THE FINITE STRING

VOLUME 15   NUMBER 5

With this issue, David G. Hays completes his term as Editor of AJCL and breathes a sigh of relief. Personal matters have made the last two issues of AJCL for 1978 excessively late. The next issues of AJCL will appear on paper; but the circumstances of the moment suggest that digital magnetic recording and direct wire transmission will be suitable for experimental use shortly.

CONTENTS

COMPUTATIONAL LINGUISTICS IN THE USSR

Joyce Friedman

ABSTRACT

As part of an official U.S./USSR Science Exchange on Applications of Computers in Management, a subgroup on natural language processing visited the Soviet Union from May 28 through June 11, 1978. The group met with scientists in Moscow, Novosibirsk, Leningrad, and Kiev. There were formal meetings and presentations of technical material, and also many informal discussions. This report presents a view of Soviet computational linguistics which emerged from these discussions.

Background

The U.S./USSR Science Exchange on Applications of Computers to Management includes many subtasks. The exchange in natural language processing is one task under the topic "theoretical foundations for software in applications in economics and management". The exchange in natural language processing was to have begun in June 1977. However, a scheduled trip by U.S. scientists was cancelled at the last minute by the USSR side; the reason given was that there were no hotel rooms available in Moscow. In spite of this initial disappointment the exchange began in November 1977 when three Soviet scientists visited the United States for two weeks. The visitors were Alexander Narin'vani of the Academy of Sciences Computing Center in Novosibirsk and Victor Briabrin and Dmitri Pospelov of the Academy of Sciences Computing Center in Moscow. The trip reported in this note is the rescheduled visit by the U.S. delegation. It took place May 28 to June 11, 1978.

The members of the U.S. delegation were: Donald Aufenkamp, N.S.F., U.S. Chairman of the U.S./USSR Joint Working Group on Scientific and Technical Cooperation in the Application of Computers to Management; Sue Bogner, H.E.W.; Joyce Friedman, Department of Computer and Communication Sciences, The University of Michigan; John Makhoul, Bolt Beranek and Newman, Inc., Cambridge; Stanley Petrick, Mathematics Department, IBM T. J. Watson Research Center, Yorktown Heights; Sally Sedelow, Departments of Linguistics and Computer Science, University of Kansas; and Walter A. Sedelow, Departments of Sociology and

Computer Science, University of Kansas. The U.S. delegation was accompanied throughout the trip by A. S. Narin'yani of Novosibirsk.

This report groups together similar work done in different locations. The main patterns of the natural language processing and theorem-proving systems can be viewed as based on (1) linguistics, (2) artificial intelligence, or (3) logic, although the distinctions are to some extent arbitrary. We also give in overview of the computers and programming languages available for work in computational linguistics. Work on lexicography, thesauri, and speech recognition was also discussed on the visit, but is not covered in this report.

(1) Linguistically-based Work on Natural Language

The main roots of the linguistically-based work are the meaning-text model of Mel'chuk, dependency grammar, and transformational grammar. They are variously interpreted by different systems.

Zoya Shalyapina, Laboratory of Machine Translation, Institute of Foreign Languages, described an English to Russian machine translation system under development since 1972 and based primarily on the meaning-text model. The representation is a dependency tree, with word order information, morphology and semantic/syntactic valencies. This structure preserves all the surface data but is also close to a semantic representation of the text. There is a dictionary and a grammar for each language.

The grammar rules are of the two forms: If <structure> then <condition>, and if <stucture> then <transformation>. Semantic information includes semantic descriptions of lexical and morphological units and the semantic acceptability of word pairs. There is a dictionary of 10,100 lexemes, described in terms of 30 semantic primitives The syntactic and semantic structures are compatible, so analysis goes only as deep as is necessary for a given sentence. Shalyapina's group works on linguistic aspects only; there is no computer implementation.

Uri Apresyan also works with the meaning-text model and with machine translation as the goal. His work is primarily on French to Russian translations, but he also works on English. His English grammar is said to be the most complete ever published: the Russian grammar will soon appear. The linjuistic model will have four parts: morphology, deep syntax, surface syntax, and semantics; however, the current reduced model lacks semantics. A dictionary gives for each ord its morphology, its syntactic and semantic features (there are 150 syntactic features; 500 semantic features), the semantic criteria for possible governing words, and selectional restrictions. Rule schema or "syntagmas" go from morpheme structure to a surface syntactic structure that is an unordered dependency tree. There are about 200 syntagmas for Russian, each representing 20 rules. A syntagma allows a tree with X over Y to be constructed from a string containing X and Y under various complex conditions. The lexical information and the syntagmas determine the transformation from word strings to surface-syntactic structure. A deep structure is then defined by "paraphrastic" rules, which convert, for example, strike to

deliver when the object is a blow. The deep structure is no longer language-specific but is universal, and serves as the basis for translation between languages. Apresyan stressed the value of continuing to work on the same linguistic model in order to complete its development; he contrasted this with the attitude of some current American linguists.

The linguist Iakalev, of the Economics Institute is developing a natural language interface for a data base system. This work has computer support and is to be running soon in a large factory. The natural language subset has sentences such as "what is the number of workers of <type> in <place>?" and is said to be easy for economists to learn. The system is based on very recent models of transformational grammar: Iakalev mentioned "traces" and some of Jackendoff's theories. The system goes from input to a deep structure from which it constructs a formula for the computation of a numerical result.

(2) Artificial Intelligence Work in Natural Language

AI-based systems are being developed at the Computing Center of the Academy of Sciences at Moscow, under the direction of Victor Briabrin and at the Computing Center of the Siberian Division of the Academy of Sciences, Novosibirsk under the direction of Alexander Narin'yani, in Ershov's group.

The system demonstrated to us in Moscow was DILOS (Dialogue Information Logical System). This work is heavily influenced by artificial intelligence work in the U.S. (Briabrin spent seven

months at L.I.T. working with William Martin and with Carl Hewitt.) DILOS is written in LISP and runs on the BESM-6 computer in Moscow, as well as on a PDP-11/45 at the International Institute for Applied Systems Analysis in Laxenburg, Austria. The system is intended both to test various approaches to natural language processing and for practical applications. It contains an ATN linguistic processor and a semantic processor based on frames. The current applications area is airline ticket reservations; the demonstration was however on a very small data base of AI. Natural Language Systems (including DILOS, SUS, RPL, OWL, and LUNAP). The system was able to answer simple natural language questions from the data base but it was not possible from the demonstration to get a good feeling for the actual range of language accepted.

Narin'yani's group in Novosibirsk has 17 people, including 6 linguists and 9 mathematicians and programmers. Until a few years ago, the work followed Mel'chuk's model. This has now been abandoned here and work proceeds along four lines, so far relatively independently: (1) Narin'yani is developing a _formal linguistic model_ which combines dependency and constituent structure in a mixed multi-level representation. Analysis proceeds by local modification of the graph structures, expanding and compressing case frames at various levels. The linguistic model so far includes formal description of adverb groups and adjective groups. This formal model has now been written up, but so far is not implemented. (2) The semantic question-answering system VOSTOK-0 contains a formal model of time. On the basis of texts of sentences such as "From the 3rd up to the 10th of March

Mike was in Moscow" it answers questions'like Where was Mike at noon on the 17th of March?". The system is coded in SETL and was demonstrated to us. While the natural language fragment is still small, even for a model of time, (e.g. no time adverbials), the inferencing scheme worked successfully. (3) Several "applicational" systems are being developed. The first of these, the PL-1 "mini" or "toy" system ZAPSIB-0 uses essentially no syntactic analysis (though it relies heavily on word order). It has a well-defined subject domain, a data base of personnel information, and answers questions such as "who under 30 earns more than average?" (Salary information is public in the USSR.) In this very limited subject domain, the approach works well. The "midi" applicational system is under development and is more syntactically oriented. It will contain a nondeterministic bottom-up parser for a binary context-sensitive grammar with discontinuous constituents. (4) The final subgroup is the programming language group; it has implemented SETL on the BESM-6.

(3) Logic-based work in Natural Language.

In Moscow, at VINITI, the linguist E. B. Paducheva and the mathematician T. D. Korelskaya are developing jointly an approach to natural language processing based on both transformational grammar and first-order logic. The current domain is converse theorems in geometry. The system is able to process geometry theorems and produce their "converse theorems". In this system the semantic representation language is first-order logic. Algorithmic procedures for analysis and synthesis have been

developed, as well as processing procedures within the logic. The linguistic part of the method is based on transformational grammar. As is the case with most of the Soviet work on transformational grammar, the deep structure uses dependency grammar rather than constituent structure grammar. The transformations are originally written in the forward direction, i.e. from deep to surface structure. Analysis is done using a "reversed" version of each transformation (not obtained automatically). While the forward transformations are independent of order the reversal rules are strictly ordered, for efficiency. There are 30-35 transformations, each expressed as a structural description, given as a template, and a structural change, given as a sequence of elementary operations. The work is developed in detail, but has no computer implementation. The system is said to contain interesting solutions to problems of quantification, negation, and conjunction reduction. The authors reported, with some amusement, that the description of the work was printed in 42,000 copies.

The current work at the University of Leningrad under G Tseitin, Faculty of Engineering and Mathematics, was described to us by others as based on logic, but Tseitin himself took a philosophical approach in his discussions with us. His remarks were more suggestive than descriptive. He indicated that his approach to natural language was by analogy to programming languages, using macros as in operating systems. He claimed "that there is no such thing as meaning", but said that his approach did use procedural semantics. His previous work on

complexity and theorem-proving is not related to his work on natural language. However, heid & argue 'that a natural language system for computers should reflect the fact that natural language performance by people does not require exponential time. Tseitin's own current work is not on natural language, as he is busy writing an ALGOL68 implementation.

Tseitin and Liakina, formerly of the Facdlty of Philology, also talked about several earlier natural language systems which I am unable to distinquish. They are described in a number of publications from 1966 on. In gereral, they employ dependency grammars, and use transformations during syntactic analysis. Restrictions on the grammar are stated in the predicate calculus and resolution theorem-proving is used. The goal is English to Russian translation of scientific prose.

The system of J. Kapitonova, Head of the Laboratory of Applied Cybernetics at the Institute of Cybernetics at Kiev, is an interactive theorem-proving system for mathematical texts. The objective is to be able to fill in the standard gaps in proofs, as indicated by "it is obvious that" or "as in the proof of the previous Theorem". The text is first processed manually into a highly stylized mathematical language. Only the formal material, theorems and proofs, is analyzed; discussion is treated as comment and is ignored by the programs. Several large texts, including Curtis and Reiner <u>Algebraic Theory of Groups</u>, have been preprocessed. The theorem-prover is \tailored to the specific mathematical domain. It uses resolution theorem-proving, heuristic techniques, as well as special mathematical and logical

techniques. The system has been programmed and is about to be tried out on a recent thesis. This project is of ten years duration, and has had a minimum of 12 people.

Interest in Montague grammar was considerable. My talk in Moscow was very well attended, and there were many good questions. The audience was generally familiar with Montague's work and with recent papers on the topic in <u>Artificial Intelligence</u> and <u>Theoretical Linguistics</u>. The interest seemed to come from a more general interest in logic as a knowledge representation in natural language systems. Agafanov in Novosibirsk is also interested in the possible applications of Montague grammar to programming languages.

Computing in Computational Linguistics

Computer access appears to be much more difficult to obtain for computational linguists in the Soviet Union. Many of the projects had no computer support, even though they were in areas where computer testing of grammars or theories could be very useful. Most of the computing was on the second-generation computer BESM-6, although there are more recent computers, e.g., the ES-EBM (Ryad) series, available for other purposes. U.S. computers were on order from Hewlett-Packard, CDC, and Burroughs. The terminals we saw were mainly graphics terminals from Eastern Europe, with both Roman and Cyrillic character sets, and seemed fine in use.

There is much interest in advanced programming languages.

SETL is implemented in Novosibirsk. (This is with the aid of the U.S./USSR Science exchange.) In Moscow, PASCAL is implemented. In Leningrad, Tseitin is implementing ALGOL68 for the Ryad series of computers, compatible with the IBM 360.

We did have occasion to see some interactive systems in operation. The languages were impressive, but the programmer support was not. There seemed to be few error diagnostics. When there were crashes it was not possible to tell which were due to the computer and which to the programs.

Conclusions

Work on natural language processing in the USSR seems to be along three major lines. The work by linguists is motivated by machine translation. It relies on versions of Mel'chuk's meaning-text model, with some type of transformations on a dependency base. It is characterized by a great deal of sophisticated development of large grammars, by large groups of linguists, but is without computer support. The artificial intelligence work is directed toward data base information systems, is at an earlier state of development, and is heavily based on U.S. work. It is carried out in Computing Centers and has good programming and computer support. The logic-based work in carried out by individuals or small groups in several locations without computer support, and by one large group with computers.

CONTENTS

CONTENTS

Appendix II. Properties of Lexical Relations.

a. Reflexivity, Symmetry, Transitivity.

Certain properties of lexical-semantic relations can be very useful in deductive inference. For instance, if we know that a cheetah is a kind of mammal and a mammal is a kind of vertebrate then we can deduce that a cheetah is a kind of vertebrate. Writing T for the taxonomy relation, we can abbreviate this sentence: if cheetah T mammal and mammal T vertebrate then cheetah T vertebrate. Whenever bTc and cTd, it follows that bTd. This fact can be described much more efficiently by the statement that the taxonomy relation is transitive. Two other commonly mentioned properties of relations are reflexivity and symmetry. These properties may apply to predicates formed from lexical entries as well as to lexical-semantic relations.

To be precise, a relation R defined on a set S is said to be a *transitive* relation if whenever b and c are R-related and also c and d are R related then b and d stand in a relation R also. Synonymy is a transitive relation just as transitivity is. The preposition *in* behaves in the same way. If Sam is in the kitchen and the kitchen is in the hotel, then we know that Sam is in the hotel. The time interrelation *before* behaves like this, too. If Zorro arrived before the posse did and the posse arrived before the explosion, then we know that Zorro arrived before the explosion.

A relation R defined on a set S is said to have the *reflexive* property if all the elements of S are R-related to themselves, that is, if mRm is true for all members m of the set S. The synonymy relation has this property: a word means the same thing as itself. The antonymy

relation ANTI does not have this property. It is not true that *hot* ANTI *hot*, for example.

A relation R defined on a set S is said to be *symmetric* if whenever, b and c are R-related then so are c and b; that is, R is symmetric if and only if bRc always implies cRb. Synonymy also has this property. If b is synonymous with c, then c is synonymous with b. So has antonymy. Given that *hot* ANTI *cold*, we immediately know that *cold* ANTI *hot*. Taxonomy is not symmetric, however. A lion is a kind of mammal, but a mammal is not a kind of lion.

In question answering we may be just as interested in drawing negative conclusions as positive ones. Thus it may be important to know that if bRc is true then cRb must be false. The term *asymmetric* is used to describe a relation R for which bRc and cRb are never both true, at least when b and c are different elements of the set S. Taxonomy is asymmetric and so is the time interrelation *before*. If the question asks, "Did c happen before b?" and we know that b happened before c, we can answer with a confident no. For want of a better term we will say that the relation R is *non-symmetric* if it is neither symmetric or asymmetric. In this case bRc and cRb are sometimes both true and sometimes not. Similarly, the term *irreflexive* is used for the case in which mRm is never true, while the term *nonreflexive* is used for the case in which mRm is sometimes true and sometimes not. In the same way *intransitive* is taken to mean that if bRc and cRd, we can conclude that b and d are not R-related, while *nontransitive* will mean that bRd is sometimes true if bRc and cRd, but not always.

Each lexical relation itself has a lexical entry. The reflexivity, symmetry, and transitivity properties of the relation are listed in this

entry, as they are in the entries for interrelational operators and prepositions and other lexical items for which they are relevant. There are also lexical entries under the property names, *reflexive, irreflexive,* etc. listing the appropriate axioms. The motivation behind lexical entries for properties is first of all greater generality. Secondly, it makes it much easier to add lexical relations and to add other properties which turn out to be useful.

At this stage of development there are several transitivity axioms:

For lexical relations Rel, like taxonomy

$$b \text{ Rel } c \wedge c \text{ Rel } d \rightarrow b \text{ Rel } d$$

For interrelations J, like *before*

$$Holds(I(J,Z_1,Z_2)) \wedge Holds(I(J,Z_2,Z_3)) \rightarrow Holds(I(J,Z_1,Z_3))$$

For prepositions Q like *in* or *above*

$$Holds((F(location,Z_1,Prep(Q,Z_2))) \wedge Holds(P(location,Z_2,$$

$$Prep(Q,Z_3))) \rightarrow Holds(P(location,Z_1,Prep(Q,Z_3)))$$

Intuitively these are all instances of the same concept, transitivity. There should be some single way of expressing it. It is a defect of this representation system that there is not.

A relation that is reflexive, symmetric, and transitive is called an equivalence relation. The synonymy relation is an equivalence relation since it has all three properties. If R is an equivalence relation, then a subset consisting of all the elements which are R-related to a particular element x by the equivalence relation is called an equivalence class. In an equivalence class all the elements are R-related to each other. An equivalence relation partitions a set into equivalence classes; each element

of the set belongs to exactly one equivalence class. The synonymy relation partitions the items in the lexicon in just this way. There is a class consisting of *suspicion* and all the words synonymous with *suspicion*, like *mistrust* and *doubt*. These synonymy classes are disjoint; each word sense in the lexicon belongs to exactly one of them (cf. Edmundson and Epstein 1972, Palmer 1976).

With this as a basis an equivalence relation of paraphrasability between sentences can be established. Sentence S_1 is a paraphrase of sentence S_2 if one is obtained from the other by substituting synonyms for each other.[1]

> Mr. Kennedy viewed Lady Laura with suspicion.

> Mr. Kennedy regarded Lady Laura with mistrust.

We might also allow substitution of conversives, nominalizations, etc.

> Nancy was Sally's student.

> Sally was Nancy's teacher.

