NRL Report 8893

Automatic Processing of Navy Message Narrative

ELAINE MARSH, JUDITH FROSCHER, RALPH GRISHMAN,
HENRY HAMBURGER, AND JOAN BACHENKO

Navy Center for Applied Research in Artificial Intelligence
and
Computer Science and Systems Branch
Information Technology Division

August 21, 1985

NAVAL RESEARCH LABORATORY
Washington, D.C.

REPORT DOCUMENTATION PAGE

1a. REPORT SECURITY CLASSIFICATION UNCLASSIFIED	1b. RESTRICTIVE MARKINGS
2a. SECURITY CLASSIFICATION AUTHORITY	3. DISTRIBUTION / AVAILABILITY OF REPORT Approved for public release; distribution unlimited.
2b. DECLASSIFICATION / DOWNGRADING SCHEDULE	
4. PERFORMING ORGANIZATION REPORT NUMBER(S) NRL Report 8893	5. MONITORING ORGANIZATION REPORT NUMBER(S)
6a. NAME OF PERFORMING ORGANIZATION Naval Research Laboratory 6b. OFFICE SYMBOL (If applicable) 7510	7a. NAME OF MONITORING ORGANIZATION Naval Research Laboratory
6c. ADDRESS (City, State, and ZIP Code) Washington, DC 20375-5000	7b. ADDRESS (City, State, and ZIP Code) Washington, DC 20375-5000
8a. NAME OF FUNDING / SPONSORING ORGANIZATION 8b. OFFICE SYMBOL (If applicable)	9. PROCUREMENT INSTRUMENT IDENTIFICATION NUMBER
8c. ADDRESS (City, State, and ZIP Code)	10. SOURCE OF FUNDING NUMBERS

PROGRAM ELEMENT NO.	PROJECT NO.	TASK NO.	WORK UNIT ACCESSION NO.
62721N		RF21-241-805	DN180-313

11. TITLE (Include Security Classification)
Automatic Processing of Navy Message Narrative

12. PERSONAL AUTHOR(S)
Marsh, Elaine; Froscher, Judith; Grishman,* Ralph; Hamburger,** Henry; and Bachenko, Joan

13a. TYPE OF REPORT	13b. TIME COVERED	14. DATE OF REPORT (Year, Month, Day)	15. PAGE COUNT
Interim	FROM ________ TO ________	1985 August 21	24

16. SUPPLEMENTARY NOTATION
* Also of Computer Science Department, New York University
** Also of Computer Science Department, George Mason University

17. COSATI CODES		
FIELD	GROUP	SUB-GROUP

18. SUBJECT TERMS (Continue on reverse if necessary and identify by block number)
Natural language understanding; Navy messages; Knowledge representation; Production rule systems

19. ABSTRACT (Continue on reverse if necessary and identify by block number)

This report describes an experimental system that employs techniques of computational linguistics and artificial intelligence to extract information from Navy messages. The long-term goal of this work is to develop capabilities that will enable systems to handle a broad spectrum of messages, from highly formatted messages with little English description to messages consisting entirely of English narrative.

In this report, an experimental system that takes the first step towards automated understanding of Navy messages was constructed and implemented. The system extracts informational content from reports about shipboard equipment failure and uses the content to assign a distribution list to each message and to generate a summary of the equipment failure. The system can be extended to other Navy

(Continues)

20. DISTRIBUTION / AVAILABILITY OF ABSTRACT ☒ UNCLASSIFIED/UNLIMITED ☐ SAME AS RPT. ☐ DTIC USERS	21. ABSTRACT SECURITY CLASSIFICATION UNCLASSIFIED	
22a. NAME OF RESPONSIBLE INDIVIDUAL Elaine Marsh	22b. TELEPHONE (Include Area Code) (202) 767-2382	22c. OFFICE SYMBOL 7510-011:EM:gth

DD FORM 1473, 84 MAR — 83 APR edition may be used until exhausted. All other editions are obsolete.

19. ABSTRACT (Continued)

message types and used for other military applications. For extracting message content, an approach developed by Sager et al. at New York University was adapted. This approach, called "information formatting," uses an explicit grammar of English and a classification of the semantic relationships within the domain to derive a tabular representation of the information in a message narrative. Prototype knowledge bases for distribution and summarization were implemented in **OPS5** and use the information format to automatically construct a distribution list and summary. A comparison of computer-generated summaries with those obtained manually showed good agreement, indicating that it is possible to automatically process message narrative and generate appropriate, and ultimately useful, summaries.

