



MICROCOPY RESOLUTION TEST CHART

NATIONAL BUREAU OF STANDARDS-1963-A

AD A 082119

Report No. 432



LEVEL III

Theoretical Studies in Natural Language Understanding

Annual Report

1 May 1978 to 30 April 1979

**DTIC
ELECTE
MAR 18 1980
S D
B**

Prepared for:
The Office of Naval Research

FILE COPY

DISTRIBUTION STATEMENT A
Approved for public release;
Distribution Unlimited

80 3 14 022

6 THEORETICAL STUDIES IN NATURAL LANGUAGE UNDERSTANDING.

9 Annual Technical Progress Report.
1 May 1978 — 30 April 1979,

12/21

Program Code No. ID 121408
Contract No. N00014-77-C-0371
Name of Contractor:
Bolt Beranek and Newman Inc.
Effective Date of Contract:
1 May 1977
Amount of Contract:
\$250,350

Short Title of Work:
Language Understanding 11 30 Apr '79
Contract Expiration Date:
30 April 1980
Principal Investigator:
William A. Woods
(617) 491-1850, ext. 4361
Scientific Officer:
Marvin Denicoff

14 BEN-4332

Sponsored by
The Office of Naval Research

This research was supported by the Office of Naval Research
under Contract No. N00014-77-C-0371.

060100

Unclassified

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

REPORT DOCUMENTATION PAGE		READ INSTRUCTIONS BEFORE COMPLETING FORM
1. REPORT NUMBER BBN Report No. 4332 ✓	2. GOVT ACCESSION NO.	3. RECIPIENT'S CATALOG NUMBER
4. TITLE (and Subtitle) THEORETICAL STUDIES IN NATURAL LANGUAGE UNDERSTANDING Annual Report 1 May 1978 - 30 April 1979 A060332	5. TYPE OF REPORT & PERIOD COVERED Annual Report 5/1/78 - 4/30/79	
	6. PERFORMING ORG. REPORT NUMBER BBN Report No. 4332	
7. AUTHOR(s) W. A. Woods	8. CONTRACT OR GRANT NUMBER(s) N00014-77-C-0371 ✓	
9. PERFORMING ORGANIZATION NAME AND ADDRESS Bolt Beranek and Newman Inc. ✓ 50 Moulton Street Cambridge, MA 02238	10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS	
11. CONTROLLING OFFICE NAME AND ADDRESS Office of Naval Research Information Systems Program Arlington, VA 22217	12. REPORT DATE 30 April 1979	
	13. NUMBER OF PAGES 32	
14. MONITORING AGENCY NAME & ADDRESS (if different from Controlling Office)	15. SECURITY CLASS. (of this report) Unclassified	
	15a. DECLASSIFICATION/DOWNGRADING SCHEDULE	
16. DISTRIBUTION STATEMENT (of this Report) Distribution of this document is unlimited. It may be released to the Clearinghouse, Department of Commerce, for sale to the general public.		
17. DISTRIBUTION STATEMENT (of the abstract entered in Block 20, if different from Report)		
18. SUPPLEMENTARY NOTES		
19. KEY WORDS (Continue on reverse side if necessary and identify by block number) English-like language, JARGON, KL-ONE, knowledge acquisition, knowledge representation, natural language understanding, parsing, semantic networks, structured inheritance networks.		
20. ABSTRACT (Continue on reverse side if necessary and identify by block number) This report gives a brief overview of the ONR project in natural language understanding at BBN, and a detailed presentation of a language, JARGON, which was developed under that project. JARGON is a formalized and somewhat stylized subset of natural English designed to serve as a lexical language for a KL-ONE structured inheritance network and to permit experimental investigation of the growth and evolution of a complex		

Unclassified.

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

20. Abstract (cont'd.)

knowledge base. The JARGON language has been used to explore the processes of building up and evolving knowledge structures, and has given us several insights into the nature of knowledge representation and acquisition.

Unclassified.

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

TABLE OF CONTENTS

1. INTRODUCTION 1

2. THE JARGON LANGUAGE 5

 2.1 JARGON Syntax 6

 2.1.1 Possessives, Roles and the verb HAVE 7

 2.1.2 Noun Phrases 8

 2.1.3 The verb BE 10

 2.1.4 Naming and Describing Concepts 11

 2.1.4.1 Relative Clauses 12

 2.1.4.2 Conjunctions 15

 2.1.5 Describing Roles 16

 2.1.5.1 Disambiguating Role Descriptions 20

 2.1.6 Editing and Browsing 21

 2.1.7 Structural Conditions 22

 2.1.8 Quantification 23

 2.1.9 Anaphoric Reference 24

 2.2 Uses of JARGON 27

3. CONCLUSIONS 30

4. REFERENCES 32

ACCESSION for	
NTIS	White Section <input checked="" type="checkbox"/>
DDC	Buff Section <input type="checkbox"/>
UNANNOUNCED	<input type="checkbox"/>
JUSTIFICATION _____	
BY _____	
DISTRIBUTION/AVAILABILITY CODES	
Dist. AVAIL and/or SPECIAL	
A	

1. INTRODUCTION

The ONR project on theoretical studies in Natural Language Understanding is concerned with developing the theoretical underpinnings for a system that can understand and deal with an open-ended range of natural concepts. Such capabilities are needed for a variety of natural language understanding systems, but most crucially are needed for systems of the kind that I have been calling Knowledge Management Systems - these are systems whose underlying "data" may be fundamentally natural language information that the system is required to "understand". This includes systems such as a military debriefing system that would be able to assimilate and understand military debriefing reports in sufficient detail to look for similarities and patterns, sophisticated command information systems that are able to represent models of intentions of enemy and friendly forces as well as current and past positions, and sophisticated command decision aids that are capable of representing and displaying complex alternative hypothetical courses of action and projecting likely outcomes of alternatives.

