American Journal of Computational Linguistics *Microfiche 34 : 34*

SEMANTIC PROCESSING FOR SPEECH UNDERSTANDING

GARY G. HENDRIX

Artificial Intelligence Cente
Stanford Pesearch Institute
Menlo Park, California 94025

ABSTRACT

The semantic component of the speech understanding system being developed jointly by SRI and SDC rules out phrase combinations that are not meaningful and produces semantic interpretations for combinations that are. The system consists of a semantic network model and routines that interact with it. The net is partitioned into a set of hierarchically ordered subnets, facilitating the encoding of higher-order predicates and the maintenance of multiple parsing hypotheses. Composition routines, combining utterance components into phrases, consult network descriptions of prototype situations and surface-to-deep-case maps. Outputs from these routines are network fragments consisting of several subnets that in aggregate capture the interrelationships between a phrase's syntax and semantics.

Acknowledgement

This research was supported by the Defense Advance Research Projects Agency of the Department of Defense and monitored by the U.S. Army Research Office under contract No. DAHC04-75-C-0006.

OVERVIEW

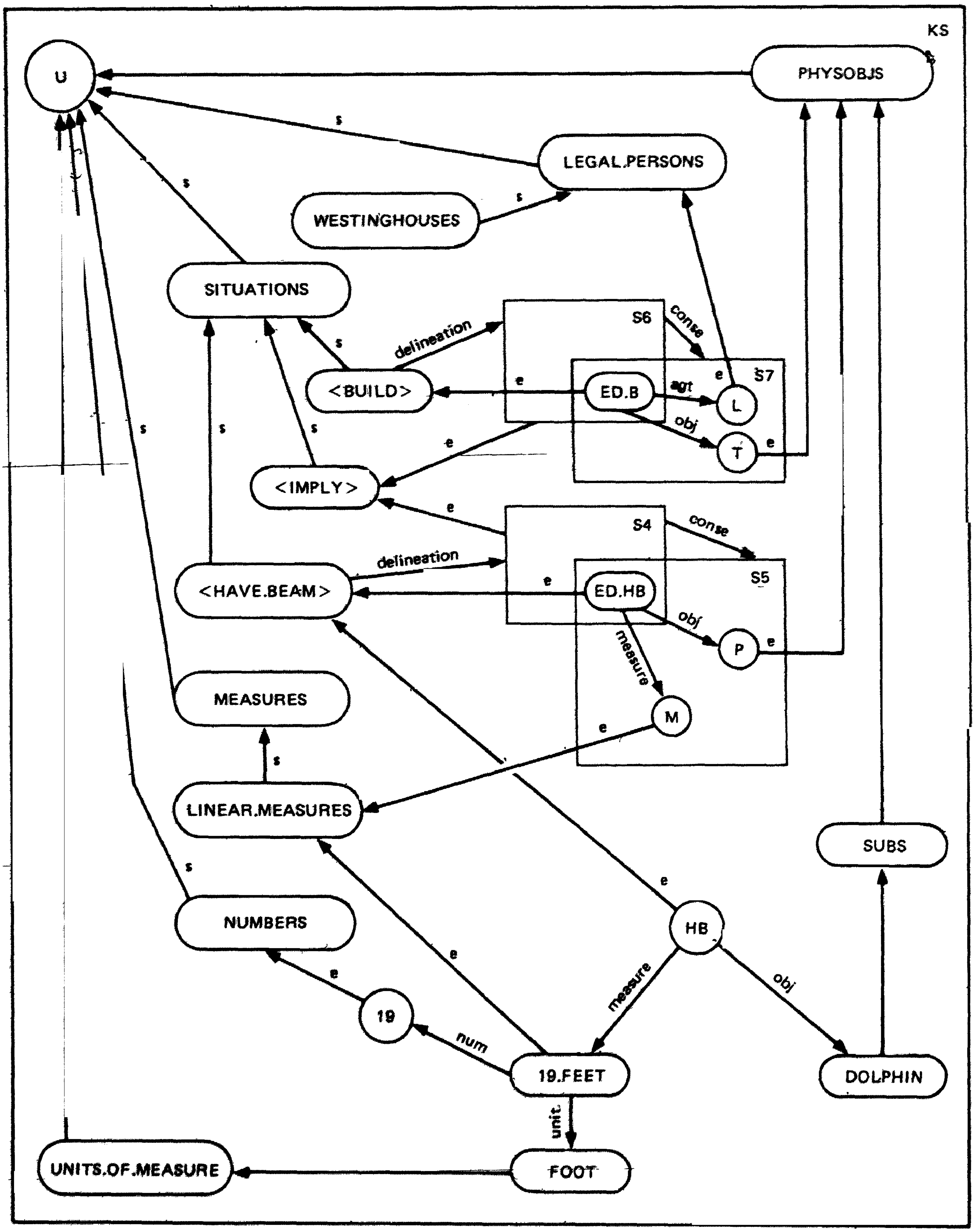
This paper describes aspects of the semantic component of the speech understanding system currently being developed jointly by SRI and SDC. (For a comprehensive discussion of nonacoustic portions of this system, see Walker et al., 1975.) The semantic component consists of two major parts: a semantic network coding a model of the task domain and a battery of semantic composition routines (SCRs) that are coordinated with the language definition (roughly, the "grammar" for the speech understanding system; see Paxton and Robinson, 1975, and Robinson, 1975). This paper concentrates exclusively on the interplay between these two major parts during parsing. However, the semantic component also plays important roles in knowledge management, discourse analysis, prediction, and question answering.

