plant
(t3(APPEAR T X1, S GROUND))	and appears from the ground
(R1 TOK Appear, T X1, LOC X3, MD POSSIBLE)	<u>to appear on</u>
(ti(SEE T X1, LOC X3, MD NOT)	Initially the plant could not be seen on X3
(t2(OR (MOVE T X1) (MOVE T SOIL)))	Then the plant or the soil moved
(tf(SEE T X1, LOC X3, MD POSSIBLE))	Finally the plant can be seen at X3
(X3 TOK ISLAND, Q 1, LOC TROPIC, VEGETATION LUSH, TEMP HOT, RAINFALL HEAVY)	<u>a tropical island</u> - the island is in the tropics with lush vegetation, high temperatures and heavy rainfall
(ti (Form G X3, S SEABOTTOM, I (OR(VOLCANO, CORAL))))	Initially an island was formed by coral or volcano from the seabottom.
(R2 TOK FORM, G X3, S SEABOTTOM, I (OR(VOLCANO CORAL)) MD PAST, NEW))	<u>newly formed</u>
(ti(ACT T SEABOTTOM, I VOLCANO, MD PAST))	Initially a volcano acted on the seabottom
(t2(CHANGE S SEABOTTOM, G X3, I VOLCANO, MD PAST))	Then the seabottom changed into an island
(tf(APPEAR T X3, S SEABOTTOM, MD PAST, TIM, NEW))	Finally the island appeared from the seabottom.
(R3 TOK SPECIES, T1 X1, T2 X4, MD PRES, INDIC)	<u>One...is the stately and graceful coconut palm</u>
(X4 TOK COCONUT PALM, MOD (AND (STATELY GRACEFUL))) (SUP X4 TREE) (SUP TREE PLANT) (HEIGHT X4 100ft) (GIRTH X4 18ins) (DURATION X4 100yrs) (APPEARANCE X4 GRACEFUL)	The plant is a species of coconut palm which is a tree with a height of 100 ft, a girth of 18 inches and a lifetime of 100 years. It is tall and graceful.
(R3.5 TOK HASPRT, T1 X1, T2 X5)	The coconut palm has seeds
(X5 TOK SEED, T2* R3.5, SUP(X10 TOK COCONUT))	The seed is the coconut
(ti (GROW S PLANT, G BLOSSOM, SEED)))	which initially is grown from a coconut palm,
(t2 (DROP S PLANT, T SEED, G EARTH))	then dropped to earth,
(t3 (MOVE A UNK, T SEED, S PLANT, G UNK, I (WIND, WATER)))	moved away from the plant

(tf(GROW S SEED, G PLANT, I (SOIL, SUN, WATER)))	then grows a new plant
(R4 TOK EXCHANGE, T X5, S 'person', G-'person', I-'money')	The coconut people exchange for money
(R5 TOK PROTECT, T X5, S R6, I X7, MD SECURELY))	The coconut is securely protected by its husk and shell from the action of sea-water
(ti(THREATEN S R6, T X5))	Initially seawater threatens the coconut
(tf(PREVENT I X7 T (ACT S R6, T X5)))	then the husk and shell prevent sea water from acting on it.
(R6 TOK ACT S X6, T X5, MD POSSIBLE)	Seawater might act on the coconut
(ti(CONTACT T1 X6, T2 X5))	Initially seawater might contact the coconut
(tf(CHANGE I X6, T2 X5))	then the seawater might change it into something unknown
(x6 TOK SEAWATER)	Seawater
(ISPRT SEAWATER OCEAN)	Seawater is part of the ocean
(ti(FORM G Seawater, S (HYDROGEN, OXYGEN, SALTS), I HEAT))	Initially seawater is formed from H, O, and salts by heat.
(t2(OXIDIZE I SEAWATER, T SEED))	Then seawater oxidizes objects
(tf(FORM S Seawater, G(HYDROGEN, OXYGEN, SALTS), I HEAT))	Finally heat forms H, O, and salts from seawater
(x7 TOK AND, ARGS (X8, X9), ISPRT X5, I* X7)	The coconut has a husk and shell
(x8 TOK HUSK, THICKNESS THICK, MATERIAL FIBROUS...)	The husk is thick and fibrous
(x9 TOK SHELL, HARDNESS HARD...)	The shell is hard