A principal component of a knowledge management system is a representational system - a set of conventions for structuring and storing knowledge inside the machine. This representational system must be such that it is adequate to encode any piece of knowledge that one would like to store and also structured so as to

facilitate the inferential processing that must take place on this knowledge in order to understand sentences, assimilate new knowledge, and answer questions. In particular, it must support the associative accessing of pieces of knowledge related to a given piece, and it must be able to capture generalities - i.e., represent generic facts at the appropriate level of generality so that a single fact need be learned and stored only once and not separately as many individual cases.

Critical to being able to construct systems of the kind we envisage, is the development of a representational system that is able to handle open-ended domains of knowledge. Work in artificial intelligence has produced a number of "semantic network" notations which purport to represent information in various domains, but the semantics of such notations has for the most part been undefined (see Woods [1975]). Recent interest in knowledge representation has increased the activity in this area, but in most such systems the semantics of the notation is still unclear. In most such representations there is no consistent understanding of the semantic import of the various types of connections between concepts that is "understood" and enforced (and hence usable) by the representation system itself. A consequence of this is that as new kinds of knowledge are added to such a network, new link types need to be created and their meaning has to be somehow inserted into the

various inference functions that operate on the network. This turns out to be a severe restriction on the extensibility of a knowledge base expressed in such a system.

What is required is a uniform framework for representing knowledge that is independent of specific domain content and within which domain-specific concepts can be defined. Consequently, the natural language research group at BBN, under partial ONR support has been working on the development of a knowledge representation system, KL-ONE, whose emphasis is on the development of a universal set of "epistemological primitives" whose semantics is well understood and which are sufficient to construct the representations of any domain-specific concepts that one might want to define. This representation system takes the form of a structured inheritance network [Brachman, 1977] and is described in some detail in our last annual report [Brachman, 1978].

A major emphasis of our ONR project has been on the problems of acquiring and evolving a knowledge base in a complex domain. An important component of this work is an effort to extend our understanding of theoretical aspects of the semantics of knowledge representation and of algorithms that would be used in mechanical inference and perception with respect to such a knowledge base. Another is an attempt to understand the nature of a conceptual knowledge base and the way such knowledge bases can grow and

evolve. To support this latter effort, we have begun to develop a language within which a human experimenter can build up and modify a knowledge base and thereby explore the nature of evolutionary knowledge acquisition. This language, called JARGON, has enabled us to gain an understanding of the process of knowledge base evolution and also appears to be stimulating the understanding of KL-ONE concepts. (It does the latter by suggesting natural conceptualizations that should be concisely representable and by giving natural language glosses for the different KL-ONE structures which aid in the intuitive understanding of their use and meaning.)

During the past year, we have continued to advance our understanding of the epistemological foundations for knowledge representation and have made some improvements to the knowledge representation system KL-ONE. In addition, as mentioned above, we have begun the development of the JARGON language for evolving a complex knowledge base. This language is an English-like lexical notation for KL-ONE which is a combination of an input/output language and a knowledge structure editor. We have implemented an initial parser and interpreter for this language which allows a person to conveniently build up and modify a knowledge structure in KL-ONE and to gradually evolve that structure into an increasingly more adequate model of a knowledge domain. The JARGON interface has served to drive our knowledge representation research in interesting ways.

JARGON performs both editing and input functions with syntactic constructions that follow closely the form and structure of natural English. Although JARGON is similar to other "English-like" languages in the sense that it makes some radical simplifications in the range of syntax that it permits, it differs from many such languages in that it preserves rather faithfully the underlying conceptual structures of the sentences that it understands. The JARGON language has been used as a tool for building a number of knowledge networks for various domains of knowledge to gain some experience with the processes of acquiring a model of a knowledge domain. In this report, we will give an overview of the JARGON language and a discussion of its use.

2. THE JARGON LANGUAGE

JARGON is a formalized, stylized subset of natural English which is intended to be (or become) an epistemologically complete and relatively natural language in which to express the conceptual structures that underly natural language use and human thought. It is intended to serve as a surface (lexical) language for an underlying structured inheritance network. Its major verbs are natural formalizations of the English words "be", "have", and "satisfy", which together with a few other verbs (such as "called") appear to constitute the bulk of an epistemologically complete

foundation for an open-ended range of natural concepts. JARGON aspires to be able to represent the storable characteristics of any natural concept using this foundation of basic epistemological operators and references to other concepts defined in terms of them. It is unlikely that in its current state JARGON fully achieves this goal, but it has made significant progress in that direction.