An SCR is called with network representations of components that the associated language definition rule has found to be syntactically capable of combining to form a larger phrase. Using knowledge from the semantic net, the SCRs eliminate combinations that, although syntactically acceptable, do not meet semantic criteria for meaningful unification. For combinations that are acceptable, the SCRs build network structures to represent the meaning of the composite phrase, using the network structures of the components as building blocks. These net structures are constructed so that (1) multiple hypotheses concerning the proper incorporation of a given utterance constituent in larger phrases may be encoded simultaneously in one net, (2) competing users of a constituent may share a single network structure representing the constituent, and (3) the association between each syntactic unit of an input and its translation image in the network is explicitly encoded for use in discourse analysis.

THE SEMANTIC NETWORK

The semantic network is the principal information source for SCRs, encoding such diverse entities as objects, situations, categories, taxonomies, definitions, and quantified statements. Network structures indicating possible relationships between objects are used to determine the meaningfulness of phrase combinations, while the network itself serves as the medium for recording interpretations of utterance fragments during parsing. The structure of this network differs from that of conventional nets in that nodes and arcs are partitioned into "spaces". These spaces, playing in networks a role roughly analogous to that played in strings by parentheses, group information into bundles that help to condense and organize the network's knowledge. An introduction to net partitioning is provided elsewhere (Hendrix, 1975).

An illustrative portion of the permanent knowledge section of the semantic network is depicted in Figure 1. In the upper left corner is node 'U', representing the universal set U. To the right is node 'PHYSOBJS', representing the set PHYSOBJS of



SA-3804-33

FIGURE 1 A SAMPLING FROM THE GENERAL KNOWLEDGE NET physical objects. That PHYSOBJS is a subset of U is indicated by the s-arc from 'PHYSOBJS' to 'U'. A subset of PHYSOBJS is SUBS, the set of all submarines. A particular element of SUBS, as indicated by the e-arc from 'DOLPHIN' to 'SUBS', is the DOLPHIN.

The DOLPHIN is a participant in a particular situation, HB, the situation in which the DOLPHIN has a beam of 19 feet. HB is an element of <HAVE.BEAM>, the set of all situations in which a physical object is characterized by a measure of its breadth. Certain outgoing arcs from a node representing a situation are used to specify situation attributes through deep semantic cases. For example, the outgoing obj-arc from 'HB' specifies the value of the "obj" (object) attribute of HB to be DOLPHIN. Hereafter the notation "#@obj" will be used to indicate "the value (#) of the attribute (@) obj."