CONTENTS

AUTOMATIC PROCESSING OF NAVY MESSAGE NARRATIVE

1. INTRODUCTION

The Navy's future message systems will perform many tasks, e.g., message dissemination and retrieval, that require computer interpretation of message content. At the Naval Research Laboratory, we have built an experimental system that employs techniques of computational linguistics and artificial intelligence to extract information from Navy messages. The long-term goal of this work is to develop capabilities that will enable systems to handle a broad spectrum of military messages from highly formatted messages with little English description to messages consisting entirely of English narrative. For the initial investigation, we have limited the problem to Navy operational reports; these messages obey strict formatting conventions but also contain important narrative descriptions. However, the message processing system can be extended to other message types with minimal modifications.

The system we describe extracts content from messages about shipboard equipment failure; these messages form a class of operational reports called "CASREPs" (CASualty REPorts). The system uses message content to assign a distribution list to each message and to generate a summary of the equipment failure [1]. In constructing the system, we have adapted an approach developed by Sager et al. at New York University (NYU). This approach, called "information formatting," uses computational linguistic techniques to construct a tabular representation of the information in a message narrative. One advantage of information formatting is that the same techniques can be used on messages from different domains.

In Section 2, we describe how CASREPs are used and the kinds of information they contain. Sections 3 and 4 discuss the information formatting approach and its adaptation to the CASREP data. Section 5 describes the dissemination and summary applications. Section 6 evaluates the performance of the summarization program by comparing computer-generated summaries to those obtained by manual summarization procedures. Section 7 concludes this report with a discussion of future research issues.

2. THE CASREP CLASS OF MESSAGES

In planning the system, we first observed characteristics of military messages, e.g., degree of formatting, that would facilitate their analysis in the short term [2]. CASREPs were chosen as the initial focus because their content and form are restricted, but their text still contains English narrative. These restrictions on content make the task of narrative analysis tractable. In addition, CASREPs are an important message type, providing, among other things, current information about ship readiness and equipment performance [3].

A CASREP is sent whenever an equipment malfunction cannot be repaired within 48 hours. Its purpose is to provide explicit information about the equipment that failed and the Navy unit, usually a ship, that filed the report. CASREPs inform operational and support personnel about equipment casualties that could affect a unit's ability to perform in a mission area. They also report the unit's need for technical assistance and for parts to correct the malfunction.

Manuscript approved February 19, 1985.

These reports are filed in a series. The INITIAL CASREP is sent within 24 hours of the equipment malfunction. UPDATE reports describe the current status of the problem and can note changes to the INITIAL report. CORRECT is submitted when the failed equipment is repaired or replaced. A unit files a CANCEL message if the problem can be corrected during the ship's overhaul. Together these report types constitute the CASREP class of messages; we will use the term CASREP to refer to any or all members of this class.

As in other operational reports, the text of a CASREP is formatted and consists of a sequence of data sets. Among other things, these data sets specify an equipment identification, an estimated time of repair, and an itinerary, all in accordance with format conventions. An example is the CASREP in Fig. 1.

The first data set in the text of Fig. 1 is labeled MSGID; this identifies the message as belonging to the CASREP class and the sender as the USS XXXXXXXX. The second data set labels the message as an INITIAL CASREP, identifies the equipment that failed (a high-frequency transmitter), and rates the effect of the failure as CAT(egory) 2, i.e., substantially combat ready, with only minor deficiencies. The ESTIMATE data set gives the expected time when the repair will be completed; for this CASREP, the value is 25 Aug 1982, at 11:59 p.m. The ASSIST data set, with the value "TECHNICAL," states that the ship will need outside assistance to correct the problem. The ASSIST data set is augmented by the AMPN (amplification) line following it. The RMKS section, which is optional and found at the end of the message when present, describes the equipment malfunction and its cause.

To process such messages, the system provides a representation of message content that can be readily accessed and used by application systems. This is accomplished by a message interpreter that comprises two interacting components: message decomposition, which determines the overall structure of a message, and narrative analysis, which generates the structures that enable the computer system to interpret the English narrative. For messages like CASREPs, message decomposition is straightforward because it can use the formatting conventions to extract the *pro forma* (strictly formatted) information and pass the information on to the application systems.* However, information extracted just from *pro forma* data sets is not sufficient for some applications. Information from the narrative portions of the text is also required, although narrative analysis is not straightforward. In the section that follows, we will describe a linguistically motivated approach to the analysis of CASREP narrative.