> Sally taught Nancy.

The equivalence classes of this relation, each one of which is the set of all paraphrases of a given sentence have a definite theoretical importance and some practical significance in question answering. One member of a class might well be part of the story; another the right answer to a question.

---

[1] This representation system can be viewed as defining a relation P such that S_1 P S_2 if and only if S_1 and S_2 have the same representation. If the representation system is well defined, then P should define the same equivalence classes as the paraphrasability relation

b. *Inverses.*

The inverse R of the relation R is the relation which "goes in the opposite direction" from R; that is, bRc if and only if cRb. Thus, *bake* T *make* and *make* T *bake* are two ways of saying the same thing. Both pieces of information are stated in the lexicon. However, the lexical entry for *bake* includes T *make;* the lexical entry for *make* includes T *bake.* Why bother to say the same thing in different places? There are two reasons for this. First of all, the inverse relation may be a relation that is commonly and easily verbalized, worth naming in its own right. This is certainly true of the CHILD relation, as in *puppy* CHILD *dog.* Instead of asking "What is a baby dog called?", we could ask "What is a grownup puppy called?" or "What does a puppy grow up to be?" The second reason is that putting this information in both entries can make searches easier and much faster. We may only have one half of the pair and need the other. We may have *dog* and *puppy.* This is easy if we have the information CHILD *puppy* in the *dog* entry. Otherwise we might have to search the whole lexicon. In other situations we have two words but no direct connection between them. For example, suppose the system knows *lion* T *mammal* and *mammal* T *vertebrate* and is then asked, "Is a lion a vertebrate?" The connection between *lion* and *vertebrate* can be found much more quickly if the search starts from both the *vertebrate* end and the *lion* end of the chain at the same time, but to do this there must be a pointer to *mammal* in the *vertebrate* entry. Another question comes to mind. Why call the inverse relation to CHILD by the clumsy name $\overline{\text{CHILD}}$ instead of its proper name PARENT? The ECD uses two different names for

a relation and its inverse (S_0 and V_0 are inverses, for example). If this were done here, two versions of the appropriate axiom schemes would be needed, one in the CHILD entry and one in the PARENT entry.

Since a relation R is called symmetric if bRc always implies cRb, it follows that a symmetric relation is its own inverse. The synonymy relation S and antonymy relation ANTI are both self-inverse in this sense. For this reason we never need the symbol $\overline{\text{ANTI}}$, etc. $\overline{\text{ANTI}}$ is ANTI   The entry for *hot* includes ANTI *cold,* the entry for *cold* includes ANTI *hot.*

c. Unique Linkage.

Raphael (1968) has proposed a property which seems extremely useful. He calls it *unique-linkage* (U). Mathematicians usually refer to such relations as one-to-one. A relation R has the unique-linkage property if whenever xRy then bRy is false for any $b \neq x$ and xRc is false for any $c \neq y$, i.e. any object is R-related to at most one other. Raphael's example of unique-linkage is the relation "just to the right of". The behavior is especially characteristic of the queuing relation, e.g. with days of the week, Monday Q Tuesday, etc. Some relations may be uniquely linked on one side only, e.g. mother-child is uniquely linked on the left. We can define U_L unique-linkage on the left and U_R unique linkage on the right. (A relation which is U_R is a single-valued function. If R has the U_L property, then its inverse is a single-valued function.)

Raphael also proposed for SIR-1 (ibid, p. 101) a property which he calls irreflexive. R is set-nonreflexive if

$$(\forall X \quad M) \sim (\forall \beta \subset X) \; (\exists \alpha \subset X) \; (\alpha \; R \; \beta)$$

In the SIR model both the 'X is a part of Y' and the 'X is owned by Y'

relations have this property. What it says is that every set in the model has a minimal element with respect to the relation R. A simpler version of this property is sufficient for our purposes.

Minimum $(\forall X \subseteq M) \sim (\forall Y \subset X)(\exists Z \in X)(ZRY)$
Condition   Every nonempty subset has a minimum.

Maximum $(\forall X \subseteq M) \sim (\forall Y \subset X)(\exists Z \in X)(YRZ)$
Condition   Every nonempty subset has a maximum.

The part-whole relation has both properties in our model. In any non-empty subset in the model there is something in it that is not a proper subpart of anything else in that subset, and also something that has no proper subpart. A relation that has this property stops somewhere. It is not reflexive and not circular. A search that goes on looking for links of this kind will stop somewhere. The relation 'is an ancestor of' has this property. We will eventually run out of ancestors in one direction and descendants in the other, at least, inside a finite model.

The properties of relations are summarized in Table 4.

Table 4. Properties of Relations

Property	*Definition*
symmetric	$(\forall X \in M)(\forall Y \in M)(XRY \rightarrow YRX)$
asymmetric	$(\forall X \in M)(\forall Y \in M)(XRY \rightarrow \sim YRX)$
reflexive	$(\forall X \in M)\ XRX$
irreflexive	$(\forall X \in M) \sim (XRX)$
transitive	$(\forall X \in M)(\forall Y \in M)(\forall Z \in M)(XRY \wedge YRZ \rightarrow XRZ)$
intransitive	$(\forall X \in M)(\forall Y \in M)(\forall Z \in M)(XRY \wedge YRZ \rightarrow \sim (XRZ))$
uniquely linked	$(\forall X \in M)(\forall Y \in M)(XRY \rightarrow (\forall Z \in M)((ZRY \rightarrow X=Z)$ $(XRZ \rightarrow Y=Z)))$
uniquely linked on the left	$(\forall X \in M)(\forall Y \in M)(XRY \rightarrow (\forall Z \in M)(ZRY \rightarrow Z=X))$
uniquely linked on the right	$(\forall X \in M)(\forall Y \in M)XRY \rightarrow (\forall Z \in M)(XRZ \rightarrow Z=Y))$

94

d. *Partial Ordering.*

Any transitive relation defines a partial ordering. Several of the lexical relations discussed earlier are transitive; many lexical items are transitive too. One important reason for representing time in terms of the transitive interrelation *before* is to allow one to make the same kinds of simple deductions about time that one can make about taxonomy. Some transitive relations, like taxonomy, are also reflexive. In this case we talk about a *weak ordering.* ($X \le Y$ for numbers is a weak ordering.) Some are not reflexive, these are called *strong* ordering relations. ($X < Y$ for numbers is a strong ordering.) The time relation *before* is a strong ordering relation. For any weak ordering there is a strong ordering and conversely. Starting with the taxonomy relation T, for example, a relation T_1 or "proper taxonomy" can be defined consisting of the pairs x and y for which xTy but x and y are different. Then xT_1y means that x is a kind of y but different from y. If instead one starts with a strong ordering relation *before*, one can define a weak relation "$before_1$" for which x $before_1$ y means that either x before y or x cooccured with y.

The queuing relation Q is not itself a partial ordering but a partial ordering can be derived from it. Monday Q Tuesday and Tuesday Q Wednesday, but it is false that Monday Q Wednesday. Queuing is an 'immediate successor relation like the relation between a natural number n and the next number n+1. A relation Q' can be defined such that xQ'y if either xQy or there are some objects $z_1, z_2, \ldots, z_n$ such that xQz_1, z_1Qz_2, $\ldots z_nQy$. It follows immediately that if bQc and cQd then bQ'd. Q', the 'successor' relation,

is now transitive, for if bQ'c and cQ'd, then one can find a chain of
Q-related objects linking b and d just by concatenating the chain
linking c and d.   Raphael's pair of relations jright and right behave
this way.   The relations "is a child of" and "is a descendant of" are
also paired in this way.

MODELS OF THE SEMANTIC STRUCTURE OF DICTIONARIES

Kenneth C. Litkowski
16729 Shea Lane
Gaithersburg, Maryland 20760

CONTENTS

Figures

Table

3

SUMMARY

Ordinary dictionaries have not been given their due, either as sources of material for natural language understanding systems or as corpora that can be used to unravel the complexities of meaning and how it is represented. If either of these goals are ever to be achieved, I believe that investigators must develop methods for extracting the semantic content of dictionaries (or at least for transforming it into a more useful form).

It is argued that definitions contain a great deal of information about the semantic characteristics which should be attached to a lexeme. To extract or surface such information, it will be necessary to systematize definitions and what they represent, probably using semantic primitives. In this paper, I describe procedures which I have developed in an attempt to accomplish these objectives for the set of verbs in Webster's Third New International Dictionary (W3). I describe (1) how I have used the structure of the dictionary itself in an attempt to find semantic primitives and (2) how it appears that the systematization must incorporate a capability for word sense discrimination and must capture the knowledge contained in a definition.

The body of the paper is concerned with demonstrating that semantic information can be surfaced through a rigorous analysis of dictionary definitions. The first step in this process requires a comprehensive framework within which definitions can

be analyzed. In developing this framework, we must remember
that each word used in a definition is also defined in the dic-
tionary, so that we must be able to uncover and deal with vi-
cious circles. The framework must also be capable of represent-
ing traditional notions of generative grammar to deal with the
syntactic structure of definitions. suitable framework ap-
pears to be provided by the theory of labeled directed graphs
(digraphs).

Using points to represent dictionary entries and lines to
represent the relation "is used to define", two models of the
dictionary are described. From these models and from digraph
theory, we can conclude that there may exist primitive units of
meaning from which all concepts in the dictionary can be
derived.

To determine primitive concepts, it is necessary to sub-
ject definitions to syntactic and semantic parsing in order to
identify characteristics that should be attached to each defi-
nition. Syntactic parsing such as that implemented for systemic
grammar by Winograd is the first step. semantic parser must
next be developed. It appears that definitions themselves, and
particularly definitions of prepositions (which are used to ex-
press sense relations), will be of significant help in develop-
ing such a parser. Further work is necessary to develop proce-
dures for surfacing from definitions information about the con-
text which must be associated with each sense. It appears as if
this parser will have more general use for ordinary discourse.

5

These notions lead to the ultimate model of a dictionary, where points represent concepts (which may be verbalized and symbolized in more than one way) and lines represent relations (syntactic or semantic) between concepts.

Based on these models, procedures for finding primitive concepts are described, using the set of verbs and their definitions from W3. Specific rules are described, based on some elementary graph-theoretic principles, structural characteristics of dictionary definitions, and the parsing of the definitions. These rules have thus far reduced the initial set of 20,000 verbs to fewer than 4,000, with further reduction to come as all rules are applied.

It is argued that this approach bears a strong relationship to efforts to represent knowledge in frames. Although much work is needed on the parser and on a computerized version of this approach, there is some hope that the parser, if expectations are borne out, will be capable of transforming ordinary discourse into canonical frame representations.

6

1. INTRODUCTION

During the past 15 years, scientists in many fields have
been building a reservoir of knowledge about the semantic char
acteristics of natural language. Perhaps somewhat inexplicably
these developments have for the most part ignored the semantic
content of dictionaries, despite the fact that even a small one
contains a vast amount of material. Some attempts have been
made to dent these repositories, but the steps taken have been
tentative and have not yet borne significant fruit, perhaps be-
cause the sheer volume and scope of a dictionary is so over-
whelming. As a result, most studies have dealt with only a few
definitions without a comprehensive assault on the whole. While
such studies have led to many insights, it seems that the full
usefulness of a dictionary's contents will be realized only
when a comprehensive model of its semantic structure is devel-
oped.

Any system intended to provide natural language under-
standing must necessarily include a dictionary. If any such
system is to achieve broad applicability, its dictionary must
cover a substantial part of the natural language lexicon. For
this to occur, the developers of a system must either create a
dictionary from scratch or be able to incorporate an existing
dictionary. Given the amount of effort that usually goes into
development of an ordinary dictionary, the former alternative
is rather impractical. However, little has been done toward
meeting the latter alternative; with what follows, I will

describe the approach which I believe must be followed in transforming the contents of an ordinary dictionary for use in a true natural language system.

In order to be used in a language understanding system, a dictionary's semantic contents must be systematized in a way that the sense in which a word is being used can be identified. Before this can be done, it is necessary to characterize what is already contained in each definition. To do this, it seems necessary to write the meaning of each definition in terms of semantic and syntactic primitives. My purpose in this paper is (1) to describe how to use the dictionary itself to move toward identification of the primitives, at the same time (2) showing how this process can be used (a) to provide the capability for discriminating among word senses (i.e. characterizing the frames into which a given word sense will fit) and (b) to characterize knowledge contained or presupposed in a definition.

Before embarking on the description, it is necessary to point out some limitations which should be kept in mind as the reader proceeds. First, in trying to present an overview of my approach, I have had to forgo describing the detailed steps which I have followed to date. Second, even had I presented a full description, I would still have been short of providing sufficient details to enable computer implementation of any procedures. Third, since the approach presumes that concepts represented by the lexicon are tne realizations of many as yet unknown recursive functions to be discovered by stripping away

one-layer at a time, results other than procedures to be used
in stripping will not emerge until all layers have been re-
moved. (However, I do argue that the "stripping" procedures are
inherently useful, in that they will constitute a parser even
in the intermediate stages.) Fourth, since I have not had ac-
cess to a computer, which has become essential for significant
further progress, I have been unable to determine how far the
procedures I have developed would take me, so there is an in-
herent uncertainty as to how much further development is needed.
Notwithstanding these limitations, I am hopeful that what is
presented will provide a satisfactory framework for further in-
vestigations into the contents of dictionaries. I will comment
further on these limitations and how they might be overcome at
the end of the paper.

2. ATTITUDES TOWARD DICTIONARIES

Many of the significant contributors to the present under-
standing of meaning (such as Katz and Fodor 1963, Fillmore 1968
and 1971, Chafe 1970, Jackendoff 1974, Winograd 1972, and
Schank 1972) have generally ignored dictionaries. Yet, each has
presented a formulaic structure for lexical entries to serve as
a basis for the creation of a new dictionary Although their
perceptions about the nature of language are well-established,
their formalisms for lexical entries have not taken advantage
of the equally well-established practices of lexicography.

The rationale underlying the development of new formalisms'
expressed in some cases and implicit in others, is that lexical

entries in dictionaries are unsatisfactory because they do not contain sufficient information. These formalisms thus require that semantic features such as "animate" or "state" be appended to particular entries. While it is true that ordinary dictionary entries do not overtly identify all appropriate features, this may be less a difficulty inherent in definitions than the fact that no one has developed the necessary mechanisms for surfacing features from definitions. Thus, for example. "nurse" may not have the feature "animate" in its definition, but "nurse" is defined as a "woman" which is defined as a "person" which is defined as a "being" which is defined as a "living thing"; this string seems sufficient to establish "nurse" as "animate". In general, it seems that, if a semantic feature is essential to the meaning of a particular entry, it is similarly necessary that the feature be discoverable within the semantic structure of a dictionary. Otherwise, there is a defect in one or more definitions, or the dictionary contains some internal inconsistency. (Clearly, it is beyond expectation that any present dictionary will be free of these problems.)

The possibility of defective definitions has also generated criticisms, more direct than above, on the potential usefulness of a dictionary. On one hand definitions are viewed as "deficient in the presentation of relevant data" since they provide meanings by using "substitutable words (i.e. by synonyms), rather than by listing distinctive features" (Nida 1975:172). On another hand, the proliferation of meanings

attached to an entry is viewed as only a case of "apparent

polysemy" which obscures the more general meaning of a lexeme

by the addition of "redundant features already determined by

the environment" (Bennett 1975:4-11). Both objections may have

much validity and to that extent would necessitate revisions to

individual or sets of definitions. However, neither viewpoint

is sufficient to preclude an analysis of what actually appears

in any dictionary. It is possible that a comprehensive analysis

might more readily surface such difficulties and make their

amelioration (and the consequent improvement of definitions)

that much easier.

Even though dictionaries are viewed somewhat askance by

many who study meaning, it seems that this viewpoint is influ-

enced more by the difficulty of systematically tapping their

contents than by any substantive objections which conclusively

establish them as useless repositories of semantic content.

However, it is necessary to demonstrate that a systematic

approach exists and can yield useful results.

3. PREVIOUS RESEARCH ON DICTIONARIES

Notwithstanding the foregoing direct and indirect criti-

cisms. some attempts have been made to probe the nature and

structure of dictionary definitions. A review of relevant as-

pects of two such studies will help the material presented here

stand out in sharper relief.

Olney 1968 describes the conceptual basis of many project-

ed routines for processing a machine-readable transcript of

<u>Webster's Seventh New Collegiate Dictionary</u> (W7). The primary
objectives of these routines were the development of

> "(a) rules for obtaining certain of the senses described
> for W7 entries from other senses described for the
> same entries or from senses described for other W7
> entries from which the first (at least in typical
> cases) were derived morphologically; and
>
> (b) semantic components and rules for combining them to
> yield specifications of senses that cannot convenient-
> ly be obtained by rules referred to in (a) above."
> (ibid.:6)

Although these objectives are reasonable, they do not take ad-
vantage of the possibility that the semantic structure of a
dictionary might be a unified whole. As a result, any routines
that are developed seem to require the serendipitous perception
of patterns. Further, if a dictionary does have a unified se-
mantic structure, it is not clear that a rule relating meaning
to form will be relevant to a model of the semantic structure
even though interesting results might emerge. It seems neces-
sary to have some comprehensive view that will permit us to
know whether a particular rule is well-formed. This lack of ob-
jective criteria also imperils any analysis that selects a sub-
set of definitions for detailed analysis. The selection of a
subset of the dictionary should arise from well-defined a pri-
ori considerations rather than an intuition that a particular

subset seems to be related. An example of this intuitive approach appears in Simmons 1975 and 1976.