The network of Figure 1 has been divided into five spaces, KS, S4, S5, S6, and S7. Pictorially, each of these spaces is represented by a box. The most global information in the network is encoded in space KS (the outermost box, sometimes called the "Knowledge Space") which includes such entities as nodes 'U' and 'PHYSOBJS' and the s-arc connecting them. The boxes representing spaces S4 through S7 may be thought of as holes in the box of KS. Paralleling the relationship between an inner and an outer block of an ALGOL program, each of these spaces specifies a more local area of the net than is specified by KS. From the perspective of S5, for example, it is possible to access both local node 'P' and (relatively) global node 'PHYSOBJS'. However, from KS the nodes and arcs inside S5 are not accessible. The hierarchy of space localization may be represented by a partial ordering such as that of Figure 2. From any space S, the nodes and arcs are accessible that lie in S or in any space S' above S in the hierarchy. For example, from S3 only nodes and arcs in S3, S2, S1, and KS are accessible.

Pictorially, it may be necessary to draw an arc crossing box boundaries. In such cases, the arc belongs to the space (or spaces) in whose box the arc label is written. Spaces may overlap. For example, in Figure 1, node 'ED,HB' lies in both space S4 and space S5. Further, a space may serve as a node in a more global space. Both S4 and S5 behave as nodes in KS and are connected by a conse-arc (consequence).