2.1 JARGON Syntax

JARGON makes some considerable simplifications to the syntactic structure of ordinary English in order to avoid the necessity of having extensive lexical knowledge about the open-ended vocabulary that it will use to name concepts and roles. The most dramatic example of this is the introduction of an explicit and separate plural morpheme "PL" to be used preceding the singular form of a noun to express plurality rather than using the English conventions of adding "s", "es", "ies", or using irregular plurals. Thus, in JARGON one uses the locution "PL HORSE" instead of "HORSES", "PL TAX" instead of "TAXES", "PL STICK" instead of "STICKS", and "PL CHILD" instead of "CHILDREN". JARGON also completely ignores gender constraints on the use of pronouns and determiners (e.g., "his", "her", and "its" can be used indiscriminately refer to things regardless of gender). These

simple differences between JARGON syntax and that of ordinary English free JARGON from having to have specialized knowledge of English irregular forms such as "people" (PL PERSON) and "children" (PL CHILD). Moreover, they can be stated as a few simple rules that are not difficult for people to become accustomed to using.

2.1.1 Possessives, Roles and the verb HAVE

Jargon has a fairly extensive range of English noun phrase structure, including a model of possessives. In general, its use of possessives and the verb "have" correspond to the KL-ONE notions of "possession" of "generalized attributes" (called "Roles"). The Roles in KL-ONE Concepts represent generalizations of the ordinary notion of an attribute of an object which include not only attributes such as color, but also things like arguments of functions, cases of verbs, parts of composite objects, and objects, facts, and measurements associated with a concept. JARGON permits statements about Concepts having Roles, the fillers of Roles having Roles, etc. For example, JARGON understands the noun phrases:

JOHN'S MOTHER'S SISTER'S SON

THE SON OF JOHN'S MOTHER'S SISTER

THE SON OF THE SISTER OF JOHN'S MOTHER

THE SON OF A SISTER OF JOHN'S MOTHER

THE SON OF A SISTER OF THE MOTHER OF JOHN

and the sentences:

JOHN HAS A MOTHER

JOHN'S MOTHER HAS A SISTER.

(Notice the ability of the various locutions above to specify or not specify determiner information for intermediate steps along a chain of roles.)

2.1.2 Noun Phrases

JARGON's noun phrases come in basically two types - **Concept designations** and **Role chains** (the latter being indicated by chains of possession using "'S" or "OF"). Role chains are always anchored to a Concept (called the "root" of the chain). Such expressions in JARGON are interpreted as focus-subfocus chains a la Brachman [1977] in the underlying KL-ONE representation. Such chains consist of a sequence beginning with a Concept and followed by pointers successively to a Role of that Concept, a Role of a value restriction of the previous Role, a Role of a value restriction of that Role, ..., etc. This interpretation turns out to be a natural and intuitive correspondence between a basic KL-ONE structure and a common form of natural language reference. The naturalness of this correspondence appears to support the belief that the role chains of KL-ONE have a correspondence in reality to the way people

structure knowledge. Moreover, this correspondence gives a motivated model for what otherwise seems a special set of locutions in ordinary English.

JARGON's syntax permits most of the normal determiner structure of English noun phrases, with the one non-English syntactic convention that plural nouns in JARGON are represented with a prefix determiner PL (As mentioned earlier, this eliminates the need for JARGON to have morphological rules and a dictionary for irregular pluralization of nouns.) Thus, JARGON permits the noun phrase constructions:

A PERSON

PL PERSON (i.e., "people")

ALL PL PERSON

NONE OF THE PL PERSON

JOHN'S PL FINGER

JOHN'S PL ARM'S PL FINGER.

Determiners in noun phrases that specify roles can indicate numerical ranges for the number facet of the role as in:

A PERSON HAS 2 PL PARENT

A PERSON HAS AT MOST 2 PL EYE

JOHN HAS AT LEAST 5 PL CHILD

A SUBURBANITE HAS FROM 2 TO 5 PL RECREATIONAL-ACTIVITY

2.1.3 The verb BE

JARGON understands senses of the verb "be" corresponding to the KL-ONE superConcept relationships between generic concepts, the individuates relationship between an individual concept and a generic, the value restriction relationship between a role and a concept, and the mods, diffs, and sats relationships between roles. It also understands senses of "be" as a copula linked to predicates such as "be called", "be like", "be the same as", etc. Examples of the use of "be" in JARGON are:

JOHN IS A PERSON (individual concept individuates a generic)

A PERSON IS A LEGAL ENTITY (generic concept is a subconcept of another)

A PERSON'S PL ARM ARE SOME OF IT'S PL APPENDAGE (differentiation relationship between a role and another role)

A PERSON'S FATHER IS THE SAME AS HIS MOTHER'S HUSBAND (structural condition relating two roles or role chains from a concept)

The first of these creates an individuates link between the concepts JOHN and PERSON. The second creates a superconcept link between PERSON and LEGAL-ENTITY. The third creates a differentiation link between a role named ARM attached to PERSON and a role named APPENDAGE which is either attached to or inherited by PERSON. (If PERSON inherits a role ARM rather than having one attached, then a local role is created which will be a modification

of the inherited role. If PERSON has no role ARM, then one will be created.) The last sentence creates a special kind of structural condition (called a "Role-value map") between the FATHER Role of PERSON and a Role chain that starts with the concept person and contains pointers to the MOTHER Role of PERSON and the HUSBAND Role of some value restriction of the MOTHER Role. (This construction assumes that the value restriction of PERSON's MOTHER Role is known and that it in turn has a Role named HUSBAND. This would be true if for example one had previously said A PERSON'S MOTHER IS A MARRIED-WOMAN and A MARRIED-WOMAN HAS A HUSBAND.)