In Quillian 1968, the analysis of dictionary definitions was part of a study of semantic memory, and for that reason was not concerned with the full development of a dictionary model. In that study, a person determined the meaning of a concept when he "looked up the 'patriarch' word in a dictionary, then looked up every word in each of its definitions, then looked up every word found in each of those, and so on, continually branching outward until every word he could reach by this process had been looked up once." This process was never actually carried out because (1) not all words in a dictionary were used in the computer files, (2) the process was terminated when a common word was found in comparing the meanings of two words, and (3) there was a belief that there are no primitive word concepts. The termination of a search as designed was necessary in any event since, without any restrictions, it is likely that a large part of the dictionary would have been reached on every occasion. More importantly, Quillian did not fully consider what was happening when branching led to a word already encountered, namely, that a definitional circularity was thereby uncovered Such circularities which might be vicious circles, must be treated specially (as will be shown below), and hence, Quillian's unrestricted branching should have been modified. Quillian also overlooked the possibility that a concept common to two patriarchs is more primitive than either. The continued

comparison of more and more primitive concepts, along with re-
strictions on the outward branching, implies that primitive
concepts actually do exist.

Based on these observations, I take, as a working hypoth-
esis, the assumption that a dictionary may be a unified whole
with underlying primitive concepts.[1] With this beginning, it is
necessary to articulate a model of the dictionary which will
permit an identification of the primitive concepts through the
application of well-defined rules or procedures. It is proposed
that what follows constitutes the first steps toward meeting
this objective.

4. DESCRIPTION OF DICTIONARY CONTENTS

Since a dictionary contains much material, it is first
necessary to delineate exactly what is to be modeled.[2] For this
purpose, it is assumed that the semantic content of a dictio-
nary essentially resides within its definitions, thereby ex-
cluding from formal analysis such things as the pronunciation,
the etymology, and illustrative examples. s presently con-
ceived, the analysis will focus on the word being defined
(hereafter called the main entry), the definitions (including
sense numbers and letters used as delimiters), part-of-speech

---

[1] No dictionary is likely to satisfy this assumption, which is
only a theoretically desirable characteristic. The assumption
enables us to exclude the definienda from the models.

[2] In the interests of space, I have glossed over a large number
of intricacies that would have to be dealt with in arriving
at a machine-readable transcript suitable for analysis.
Several pages would be required to describe them fully.

labels, status or usage labels, and usage notes. The manner in which these features will be employed will be made clear as the analysis proceeds.

The hypothesized unified nature of a dictionary arises from the fact that definitions are expressed by words which are also defined[3] (i.e., there is no semantic metalanguage). If we wish to understand the meaning of a given definition, then we must first understand the meanings of its constituent words. Since each constituent corresponds to a main entry, then, in order to understand the meaning of the given definition, we must understand the meaning of the constituent words' defini- tions. Continued repetition of the process is nothing more than the outward branching process described by Quillian; however, as mentioned before, we must make this branching more disci- plined in order to deal with vicious circles and avoid unwanted circularities.

If we are to have a fully consistent dictionary, its model must show how each definition is related to all others. Thus, for each definition, X, the model should enable us to identify (1) those definitions of the constituent words of X that apply and those that do not apply, and (2) the production rules that generated X from these definitions. For example, in the defini- tion of the noun broadcast, "the act of spreading abroad",[4] it

---

[3] There are some exceptions to this assertion, such as proper names, biological taxa, and other special symbols, as pointed out by the Journal's referee.

is necessary that the model indicate (1) which of the defini-
tions of <u>the</u>, <u>act</u>, <u>of</u>, <u>spread</u>, and <u>abroad</u> apply, and (2) the
production rules by which <u>the</u> and <u>act</u> and all other colloca-
tions) occur together. If this can be done for each definition
in the dictionary, and if any inconsistencies are reconciled,
then, as will be shown, it should be possible to find the prim-
itive concepts in the dictionary and to transform each defini-
tion into a canonical form.

5. <u>BASIC MODEL</u>

The theory of (labeled) directed graphs (digraphs)[5] is
used as the formalism for the models. Digraph theory deals with
the abstract notions of "points" and "directed lines"; its
applicability to the problem before us therefore depends on how
these notions are interpreted. In this respect, it is important
to distinguish the manner in which this theory is used here
from the manner in which it previously has been used in seman-
tics and linguistics. The two most common uses are (1) where
trees display phrase and syntactic structures (cf. Katz and
Fodor 1963), or (2) where directed graphs portray the sequen-
tial generation of words in a sentence or phrase (cf. Simmons
1972). In these cases and others (cf. Quillian 1968 and Ben-
nett 1975) graphs are used primarily as a vehicle for display

---

[4] All definitions used in this paper are taken from <u>Webster's
Third New International</u> <u>Dictionary</u>, Encyclopaedia Britannica,
Chicago, 1965.

[5] Terminology for digraphs follows Harary 1965.

and no results from graph theory are explicitly employed to draw further inferences. However, as used here, graphs consti tute an essential basis for the analysis and hence will play an integral role in a number of assertions that are made.

In the simplest model, a point can be interpreted as representing all the definitions appearing under a single main entry; the main entry word can be construed as the label for that point. The part-of-speech labels, status or usage labels, and usage notes are considered integral to the definitions and may be viewed as part of a set of characteristics of the individual definitions. A directed line from x to y will be used to represent the asymmetric relation "x is used to define y"; thus, if the main entry x appears exactly or in an inflected form in a definition of y, then xRy. (This does not preclude a distinct line for yRx or xRx.) Therefore, we can establish a point for every main entry in a dictionary and draw the appropriate directed lines to form a digraph consisting of the entire dictionary. (This digraph may be disconnected, but probably is not.) An example, which is a subgraph of the dictionary digraph, is shown in Figure 1 on the next page. Except for broadcast, only the labels of each point are shown, but each represents all the definitions appearing at its respective main entry. The directed line from act to broadcast corresponds to the fact that "act is used to define broadcast", since its token appears in "the act of spreading abroad". In this model, the token "spreading" is not represented by a point, since it is not a main entry.

17

broadcast   (the act of
                                        spreading abroad)

the   act   of   spread   abroad
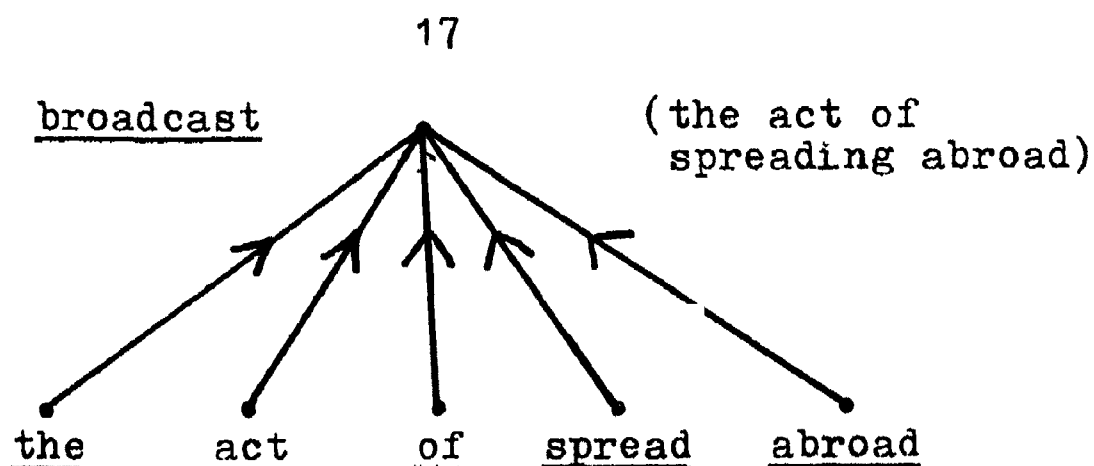
Figure 1.   A typical subgraph of the dictionary
digraph using the basic model.

Since the definition shown is not the only one for broadcast,
this point has additional incoming lines which are not shown.

The resultant digraph for even a small dictionary is ex-
tremely large, perhaps consisting of well over 100,000 points
and 1,000,000 lines. Clearly, such a digraph provides little
fine structure, but even so, it does have some utility. The
manner in which it can be used is described in Section 9.

6. EXPANSION OF THE MODEL: POINTS AS DEFINITIONS

Letting each point in the basic model represent all the
definitions of a main entry provides very little delineation of
subtle gradations of semantic content. As a first step toward
understanding this content, it seems worthwhile to let each
point represent only one definition. However, the basic model
will not trivially accommodate such a specification (primarily
because of the interpretation given to the directed line), and
thus it must first be modified.

In the basic model, the existence of a line between two
points, x and y, asserts that xRy, i.e., "x is used to define
y". Since the points represent all the definitions under the

main entries, the existence of a line arises from the simple
fact that x appears in at least one of y's definitions. If the
point y represents only one definition, say y_j, there is no
difficulty in saying that xRy_j. However, if we wish every point
to represent only one definition, then we must find the defini-
tion of x, say x_1, for which x_1Ry_j is true. Referring to the
subgraph in Figure 1, this amounts to determining, for example,
which definition of <u>abroad</u> is used to define the token "abroad"
in "the act of spreading abroad", that is, finding the i such
that "$\underline{abroad}_1$Rthe act of spreading abroad" or
"$\underline{abroad}_1R\underline{broadcast}_j$".

It should be intuitively clear that interpretation of
points as single definitions is desirable. However, there are
no a priori criteria by which the appropriate value of i can be
determined, and hence there is no immediate transformation of
the basic model into a model where each point represents one
definition. Since this objective is worth pursuing, it is there-
fore necessary to develop criteria or rules according to which
the desired transformation can be made.

In the application of rules that may be developed, it will
be convenient to make use of a model intermediate between the
basic one and the one with points as definitions. For this pur-
pose, we can combine the two models by employing a trivial re
lation, x_1Rx, which says that the ith definition of x is used
to define x; this holds for all definitions of x. The line re-
flecting xRy would remain in the model, so that the digraph

broadcast   (the act of
spreading abroad)

the   act   of   spread

abroad

$\underset{\substack{(\text{over a} \\ \text{wide area})}}{\text{abroad}_1}$ $\underset{\substack{(\text{at} \\ \text{large})}}{\text{abroad}_2}$ $\underset{\substack{(\text{widely} \\ \text{apart})}}{\text{abroad}_3}$ • • • abroad_n
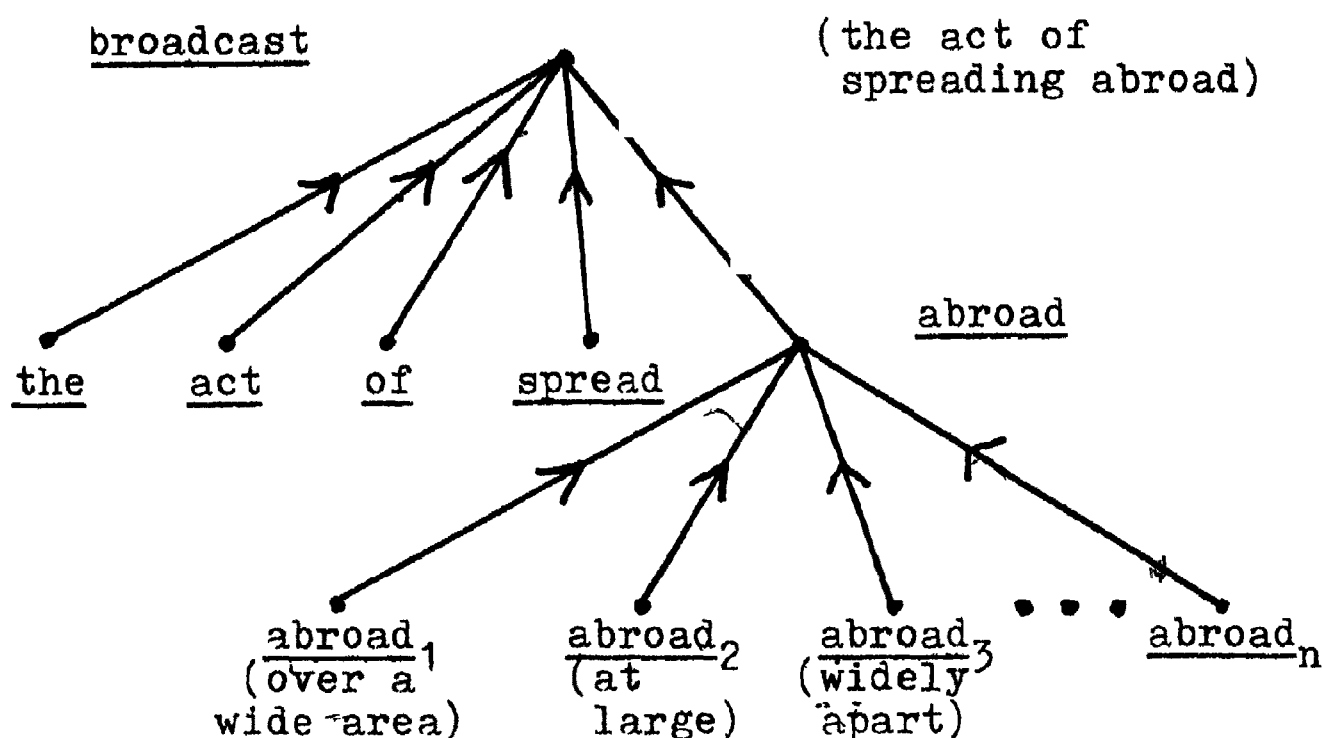
Figure 2.   Subgraph of model with points representing
both single and multiple definitions.

would show both $x_1 Rx$ and xRy_j and x would be a carrier, as il-
lustrated in Figure 2. In this case, the unsubscripted abroad
represents all the definitions of abroad (only some of which
are shown). If and when suitable criteria establish, for ex-
ample, that abroad_1, but not abroad_2, abroad_3,..., fits the
context of the token "abroad" in the definition of broadcast,
it would then be possible to draw a line directly from abroad_1
to broadcast without the intermediation of the unsubscripted
point abroad, thus eliminating paths from abroad_2, abroad_3,...,
to broadcast.

This model thus includes the points of the basic model and
adds points to represent each individual definition in the dic-
tionary. The lines between these points ensure that no relation
in the basic model is lost. As described in the example, it is
necessary to develop rules according to which the points repre-

senting more than one definition can be eliminated or bypassed, so that the only relations, xRy, that remain are such that x and y are points which represent one definition.

It may happen during the application of rules that some lines to a carrier will be eliminated with more than one still remaining. In such a case, it will still be useful to modify the digraph as much as possible. For example, if xRy in the basic model, where x has m definitions and y has n, and xRy_j in the expanded model, then $x_1, \ldots, x_m Ry_j$. It may be that some criterion indicates that, say $x_1, x_2 Ry_j$ but that $x_3 \ldots, x_m Ry_j$. When this occurs, we can create two points x_a and x_b such that $x_1, x_2 Rx_a \ x_a Ry_j$, and $x_3, \ldots, x_m Rx_b$, but with no line from x_b to y_j, as illustrated in Figure 3. The utility of this type of

broadcast

abroad$_a$

abroad$_b$

abroad$_1$ abroad$_2$ abroad$_3$ abroad$_4$ abroad$_n$
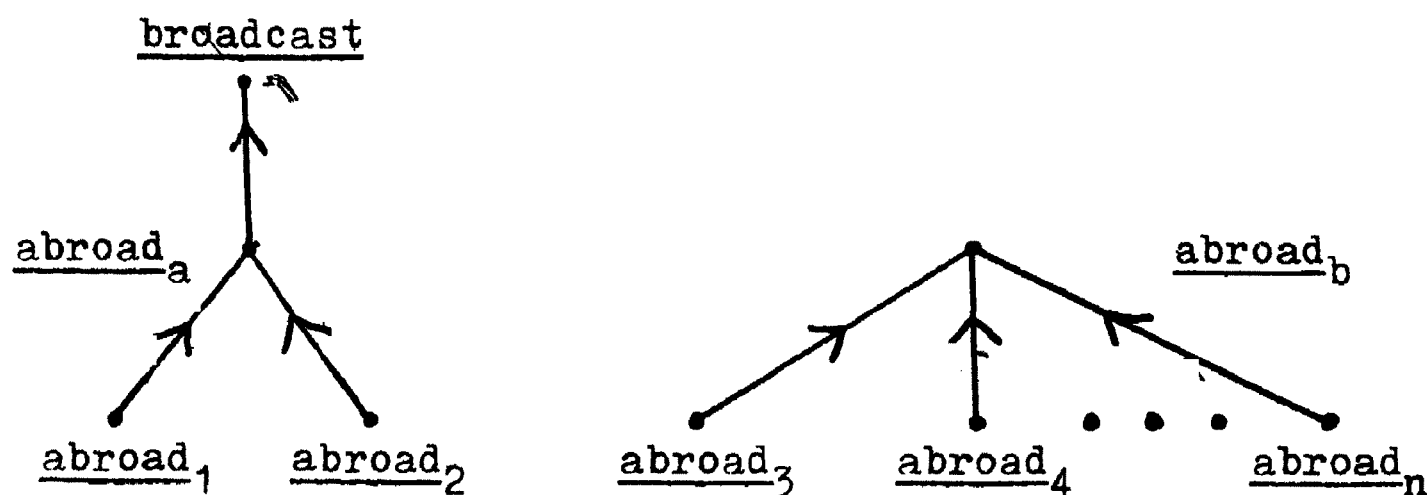
Figure 3.   Subgraph of expanded model
with grouping of definitions.

grouping will be demonstrated in Section 9. In any event, since many criteria will eventually be required in the elimination of points representing two or more definitions. this ability to group definitions is a necessary mechanism for modeling intermediate descriptions of the dictionary. (It should be noted

here that all such points will not be eliminated; those that re-
main will indicate an essential ambiguity in the dictionary;
this is further discussed in Section 8.)

7. <u>SEMANTIC, STRUCTURAL, AND SYNTACTIC PARSING OF DEFINITIONS</u>

The basic and expanded models, exampled in Figures 1, 2,
and 3, do not portray any of the meaning of the dictionary, but
rather indicate where particular relationships exist. In fact,
these two models portray only the relation "is used to define"
as if there is no other relation between definitions. This ap-
proach does not capture some very important elements that go to
make up a definition.