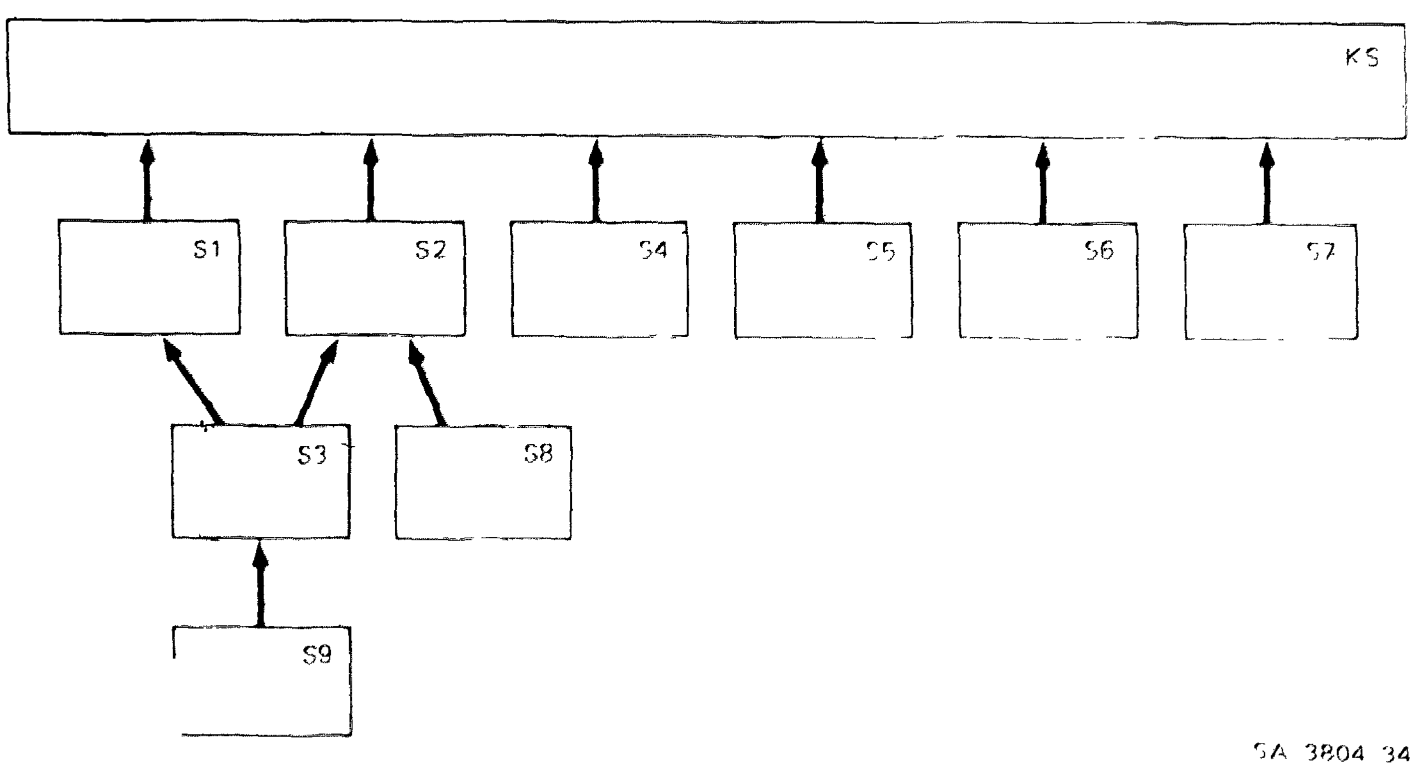


FIGURE 2 SPACE LOCALIZATION HIERARCHY

Typically, localized spaces such as S4 and S5 are used to encode higher-order "predicates," such as quantifiers, logical connectives, and hypothetical data. Here, S4 and S5 are used to encode an implication. The space S4, doubling as a node in space KS, is connected by an e-arc to '<IMPLY>' and by a conse-arc to 'S5'. The interpretation of any element of set <IMPLY> is that if entities can be found matching the structure of the element space, then the existence of entities matching the structure of the associated conse space may be inferred. The only structure encoded in element space S4 is a node ('ED.HB') with an e-arc to '<HAVE.BEAM>'. This structure matches any concrete instance of <HAVE.BEAM> (such as HB). Thus, for any instance of <HAVE.BEAM>, entities matching the structure of S5 must exist. The structure of S5 indicates that the element of <HAVE.BEAM> will have a #@obj, which is an element of PHYSOBJS, and a #@measure, which is an element of LINEAR.MEASURES.

The implication encoded by S4 and S5 serves to delineate the set <HAVE.BEAM>. That is, the implication indicates all the attributes (deep cases) of a <HAVE.BEAM> situation and their ranges of acceptable values. This delineation may be used during parsing to test the plausibility of a given group of entities being united in a <HAVE.BEAM> situation or, in a predictive mode, to suggest possible sentence participants. Such delineations are encoded for every situation and event set known to the system, a second example in Figure 1 being the delineation of set <BUILD>.

THE SYSTEM IN ACTION

The use of the SCRs and semantic network in translation may be seen by considering the parsing of

"The power plant of the sub was built by Westinghouse."

The ultimate result of the translation process for this utterance is the network structure recorded in the SCRATCH space of Figure 3. Structures representing new inputs are constructed in a scratch space (or spaces) to prevent them from becoming confused with the system's permanent knowledge (recorded in KS). Since the system understands new inputs by appealing to previous knowledge, there are many links, in the form of e-arcs, from the SCRATCH space into KS. (Note: Only a fragment of KS is shown in the various figures of this paper.)