TABLE 2 ANALYSIS OF TWO SENTENCES

DISCUSSION AND CONCLUSIONS

This paper has not been directly concerned with English parsers, generators or question-answering programs which are the systems that produce and use the data structures that have been presented. Each such system adds requirements to the lexical structure in the form of additional information such as syntactic word-classes, features, and rules of inference, but all can work with the same attribute-value list structure. In our uses each such system consults the lexicon under the control of a program-grammar written in a Woods (1970), Agumented Transition Network, (SEE Simmons, 1972, for detail). The data management problems are considerable. The lexicon must eventually comprise 10-20 thousand entries; the grammar-programs will include thousands of statements; storage for analyzed text must be in hundreds of thousands of words; and all this must be randomly and quickly accessible, and indexed from the lexicon.

In large experimental systems--which frequently run to over a hundred thousand words of program--much of the data management function is accomplished by an operating system that offers a virtual memory of several hundred thousand words and by advanced languages such as PLANNER, CONNIVER, CLISP, or QA4. These languages hash-code words, assertions, and procedures and access them with great ease and rapidity. They allow statement of property list structures in such high level forms as are illustrated in this paper. They accomplish some of the easier inference forms automatically as evidenced in the now classic Planner example cited by Winograd.

In this example, assume that some program process such as the semantic analyzer generates the question:
(QUERY FALLIBLE TURING)
which means, "Is Turing fallible?".
Previously, the Planner-type system has been given:

(THEOREM FALLIBLE X

((HUMAN X) TRUE)))

and (ASSERT HUMAN TURING)

The theorem says if X is human it is true X is fallible. The assertion states that Turing is human. A Planner-type system automatically accomplishes the deduction that Turing is fallible.

As this paper is being written, the natural language processing techniques, the design of appropriate data structures, the integration of efficient virtual memory accessing systems, and the realization of powerful languagès for deductive data management are all research fields that offer more questions than applications. Yet, as results progressively emerge from these fields, they may well revolutionize customary methods and programs for data management by lifting into the design of operating systems and programming languages those paging, interpretation and retrieval functions so invaluable now in data management systems.

REFERENCES

1. CHARNIAK, Eugene C. "Toward a Model of Children's Story Comprehension", AI TR-266, MIT, Cambridge, Mass., 1972.

2. COMPTON'S PICTURED ENCYCLOPEDIA. F. E. Compton & Co. Chicago, 1962.

3. HEIDORN, George E. "Natural Language Inputs to a Simulation Programming System", NPS-55HD, Naval Post Graduate School, Monterey, Calif., 1972.

4. HENDRIX, Gary G. "Modeling Simultaneous Actions and Continuous Processes", J. Artificial Intell. In Press.

5. HENDRIX, Gary G., Thompson, Craig and Slocum, Jonathan. "Language Processing via Canonical Verbs and Semantic Models", Proc. 3rd Int. Jt. Conference on Artificial Intelligence, Stanford Res. Institute, Menlo Park, Calif 1973.

6. RAMANI, S. "Language Based Problem-Solving", Tata Institute of Fund. Research, Computer Group, Bombay, 1969

7. SCHANK, Roger C., Goldman, N., Rieger, C. J., Riesbeck, C. Margie: Memory, Analysis, Response Generation and Inference on English", Stanford A. I Memo, Palo Alto, Calif. 1973

8. SIMMONS, Robert F., "Some Semantic Structures for Representing English Meanings", In Freedle,R.O. and Carrol, J. B. (Eds.), Language Comprehension and the Acquisition of Knowledge, Winston & Sons Inc., N.Y. 1972.

9. SIMMONS, Robert F., "Semantic Networks: Their Computation and Use for Understanding English Sentences", to appear in Schank, R. and Colby, K. (Eds.), Computer Simulation of Cognitive Processes, Prentice Hall, N.Y. 1973.

10. SIMMONS, R. F. & Slocum, J. "Generating English Discourse from Semantic Nets", Comm. ACM v15, #10, Oct. 1972.

11. WINOGRAD, Terry, Understanding Natural Language. New York: Academic Press, 1972.

12. WOODS, Wm. A. "Transition Network Grammars for Natural Language Analysis" Comm ACM V13, #10, Oct. 1970.