Instead of being analyzed directly into its ultimate con-
stituents, as in Figures 1 and 2, the definition, "the act of
spreading abroad", should first be broken down into subphrases
and then into its ultimate constituents, as in Figure 4, shown
on the next page. A-desirable property of the new points is
that they have the syntactical structure of derinitions. Thus,
"the act" and "spreading abroad" have the form of noun defini-
tions; "spread abroad" has the form of a verb definition; and
"of spreading abroad" (not shown, but feasible under a differ-
ent parsing) has the form of an adjective definition. This
would eliminate such combinations as "act of" or "of the". The
points representing phrase constituents of a definition thus
have the form of definitions, but lack a label.

The absence or presence of a label seems to make no dif-
ference in understanding the definition represented. In fact,

22
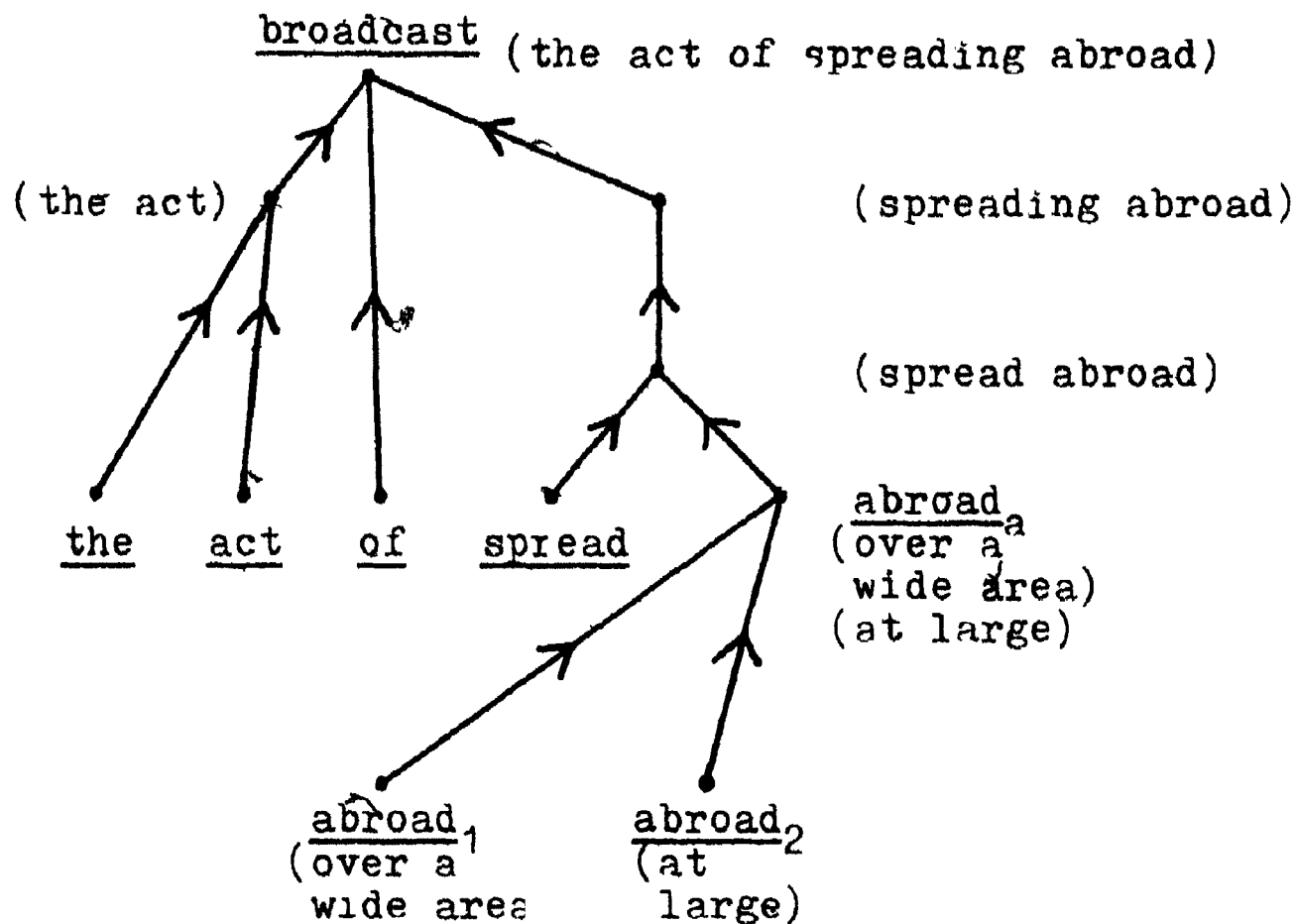
Figure 4. Subgraph of a model incorporating a parsing system.

it seems valid to represent identically worded definitions or
phrase constituents, regardless of the number of main entries
under which they appear, by a single point with multiple labels.
Thus, if each of the main entries disperse, scatter, and dis-
tribute has a definition verbalized as "spread abroad", these
three words can be labels of the point "spread abroad" in Fig-
ure 4. Such a construction has no effect on the analysis of the
definition "the act of spreading abroad" or "spread abroad" as
shown in Figure 4, and similarly, the analysis there would have
no effect on any analysis involving disperse, scatter, or dis-
tribute. Since there is a large number of instances where du-
plicate wording appears in a dictionary, the approach given
here would effect a substantial reduction in the size of the

digraph. (This is not to say that the words <u>disperse</u>, <u>scatter</u>, and <u>distribute</u> have the same meaning, but rather that in some instances these words can express the same concept.)

The definition, X, "the act of spreading abroad" is essentially an entity unto itself. The definitions of its component words have similar independence. However, like atoms in molecules, we need to identify those forces which hold the components together and which endow the whole with whatever characteristics it has. The definitions of the component words may require several words for their expression. but they are symbolized by one word in the definition X; even so the symbol and the definition both represent the same entity, which has certain characteristics enabling it to be acted upon by certain forces. These characteristics are the semantic, structural, and syntactic properties of definitions, and the forces are the production rules by which the entities (i.e. the component definitions or their symbols) are brought together. A definition may be viewed as the realization of such rules operating on the characteristics of other definitions. The nerculean task before us is to build a parsing system or recognition grammar which will articulate the characteristics to be attached to each definition and which will capture the production rules necessary to portray the relationships between definitions. The remainder of this section will present my ideas on how to approach this task.

The process which I have used for finding primitives entails showing that one definition is derived from another thereby excluding the former as a candidate for being primitive. Such a demonstration of a derivational relationship requires a parser. Each pattern which I observe between definitions helps to exclude further definitions and simultaneously becomes part of the parser. As a result, identification of the characteristics to be attached to each definition does not have to be accomplished all at once; as will become clear below, our purposes can be served as the components of the parser are delineated. Thus, success does not require full articulation of the parser before any parsing is initiated. The following represents general observations about the form of the parser as it has emerged thus far.

The first set of characteristics would result from the syntactic parsing of each definition. The purpose of this step would be simply to establish the syntactic pattern of each definition. The output of this step would be similar to that generated by Winograd (1972) in his parser. The 'dictionary' for the parser would be the very dictionary we are analyzing, although only the main entry, its inflectional forms, and its part-of-speech label would be used in this step. Ambiguous parsings and failures would be kicked out; the failures in particular, would provide an excellent source for refining the parser used by Winograd. Clearly, this step is not trivial, and it might even be argued that it is beyond the state-of-the-art.

However, by using a corpus as large as a dictionary and by

kicking out failures and ambiguities, I believe that this step

will significantly advance the state-of-the-art

The second set of characteristics would be determined from

a semantic parsing of the definitions, that is, an attempt to

identify the cases and semantic components present within each

definition. For this study I have found the following dis-

tinction to be useful: A case is a semantic entity which is not

intrinsic to the meaning of a word, e.g. that someone is an

agent of an action, whereas a component is an intrinsic part of

the meaning, e.g. a human being is animate It is necessary to

articulate recognition rules for determining that a particular

case or semantic component is present The little that has been

done to develop such rules has been based primarily on syntac-

tic structures or a priori assertions that a given case or com-

ponent is present. Despite the recognized deficiencies of dic-

tionaries, I believe that it is possible to bring much greater

rigor to such rules with evidence gleaned directly from the

definitions. For example, cut has a definition, "penetrate with

an instrument"; this definition would be parsed as having the

instrument case. (Note also that this definition makes the in-

strument case intrinsic to cut.) However, in most cases. it

will be necessary to examine the definitions of the constituent

words. For example, the verb knife has the definition, "cut

with a knife"; although it is quite obvious in this instance

that a knife is an instrument, rigor demands that we go to its

definitions where we find, "a simple instrument ...". A great
deal of analysis may ultimately be required to discern the in-
trinsic characteristics to be attached to a definition, but I
believe that many of these can come from the dictionary itself
rather than from intuition.

Although the number of cases and components discussed in
the literature is not very large, the number of ways in which
they may be expressed, at least in English, is significantly
larger. In addition, there is still a large amount of ambiguity,
i.e., not every form specifically indicates the presence of a
particular case. For example, a definition, "act with haste"
does not indicate that "haste" in an instrument: rather, "with
haste" expresses a manner of acting. Unraveling all these nu-
ances requires a great deal of effort. However, it appears that
a particularly good source of help in this endeavor might be
found in the definitions of prepositions (which are used pri-
marily to indicate sense relations).

Bennett 1975 found it possible to express the meaning of
spatial and temporal prepositions (a high percentage of all
prepositions) with only 25 components. However, in Webster's,
the number of their definitions is at least two orders of mag
nitudes higher. The difference seems to lie in the "apparent
polysemy" which, as Bennett says, arises from the inclusion in
prepositional definitions of "redundant features already deter-
mined by the environment". In other words, many prepositional
definitions contain information about the context surrounding

the preposition, particularly what sort of entities are related
by the prepositions. My examination of verb defintions contain-
ing prepositions has led to the observation of many noticeable
word patterns, i.e. collocations, which appear to be useful in
the recognition of cases. For example, one definition of <u>of</u>
states that its object indicates "something from which a person
or thing is delivered". In examining verb definitions, there
appears to be a distinct set of verbs with which this sense is
used in the following frame "(transitive verb)(object) of (some-
thing)". The verbs that fit the slot are exemplified by <u>free</u>,
<u>clear</u>, <u>relieve</u>, and <u>rid</u>. Thus, if this pattern appears, the ob-
ject of the preposition can be assigned the meaning "something
from which a person or thing is delivered". Through the use of
prepositional definitions in this way, I have therefore been
able to articulate some semantic recognition rules by which the
sense or case of a noun phrase the object of a preposition)
can be identified. My use of this technique has barely begun,
so that it is presently unclear whether this approach will suf-
fice to disclose all the case information that we wish to iden-
tify with a semantic parser, but if not it will certainly make
significant strides toward this objective.

Parsing of a definition according to the preceding notions
is still not sufficient to identify the semantic components
which should be attached to a main entry, since much of the se-
mantic content is only present by virtue of the definition's
constituent words. Thus, a complete rendering of a definition's

semantic content must be derived from the semantic characteristics of its constituents, in a recursive fashion, all the way down to the primitives. Although identification of these primitives is the primary goal of the approach being presented here, and hence, intrinsically incomplete until the analysis is completed, the set of semantic characteristics for a particular definition can be developed as we proceed toward our goal. To do this, it will be necessary to articulate rules which indicate how semantic characteristics may be transmitted from one definition to another. An example of such a rule is: If the noun X possesses the semantic component "animate", and if X is the core noun (i.e. genus) in definition y_f of the noun Y, then Y will also have the component "animate". Another example is: If a verb X has a definition x which has been parsed as having an instrument case, and X is the core verb of a definition y_j of Y, and y_j also has been parsed as having the instrument case, then the instrument in y_j is "a type of" the instrument in x_i. It will also be necessary to articulate other derivational (such as the application of a causative derivation to a state verb) and transformational (such as the application of a gerundial transformation to any verb) rules. This process of delineating how semantic characteristics are transmitted will at the same time give more meaning to the lines of the dictionary digraph than simply "is used to define".

The third, and final, set of characteristics that must be attached to a definition is a specification of the context that

must be present if that definition   intended. The context re-
strictions may require that the definiendum must be used in a
particular syntactical way, for example, as a transitive or in-
transitive verb. Usage restrictions may specify the presence of
particular words such as particles or objects. For example,
there is a distinct set of definitions for the idiom take out,
which thus requires the presence of the particle "out" in addi-
tion to the verb. One definition of the transitive verb chuck
requires the object "baseball". Other definitions may require a
specific subject. Finally, there are semantic restrictions that
may be discernible only from the definition itself. For example
two definitions of the verb cheer are: "to give new hope to"
and "lift from discouragement, dejection, or sadness to a more
happy state"; if the second definition is intended, it seems
necessary that the context indicate the prior state of discour-
agement, dejection, or sadness, since we cannot presume such a
state, for someone might have been in a happy or non-sad state
and simply received some new hope. In the absence of the neces-
sary context, we would default to the first definition.

Thus far in my research, I have not devoted any effort to-
ward developing procedures for prescribing the context based on
the definition. I expect that initiation of this step will ben-
efit from further results of the first two steps.

Although the parsing system outlined in this section may
appear to be exceedingly complex, such an eventuality is not
unexpected. The characteristics to be attached to each defini-

tion are not significantly different from those proposed by
Fillmore 1971. It is also important to note that some of the
goals of analyzing the contents of a dictionary are to reduce
the amount of redundancy, to remove vicious circles, and to
represent the meaning óf a word in a more efficient way. Hope-
fully, this type of analysis would eventually lead to a sub-
stantial reduction in the size of a dictionary; the prospects
for this are considered further in the next section.

8. THE ULTIMATE MODEL: POINTS AS CONCEPTS

At this juncture, it is necessary to ask whether the
points of the digraph models sufficiently correspond to meaning
as we wish it to be represented. In the two models described
thus far, the analysis of a definition was deemed complete when
the appropriate definitions of the constituent words had been
identified. This situation is not entirely, satisfactory, since,
if a constituent word has more than one definition that applies,
the definition being analyzed is subject to more than one in-
terpretation and hence may be called ambiguous with respect to
that constituent. For example, if the two definitions of abroad,
"over a wide-area" and "at large", fit the definition of broad-
cast to yield either "the act of spreading over a wide area" or
"the act of spreading at large", it is not legitimate to ex-
clude one. This situation is only a reflection of the fact that
natural language is almost always somewhat ambiguous. However,
in accepting this fact, it is necessary that we incorporate it
into our models.

Parts of the parsing system described in the last section
will help to discriminate and select those definitions of a con-
stituent word which fit a given context. As the parser is re-
fined, the candidates for a particular context will be narrowed
as described in Section 6, but many instances will remain where
more than one definition fits the context. We might say that
any point representing more than one definition thus consti-
tutes an ambiguity. Viewed differently, we might also say that
the context is not sufficient to distinguish among all the def-
itions of a word. In other words, we can 'blame' the ambigu-
ity on the context.

We must expect that ambiguity will be present in the dic-
tionary and deal with it on that basis. For purposes of illus-
tration, let us say that __abroad__ shown in Figure 4 is one such
point. To remove such points from the d graph, we must make two
points for the definition of __broadcast__, one representing "the
act of spreading abroad$_1$" and one representing "the act of
spreading abroad$_2$". These two points use the same words for ex-
pressing a definition and will be distinguishable only by the
fact that their underlying definitions are different. Because
of this situation, it is no longer valid to say that a point of
the model represents a definition: rather, we will say that a
point represents a "concept".

It is also possible that the concepts represented by two
or more points can be shown to be equivalent. The concept, "the
act of spreading abroad", has been shown to be equivalent to

"the act of spreading over a wide area". If the latter phrase-
ology appears under some main entry, say <u>distribution</u>, then
both it and the definition of <u>broadcast</u> would eventually be an·
alyzed in the same way. We will say that both expressions may
represent the same concept and hence are equivalent at least to
this extent. (Since the-other definitions of these words would
be different they are not totally equivalent.) This concept
will thus be represented by one point, labeled by either <u>broad-
cast</u> or <u>distribution</u> and equivalently verbalized as "the act of
spreading abroad" or "the act of spreading over a wide area".
This interpretation is a reflection of the fact that in ordi-
nary speech a single concept may be verbalized in more than one
way.

The observations in this section lead to the following de-
scription of the 'ultimate' model: The semantic content of a
dictionary may be represented by means of a digraph in which
(1) a point represents a distinct concept, which may be verbal-
ized in more than one way and may have more than one label, and
to which is appended a set of syntactic, semantic, and usage
features, and (2) a line represents an instance of some one of
a set of operators which act on the verbalizations or labels of
a point according to the features of that point to ield the
parametric values of another point. It should go without saying
that the complete portrayal of a dictionary according to this
model requires a considerable amount of further work; nonethe-

less, I believe that the model provides the appropriate framework for describing a dictionary.

9. PROCEDURES FOR FINDING THE PRIMITIVES

In Section 3, I stated that the model of a dictionary should permit the transformation of each definition into its primitive components. Based on the preceding descriptions. it is suggested that the full articulation of the ultimate model will satisfy this objective for the following reasons: (1) An elementary theorem in the theory of digraphs asserts that every digraph has a point basis, that is, a set of points from which every point in the digraph may be reached. Since points represent concepts in the ultimate model, it seems reasonable to assert that the point basis of its digraph represents the set of primitive concepts out of which all others in the dictionary may be formed. Based on the characteristics of the points in that model, it is possible (and perhaps even necessary) that each primitive concept would be <u>verbalized</u> in several ways and <u>symbolized</u> in several ways (as will be shown below) (2) Since the digraph has a finite number of points and lines, the sets of primitive concepts and operators are also finite.

It only remains to find the primitive concepts; this will be done by applying rules, based on the models and the parsing system, to identify words and definitions which cannot be primitives. Essentially, the assertion that a word or definition is non-primitive requires a showing that it is derived from a more primitive concept and that a primitive cannot be derived from

it. These non-primitives can be set aside and their full syn-
tactic and semantic characterization can be accomplished after
the primitives have been identified. Although no primitives
have yet been identified (since the described procedures have
not been fully applied), their form and nature will be delin-
eated.