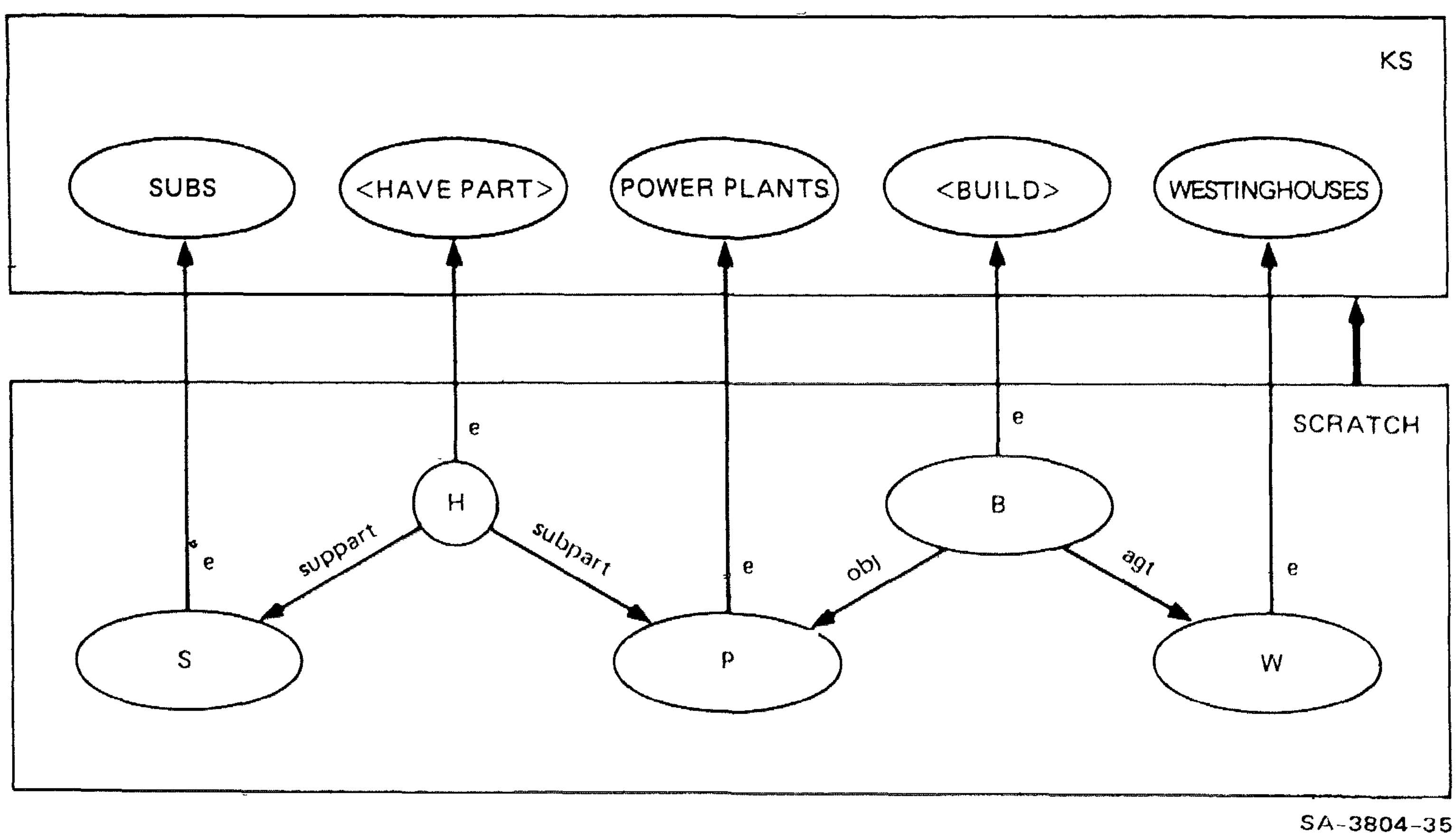


FIGURE 3 PARSE TARGET STRUCTURE FOR "THE-POWER-PLANT OF THE-SUB WAS-BUILT BY WESTINGHOUSE"

The interpretation of the network in the SCRATCH space is as follows: Node 'B' represents an element of the set <BUILD> of building events in which a #@agt W built a #@obj P. The agent W of the building event is an element of the set of WESTINGHOUSES. The #@obj built by W is P, an element of the set POWER.PLANTS. According to node 'H', this power plant is the #@subpart in a <HAVE.PART> relationship in which S, the particular member of SUBS currently in context, is the #@suppart (superpart). Discourse analysis mechanisms discussed in Deutsch (1975) and, more fully, in Walker et al. (1975) will be used to associate W with the unique Westinghouse Corporation known to the semantic net in space KS. The other definite NPs ("the sub" and "the power plant of the sub") will likewise be resolved.

To suppress secondary details while considering the building of this structure, assume the highly simplified language definition:

Grammar		Lexicon
R1:	S => NP VP	NP: the-power-plant,
R2:	NP => NP PREPP	the-sub, Westinghouse
R3:	VP => VP PREPP	VP: was-built
R4:	PREPP => PREP NP	PREP: of, by

(Note: "the-power-plant" is not treated as an NP in the actual system. Rather, NOM "power plant" is first combined with PREPP "of the sub" and only afterward is "the" appended to produce the NP "the power plant of the sub".)

In the translation process, spaces are created to represent the semantics of each grammatically defined constituent of the total utterance. These spaces are shown in Figure 4 with heavy arrows indicating the space hierarchy.