3. DEFINING THE INFORMATION STRUCTURE

The central task of narrative analysis is the extraction and representation of particular types of information contained in the narrative portions of a message. This task is difficult because the structure of the information, and often much of the information itself, is *implicit* in the narrative. An example is the ambiguity of the word *IF* in sentences (1a) and (1b) below. In (1a), *IF* is a noun that stands for "intermediate frequency"; in (1b), *IF* is a subordinating conjunction that introduces a clause:

(1) a. APC-PPC VOLTAGES TO T-827 IF STAGE ARE IN EXCESS OF 10 VOLTS.

b. PRINTER RUNS OPEN IF TTY IS PATCHED WITH KG-14.

In such cases, the particular meaning of the word is determined by the structure of the sentence. For example, in (1a), *IF* can only be a noun, and not a subordinating conjunction, as it is in (1b). If *IF* were a subordinating conjuction, then the head noun of the subject, *STAGE*, and the verb, *ARE*, in the string *IF STAGE ARE IN EXCESS OF 10 VOLTS* should agree in number, and they do not. *STAGE* is singular, and *ARE* is plural. This lack of agreement makes it impossible to analyze the string as a subordinate

*While both the *pro forma* and the narrative data sets may contain typographical errors, ungrammatical forms, and other types of ill-formed input, we have not been concerned with them here, besides characterizing and adding to the grammar those "ill-formed" constructions that are consistently used in the messages. Such consistent forms include different types of sentence fragments and Navy-specific constructions, such as date-time groups.

```
P 162305Z AUG 82
FM  USS XXXXXXXX
TO  RUCLBDA/COMINEGRU TWO
    RUCBSAA/COMNAVSURFLANT NORFOLK VA
    RUCBSAA/CINCLANTFLT NORFOLK VA
    RUENAAA/CNO WASHINGTON DC
    RUEOALA/NAVSAFECEN NORFOLK VA
    RULSSAA/COMNAVSEASYSCOM WASHINGTON DC
    RULSSAA/CHNAVMAT WASHINGTON DC
    RUEBBSA/NSC NORFOLK VA
    RUEDNAA/SPCC MECHANICSBURG PA
    RUCLFEA/MOTU TWELVE
INFO RULSSAA/COMNAVELEXSYSCOM WASHINGTON DC
BT
MSGID/CASREP/MSO### XXXXXXX/4//
POSIT/8204W4-2443N3/161500ZAUG82//
CASUALTY/INITIAL-82004/AN-URT-23V HF TRANSMITTER/EIC:QEIN/CAT:2//
ESTIMATE/252359ZAUG82/RECEIPT OF PARTS NLT 24AUG82//
ASSIST/TECHNICAL/PORT EVERGLADES FL//
AMPN/REQUEST ASSISTANCE FROM MOTU TWELVE MAYPORT//
RMKS/SHIPS SCHEDULE: 16AUG-19AUG OPEVAL KEYWEST OPAREA. 20AUG-23SEP
OPEVAL FT. LAUDERDALE FL//
RMKS/SHIP WILL BE IN PORT EVERGLADES IN THE EVENINGS AND ON
WEEKENDS UNTIL 23 SEP.//
PARTSID/APL:58557823CL/CID:1A1A3/JCN:N07973-OE06-7545//
TECHPU/NAVELEX 0967-LP-879-50X10//
1PARTS
/DL NATIONAL STOCK NO.   RQD   COSAL   ONBD  CIRCUIT
/8$&1H5820-00-988-8033   001    000     000   1A1A3
/02 1H5820-00-988-3043   001    000     000   1A1A6//
AMPN/REASON ITEM NOT ONBOARD - NO ALLOWANCE.  ALL PARTS LISTED
IN PARTSID APL//
1STRIP
/DL DOCUMENT ID        QTY  PRI  RDD  ACTIVITY  REQUISITION STATUS
/01 V07973-2228-W542   001   06  236      NNZ  162300Z AUG 82
/02 V07973-2228-W543   001   06  236      NNZ  162300Z AUG 82
RMKS/APC-PPC CIRCUIT IS INHIBITING EXCITER AND PA DRIVER IN ALL
OPERATE MODES.  RADIO SET WILL TUNE USING TUNE KEY, LOCAL KEY
AND REMOTE KEY.  DRIVER AND PA CURRENTS GOOD DURING TUNING.
IN OPERATE MODES DRIVER CURRENT AND RF POWER OUT ARE ZERO,
AS IS INPUT TO PA.  APC-PPC VOLTAGES TO T-827 IF STAGE ARE
IN EXCESS OF 10 VOLTS.  PPC IS NOT ADJUSTABLE.  APC CAN BE
ADJUSTED TO 8 VOLTS MIN. WHICH ALLOWS EXCITER TO OVERDRIVE
IN TUNE.  SYSTEM KEYLINE APPEARS GOOD IN THAT ALL ESSENTIAL
RELAYS SWITCH WHEN KEYED AND COUPLER CONTROLLER STANDBY
LIGHT GOES OUT.  PA CURRENT OK WHEN SYSTEM IS KEYED IN
OPERATE MODE.//
DCLAS/DECL 30NOV82//
BT
#0675
```

Fig. 1 — A sample CASREP

clause. However, in (1b), it is possible to analyze *IF* as a subordinating conjuction, because the subject and verb of the clause agree in number: the subject, *TTY*, is singular, and the verb, *IS*, is also singular.