2.1.4 Naming and Describing Concepts

In JARGON, Concepts can be named either by unique KL-ONE names or by descriptions. When a description is used, if the Concept named by the description already exists as a Concept in the KL-ONE network, then that concept is referred to. If such a Concept does not exist, then one is created and it is automatically linked by superConcept links to all existing Concepts in the network which subsume it (and it is linked by inverse superconcept links and inverse individuation links to all Concepts which it subsumes). For named Concepts, JARGON employs a fairly natural Convention for indicating which ones are individual concepts and which are generic: generic concepts are referred to by noun phrases whose

determiner is A or PL, while individuals are referred to by single names without determiner (as in English proper names). For individuals introduced by description, a special use of the adjective INDIVIDUAL is used to distinguish individuals from generics (the default interpretation is generic if the adjective INDIVIDUAL is not used). Examples of the four major kinds of noun phrases that refer to concepts are:

JOHN (a named individual)

A PERSON (a named generic)

A PERSON WHOSE SEX IS MALE (a generic description)

AN INDIVIDUAL PERSON (an unnamed individual)

2.1.4.1 Relative Clauses

As indicated in one of the examples above, JARGON permits relative clauses to be used in descriptions of concepts. JARGON's syntax for relative clauses is virtually the same as ordinary English usage, permitting relative clause modifications such as:

A PERSON WHOSE HOBBY IS GOLF

A PERSON ONE OF WHOSE PL HOBBY IS GOLF

A PERSON SOME OF WHOSE PL HOBBY ARE PL SPORT

A PERSON WHO IS A PHILOSOPHER

The first of these will create a concept which is a subconcept of PERSON and which has a Role whose name is HOBBY and whose value is

(the individual) GOLF. If PERSON already has a HOBBY Role, then the new role attached to this new Concept will be linked to the HOBBY Role of PERSON with a MODS link (unless by chance the HOBBY Role of PERSON already has GOLF as its value, in which case no new Role need be created, and in fact no new Concept will be needed, and hence the entire expression will refer to the named Concept PERSON). The second two descriptions assume that it is already known that PERSON has (or inherits) a Role HOBBY and will create a subconcept with a Role which differentiates the HOBBY Role of PERSON into a subrole with the indicated restriction. The last description describes a concept which has superconcept pointers to both PERSON and PHILOSOPHER.

Descriptions of concepts can be used in sentences whose verb is BE to add concepts to the network and/or to create connections between them. Both the subject and the object of the BE assertion will be created if they do not already exist, and then a superconcept link will be added between them. Examples of such sentences are:

A PERSON IS A PHYSOBJ

A MAN IS A PERSON WHOSE SEX IS MALE

A PERSON WHOSE HOBBY IS GOLF IS AN UPPERCLASS-PERSON

A PERSON WHOSE HOBBY IS GOLF IS A PERSON WHOSE STATUS IS HIGH

An interesting ambiguity occurs in the second sentence above that JARGON users must be aware of. This sentence would tend to be read by most English users as a definition of MAN, but it is not so treated by JARGON. Rather, JARGON creates a concept for PERSON WHOSE SEX IS MALE and then links the named concept MAN to it by a superconcept link. The reason for this is that not all such sentences (i.e., sentences whose subject is a named concept and whose object is a modified description) are properly taken as definitions. (E.G. AN OSTRICH IS AN AUSTRALIAN BIRD does not imply that every AUSTRALIAN BIRD is an OSTRICH). The interpretation of such sentences in ordinary English is done by pragmatic considerations of whether the statement is plausible as a definition or not. Since we would like the JARGON syntax to be unambiguous syntactically and not to mean different things depending on what the system knows or believes about the plausibility of definitions, the most general (non-definitional) interpretation of such sentences is taken. If one wants to define MAN to be A PERSON WHOSE SEX IS MALE, he can use either of the locutions:

A PERSON WHOSE SEX IS MALE IS CALLED A MAN

A MAN IS A PERSON AND HAS SEX MALE.

2.1.4.2 Conjunctions

One of the above examples contains a use of the conjunction AND. JARGON permits the use of conjunctions in a variety of ways that all faithfully follow ordinary English usage. However, in order to eliminate ambiguity, JARGON does not permit certain conjunction constructions that ordinary English would allow. Basically JARGON does not allow simple conjunction of complete sentences (e.g., JOHN IS A PERSON AND MARY IS A PERSON) since the initial portions of such sentences could be ambiguous and their parsing would thereby be less efficient. For example, JOHN IS A PERSON AND A PHILOSOPHER... could be followed either by ...AND A LOVER or ...IS A SCIENTIST. Since such simple conjunctions can be said just as easily (with one fewer word in fact) as two separate assertions, JARGON permits only the conjoined noun phrase interpretation of two noun phrases that occur separated by a conjunction. (This restriction may be eliminated in a later version of JARGON by using a more clever parsing function, but for the moment it is not a serious restriction. As a result of such simplifications in its syntax, the JARGON parser is extremely fast, which is a more desirable property for its intended uses.) In general, JARGON permits conjoined noun phrases, conjoined verb phrases, reduced conjoined verb phrases, and conjoined relative clauses as in:

DAVID AND RON ARE PL SCIENTIST (conjoined subject NP's)

A PERSON HAS AN OCCUPATION AND IS AN ANIMATE OBJECT
(conjoined verb phrases)

A PERSON HAS AN OCCUPATION AND PL HOBBY (reduced
conjoined verb phrases - i.e., the second HAS is
implicit)

A PERSON WHOSE SEX IS MALE AND WHOSE HOBBY IS GOLF
(conjoined relative clauses)

Scope ambiguity of multiple conjunctions is resolved by a default convention that a phrase following a conjunction is conjoined with the immediately preceding thing it could be conjoined with. Parentheses can be used in some cases to force scoping that would violate this convention. Examples of conjunction scoping are:

A PERSON WHOSE HOBBY IS A SPORT WHOSE PACE IS VIGOROUS
AND WHOSE OBJECT IS A BALL (the object of the sport is a
ball, not the object of the person)

A PERSON WHOSE HOBBY IS A SPORT (WHOSE PACE IS VIGOROUS)
AND WHOSE WIFE IS A SOCIALITE (the person has the wife,
not the sport).

2.1.5 Describing Roles

Unlike Concepts, Roles in JARGON are never referred to uniquely by name. This corresponds to the fact that Role names in the underlying KL-ONE are not necessarily distinct for distinct Roles. In the underlying representation Roles are specified by their internal handles and not by name. It is even possible to

have a Role without a name. Role descriptions in JARGON come in two varieties - noun phrases occurring as objects of the verb HAVE and noun phrases containing possessive elements 'S or OF. Descriptions of the first kind are used to construct Roles in statements of Role ownership, while the second are used to refer to (presumably) previously defined roles or to chains of previously defined roles. References of the second kind are used in statements of structural conditions attached to concepts (discussed below).

In a Role description consisting of a noun phrase object of the verb HAVE, the head noun is usually taken as the name of the Role being described. (Only for certain specific head nouns such as "role" is this not true.) In many cases, the Role name is the same as the name of a Concept that occurs in the network. Using ordinary English intuitions, the fact that a Role has a name that is the same as the name of a Concept would be suggestive that its values are restricted to be instances of that Concept. For example, the Concept EMPLOYEE may have a Role named COMPANY, which would suggest that the filler of that role would also be a COMPANY. However, JARGON does not enforce this interpretation even when there is a Concept with the same name as a Role. Instead, JARGON forces the user to explicitly indicate any value restrictions that he wants attached to Roles. (The above example will leave the

value restriction of the COMPANY Role of EMPLOYEE unspecified.) JARGON provides ways to specify the value restriction of a Role by means of locutions such as:

AN EMPLOYEE HAS AN EMPLOYER WHICH IS A COMPANY

A MAN'S SEX IS MALE

A MAN HAS SEX MALE

The first of these sentences creates a role attached to EMPLOYEE whose name is EMPLOYER and whose value restriction is COMPANY. The second apparently presupposes that a man has a sex and adds the value restriction MALE (although it will create a sex role if one does not already exist). If MAN inherits its sex role from a superConcept (e.g., from PERSON) rather than having it attached directly, then this sentence will create a local role which is a modification of the higher role and attach the new value restriction to the local one. The third sentence is a special locution which is available for specifying values of Roles whose value can be named with a single word.

In some cases one may want to create a Role without specifying a name. JARGON provides a specialized syntax for expressing this case by use of the adjective "associated" as in:

A PERSON HAS AN ASSOCIATED COMPANY

This indicates that the concept PERSON has a Role whose name is unstated, but whose value restriction is COMPANY. This is distinct from:

A PERSON HAS A COMPANY which (as discussed above) indicates that PERSON has a Role named COMPANY but does not indicate a value restriction for that Role. The full citation of both Role name and value restriction could be expressed variously as:

A PERSON HAS A COMPANY WHICH IS A COMPANY

A PERSON HAS AN ASSOCIATED COMPANY WHICH IS CALLED HIS COMPANY

The latter locution can be used to attach a name to a Role that does not have one or to change the name of a Role or a concept as in the sequence:

A PERSON HAS AN ASSOCIATED COMPANY

A PERSON'S ASSOCIATED COMPANY IS CALLED HIS COMPANY

A PERSON'S COMPANY IS CALLED HIS EMPLOYER A PERSON IS CALLED AN EMPLOYEE

This sequence adds a Role to PERSON whose value restriction is COMPANY, then adds the Role name COMPANY to that Role, then changes that Role name to EMPLOYER (in a later implementation this may add an alternative name rather than changing the name), and then finally changes the name of the Concept PERSON to EMPLOYEE. The ability to change the names of Roles and Concepts without changing

their structure provides a very useful tool for the evolution of a satisfactory knowledge representation of a given domain. When one discovers that the attributes that he has been associating with a concept indicate that the concept is more properly named something different (as in the above example where the attributes attached to person were actually attributes of employed people and not people in general) then the simplest change may be to simply rename the concept rather than to transfer all of its properties to a new one.