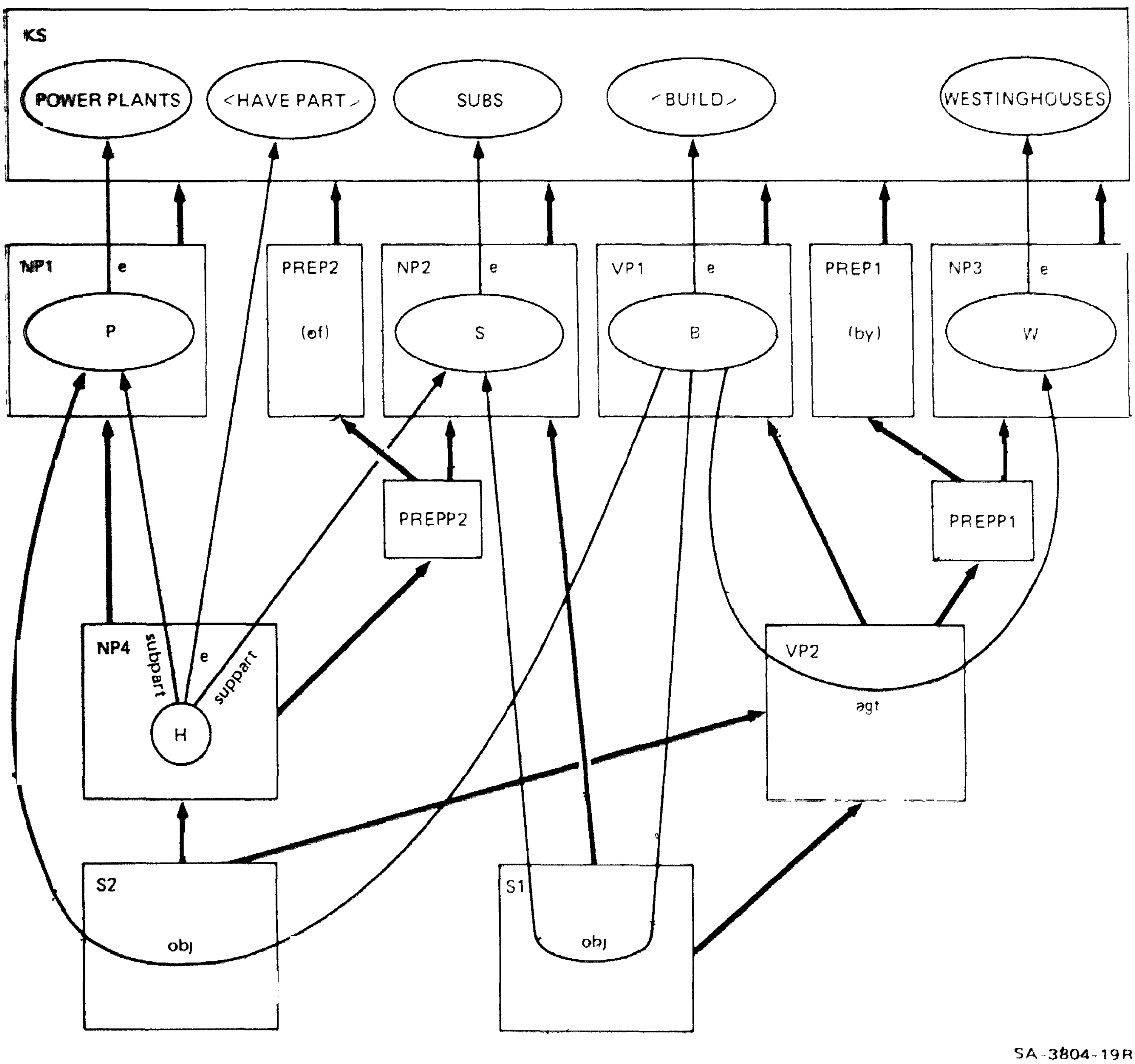


FIGURE 4 MULTIPLE SCRATCH SPACES FOR "THE-POWER-PLANT OF THE-SUB WAS-BUILT BY WESTINGHOUSE"

At the start of processing, space KS contains knowledge about power-plants, <HAVE,PART> relationships, submarines, <BUILD> events, and Westinghouse. On spotting the noun phrase "the-power-plant", an SCR is called to set up a space, NP1, below KS in the partial ordering. Within this space, a structure is created representing the meaning of "the-power-plant". Similarly, new spaces are set up to encode the other sentence constituents that correspond to explicit lexical entries.