To demonstrate the validity of my approach, I have been
applying rules developed thus far to the set of verbs in Web-
ster's Third New International Dictionary (20,000 verbs and
their 111,000 definitions). This set was chosen because of
their importance (cf. Chafe 1970) and the (bare) feasibility of
coping with them manually (although it may be another 3-4 years
before I am finished, at my current rate of progress). I have
attempted to formulate my procedures with some rigor, keeping
in mind the ultimate necessity of computerization. I have de-
veloped some detailed specifications for some of my procedures,
envisioning the use of computer tapes developed by Olney, but
have not completed these since I do not presently have access
to a computer.

Despite the focus on verbs, it will become clear that
words from other parts of speech are inextricably involved in
the analysis. Also, the rules that are presented can, for the
most part. be applied to other parts of speech. Notwithstanding
the fact that the meaning of many verbs is derived in part from
nouns and adjectives, I believe that each verb definition also
contains a primitive verb constituent.

Each verb definition consists of a core verb (obligatory) and some differentiae (optional). (The definitions of other parts of speech have a similar structure, i.e. a core unit from the same part of speech and some differentiae.) The subgraph of the total dictionary digraph formed by core verbs accords fully with the models described in Sections 4, 5, and 7. Therefore, any rules developed on the basis of those models will apply equally to the verb subgraph. We need only keep in mind that the differentiae come from other parts of speech and become embodied in the core verb. This is how the verb <u>cut</u> comes to have the instrument case intrinsically. To begin the analysis, we will let E represent the set of those verb definitions which have been identified as non-primitive; initially, this set is empty.

<u>Rule 1. If a verb main entry is not used as the core unit of any verb definition in the dictionary,</u> then all its definitions may be placed in E.</u> (This rule applies to points of the basic model which have outdegree 0, i.e. no outgoing lines.) Since no points can be reached from such a verb, it cannot be
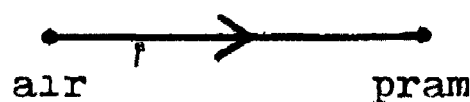

air   pram

Figure 5. Basic model, verb subgraph
example subject to Rule 1.

primitive. In Figure 5, the point labeled by <u>pram</u> represents the definition "to air (as a child) in or as if in a baby carriage"; since <u>pram</u> is the core unit for no definition in the

dictionary, all its definitions mav be excluded as non-primitive. In W3, this rule applies to approximately 13,800 verbs out of 20,000; the number of definitions in the verbs excluded is not known.

Rule 2. If a verb main entry is used only as the core unit of definitions already placed in E, then all its definitions may also be placed in E. (This rule applies to points of the basic model with positive outdegree. The uses of such verbs as core units follow definitional paths that dead-end; hence, they cannot be primitive. Figure 6 shows a portion of the dictionary

cover   cake   barkle
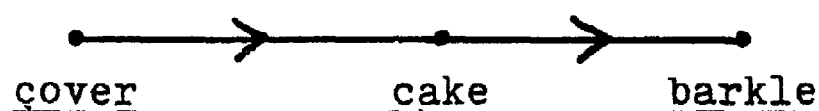
Figure 6.   Basic model, verb subgraph
            example subject to Rule 2.

digrapn where the verb cake defines only barkle, which in turn is not used to define, any verb. Thus, the definitions of cake may be included in E after the definitions of barkle have been entered. In W3, this rule applies to approximately 1400 of the 6200 verbs that remained after application of Rule 1.

Rule 3. If the verbs forming a strong component are not used as core units in any definitions except those in the strong component or in definitions of verbs already placed in E by Rules 1, 2, or 3, then the definitions of all verbs in the strong component may be placed in E. (This rule applies to points of the basic model which constitute a strong component, i.e. a maximal set of points such that for every two points, u

and v, there are paths from u to v and from v to u. This rule does not apply when the strong component consists of all points not yet placed in E.) A strong component consisting of the verbs _aerate_, _aerify_, _air_, and _ventilate_ is shown in Figure 7.
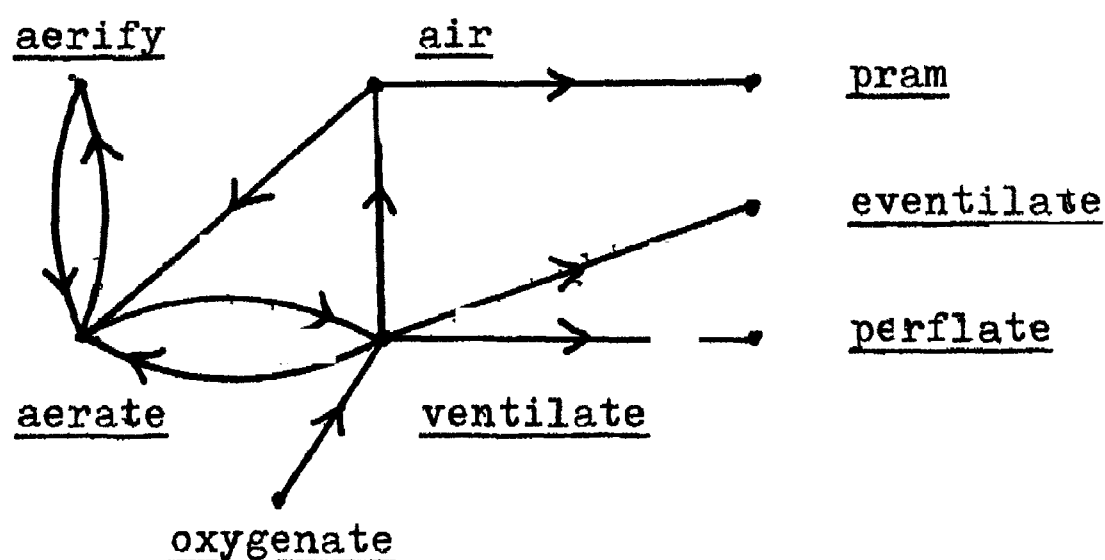
Figure 7.   Basic model, verb subgraph
example subject to Rule 3.

Except for _oxygenate_, the other verbs defining the set constituting the strong component are not shown. Since it is possible to start at any of the four and follow a path to any other of the four, there is no real generic hierarchy among them. It is possible to emerge from the strong component and follow paths to _pram_, _eventilate_ and _perflate_, to which, however, Rule 1 applies. If we follow a definitional path that leads into this strong component, we can never get out again or if we do we will only dead-end. Hence, the definitions of all the verbs in the strong component are not primitive and may be placed in E. In W3, this rule applies to approximately 150 of the 4800 remaining after the application of Rule 2. Actually, Rules 2 and 3 may be applied in tandem; based on those placed in E. Thus,

after Rule 3 places the definitions of <u>aerate</u>, <u>aerify</u>, <u>air</u>, and <u>ventilate</u> in E, it so happens that Rule 2 then applies to the definitions of <u>oxygenate</u>.

After Rules 1,2, and 3 are applied to the digraph of the basic model, the remaining points constitute a strong component of approximately 4500 points. This differs from those to which Rule 3 applies in that there would be no points left if we placed all its points in E. This final strong component is the basis set of the basic model, that is, any point of the basic model (i.e. any main entry in the dictionary) may be reached from any point in the final strong component (but not conversely).

At this juncture, we can proceed no further with the basic model alone; it is necessary to expand the points of the final strong component into two or more points each representing a subset of the definitions represented by the original point, as previously shown in Figure 3. In part, this can be accomplished by identifying individual definitions which are not used.

<u>Rule 4. If any definition can be shown to be not used as the sense of any core unit (or only those already in E), it may be placed in E.</u> This rule is essentially a restatement of Rule 1 for individual definitions and includes the following two subrules, among others not presented.

<u>Rule 4a. If all the remaining uses of a verb are transitive (intransitive), then its intransitive (transitive) definitions are not used and may be placed in E.</u> The expansion of a

point into transitive and intransitive uses is a good example
of how the points of the basic model are transformed into
points of the expanded model.

Rule 4b. If a definition is marked by a status label
(e.g. archaic or obsolete), a subject label, or a subject guide
phrase, it may be placed in E. Lexicographers creating W3 were
instructed not to use such marked definitions in defining any
other word..

Other rules have been developed in an attempt to identify
the specific sense of the core verb, or those senses of a verb
which have not been used in defining other verbs, but are not
presented here. However, there are too many instances where the
differentiae of a definition do not provide sufficient context
to exclude all but one sense (for example, many senses of move
fit into a definition phrased "move quickly"). In order to con-
tinue toward the primitives, we must shift gears slightly and
ask whether a definition can be characterized as "complex",
that is, derived from more primitive elements. For example, one
definition of make is "cause to be", which can be labeled as
complex because it consists of a causative component and a
state component, each of which is more primitive by itself than
"cause to be".

The importance of the notion of a complex definition be-
comes evident when we try to visualize how a primitive concept
will be identified. To understand this, we must consider some
further properties of the digraph. After the application of

Rule   and any subsequent rule), the remaining graph is a final strong component. (Recall that in a strong component, for each two points, u and v, there is a path from u to v and one from v to u.) Assuming that each point represents a concept (as in the ultimate model), the fact that two concepts are in the same strong component means that they are equivalent. In more traditional terms, what we have is a definitional vicious circle,.that is, a definitional chain which adds nothing to our understanding of the meanings involved.

Using the digraph of the final strong component, we can identify (and examine one by one) all putative definitional cycles or vicious circles; these will fall into three classes. The first class will consist of improper cycles, which can be removed by determining that one point is more complex (and hence not equivalent to the definition from which it is derived) Further rules for_ characterizing a definition as complex are given below. The second class of cycles will be real vicious circles, which fortunately can be removed, but only under certain conditions. For example, one definition of _jockey_ is "maneuver for advantage", while one definition of _maneuver_ is "jockey for position"; these two definitions constitute a vicious circle. In order to remove it. there must be some other definition of either verb which constitutes its meaning; in this case, it is found under _maneuver_, specifically, "shift tactics". Thus, in order to remove a vicious circle, we must find some way out. If we cannot, we have the third class of

cycles; this class will comprise the set of basic concepts. If there had been no way out for the example of _jockey_ and _maneuver_, we would have said that no meaning was conveyed by either verb, but rather that the meaning was established by use. This third set of cycles is what is sought by the procedures described in this paper.

As mentioned above, the crux of the analysis after the application of Rules 1 to 4 is the identification of complex concepts. Essentially this entails a showing that, for any definition y_j of verb Y, with Y as the core verb of definition x_j of verb X, the differentiae of x_j make y_i generic to x_j. For example, all transitive definitions of _cut_ would be generic to a definition in which "cut" is used with an object, even without narrowing down to one definition. The general rule may now be stated.

Rule 5. If any definition is identified as complex, it may be placed in E. The net effect of this rule is to break one or more putative cycles of equivalent definitions or concepts, enabling them to be transformed into a strict hierarchical order which will eventually be subject to Rule 4. Thus, the complex definition and all definitions that can be shown to be derived therefrom can be placed in E, because they cannot be part of a primitive cycle.

Rule 5 is implemented only by very specific recognition rules, which are essentially part of the parser. The specific rules entail a showing that some component has teen added in

the differentiae of a definition that is not present in the meanings of its core verb. For example, the "manner" component is not intrinsic to the meaning of the verb move; therefore, when a definition has the core verb "move" with an adverb of manner, it can be marked as complex. In establishing a component as non-intrinsic, it is necessary to articulate rules for recognizing the presence of the "manner" component (such as a phrase in a _____ manner" or an "-ly" word with a definition "in a _____ manner") and then to determine if that component is present in any definitions of a particular verb. If not, then the verb can be labeled as complex whenever it is used as the core verb in a definition with differentiae that fit the recognition rule. In addition to move, I have determined that, for the manner component, the verbs act, perform, utter, speak, express, behave, and many others follow the rule. Table 1, on the next page, identifies some specific components, a brief description of how they are recognized, some of the verbs to which the particular rule applies, and an example of a definition labeled as complex by the rule and hence placed in E.

If a definition has a core verb whose applicable sense is one which has been marked as complex, it too can be so marked, since it is derived from a complex definition. For example, all definitions of the form "make adjective", i.e. with an adjective complement, are derived from the definition of make, "cause to be or become" and hence can be marked as complex. In addition, if all definitions of a verb have been marked as

Table 1

Recognition Rules for Semantic Components

Name of Component	Recognition Rule	Applicable Verbs	Examples of Definitions
1. Aspect	Verb + Infinitive	cease, begin, strive, continue	commence vi 2, "begin to be"
2. Causative	Causative verb + Infinitive	cause, force, compel, induce	confront vt 2a, "compel (a person) to face, take account of, or endure" make vt 10a, "cause to be or become"
3. Instrument	Verb + "with" + noun defined as instrument, device, etc.	apply, fasten, cut, beat	knife vt 2a, "cut with a knife"
4. Means (Process)	Verb + "by" + Gerund	make, prepare, form, shape	draw vt 4e4, "shape (glass) by drawing molten glass from the furnace over a series of automatic rollers"
5. State Entry	Verb + "into" + noun defined as "the state of ..."	bring, put, throw, fall	disorder vi, "fall into confusion"
6. Deliverance	Verb + "of" or "from" + noun	free, relieve, rid, empty	clear vt 2g2, "rid (the throat) of phlegm"

complex then all definitions in which it appears as a core verb can be similarly marked and placed in E.

Through the development and application of further parsing rules under Rule 5, I am hopeful that I will eventually arrive at the set of primitive verb concepts (i.e. cycles or vicious circles with no way out). I have already reduced the number of verbs from 20,000 to less than 4,000. This number would be much lower, but for the fact that I am applying the rules manually and I must exercise time-consuming care to ensure correctness.

After the primitive concepts have been identified, it will be necessary to go back to all the definitions that were set aside in the process of finding the primitives, so that their semantic characteristics can be articulated. I fully expect that the parsing system which will have been developed will be able to accomplish much of this task I also expect that the parsing system will have equal applicability as a general parser capable of formally characterizing ordinary discourse in a canonical form. Of course, verification of this expectation will have to await a full presentation of the parser.

10. RELATIONSHIP TO EFFORTS TO REPRESENT KNOWLEDGE IN FRAMES

The process which has been outlined in the preceding sections is closely akin to current efforts to represent knowledge in frames. (Cf. Winston 1977 for an elementary presentation of this notion.) Briefly, a frame consists of a fixed set of arguments, some of which may be specifically related to others, and some of which may have specific values. frame is intended to

represent a stereotyped situation, with the arguments identify-
ing the various attributes which the situation always possesses.
In terms of case grammar, for example, a movement frame will
contain arguments or slots for an agent, an instrument, and a
destination.- By tying frames together in specific relationships,
we can build larger and larger frames to represent more and
more knowledge, perhaps constructing a series of events, an in-
ference structure, or a description of a scene.

Before building these large structures, it is necessary to
represent very small pieces of knowledge. Heretofore, this has
been done by postulating the components of frames to represent
such things as actions and state changes. But this can be ac-
complished on a more rigorous basis. For example, if we first
locate all definitions using "move" as its core verb and then
identify all the case structures in which it is used, we will
have a generalized frame which characterizes most if not all of
the possible uses of "move". (This approach is currently being
followed by Simmons 1977.) Each definition in which "move" is
used could then be represented by the generalized frame with
some of its slots filled. This process can be followed for any
word for which we wish to develop a frame.

If, in addition, we analyzed the definitions of move, we
will find that they, in turn, represent instantiations of still
other frames, which will be even more generalized than those
developed for the uses of "move". The difference between the
frames representing the definitions of move and those represent-

ing the uses of "move" is that the latter are the same as the former with some slots filled. Within the bounds of the ambiguity present in the dictionary, this slot-filling will identify which definition of <u>move</u> are employed in which uses of "move". It seems to me that this is nothing more than the process which has already been described using a graph-theoretic approach, except that the generalized frame for each verb will not be, carried along through each step. Moreover, since the semantic parsing system which has been described will be based largely on the relationships derived from the definitions of prepositions, and these comprise most of the case relationships, the parsing system will effectively circumscribe the permissible elements (i.e. slots) which can be present, given any particular. context. Thus, although the phraseology is different, the effect is the same.

If there is an essential equivalence between these two approaches, then. since frames purport to represent knowledge, the process described, if successful, will result in an articulation of whatever knowledge is contained in a dictionary. What this implies is that the lexicon contains a great deal of knowledge about the world and not just information which will enable us to understand such knowledge.

Frames provide a great deal of insight to the approach which has been described here, but the reverse also seems to hold true. If the semantic content of each definition can be captured, then it may be possible to articulate the frame for

47

any utterance by combining the characteristics of the definitions of its constituent words within what is permitted by the parsing system.

11. FINAL REMARKS

In Section 1, I described some limitations of this paper and my research. This paper suffers from a lack of sufficient detail to enable a reader or researcher to replicate what I have done or to take the next steps of computerizing the procedures which I have developed. I will provide further details on the specific steps I have followed in reducing the set of verbs from 20,000 to 4,000 to anyone requesting. With respect to computer specifications, I have prepared some, but stopped because I have no access to a computer. However, if any researcher is interested in pursuing this (or setting graduate students to work), I am prepared to develop the necessary specifications and to work hand-in-hand for the further advancement and refinement of this methodology.

I also indicated in Section 1 that my research presently shows no final results and that I do not even know how much further effort will be necessary to explicate the parsing system which has been described. Clearly, there are great distances yet to be covered toward a goal of being capable of transforming ordinary discourse into a canonical form. I believe that characterization of the contents of an ordinary dictionary is an essential step in attaining this goal, and I am hopeful that my approach can be used to develop such a characterization.