2.1.5.1 Disambiguating Role Descriptions

The fact that Role names are not unique permits situations in which a single Concept will inherit two Roles of the same name from different inheritance paths. In order to enable a JARGON user to refer unambiguously to the different Roles in such cases, JARGON permits Role names to be modified with an "as" prepositional phrase indicating the concept to which the named Role is attached. For example, John's title as an employee is different from his title as an officer of his sailing club. Thus one can say such things as:

A PERSON'S JOB-TITLE IS HIS TITLE AS AN EMPLOYEE If one should encounter a situation where several Roles of the same name are attached to the same concept, then one can disambiguate them by means of the ORDINAL determiner discussed below.

2.1.6 Editing and Browsing

We discussed above the use of the locution "be called" to change the name of a Concept or a Role. This is only one example of a general capability of JARGON to serve as an editing language for KL-ONE structures as well as a language in which such knowledge structures can be expressed. The JARGON syntax permits an English-like sentence or noun phrase of the forms discussed above to be prefixed by any of a number of imperative operators such as CREATE, ADD, FIND, SHOW, KILL, and DELETE. The default interpretation of an ordinary sentence or noun phrase without such a prefix operator (as in all of the above examples) is ADD, which causes new structure to be created and added to the network (CREATE is effectively a synonym for ADD, although a later implementation may introduce a difference). Conceptual structures introduced with the ADD operator will be constructed by the addition of any necessary new structure (e.g., a new Role, a new value restriction, or a new superconcept link) but will make use of whatever structure already exists. If the Concept described already exists, then no new structure is created.

The FIND operator will find a Concept from its description if it exists, but will return NIL otherwise. The operator SHOW will print out the local structure of a Concept in a prettyprint format which can be used for browsing the network. Finally, KILL when

applied to a concept or a Role description will delete that Concept or role from the network. KILL, when applied to an assertion such as A MAN IS A PERSON will make the assertion false by removing its main connective (in this case removing the superconcept connection between MAN and PERSON). (DELETE is effectively a synonym for KILL, although a difference may be introduced in a later implementation.) Note that KILL does not fully invert the effect of ADD in such cases since it may leave structure that was constructed when the assertion was originally made. KILL when applied to expressions of Role ownership will delete the resulting detached Role. The effect of a KILL on an expression of satisfaction (i.e., a clause whose verb is "satisfy") will delete the structural condition corresponding to that expression, but has not yet been implemented.

2.1.7 Structural Conditions

In order to describe the structural conditions of Concepts and to handle certain meta statements about KL-ONE Concepts, the JARGON syntax permits the use of the verb "satisfy" in constructions such as:

A PERSON'S INCOME (I) AND HIS INCOME-TAX (T) SATISFY A
LESS-THAN WHOSE LEFTARG IS (T) AND WHOSE RIGHTARG IS (I)

JARGON also permits the use of tag variables indicated in parentheses as above in order to make the linkages between the coreferential Roles of the structural condition and the Roles of the Concept which is being modified. In the above case, a ParaIndividual LESS-THAN will be attached as a structural condition to PERSON with its LEFTARG and RIGHTARG Roles linked to the INCOME-TAX and INCOME Roles of PERSON via Coref Roles. Other variations of the verb "satisfy" which one may want to use and are therefore provided in the JARGON syntax are:

A PERSON'S MOTHER SATISFIES MOTHER-PREDICATE

JOHN SATISFIES FATHER-PREDICATE

At the current time, the jargon interpreter has not fully implemented the semantic interpretation of satisfaction statements into KL-ONE structures. Such implementation will await further stabilization of the KL-ONE structures for structural conditions.

2.1.8 Quantification

One of the most significant characteristics of the JARGON syntax, but another feature that has not yet been fully implemented in the interpreter, is the specification of quantification in structural Conditions attached to concepts. This syntax is well suited to a kind of abstract quantificational structure which is

being developed in KL-ONE, but which has not yet been fully codified and implemented. The permitted syntax for such quantification can be illustrated by the following schematic example:

A CONCEPT1'S ROLE1 (X) ITS ROLE2 (Y) AND ITS ROLE3 (Z) SATISFY PL CONCEPT2 WHOSE ROLE21 IS (Z) WHOSE ROLE22 IS (X) AND WHOSE ROLE23 IS (Y) FOR EACH (X) AND FOR SOME (Z) WHERE FOR EACH (X) THERE ARE EXACTLY 3 (Y) AND VICE VERSA AND FOR EACH (Z) THERE IS AN (X).

In this example, the coreferentiality of Roles between the Concept being modified and the ParaIndividual CONCEPT2 is specified by the syntax discussed above for the verb SATISFY. The phrases FOR EACH (X) AND FOR SOME (Z) assert that there is one such instance of CONCEPT2 for each value of CONCEPT1'S (X) Role and there is such an instance for some value of its (Z) Role. The phrases following the WHERE indicate the dependencies (if any) between the different variables in the quantification. Notice that the dependencies specified here could not normally be expressed in ordinary predicate calculus notations without multiple copies of the matrix predicate being quantified.