As the parser groups subphrases into larger units, SCRs are called to aid in the process. Using rule R4, PREPP1 ("by") and NP3 ("Westinghouse") are combined to form PREPP1 ("by Westinghouse").[1] PREPP1 is allocated its own space, but no new structures are created within it.

When syntactic considerations suggests combining VP1 ("was-built") with PREPP1, the appropriate SCR is called. Consulting a surface-to-deep-case map associated with the lexical entry for the verb "build", the SCR determines that a "by" PREPP following the verb often signals the deep agt case in a passive construction. Operating under this hypothesis, the SCR checks the voice of VP1. Passing this test, the SCR next checks the semantic feasibility of the NP of PREPP1 serving as the #@agt in a <BUILD> event. To do this, the SCR consults the #@delineation of <BUILD> in space KS (see Figure 1). The delineation is encoded as an <IMPLY> situation in terms of spaces S6 and S7. As discussed earlier, this delineation indicates that any #@agt of a

<BUILD> situation must be an element of LEGAL.PERSONS. The candidate for the #@agt position is W of space NP3. Since W is an element of WESTINGHOUSES and WESTINGHOUSES is a subset of LEGAL.PERSONS, W is accepted. A construction such as "built by the submarine" would have been rejected.

Once VP1 and PREPP1 have passed the acceptability tests, a new space, VP2, is constructed to encode the resultant VP. This new space links node 'B' of VP1 with node 'W' of NP3 via an agt-arc. This new arc is accessible only from space VP2 (and lower spaces in the hierarchy) and is not accessible from either VP1 or NP3. This leaves the components encoded in VP1 and NP3 free to combine in alternatives to VP2 if need be.

Continuing the parse, NP2 ("the-sub") is combined with VP2 ("was-built by Westinghouse") to form S1, after passing tests similar to those above. The obj-arc linking the constituent phrases of S1 is contained in space S1 and hence is inaccessible from the spaces of the constituents. Notice that the construct "the-sub was-built by Westinghouse" which is encoded by S1 is a spurious interpretation of utterance components.

Using rule R4, PREP "of" may be combined with NP2 to form PREPP2. The network structures accessible from PREPP2 do not include the (spurious) obj-arc from 'B' to 'S' that lies in space S1.

When the syntax of rule R2 suggests combining NP1 and PREPP2 to form a new NP ("the-power-plant of the-sub"), an SCR is called. This SCR checks NP1 to see if it is relational in nature (as is "beam" in "beam of the Dolphin") and hence expecting an argument to be supplied. Since NP1 fails this test, the SCR checks the properties of the PREP "of" and discovers that it may be used to encode <HAVE,PART> situations. Calling upon the delineation of <HAVE,PART> and appropriate surface-to-deep-case maps, the SCR determines this to be a legitimate interpretation and hence builds space NP4 with a node 'H' and three arcs as shown. While these new constructs are accessible from space NP4, they are inaccessible from constituents NP1 and PREPP2 (and NP2). Furthermore, they cannot be accessed from spurious space S1; hence the construction of NP4 has not altered the view of the net from S1.

Using rule R1, S2 is constructed from NP4 and VP2. In addition to the obj-arc contained in space S2 itself, the view of the net from S2 includes all the information accessible from either space NP4 or space VP2 and hence is identical to the view from space SCRATCH of Figure 3. Since the parse corresponding to space S1 does not successfully account for the fragment "the-power-plant of", it is rejected, and S2 is accepted as expressing the meaning of the input.

The partial ordering of spaces from S2 to KS indicated in Figure 4 is identical to that represented more clearly in Figure 5, which, because of the choice of space labels, may be recognized as the parse tree of the input sentence. The syntax of the input and the association between each syntactic unit and its corresponding semantics has therefore been captured in the structures built by the SCRs.