If it seems worthwhile to pursue this approach, despite the limitations, I believe the best way to do so would be to establish a single computer-based repository for a dictionary, preferably W3, with øn-line access to researchers across the country, and to build the parser and definitional· characterizations piece by piece. (I have noted how the parsing system which I have described can be built incrementally.) The magnitude of this effort precludes much progress by individual researchers. Olney tried to do something similar with the collegiate dictionary based on W3, but by distributing bulky computer tapes. He was unfortunately premature; it may be that now is the time to try again.

49

REFERENCES

1. Bennett, David C. Spatial and Temporal Uses of English Prepositions: An Essay in Stratificational Semantics (Longman Linguistics Library, 17), Longman, New York 1975

2. Chafe, Wallace. L. Meaning and the Structure of Language, University of Chicago Press, Chicago, 1970.

3. Encyclopaedia Britannica. Webster's Third New International Dictionary, Chicago, 1966.

4. Fillmore, Charles J. The case for case. Bach, E. and Harms, R.T. (Ed.) Universals in Linguistic Theory, Holt, Rinehart and Winston, New York, 1968, 1-88.

5. Fillmore, Charles J. Types of lexical information. Semantics: An Interdisciplinary Reader in Philosophy, Linguistics, and Psychology, Cambridge University Press, Cambridge, 1971, 370-392.

6. Harary, F., Norman, R.Z., and Cartwright, D. Structural Models: An Introduction to the Theory of Directed Graphs, John Wiley and Sons, Inc., New York, 1965.

7. Jackendoff, R. Morphological and Semantic Regularities in the Lexicon, Indiana University Linguistics Club, Bloomington, Indiana, 1974.

8. Katz, J.J. and Fodor, J.A. The structure of a semantic theory. Language 39(1963), 170-210.

9. Nida, E.A. Componential Analysis of Meaning, Mouton, The Hague, 1975.

10. Olney, J., Revard, C., and Ziff, P. Toward the Development of Computational Aids for Obtaining a Formal Semantic Description of English, SDC Document SP-2766/001/00, System Development Corp., Santa Monica, 1968.

11. Quillian, M.R. Semantic memory. Minsky, M.(Ed.) Semantic Information Processing, MIT Press, Cambridge, 1968, 216-270.

12. Schank, R.C. Conceptual dependency: A theory of natural language understanding. Cognitive Psychology 3(1972), 552-631

13. Simmons, R.F. and Slocum, J. Generating English discourse from semantic networks. Communications of the ACM 15(1972) 891-905.

14. Simmons, R.F. and Amsler, R.A. Modeling Dictionary Data, University of Texas Department of Computer Sciences, Austin, 1975.

15. Simmons, R.F. and Lehmann, W.P. A Proposal to Develop a Computational Methodology for Deriving Natural Language Semantic Structures via Analysis of Machine-Readable Dictionaries, University of Texas, Austin, 1976. (Research proposal submitted to the National Science Foundation, September 28, 1976)

16. Winograd, T. Understanding natural language. Cognitive Psychology 3(1972), 1-191.

17. Winston, P.H. Artificial Intelligence, Addison-Wesley, Reading, Mass., 1977.

Vol. V, No. 2

February, 1979

PRESIDENT TO SUPPORT LIMITED PRIVACY INITIATIVE

Consistent with the selective approach of the U S to privacy regulation
(versus the omnibus approach of the Europeans on the subject), the
Carter Administration is expected to support a limited legislative
program in the 96th Congress on privacy issues. The President's response
to the recommendations of the Privacy Protection Study Commission and
previous legislative efforts, termed the privacy initiative, is emerging
from a year-long study by an *ad hoc* group within Mr. Carter's Domestic
Policy Staff. The study called "Baby Blue" (compared with a larger
,supporting blue-colored document called "Big Blue") was delivered to
the President last December. The group, known as the White House Privacy
Study Coordinating Committee, is headed by Stuart E Eizenstat, Assistant
to the President for Domestic Affairs, and Juanita M Kreps, Secretary
of Commerce.

Administration Proposals   It is reported that Mr. Carter may mention
the privacy initiative in his State of the Union Address in January
The Administration's proposals are expected to center on limiting Federal
access to data in the private sector, *i.e.*, in the area of medicine,
credit and insurance. The Privacy Coordinating Committee recommended
that these limits on access should apply equally, to state and local
governments The Committee endorsed Federal legislation leaving states
to adopt laws "that meet certain minimum standards."

The privacy proposals would give individuals the right of "ownership" to
personal data maintained in the medical, credit and insurance sectors

Thus, individuals would be entitled to review information in order to correct errors. (Aetna Life & Casualty Co. has initiated a similar policy, at the urging of William O. Bailey, Aetna Life president, and former Privacy Protection Study Commission member.) It is possible that this right of "ownership" will be incorporated into legislation amending the *Fair Credit Reporting Act*. The proposals would also forbid disclosure of information where there is an expectation of confidentiality." The Committee agreed to exclude a recommendation that would encompass computerized telephone records. The Administration's privacy agenda seems to coincide with that of Rep. Richardson Preyer (D-N.C.) who predicts the Congress will consider measures concerning medical records, banking records and third-party records.

'Administrative Steps. Besides the legislative proposals on privacy, the President is expected to take some "administrative steps," using executive authorization (see *Washington Report*, 12/78, p. 11).

'International Information Issues.' The privacy initiative precedes expected future Administration proposals on so-called "international information issues," such as overseas restrictions on transborder data flow, the transmission of data across international boundaries. Henry Geller, Assistant Secretary of Commerce for Communications & Information, has noted it is time for the U.S. to "bring . . . [its own] house in order" on privacy issues (see *Washington Report*, 12/78, p. 11).

Role of the Computer. Recognizing the role of the computer in facilitating the collection and dissemination of information, Carter officials state that legal protection against the indiscriminate use of data has not developed as rapidly as the technology. In one draft of the report prepared for the President by the Privacy Coordinating Committee, the group noted that, "We are faced by a slow but steady erosion of privacy which if left unreversed, will take us (in another generation) to a position where the extent of our human rights and vitality of our democracy will be jeopardized."

Previous Privacy Legislation. The Presidential Privacy Initiative follows passage of the *Privacy Act of 1974* and the *Right to Financial Privacy Act* (*Washington Report*, 12/78, p. 1). The *Privacy Act* limits Federal agencies' access to personal information held by other Federal agencies. The *Right to Financial Privacy Act* limits Federal access to personal information in the financial sector. Cited as a major achievement by the Carter Administration, the *Financial Privacy Act* has been criticized by certain individuals for increasing the potential number of bank examinations conducted by Federal investigators; for lacking sufficient legal grounds to challenge unreasonable access to data; and for exempting political action groups. [An internal audit, made public recently by the U.S. Postal Service criticizes the Post Office for inadequate implementation of the *Privacy Act of 1974*.]

Effect of Congressional Elections on Privacy Issues. The surprise defeat of Rep. Edward W. Pattison (D-N.Y.) in the November Congressional elections removes a staunch defender of financial privacy legislation from the House Banking Committee. Also, on the Senate side, Sen. Thomas J. McIntyre's (D-N.H.) loss is expected to change the character of the

Financial Institutions Subcommittee which the Senator chaired. However,
strong privacy advocates were elected to the House of Representatives in
California: a Democrat, Vic Fazio, sponsor of a Fair Information Practice
Bill enacted in California in 1977; and a Republican, Jerry Lewis
(no relation to the entertainer), sponsor of additional state-wide
privacy legislation.

AFIPS IN WASHINGTON

WITNESS STATEMENTS AVAILABLE THROUGH WASHINGTON OFFICE DETAILED

The AFIPS Washington Office has compiled numerous witness statements
made before the Executive and Legislative Branches of Government on
information policy issues as part of a Witness Statement Exchange initiated
last year (*Washington Report*, 11/78, p. 6). For participants in the
witness statement exchange (rules for participation described below),
the following witness statements may be obtained:

H.R. 214, The Bill of Rights Procedures Act. Philip B. Heyman, appearing
July 13, 1978, before the House Subcommittee on Courts; Richard J. Davis,
Assistant Secretary of the Treasury, Enforcement and Operation, Department
of the Treasury, appearing July 20, 1978; and Paul G. Coe, Assistant
Chief Postal Inspector, Criminal Investigations, U.S. Postal Service,
appearing July 20th.

H.R. 13015, The Communications Act of 1978. Tyrone Brown, commissioner,
Federal Communications Commission (FCC), appearing July 18, 1978, before
the House Subcommittee on Communications; Margita E. White, commissioner,
FCC, appearing July 18, 1978; James H. Quello, commissioner, FCC, appearing
July 18th; Philip S. Nyborg, vice-president and general counsel Computer
& Communications Industry Association (CCIA), appearing August 3, 1978;
Charles D. Ferris, chairman ,FCC, appearing August 9, 1978; Joseph R. Fogarty,
commissioner, FCC, appearing August 9th; Margita E. White, commissioner,
FCC, appearing August 9th; L. C. Whitney, president, National Data
Corp., appearing August 10, 1978; and Herbert N. Jasper, executive vice
president, Ad Hoc Committee for Competitive Telecommunications. appearing
August 10th.

S. 2096, The Right to Financial Privacy Act of 1977, and *S. 2293, The
Electronic Funds Transfer Act of 1977.* Robert Ellis Smith, publisher,
Privacy Journal, appearing May 19, 1978, before the Senate Subcommittee
on Financial Institutions.

S. 3270, The Justice System Improvement Act of 1978. Jeffrey A. Roth,
senior economic analyst, Institute for Law & Social Research, appearing
August 23, 1978, before the Senate Subcommittee on Criminal Laws &
Procedures; also, James Duke Cameron, chairman, Conference of Chief
Justices, appearing August 23, 1978; Patrick V. Murphy, president,
Police Foundation, appearing August 23rd; and Glen D. King, executive
director, International Association of Chiefs of Police, appearing
August 23rd.

'Confidentiality of Medical Records.' Richard I. Beattie, deputy general counsel, Department of Health, Education & Welfare, appearing May 23, 1978, before the House Subcommittee on Government Information & Individual Rights.

Exports., Juanita M. Kreps, Secretary of Commerce, appearing September 28, 1978, before the Senate Committee on Commerce, Science & Transportation.

'Future of Small Business in America.' John H Shenefield, assistant attorney general, Antitrust Division, Department of Justice, appearing July 20, 1978, before the House Subcommittee on Antitrust, Consumers & Employment; and A. G. W. Biddle, president, CCIA, appearing July 20, 1978.

High Technology Businesses. Jean N. Tariot, chairman, Incoterm Corp., appearing July 20, 1978, before the Joint Senate Committee on Small Business and House Subcommittee on Antitrust, Consumers & Employment; and Lester A. Fettig, administrator, Federal Procurement Policy, Office of Management & Budget, appearing August 10, 1978.

Rules for Participation. To participate in the exchange of statements made before the Executive and Legislative Branches of Government on information policy issues, one recent witness statement concerning information policy should be sent to: Pender M. McCarter, Research Associate, AFIPS Washington Office, 1815 North Lynn Street, Suite 805, Arlington, Virginia 22209. Thus enrolled in the program, specific witness statement requests can be made (based on the above list), by mail only, enclosing a stamped, self-addressed envelope. For each requested witness statement, one statement should be included, in addition to the first establishing participation in the program. It is not necessary to be a witness in a hearing; having access to such statements is sufficient. Updated listings of available witness statements will be issued periodically.

SPECIAL REPORT

EUROPEANS SEE 'WIDER CONCERN' IN RESTRICTIONS ON TRANSBORDER DATA FLOW: 'PROTECT PEOPLE AGAINST COMPUTERS & COMPUTERS AGAINST PEOPLE'

Citing a long privacy tradition, concern was expressed for the protection of individuals, not nations, "whoever and wherever they are," in an *International Conference on Data Regulation: European & Third World Realities*, convened in New York City, November 28-30.

'More Than Privacy Interests . . . Involved.' Frits Hondius, chief, Judicial Affairs Directorate, Council of Europe (CE), told the Online Conferences Ltd.-sponsored conference that more than privacy interests are involved in European restrictions on the transmission of data across international boundaries. (The CE is preparing a 1980 treaty concerning transborder data flow.) According to Mr. Hondius, such bodies as the 20-member CE (in which the United States is only a non-voting member) are seeking to protect "people's rights and interests." He added that the European goal is to "protect people against computers and computers

MATT NILSON, DIRECTOR, BUSINESS PLANNING, INTELSAT (LEFT), AND
BRIAN JUDD, SPECIAL ASSISTANT FOR CCIS, NATO
(AFIPS/ P. M. McCarter)

against people." Hondius also noted the lack of U.S. Government attendance
at a recent CE session as well as at the Online Conference.

Jan Freese, director-general, Data Inspection Board, Sweden, agreed that
his country's *Data Protection Act* was designed to insure "the use of
computer technology on human terms." Mr. Freese added that it was his
philosophy to "try to solve problems before they occur."

General Principles of Data Protection Cited. Hondius outlined some
general principles of data protection laws already in effect in some
seven countries. (Approximately seven more nations are expected to
follow these countries with their own privacy legislation.)

The three principles are: (1) Publicity: "People should know what is
going on in general"; (2) Propriety: "Data systems should be proper";
and (3) Control: "Recordkeeping should observe norms."

U.S. Privacy Policy Criticized. While stating that U.S. laws such as
the *Privacy Act of 1974* did represent "a legislative step forward,"
Professor David F. Linowes, former chairman, Privacy Protection Study
Commission, said that the *Privacy Act* provides "no benefits for the
general public"; contains too many exceptions and too few penalties; and
disregards accountability.

Computer users from large multinational corporations attending the
conference criticized the U.S. for a lack of leadership in formulating a
position on issues involved in transborder data flow. According to one

PROF. LINOWES, POLITICAL ECONOMY & PUBLIC POLICY, UNIVERSITY OF ILLINOIS
(AFIPS/ P. M. McCarter)

account of "an informal, not-for attribution meeting," held after one of
the conference sessions, the users formed an *ad hoc* committee to lobby
on transborder data flow issues.

U.S. Industry Criticized. Administration officials appearing at the
conference reiterated their criticism of industry for not becoming involved
in the issues, and implored industry to provide specific instances of
economic harm caused by restrictions on transborder data flow. Attending
the conference and named as primary contacts for industry were: William
Fishman, deputy associate administrator for Policy Analysis and Development,
National Telecommunications & Information Administration (NTIA), U.S.
Department of Commerce; and Morris H. Crawford, Bureau of Oceans &
International Environmental & Scientific Affairs, U.S. Department of
State.

OECD Drafting Group Meeting Held. The Drafting Group of the Organization
of Economic Cooperation & Development (OECD) met December 6-8 in Paris
to consider a new draft of Transborder Data Flow Guidelines prepared by
Peter Seipel, consultant to the OECD Secretariat (*Washington Report*,
January, 1979, p. 1). Attending the meeting as U.S. representatives
were: Lucy Hummer, Esq., Department of State; William Fishman, NTIA;
and James Howard, NTIA.

Inclusion of Manual Files, 'Legal Persons' Debated. At the OECD meeting,
there was substantial disagreement on including manual files as well as
computer files in the draft guidelines. In addition, the delegations were
divided on extending privacy protection to "legal persons" (*i.e.*, business

corporations' and various other organizations) as well as individuals.
The Europeans favor a more comprehensive approach to privacy legislation
and generally view as ineffectual the selective approach taken by the U.S.

Consensus Said to be Supporting U.S. Position. Despite these recent
developments, a consensus is said to be growing in both the OECD and the
Council of Europe supporting the U.S. position. For example, the latest
Seipel draft has been interpreted by an Administration source as being
"very favorable" to the U.S. position.

NEWS BRIEFS

A recommendation for a Special Assistant to the President for Information
   Technology Policy, Plans & Programs, contained in a tentative
   Discussion Draft of the final *Summary Report on Information Technology
   & Governmental Reorganization* of the President's Federal Data
   Processing Reorganization Project (FDPRP) (*Washington Report*, 10/78,
   p. 5), has been dropped in a final draft; according to the most recent
   version of the consensus report [now circulating among Cabinet and
   Office of Management & Budget (OMB) officials], the FDPRP majority
   view "holds that the . . . [FDPRP]recommendation can and must be
   implemented through a strong and persevering Presidential initiative
   through the OMB. . . ."; the OMB is expected to present the consensus
   report to the President after final revisions.
A formal study "to determine the Administration's policy . . . [on] the
   future role of the U.S. Postal Service in providing services by
   electronic communications" is being initiated by the White House
   under Stuart Eizenstat, the Assistant to the President for Domestic
   Policy; an Interagency Coordinating Committee, chaired by Mr. Eizenstat,
   met December 13th to outline electronic communications' issues; the
   National Telecommunications & Information Administration, designated
   as "lead staff agency" for the study, is soliciting comments from
   "interested individuals or organizations" to be considered in the
   development of the Administration's position; Congress is expected to
   address the issue this Spring.
In December, the Postal Service Board of Governors authorized temporary
   implementation of E-COM service, an electronic message service (EMS)
   for large-volume users (see *Washington Report*, 11/78, p. 3); in November,
   Postmaster General William F. Bolger approved a four million dollar
   electronic mail experiment beginning this year; also in November,
   Xerox Corp. filed a request with the Federal Communications Commission
   to reallocate a portion of the radio spectrum for EMS.
"[B]etter information is needed . . . to make assessment and evaluation of
   the policy alternatives regarding CCH [the computerized criminal
   history file]," according to an Office of Technology Assessment (OTA)
   study released in January, the first phase of a new OTA assessment
   of the Social Implications of National Information systems; entitled
   *A Preliminary Assessment of the National Crime Information Center and
   the Computerized Criminal History System* (#--enclose $2.75), the study
   notes, "Although CCH has been the subject of numerous studies, conferences

and hearings, there is only limited information regarding the ways
in which law enforcement and the criminal justice decisionmakers,
as well as other government and private individuals and the press
make use of criminal history information, its benefits, the value
of nationwide access to information, and the value of rapid
access."