2.1.9 Anaphoric Reference

JARGON does not make the assumption that one will specify a single KL-ONE Concept with a single sentence of JARGON. On the contrary, it assumes that a given KL-ONE Concept and its

relationship to other Concepts will be built up by a sequence of JARGON sentences, each specifying a reasonable sized "chunk" of information about the Concept being defined. To this end, it provides for a very simple and stylized anaphoric reference facility (but one that approximates ordinary English usage in a natural way) to make convenient references to a previously introduced concept. Thus, JARGON permits sequences such as:

A MAN IS A PERSON

HE HAS A NAME AND PL HOBBY

HIS PL HOBBY ARE PL ENTERTAINMENT-ACTIVITY

HE ALSO HAS A SOCIAL-SECURITY-NUMBER

The rules for pronominal anaphora are very simple. There is one single pronominal focus register that at every moment contains the only thing that is available for reference by a pronoun. Because of the lexical simplifications of gender and number in JARGON, pronominal references are not differentiated according to whether the pronoun is masculine, feminine, neuter, singular, or plural. Hence the JARGON user is free to use the pronouns "he", "she", "it", "they" as is appropriate, but the system will in any case take the expression to refer to the single focused entity. The pronominal focus register is set by the subject of each sentence. When each sentence is processed, as soon as its subject is determined, the "root" of the subject becomes the focused object

for the remainder of that sentence and for the initial portion of the subsequent sentence. (The root of the noun phrase in this case refers to the root of the Role chain if the noun phrase is a Role chain or to the Concept referred to by the noun phrase otherwise.)

In addition to the register for pronominal focus, there is another focus register which contains a list of possible referents for definite noun phrase anaphora (in JARGON, definite noun phrase anaphora is indicated by noun phrases introduced by determiners "this", "that", "these", and "those"). This register contains a list of all the noun phrases referred to since the subject (the contents of the pronominal focus register) last changed. Any of these can be referred to by noun phrases whose head noun is the same as that of the noun phrase to be referred to and which is introduced by one of the above determiners (again there is no inflectional agreement test between the antecedent and the anaphoric determiner). If there are several possible antecedents that satisfy these conditions, then JARGON prints out a message to that effect and the JARGON user can rephrase his request using an "ordinal determiner" to select the one he intends. (In JARGON, the ordinal determiner is indicated by the prefix ORD or ORDINAL preceding a number rather than by an inflected number such as "first", "second", "third", etc.) Thus, the following noun phrases are possible in JARGON:

A PERSON WHO SATISFIES A LIKE WHOSE AGENT IS THAT PERSON

A PERSON WHO SATISFIES A LIKE WHOSE AGENT IS A PERSON AND
WHOSE PATIENT IS THAT ORD 1 PERSON

2.2 Uses of JARGON

The JARGON language has been used to explore the process of building up and evolving knowledge structures, and has given us several insights into the nature of knowledge representation and acquisition. For example, it has become apparent that one does not proceed to build an adequate knowledge structure for a knowledge domain in a straightforward systematic fashion. Rather one initially makes over-generalizations, fails to draw distinctions, associates attributes with more specific concepts and later discovers that they can be generalized to more general concepts, decides that the name originally given to a particular concept is more appropriately used for a different concept and that the original Concept should be renamed, etc. Features have been incorporated into the JARGON language to facilitate these kinds of changes as the evolution of a knowledge structure proceeds. JARGON thus can be viewed as a combination of a knowledge input language and a knowledge structure editor.

JARGON has served as a useful adjunct to the development of the KL-ONE knowledge representation system and the understanding of

its structures in several interesting ways. One important contribution is the provision of a lexical language for expressing KL-ONE structures. Unlike many knowledge representation languages that have been worked on elsewhere, KL-ONE has been oriented toward the abstract underlying connectivity and structure of the knowledge network; and not with an ASCII text representation. Most of the work that has been done in developing KL-ONE and in describing its knowledge structures has been done in a graphical language. This graphical language is excellent for illustrating the underlying connectivity of the KL-ONE structures and as a pedagogical tool for presenting the kinds and meanings of KL-ONE constructs. We believe that this is an important reason for some of the insights that have developed there. However, it is extremely difficult to read these graphical representations for non-trivial knowledge networks because of the proliferating tangle of connections.

A utility program for dumping KL-ONE networks onto an ASCII file and reading them back in has provided one level of lexical representation which is relatively readable. However, this notation has limitations as a convenient input language in which to build up a knowledge network. JARGON provides a lexical notation that is more suitable for input and which moreover relates KL-ONE structures to corresponding constructions in English. This latter aspect of JARGON is a potentially important one that has not yet

been fully exploited. That is, the JARGON parser and interpreter provides a formalization of a correspondence between ordinary English constructions and the way that the meanings of such constructions get represented in KL-ONE. Thus, JARGON can serve as a repository and an expository device for the codification of how KL-ONE represents certain kinds of facts. Moreover, it has already begun to have an effect in revealing phenomena for which KL-ONE does not have a convenient representation and for indicating areas in which KL-ONE structures may be generalized.

The extent to which the English meanings of the rather subtle words "be" and "have" have a natural correspondence to the inheritance and Role relationships in KL-ONE seems to give evidence that the basic structure of KL-ONE is correctly capturing a large number of generalizations of natural conceptual structure. Conversely, when one finds things that are convenient and natural to say in English but are awkward to express in KL-ONE, then this can be evidence for something missing in KL-ONE. Similarly, when one finds that English constructions which are naturally expressible in KL-ONE have a natural generalization in English which does not nicely generalize in its KL-ONE representation, then this may be evidence for a missing generalization in KL-ONE.