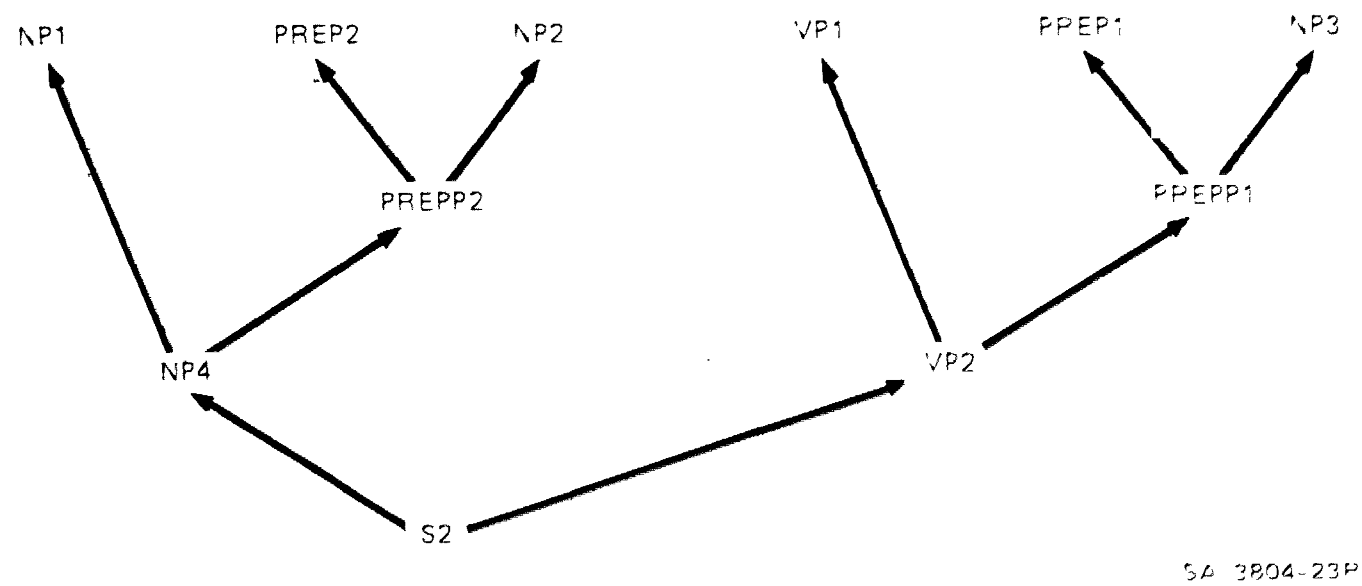


FIGURE 5 SPACE HIERARCHY ABOVE S2

DISCUSSION

Partitioning is a recent innovation in semantic networks. As shown above, this new feature enables networks to maintain alternative hypotheses (e.g., S1 and S2) concerning the use of utterance constituents and enables such competing hypotheses to share network subparts (e.g., VP2). Without partitioning, the back-linked nature of networks causes a constituent to be altered when it is incorporated into a larger unit and hence renders it unusable in alternative constructions. The highly ambiguous nature of acoustic input makes these abilities to maintain alternative hypotheses and share substructures especially important in speech understanding.

Partitioning also allows selected portions of a network to be associated with syntactic units, showing the correspondence between network entities and the syntactic structures that were used to communicate them. As discussed in the section on "Discourse Analysis and Pragmatics" in Walker et al. (1975), this association is crucial in analyzing the elliptic utterances that are so characteristic of speech.

References

Deutsch, Barbara G. Establishing Context in Task-Oriented Dialogs. Presented at the Thirteenth Annual Meeting of the Association for Computational Linguistics, Boston, Massachusetts, 30 October - 1 November 1975.

Hendrix, Gary G. Expanding the Utility of Semantic Networks Through Partitioning. Advance Papers of the Fourth International Joint Conference on Artificial Intelligence, Tbilisi, Georgia, USSR, 3-8 September 1975, 115-121.

Paxton, William H., and Robinson, Ann E. System Integration and Control for Speech Understanding. Presented at the Thirteenth Annual Meeting of the Association for Computational Linguistics, Boston, Massachusetts, 30 October - 1 November 1975.

Robinson, Jane J. A Tuneable Performance Grammar. Presented at the Thirteenth Annual Meeting of the Association for Computational Linguistics, Boston, Massachusetts, 30 October - 1 November 1975.

Walker, Donald E., et al. Speech Understanding Research. Annual Report, Project 3804, Artificial Intelligence Center, Stanford Research Institute, Menlo Park, California, June 1975.