The General Accounting Office (GAO) is preparing to release a new study
entitled *Security of Automated Information Systems of Federal Agencies*;
according to a tentative outline of the GAO report, obtained by the
AFIPS Washington Office, "organizational structures" are "inadequate"
and "comprehensive procedures" are nonexistent in current Federal
security precautions.

A research and development project to evaluate the use of data encryption
devices in protecting the Federal Reserve System's (FRS) Fedwire
operations is expected to be completed this June; Fedwire, a form
of electronic funds transfer, links FRS to member banks nationwide.

In December, the Department of Justice said it is considering computer
crime involved in counterfeit or stolen securities as well as
bribery and kickbacks.

The Federal Communications Commission (FCC) is expected to add the *Computer
Inquiry II* to its weekly agenda again, after two previous postponements;
the FCC may determine whether AT&T, a regulated communications common
carrier, can provide unregulated data processing services.

The Supreme Court is considering whether, under the *Freedom of Information
Act*, individuals can obtain confidential business data; in November,
the High Court let stand a U.S. Court of Appeals decision (*Washington
Report*, 6/78, p. 4) allowing MCI Communications Corp. to use AT&T's
local phone connection to implement Execunet, MCI's long distance
telephone service providing voice and data communications.

In December, the Office of Management & Budget (OMB) issued for comment
a directive which would require Federal agency data processing users
to account for the future cost of their DP systems; also in December,
OMB issued an annotated bibliography (#) of current laws, policies,
regulations, and "guidance documents" which are relevant to the
acquisition, management, and use of Federal data processing and related
telecommunications resources; finally, in December, OMB issued a list
(#) of Federal policies, regulations, standards, guidelines, and other
reference documents pertaining to computer security.

The "basic philosophy" of the Communications Act Rewrite "will remain the
same," according to former Rep. Louis Frey (R-Fla.), until this year
ranking member of the House Communications Subcommittee; predictions have
also been made that "significant changes" will be incorporated in the
legislation this year, previously known as the *Communications Act of
1978* (*Washington Report*, 10/78, p. 3).

A new subcommittee on "Professionalism & Malpractice of Computer Specialists"
has been formed by the Committee on Law Relating to Computers of the
American Bar Association's Science & Technology Section; heading the
subcommittee is J.T. Westermeier, Jr., member of a Washington, D.C.
law firm.

Ed.: Information for the February, 1979, *AFIPS Washington Report* is current
as of January 5, 1979, press time. Production assistance for the *Washington
Report* is provided by Linda Martin. AFIPS societies have permission to use
material in the newsletter for their own publications. Documents indicated
by the symbol "(#)" are available on request to the Washington Office. Re-
quests should specify the date(s) of the *Report* in which the document(s)
appeared. Where price is noted, make checks payable to "AFIPS."

WASHINGTON DEVELOPMENTS

PRESIDENT, CONGRESS ADDRESS INFORMATION POLICY ISSUES

Amidst predictions that the 96th Congress is concentrating on oversight
of existing Government programs, there is no dearth of information
policy-related legislation on the Congressional Calendar, sustaining the
momentum of the 95th Congress which enacted 74 new laws affecting U.S.
information policy. [Editor's Note: A House of Representatives'
Committee Print describing these laws is available on request to the
AFIPS Washington Office.]

Privacy Legislation. Much of the information policy-related legislation
centers on privacy issues. President Carter referred to planned privacy
legislation affecting Government access to records in the medical and
financial sectors (see *Washington Report*, 12/78, p. 1) in his Supplemental
State of the Union Address delivered to the Congress on January 25th.

Under the heading of "Civil Liberties: Privacy," the President said:

> Government and private institutions collect increasingly large
> amounts of personal data and use them to make many crucial
> decisions about individuals. Much of this information is needed
> to enforce laws, deliver benefits, provide credit, and conduct
> similar, important services. However, these interests must be
> balanced against the individuals right to privacy and against
> the harm that unfair uses of information can cause. Individuals
> should be able to know what information organizations collect
> and maintain about them; they should be able to correct inaccurate
> records; and there should be limits on the disclosure of
> particularly sensitive personal information.

Mr. Carter concluded defining planned administrative measures implementing privacy protections (see *Washington Report*, 2/79, p. 2), as follows:

> My Administration is developing a comprehensive privacy policy to address these concerns. Last year, legislation was enacted which established restrictions on . . . Government access to financial records. Early in 1979, I will propose privacy legislation to cover medical, financial, and other sensitive personal records. I will also take administrative actions to strengthen privacy controls for Federal agencies' records.

NTIA Proposals. The National Telecommunications & Information Administration (NTIA) is said to be preparing legislation for introduction this month (in March), implementing what is being called the President's Privacy Initiative. A principle underlying the legislation, according to an NTIA staff member, is that information collected for research and statistical purposes "should not be used [by Government] to make decisions about people."

HEW Bill. The Department of Health, Education & Welfare (HEW) is also reported to be drafting legislation on Government access to medical records. Rep. Richardson Preyer (D-N.C.), chairman of the House Subcommittee on Government Information & Individual Rights, has previously expressed interest in considering privacy measures concerning medical records (see *Washington Report*, 2/79, p. 2).

Goldwater Legislation. On January 18th Rep. Barry M. Goldwater, Jr. (R-Calif.) reintroduced privacy legislation implementing recommendations of the Privacy Protection Study Commission (*Washington Report*, 8/77, p. 1), including a bill to amend the *Fair Credit Reporting Act*. Mr. Goldwater's legislation is listed as follows:

> *H.R. 344.* A bill to amend the *Fair Credit Reporting Act* dealing with depository institutions and privacy, and for other purposes; to the Committee on Banking, Finance & Urban Affairs.

> *H.R. 345.* A bill to amend the *Fair Credit Reporting Act* dealing with consumer credit and privacy; to the Committee on Banking, Finance & Urban Affairs.

> *H.R. 346.* A bill to amend the *Fair Credit Reporting Act* dealing with insurance institutions and privacy; to the Committee on Banking, Finance & Urban Affairs.

> *H.R. 347.* A bill to amend the *Family Educational Rights and Privacy Act* to provide for the protection of the privacy of personal information, and for other purposes; to the Committee on Education & Labor.

> *H.R. 349.* A bill to amend the *Privacy Act of 1974*; to the Committee on Government Operations.

> *H.R. 350.* A bill to establish a Federal Information Practices Board to review and report on fair information and privacy practices of Governmental and nongovernmental entities; to the Committee on Government Operations.

_H.R. 354._ A bill to amend the _Internal Revenue Code of 1954_ dealing with privacy; to the Committee on Ways & Means.

_H.R. 358._ A bill to restrict the use of _Social Security Act_ account numbers as Governmental or universal personal identifiers; to the Committee on Ways & Means.

_H.R. 359._ A bill to provide for the privacy of certain public assistance and social service records used or maintained by state and private agencies under programs receiving Federal financial assistance; jointly, to the Committees on Agriculture, Interstate & Foreign Commerce, and Ways & Means.

_H.R. 360._ A bill to amend Title XI of the _Social Security Act_ to provide for the confidentiality of personal medical information created or maintained by medical care institutions providing services under the Medicare or Medicaid programs, and for other purposes; jointly, to the Committees on Interstate & Foreign Commerce, and Ways & Means.

_H.R. 361._ A bill to amend the _Social Security Act_ to provide for the protection of the privacy of personal medical information maintained by certain medical care institutions; jointly, to the Committees on Ways & Means, and Interstate & Foreign Commerce.

The California Congressman has been quoted as saying that Congress must legislate in the privacy area whenever private enterprise fails to act. Goldwater has served as a member of the Privacy Protection Study Commission

Chances for Passage of Privacy Legislation. Chances for passage of privacy legislation are unpredictable given the customary, formidable Congressional procedures as well as preoccupation with foreign relations and the domestic economy. Among the scores of privacy-related bills introduced in the 95th Congress, only the _Right to Financial Privacy Act_ (see _Washington Report_, 12/78, p. 1) passed in the early morning hours of the last day of Congress. A bemused Carter official recently noted · that a bill affecting Government access to medical records may originate in as many as four different Congressional subcommittees. Similarly, one Congressional staffer stated that information policy is "made in disparate environments." Harry M. (Chip) Shooshan III, chief counsel, House Communications Subcommittee, told a January meeting of the American Library Association that this disparity results in "contrary policies."

[At least some Congressmen are reconsidering support for one section of the _Right to Financial Privacy Act_ following a Citibank survey which estimates that compliance with the bill's notice requirements by financial institutions could cost as much as one billion dollars, recalling similar high (and, according to some privacy advocates, ultimately incorrect) estimates of costs to implement the _Privacy Act of 1974_. Sen. William Proxmire (D-Wisc.), for example, has introduced _S. 37_ repealing Section 1104(d) of the _Act_ which states that, "All financial institutions shall promptly notify all of their customers of their rights under this title." A similar bill, _H.R. 1777_, has been introduced in the House, inserting "active" after "notify all of their." _S.37_ passed the Senate last month.]

Communications Act' Rewrite. A new bill is scheduled to be introduced
the first of this month (in March) with the "basic philosophy"
intact. [Editor's Note: At least one bill is being considered,
H.R. 1580, that would "reaffirm the authority of the states to
regulate terminal and station equipment used for telephone exchange.
service in certain instances . . .," recalling the *Consumer Communications
Reform Act*, also known in the 95th Congress as the "Bell Bill."]

Federal Computer Systems Protection Act.' Reintroduced January 25th
by Sen. Abraham A. Ribicoff (D-Conn.), *S. 240* (#) provides for a
stricter financial penalty for computer crime than the previous
version, stipulating that a fine could amount to as much as two and
one-half times that of the theft. In short, the bill would make it
a Federal crime to access a computer for fraudulent purposes such
as theft, sabotage or embezzlement.

EFT Legislation. Introduced January 23rd as *S, 108* (#) and *H.R. 1289*
(#), the *Truth in Lending Simplification and Reform Act* provides
that all of the provisions of the *EFT Act* (see *Washington Report*,
12/78, p. 1) would become effective this June instead of May, 1980,
as provided in the *EFT Act*. Supporters in the House and Senate are
predicting early passage with the President's approval expected in
"late Spring." In addition, *H.R. 852* would implement additional
EFT privacy legislation.

Electronic Mail. In his Supplementary State of the Union Message,
President Carter alluded to "proposals on the role of the Postal
Service in providing electronic mail services." The House Committee
on Post Office & Civil Service is planning hearings on electronic
mail, though not in connection with any legislation, according to
Michael F. Cavanagh, staff assistant, House Subcommittee on Postal
Personnel Modernization.

Copyright Protection   *H.R. 1007* would amend the *Copyright Act of
1976* to provide copyright protection for imprinted design patterns
on semiconductor chips.

Unsolicited Commercial Telephone Calls. *H.R. 377* would amend the
Communications Act of 1934 to "prohibit making unsolicited commercial
telephone calls to persons who have indicated they do not'wish to
receive such calls."

NSF Science Education Functions. *S. 210*, a bill to establish a
separate Department of Education, would transfer to the new secretary
of the proposed department "all programs relating to science education.
of the NSF or the director of the NSF." The legislation would
exempt such functions and programs as those relating to "ethical,
value, and science policy issues" or "communicating science information
to nonscientists."

Oversight Hearings. Consistent with the observation that the 96th
Congress is concentrating on oversight of existing Government
programs, budget hearings on the NTIA, the Office of Science &
Technology Policy, the National Bureau of Standards, and the Office
of Technology Assessment have been scheduled through this month.

'Contentious' Session. Overall, a "contentious" session is predicted
for the 96th Congress. Majority leader James C. Wright (D-Tex.)
has been quoted as saying, the President "still hasn't learned to
consult [with] Congressional leaders." Primary emphasis is expected
to be on the budget and related legislation. [Editor's Note: DP
aspects of the Fiscal Year 1980 budget will be analyzed in next
month's AFIPS Washington Report.]

AFIPS IN WASHINGTON

Standards Do Not Cover Recent Developments
in Information Processing, AFIPS Panel Says

CIVIL SERVICE SHOULD REVISE
PROPOSED STANDARDS FOR COMPUTER-RELATED OCCUPATIONS

Proposed Civil Service standards (#) affecting Government recruitment of
employees in computer-related occupations, first announced in 1978, are
already several years out of date and should be revised, according to
comments (#) released last month by an AFIPS panel.

(AFIPS/ P.M. McCarter)

AFIPS PANEL MEMBERS JOHN HAMBLEN (L), EDMUND SAWYER (R)

Recent Developments in Information Processing. According to the AFIPS panel, the proposed standards do not cover such recent developments in
the information processing field as the creation of distributive networks,
advances in telecommunications the use of intelligent terminals the
widespread application of minicomputers and microcomputers, and the
existence of on-line numeric and bibliographic data bases.

Panel Recommendations. The AFIPS panel recommended that the OPM (1)
consult with outside sources to update computer occupation standards;
(2) revise classification standards for computer-related occupations at
least every five years until at least 1990; and (3) insure that the
proposed standards conform with [existing] Civil Service law and regulations.
The group notes the pervasiveness of computer technology in Government,
the interaction of citizens with computers employed by the U.S. in various
programs, and the need for highly skilled and motivated personnel to
exploit the technology.

Panel Organization. The AFIPS Civil Service Standards Review Panel was
formed in response to a special invitation by the U.S. Civil Service
Commission, now the Office of Personnel Management (OPM), to comment on
tentative standards for the Computer Spcialist Series (GS-334) and the
Computer Clerk and Assistant Series (GS-335). The Federal government
employs standards to classify employees in pay levels according to the
difficulty, responsibility, and qualifications required for the work.

The panel reflects a variety of backgrounds including curricular work in
computer science, analysis of computer occupations for personnel purposes,
and computer usage. Comments reflect the views of the panel members,
not necessarily those of AFIPS, the Federation's constituent societies,
or the employers of the individuals involved.

Panel Members. Members of the panel were: Dr. Frances Berger, Psychometrics
Los Angeles; Dr. Karen Duncan, Mitre Corp., McLean, Va.; Dr. John Hamblen,
University of Missouri-Rolla; Charles D. LaBelle, Manufacturers Hanover
Trust Co., New York; William P. LaPlant, Jr., U.S. Air Force, Arlington,
Va.; Alexander D. Roth, Esq., AFIPS, Arlington, Va.; Dr. Terry Straeter.
NASA, Hampton, Va.; Edmund Sawyer, U.S. General Accounting Office,
Washington, D.C.; and Sidney Weinstein, Association for Computing Machinery,
New York.

New Draft. OPM is expected to issue another draft of its proposed standards
incorporating comments from groups such as AFIPS.

AFIPS Subcommittee Presents Comments to Fed on 'EFT Act'

CONSUMER LIABILITY COULD BE LIMITED TO $500 IN ALL EFT TRANSACTIONS

Proposed regulations (#) of the Board of Governors of the Federal Reserve
System (FRS) may misconstrue the *Electronic Funds Transfer (EFT) Act* (#)
to provide unlimited consumer liability in cases of unauthorized EFT
transfer, according to comments (#) released last month by an AFIPS EFT
Subcommittee. Passed by Congress last year, two sections of the *EFT Act*
pertaining to liability became effective February 8th The Subcommittee
comments reflect the views of the panel members and not necessarily
those of AFIPS, the Federation's constituent societies, the AFIPS Washington
Office, or the employers of the participants.

Unlimited Liability Questioned. According to two AFIPS Subcommittee members,
a "thorough reading" of the law "gives the impression that consumer liability
in <u>any</u> case is limited to $500.00." The Board interpretation, contained in
regulations published last December in the *Federal Register*, states, "If the
consumer fails to report within 60 days of transmittal of the periodic statement
any unauthorized electronic fund transfer which appears on the statement, the
consumer may be liable for the amount of any unauthorized transfer which the
financial institution establishes would not have occurred but for the failure
of the consumer to notify the financial institution."

Subcommittee Recommendations. Citing "adverse economic consequences of
unauthorized use," a Subcommittee majority recommended that a demand
deposit account should be established for the "express purpose" of EFT.
The majority also held that the provisions of the regulations interpreting
the consumer's liability section of the *EFT Act* should require actual
notice to the consumer before any debiting in excess of $500.00   According
to the AFIPS Subcommittee majority, "Evolving constitutional doctrines
affecting prehearing remedies for creditors suggest that in . . . [extreme
cases] there may be a constitutional requirement of prehearing notice
and an opportunity for a hearing to contest the proposed debiting before
such a 'taking' may be effected." 'Finally, the Subcommittee recommended
that the issuance of "access devices" which serve as combined debit or
credit cards should be prohibited, recognizing the increased risk of
technical failure in the transaction terminal.

Subcommittee Members. The Ad Hoc Subcommittee consists of four members
chosen by the chairman of the AFIPS Special Committee on EFTS, William
Fenwick, Esq., of Davis, Stafford, Kellman & Fenwick, Palo Alto, Calif.
Subcommittee chairman is Malcolm M. Jones, First National Bank of Denver.
Members are: Dr. John L. King, University of California, Irvine; John
C. Lautsch, Esq., Davis, Stafford, Kellman & Fenwick, Palo Alto, Calif.;
and Pender M. McCarter. AFIPS, Arlington, Va.

(AFIPS/ P.M. Mc Carter)

AFIPS SUBCOMMITTEE MEMBER JOHN L. KING

NEWS BRIEFS

Obligations for general-purpose data processing activities of Executive
Branch agencies are expected to increase $651.4 million (up 15.8
per cent) from Fiscal Year (FY) 1978 to FY 1979 and $492.4 million
(up 10.3 per cent) from FY 1979 to FY 1980, according to the Office
of Management & Budget (OMB); using the OMB estimate, in the two-year
period from FY 1979 to FY 1980, the largest "absolute growth" in
data processing and telecommunications resources is expected in the
Department of Defense (up 34.4 per cent), followed by the Department
of Health, Education & Welfare, and the Department of Energy.