The aim of narrative analysis is to make explicit the structure and content of expressions such as those in (1a) and (1b). Several formalisms, such as scripts and frames, have been developed to describe such information and have been used in text analysis [4,5]. We are using the approach called "information formatting," which was first described by Sager in 1972 and has since been developed at the New York University Linguistic String Project [6-8]. In simplest terms, an information format is a large table, with one column for each type of information that can occur in a class of texts and one row for each sentence or clause in the text. We will return to a discussion of how the text is mapped onto an information format in the next section. Our concern in this section is to give a general view of CASREP information formats.

Texts in a restricted domain discuss a limited number of classes of objects and express a limited number of relationships among these objects [9,10]. For example, the objects in CASREPs about electronic equipment include the equipment items and their component parts, the signals and data operated on by the equipment, the people and organizations who operate and maintain the equipment, and the documents involved in the maintenance process. By identifying these classes of objects and relationships, we can develop data structures suitable for storing the information derived from the message narrative. The various classes of objects and relationships have their own "slots" in this data structure, so that information can be much more readily retrieved than from the original narrative.

Table 1 shows the information format we have developed for CASREPs, listing the format slots and their significance. This is a preliminary format structure; we are continuing to enlarge and refine the structure as we study additional CASREPs. To see how this format would be used, consider the following text, taken from the assist amplification and the final remarks portion of the CASREP in Fig. 1:

> *Request assistance from MOTU twelve Mayport.*
> *APC-PCC circuit is inhibiting exciter and PA driver in all operate modes. Radio set will tune using tune key, local key and remote key. Driver and PA currents good during tuning. In operate modes driver current and RF power out are zero, as is input to PA. APC-PPC voltages to T-827 IF stage are in excess of 10 volts. PPC is not adjustable. APC can be adjusted to 8 volts min. which allows exciter to overdrive in tune. System keyline appears good in that all essential relays switch when keyed and coupler controller standby light goes out. PA current OK when system is keyed in operate mode.*

Table 2 indicates how this text is transformed into a series of format entries. To simplify this table, we have included columns for only those format slots required by this message. We have also suppressed substructure within each column so that modifier-host relationships are not noted.

Information in the narrative can be extracted more readily from this table than from the original text because the information has been made explicit in the table. Those words in the text for which there are format columns are mapped directly into the format table, along with their modifiers. For example, the ORGanization whose assistance has been requested, *MOTU twelve* (a mobile technical unit), is mapped into the column ORG along with its location modifier, *Mayport.* Parts, the condition of which is being reported in the narrative, such as *APC-PPC circuit* and *PA DRIVER*, are mapped into the PART column. Actions on parts and by parts are mapped in the same way into their appropriate format columns, REPAIR and PROCESS, respectively. Those words for which there is no format column (for example, the verb *be* in the sentence *Driver current and RF power out are zero* and the sentential modifier *in operate mode* in the sentence *PA current is OK when system is keyed in operate mode*) are not formatted. *Be* has not been assigned a relevant semantic category: the word is not informationally important, and, as a result, there is no format column for it. *In operate mode* modifies an entire sentence and not a particular host that has its own format column. Therefore, it is not formatted.

Table 1 — CASREP Semantic Categories and Modifiers of Categories

CASREP Semantic Categories

ADMIN	action or request for part, *forward, report, expedite*
FUNCTION	function performed by equipment, *broadcast, communication, operate*
INVEST	investigative act, *check, isolate, troubleshooting*
MSG	message concerning part failure or request for part, *CASREP, message*
ORG	personnel or organization, *MOTU, ship, technical, originator, technician*
PART	equipment, subsystem, or part, *antenna, AN/URC-9, controller*
PROCESS	process performed by equipment on electrical signal, *encrypt, decrypt, cycle, deflection*
PROCURE	action to request, ship, receive, or hold part, *deliver, purchase, reorder*
PROPERTY	property of part, *allowance, clearance, sync, weight*
REPAIR	repair action, *repair, adjust, overhaul*
SIGNAL	electrical signal, *current, band, voltage, UHF*
STASK	ship's task, *arrival, assignment, transit*
STATUS	equipment status, *casualty, fault, malfunction, good*