An example of a phenomenon in which JARGON may influence the development of KL-ONE is the translation into KL-ONE of sentences

such as A PERSON'S PL CHILD HAVE PL BIRTHDAY, where an apparent 3 is made about a Role chain having a Role. Currently in KL-ONE one can use Role chains in certain kinds of referential ways and can make assertions regarding the referential equivalence of two Role chains, but one cannot directly make assertions about Role chains having Roles. Only Concepts can have roles. The translation of assertions of Role chains having Roles is currently handled in JARGON by creating a Concept to represent the filler at the end of the Role chain and making the appropriate assertion about that Concept. However, the naturalness of the English assertions about Role chains suggests that there may be an equivalent generalization that could be useful in KL-ONE.

3. CONCLUSIONS

We have given a brief overview of the ONR project in natural language understanding at BBN, and a detailed presentation of a language, JARGON, which was developed under that project. JARGON is a formalized and somewhat stylized subset of natural English designed to serve as a lexical language for a KL-ONE structured inheritance network and to permit experimental investigation of the growth and evolution of a complex knowledge base. The JARGON language has been used to explore the process of building up and evolving knowledge structures, and has given us several insights into the nature of knowledge representation and acquisition.

For example, it has become apparent that one does not proceed to build an adequate knowledge structure for a knowledge domain in a straightforward systematic fashion. Instead, one gradually evolves a knowledge structure that improves as experience increases, but which is probably never completed. The nature of natural conceptions of the world appear to be such that they are never fully developed -- increased learning is always possible. Consequently it is important to have a representational system that is capable of such incremental improvement without massive reorganization. Our experience in developing knowledge structures in KL-ONE via JARGON leads us to believe that KL-ONE networks have the capability to support such incremental knowledge acquisition.

JARGON provides a lexical notation that relates KL-ONE structures to corresponding constructions in English. This aspect of JARGON is a potentially important one that has not yet been fully exploited. That is, the JARGON parser and interpreter provide a formalization of a correspondence between ordinary English constructions and the way that the meanings of such constructions are represented in KL-ONE. This can potentially serve as a repository and an expository device for the codification of how KL-ONE represents certain kinds of facts. The naturalness of the correspondence between such constructions in English and the structures in KL-ONE gives some evidence that the underlying

structures in KL-ONE capture significant aspects of the structure of human conceptualizations. Moreover, the interaction between JARGON and KL-ONE has begun to have an effect in revealing phenomena for which KL-ONE does not have a convenient representation and for indicating areas in which KL-ONE structures may be generalized. The JARGON language holds promise for pushing the development of KL-ONE in interesting ways and for increasing our understanding of the ways knowledge bases grow and evolve.

4. REFERENCES

Brachman, R.J. (1977)
A Structural Paradigm for Representing Knowledge, Ph.D. Dissertation, Harvard University, Cambridge, Ma., May, 1977 Also BBN Report No. 3605, Bolt Beranek and Newman Inc., Cambridge, Ma., May 1978).

Brachman, R. J. (1978)
Theoretical Studies in Natural Language Understanding: Annual Report, 1 May 1977 - 30 April 1978, BBN Report No. 3888, Bolt Beranek and Newman Inc., Cambridge, Ma.)

Woods, W. A. (1975)
What's in a Link: Foundations for Semantic Networks, in D. Bobrow and A. Collins (eds.), Representation and Understanding: Studies in Cognitive Science, New York: Academic Press.

Woods, W. A. (1978)
Taxonomic Lattice Structures for Situation Recognition," in TINLAP-2, Conference on Theoretical Issues in Natural Language Processing-2, University of Illinois at Urbana-Champaign, July 25-27, 1978 (also in AJCL, Mf 78, 1978:3).

Official Distribution List
Contract N00014-77-C-0371

	<u>Copies</u>
Defense Documentation Center Cameron Station Alexandria, VA 22314	12
Office of Naval Research Information Systems Program Code 437 Arlington, VA 22217	2
Office of Naval Research Code 200 Arlington, VA 22217	1
Office of Naval Research Code 455 Arlington, VA 22217	1
Office of Naval Research Code 458 Arlington, VA 22217	1
Office of Naval Research Branch Office, Boston 495 Summer Street Boston, MA 02210	1
Office of Naval Research Branch Office, Chicago 536 South Clark Street Chicago, IL 60605	1
Office of Naval Research Branch Office, Pasadena 1030 East Green Street Pasadena, CA 91106	1
Office of Naval Research New York Area Office 715 Broadway - 5th Floor New York, NY 10003	1

Naval Research Laboratory Technical Information Division Code 2627 Washington, D.C. 20380	6
Naval Ocean Systems Center Advanced Software Technology Division Code 5200 San Diego, CA 92152	1
Dr. A.L. Slafkosky Scientific Advisor Commandant of the Marine Corps (Code RD-1) Washington, D.C. 20380	1
Mr. E.H. Gleissner Naval Ship Research & Development Center. Computation & Mathematics Dept. Bethesda, MD 20084	1
Captain Grace M. Hopper NAICOM/MIS Planning Branch (OP-916D) Office of Chief of Naval Operations Washington, D.C. 20350	1
Mr. Kin B. Thompson NAVDAC 33 Washington Navy Yard Washington, D.C. 20374	1
Advanced Research Projects Agency Information Processing Techniques 1400 Wilson Boulevard Arlington, VA 22209	1