Following the White House's lead (see *Washington Report*, 2/78, p. 7),
the Federal Communications Commission (FCC) has initiated an inquiry
into the legal and policy issues raised by a consideration of the
U.S. Postal Service's Electronic Computer Originated Mail (ECOM);
initial comments are due February 25th, oppositions by March 11th,
and replies by March 18th; also, in January, the Commission, as part
of its Zero-Based Regulatory Studies, has agreed to fund a report
on "Privacy and Communications Security: the FCC's Role."

The Departments of Justice and Treasury are proposing regulations which
would authorize the departments to require financial records from a
financial institution pursuant to tne formal written request procedure
established by the *Right to Financial Privacy Act of 1978* (see *Washington
Report*, 12/78, p. 1); deadline for comments to Justice is March 2nd;
Treasury, March 5th; the Federal Reserve System also sought similar
comment by February 16th.

In January, the Federal Telecommunications Standards Committee, with
representatives from numerous Government agencies, approved the
Advanced Data Communications Control Procedures (ADCCP) protocol;
also, the National Bureau of Standards is reported to be planning
to recommend interface standards for small computers and peripheral
equipment; finally, the Federal Trade Commission is also seeking
comment on a proposed trade regulation rule which reportedly would
affect the development and implementation of standards or. certification
procedures adopted by groups such as the American National Standards
Institute.

In January, the Federal Trade Commission adopted rules which would give a
pro rata refund to students who drop out of vocational schools offering
data processing-related courses; the rules become effective next
January, 1980.

Senate confirmation hearings on Anne Jones, named by President Carter to
succeed Margita White as member of the Federal Communications Commission,
are scheduled February 23rd; Ms. White is now expected to resign her
post February 28th unless Ms. Jones is confirmed beforehand.

Vol. 5, No. 4   April, 1979

Executive Branch Plans for DP Acquisitions Outlined

FY '80 BUDGET REQUESTS IN COMPUTER AREA CONTINUE TO RISE; ARPA, ADTS, NSF, NTIA REQUESTS SUMMARIZED

Despite a nearly 10 per cent drop in the overall Fiscal Year 1980 U.S.
Budget proposed by President Carter in January (from $588 billion to
$532 billion), the Administration's budget requests in the computer area
continue to rise. Nevertheless, concern has been expressed by some
professional groups, such as the Council of Scientific Society Presidents,
that scientific research budgets are "vulnerable" and that many will not
survive Congressional scrutiny.

Specific Requests. The Budget seeks $48 million for the Information
Processing Techniques Office of the Defense Department's Advanced Research
Projects Agency, reflecting a $6.2 million increase over the FY '79
budget request of $41.8 million. The Automated Data & Telecommunications
Service of the General Services Administration is asking for $8.97
million in FY'80, also representing an increase. $19.3 million is
requested for "Computer Research" by the National Science Foundation, up
from FY '79. The Commerce Department's National Bureau of Standards is
seeking $12.09 million in the area of "Computer Science & Technology,"
an increase over FY '79. The Science & Education Administration within

the Department of Agriculture is asking for $9.86 million to cover "Technical Information Systems," reflecting an increase over FY '79. In "Information Technology & Policy," Commerce's National Telecommunications & Information Administration has requested $3.9 million, slightly less than the previous fiscal year.

Executive Branch DP Plans. On January 30th, the Office of Management & Budget (OMB) released estimates compiling Executive Branch plans for major acquisitions of general purpose data processing from FY 1979 through FY 1980. These plans are outlined in the following chart.

Preliminary OMB Estimates of General Purpose Data Processing Resources in the FY 1980 Budget (Dollars in Millions)

	FY 78	FY 79 (est)	FY 80 (est)
Department of Agriculture	94 3	111 5	118 3
Department of Commerce	107 7	125 8	140 4
Department of Defense	1 936 0	2 278 6	2 602 7
Department of Energy	277 0	306 9	351 6
Department of HEW	430 0	492 2	529 0
Department of HUD	18 8	23 8	29 0
Department of the Interior	52 0	63 1	67 9
Department of Justice	39 9	47 2	50 1
Department of Labor	44 9	56 6	59 5
Department of State	12 4	18 6	21 0
Department of Transportation	69 4	85 5	98 5
Department of the Treasury	520 9	557 8	586 2
Environmental Protection Agency	35 2	38 5	40 9
General Services Administration	60 7	53 5	55 2
National Aeronautics and Space Adm	184 9	202 7	210 2
Veterans Administration	-80 5	113 4	⌐74 9
Corps of Engineers	29 1	39 3	40 6
National Science Foundation	14 2	17 4	17 9
Office of Personnel Management	12 5	11 9	12 0
Other Agencies	95 9	123 4	154 2
	4,116.3	**4,767.7**	**5,260.1**

Bill L Hommedieu

TRANSBORDER DATA FLOWS SUBCOMMITTEE DISCUSSES INCLUSION OF MANUAL FILES, LEGAL PERSONS IN OECD GUIDELINES

The Subcommittee on Transborder Data Flows of the State Department Advisory Committee on International Investment, Technology & Development met at the State Department on January 29. (The Advisory Committee has recently changed its name from the Advisory Committee on Transnational Enterprises.)

Thomas Pickering, Assistant Secretary of State for Oceans & International Environmental & Scientific Affairs, told the gathering that the State Department would publish internal papers on transborder data flow issues (issues relating to the international transmission of computer data) after the President's Privacy Initiative has been approved and announced. The papers were scheduled to be released by the Government Printing Office in early March. The State Department was also scheduled to convene a one-day seminar on transborder data flows last month.

Ambassador Herbert Salzman, of the U.S. Mission to the Organization for Economic Cooperation and Development (OECD), said that the most important unresolved issues before the OECD Drafting and Expert Groups are the inclusion of manual files, inclusion or exclusion of legal persons (i.e., corporations and certain other legal entities), handling of sensitive data, and establishment of a mechanism for the resolution of disputes.

Coverage of Manual Files. According to William Fishman, of the National Telecommunications and Information Administration (NTIA), the U.S. has taken the position that it is conceptually unsound to distinguish between automatic and manual processing when working to insure privacy protection. In some technical areas (such as microfiche technology), it is impossible to say whether the processing is manual or automatic. since it includes aspects of both. U.S. domestic law does not draw the distinction. Fishman noted that distinguishing between automatic and manual processing would weaken the "moral authority" of the guidelines. Finally, he said that restricting the guidelines to automatic processing would cause governments to retreat to manual files to evade the effect of the guidelines.

Fishman noted, on the other hand, that most European legislation only affects automatic processing. The Europeans also point out that the origin of these privacy concerns comes from computer developments. They also claim that it would be difficult for their data inspection boards to cover the many sensitive uses of manual files.

Inclusion of Legal Persons (Corporations) as Protected Parties. Fishman observed that privacy protection is a civil rights issue in the U.S. and not an issue of corporate regulation. In the U.S. view, limiting protection to natural persons would make the guidelines relate more clearly to privacy issues. The U.S. believes that the OECD is not in any event in a position to broaden the guidelines to include legal persons until it has studied the area.

Fishman conceded that some European laws cover legal persons (with some variation among them). In some cases, "smaller" legal persons would be excluded from coverage. He noted that the draft treaty of the Council of Europe would cover legal persons.

U.S. Sees Privacy Motivations, Not Trade Protection. Fishman emphasized that the U.S. sees the current effort as motivated largely by privacy concerns -- civil rights, democratic concerns. While some nations clearly want to limit foreign data processing from their markets, Fishman said that that interest is not significant in the current OECD effort.

Another Meeting Scheduled; Subcommittee Expanded. The State Department scheduled another meeting of the subcommittee for March 9th to consider the next draft of the guidelines, so public comment can be provided to the U.S. delegation for a March 12th Drafting Group meeting.

Subcommittee Chairman Hugh Donaghue announced that the subcommittee membership has been expanded by the addition of seven members. In the near future the subcommittee will form subgroups to consider economic, tariff, employment, and other issues in more detail.

-- Alexander D. Roth

AFIPS IN WASHINGTON

AFIPS PANEL FORMING ON
PROPOSED NATIONAL COMMISSION ON USE OF COMPUTERS

An AFIPS Panel is being formed to study the implications of a pending proposal for a National Commission the the Use of Computers in Education (see *Washington Report*, 11/78, pp. 5-6), Alexander D. Roth, director of the AFIPS Washington Office, announced last month. Co-chairing the AFIPS Panel are Dr E. Ronald Carruth, director for District Services, Minnesota School Districts, St. Paul; and Prof. A.A.J. Hofman, Computer Science Program, Texas Christian University, Fort Worth. Individuals interested in serving on the panel should contact Mr. Roth at (703) 243-3000. A bill to establish the commission, introduced at the close of the last Congress (see *Washington Report*, 12/78 p. 4), is expected to be reintroduced this year.
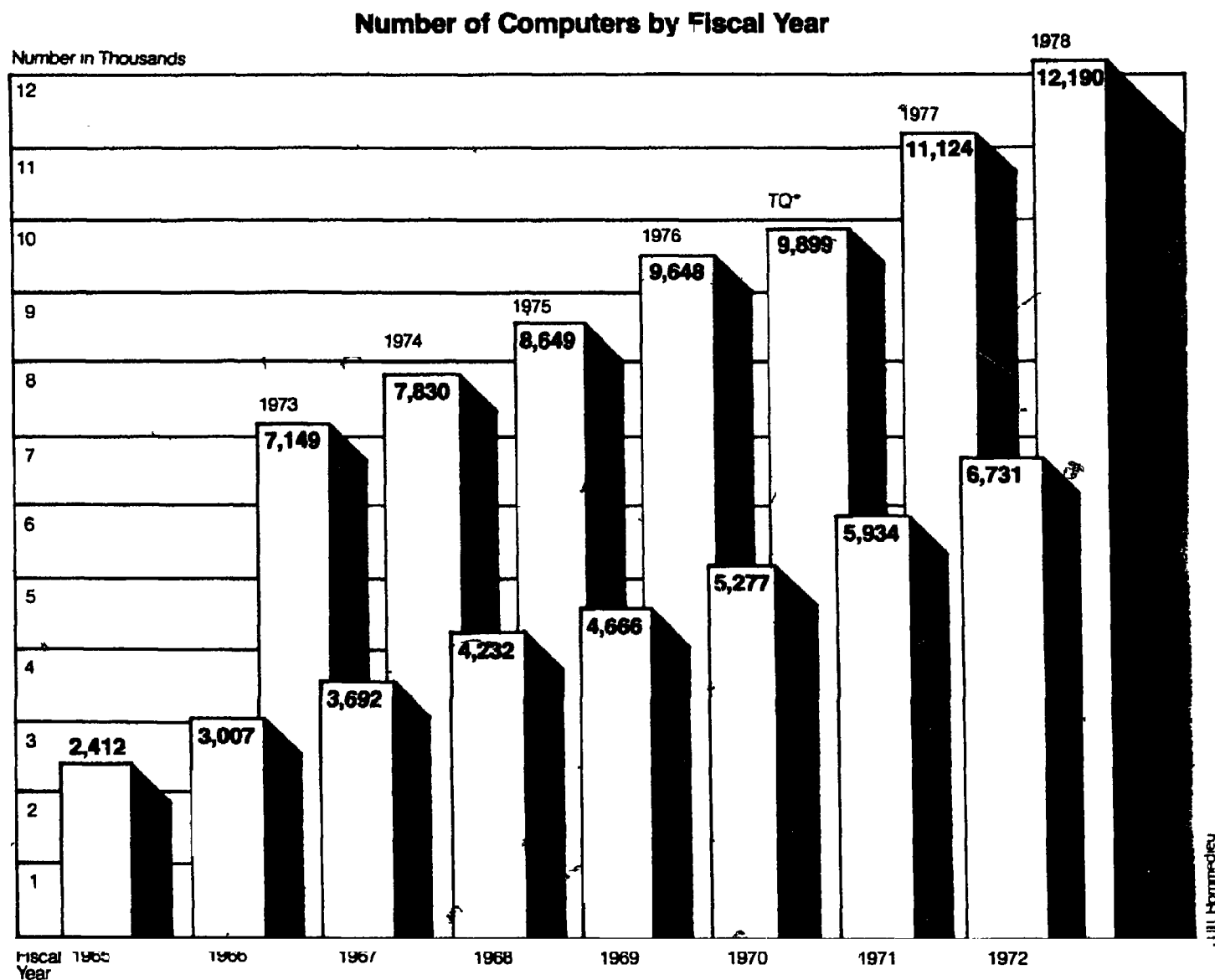
NEWS BRIEFS

The 1956 Justice Department *Consent Decree* with AT&T should be modified to permit the telephone company to engage in data communications, according to legislation introduced March 12th by Sen. Ernest F. Hollings (D-S.C.); the bill, amending the *Communications Act of 1934*, will be detailed in next month's *AFIPS Washington Report*.

In February, the House joined the Senate in passing *S. 37* (see *Washington Report*, 3/79, p. 3: repealing the notice requirement of the *Right to Financial Privacy Act*; also, in February, Rep. Richardson Preyer (D-N.C.) introduced the *Omnibus Right to Privacy Act of 1979*, H.R. 2465 (#), identical to legislation he introduced in the last Congress with the exception of a title concerning confidentiality of medical records.

In February, Secretary of Commerce Juanita M. Kreps approved adoption of the *I/O Channel Level Interface*, the *Power Control Interface*, and the *Channel Level Operational Specifications for Magnetic Tape* as Federal Information Processing Standards (FIPS); a fourth FIPS, the standard for rotating mass storage subsystems (#) was proposed by the National Bureau of Standards in January (see *Washington Report*, 10/78, p. 1).

"[T]he software development program for ACS [the Advanced Communications
   Service] will require a significant future effort previously unforeseen," '
   AT&T told the Federal Communications Commission in February, postponing
   its plans to file ACS tariffs this June as previously announced (see
   Washington Report, 12/78/ p. 6); however, AT&T repeated its request
   for a declaratory ruling which would permit the Bell System to offer
   ACS over the telephone company's existing digital facilities.

The number of Federal government computer installations (*i.e.*, including
   general-purpose computer systems and minicomputers) has risen 9.6
   per cent from 11,124 in FY 1977 to 12,190 in FY 1978, as shown in
   the accompanying chart, according to the General Service Administration's
   (GSA) *Inventory of Automatic Data Processing Equipment in the
   United States Government*, released last month by the GSA's Auto-
   mated Data & Telecommunications Service; total value of Federal
   computer installations rose from $4.77 billion in FY 1977 to $4.89
   billion in FY 1978.

Number of Computers by Fiscal Year

Number in Thousands

1978 — 12,190
1977 — 11,124
TQ* — 9,899
1976 — 9,648
1975 — 8,649
1974 — 7,830
1973 — 7,149
1972 — 6,731
1971 — 5,934
1970 — 5,277
1969 — 4,666
1968 — 4,232
1967 — 3,692
1966 — 3,007
1965 — 2,412

Fiscal Year

TQ Transition Quarter (July 1 1976—September 30 1976)   *September 30 1978*

In February, the Social Security Administration promulgated new rules
   (#) to "protect the integrity of the social security number (SSN)
   by reducing its misuse"; the rules require additional identification
   for issuing cards with SSNs as well as for issuing duplicates or
   corrected cards.

In February, the National Association of Trade & Technical Schools filed suit to contest the Federal Trade Commission's (FTC) rules (See *Washington Report*; 3/79, p. 8) which will give a *pro rata* refund to students who drop out of vocational schools; the association contends the FTC has used outdated information in formulating the rules which become effective next year.

Dr. Leland Johnson, formerly associate administrator for Policy Analysis & Development, National Telecommunications & Information Administration (NTIA), has been named chief economist, NTIA; Dale Hatfield, head of the Federal Communications Commission's Office of Plans & Policies, succeeds Dr. Johnson as associate administrator; William Fishman, formerly deputy associate administrator for Policy Analysis & Development, NTIA, has been named director of the NTIA Office of Planning & Policy Coordination.

Rep. Morris K. Udall (D-Ariz.) and newly-elected Sen. Ted Stevens (R-Ala.) succeed Sen. Edward M. Kennedy (D-Mass.) and Rep. Larry Winn, Jr. (R-Kan.), respectively, as chairman and vice-chairman of the Congressional Board of the Office of Technology Assessment (OTA); Dr. Eric H. Willis, a nuclear physicist has been appointed assistant director of OTA; Dr. Willis also heads OTA's Science, Information & Transportation Division.

In January, the President named three new members of the National Commission on Libraries & Information Science (NCLIS): Francis Keppel, director, Aspen Institute Program in Education for a Changing Society; Bessie B. Moore, executive director, Arkansas State Council of Economic Education; and Philip A. Sprague, consultant, Milton Roy Co.

Steven J. Jost, former Congressional aide, has been named director of the DPMA Washington Office.

"[T]he lack of knowledge about the dimension of the real and potential restraints on transborder data flow" is the "most serious constraint" on U.S. policymaking, according to a Carter Administration report filed with the House International Operations Committee and the Senate Commerce, Science & Transportation Committee, as required by the *Foreign Relations Authorization Act of 1979* (see *Washington Report*, 12/78, p. 2).

"[F]ederal law should allow . . . [electronic funds transfer (EFT)] to develop in an aura of consumer confidence, a pro-EFT mood rather than a negative, anti-environment, a situation which financial institutions might never be able to overcome," according to an American Bar Association (ABA) Subcommittee on EFT; in a report, completed in February, the Subcommittee on EFT of the Law and Computer Committee, ABA Section on Law & Technology, concluded that "at this stage in the development of EFT, most consumers, and even financial institution customers, do not appear to perceive statutory safeguards as a key factor in persuading them to use EFT."

Ed.: Information for the April, 1979, *AFIPS Washington Report* is current as of March 14, 1979, press time. Production assistance for the *Washington Report* is provided by Linda Martin. AFIPS societies have permission to use material in the newsletter for their own publications. Documents indicated by the symbol "(#)" are available on request to the Washington Office. Requests should specify the date(s) of the *Report* in which the document(s) appeared. Where price is noted, make checks payable to "AFIPS."