Modifiers

TIME, LOC [location], QUANT [quantity, amount, value of property], EFFORT [__man-hours], NEG [negation], MODAL [modality]

Table 2 — Information Format Table for Sample CASREP

No.	CONN	ASSIST	FUNC	ORG	PART	PROCESS	REPAIR	SIGNAL	STATUS
1.	REQUEST								
2.		ASSISTANCE		MOTU 12 MAYPORT					
3.					APC-PPC CIRCUIT				
4.	INHIBIT								
5.					EXCITER DRIVER IN ALL OPERATE MODE PLURAL				
6.	AND								
7.					APC-PPC CIRCUIT				
8.	INHIBIT								
9.					PA DRIVER IN ALL OPERATE MODE PLURAL				
10.					RADIO SET	TUNE			
11.	WHILE								
12.			USE		TUNE KEY				
13.	','								
14.			USE		LOCAL KEY				
15.	AND								
16.			USE		REMOTE KEY				
17.								DRIVER CURRENT PLURAL	GOOD DURING TUNING
18.	AND								
19.								PA CURRENT PLURAL	GOOD DURING TUNING
20.								DRIVER CURRENT	ZERO UNITS
21.	AS								
22.								INPUT TO PA	ZERO UNITS
23.	AND								
24.								RF POWER_OUT	ZERO UNITS
25.	AS								
26.								INPUT TO PA	ZERO UNITS

Table 2 (Continued) — Information Format Table for Sample CASREP

No.	CONN	ASSIST	FUNC	ORG	PART	PROCESS	REPAIR	SIGNAL	STATUS
27.								APC-PPC VOLTAGE PLURAL TO T-827 IF STAGE	IN EXCESS OF 10 VOLTS
28.					PPC				NOT ADJUSTABLE
29.					APC		ADJUST		
30.	TRNWH								
31.					APC		ADJUST		
32.	ALLOW								
33.					EXCITER	OVERDRIVE			
34.					SYSTEM KEYLINE				APPEAR GOOD
35.	IN_THAT								
36.					ALL ESSENTIAL RELAY PLURAL	SWITCH			
37.	WHEN								
38.					ALL ESSENTIAL RELAY PLURAL	KEY			
39.	AND								
40.					COUPLER CONTROLLER STANDBY_LIGHT				GO OUT
41.								PA CURRENT	OK
42.	WHEN								
43.					SYSTEM	KEY			

4. INFORMATION FORMATTING

The narrative portion of each message is automatically transformed into a series of format entries using a procedure modeled on that developed at New York University (NYU) for the formatting of medical narratives [7,8]. The procedure we are using involves three stages of processing: parsing, syntactic regularization, and mapping into the information format.

First, the text sentences are parsed using the broad-coverage Linguistic String Project English grammar [11], extended to handle the sentence fragments and special sublanguage constructions (e.g., date expressions *NLT 292300 Z SEP 82*) that appear in these messages. We have found that most of the fragment structures in CASREPs are the same as those encountered in the medical reports previously processed at NYU [12,13], so relatively little change to the grammar has been required.* The parsing procedure performs several functions. The parsing procedure disambiguates cases of lexical ambiguity, where a spelling of a word can have one meaning in one instance and a different meaning in another instance. For example, in Section 3, we discussed how *IF* as a noun (i.e., the abbreviation for *intermediate frequency*) and *IF* as a subordinating conjunction are distinguished by the parser. The parsing process also determines sentence structure. It identifies phrase and and clause boundaries. In the parse tree of the example sentence illustrated in Fig. 2, the subject of the sentence is identified as a noun phrase (NSTG). In the last sentence adjunct (SA), a subordinate clause (CSSTG) is identified as containing a subordinating conjunction (CS8), *as* and an inverted subject and verb construction (Q-INVERT). Modifier-host relations (which words modify what other words) are also identified. For example, in Fig. 2, the noun *current* is analyzed as having a left modifier *driver* that is a noun. In the first sentence adjunct (SA), the object of the preposition (NSTGO) of the prepositional phrase (PN) consists of a noun *modes* with a verb modifier *OPERATE.* Finally, the parsing process identifies the scope of conjunctions. In Fig. 2, the conjunction *and* conjoins two Left modifier of Noun (LN) + Noun (NVAR) constructions.

In the second stage, the parse trees are syntactically regularized by a series of transformations to simplify the subsequent mapping into the information format. The various types of clauses (e.g., passives, sentence fragments, inverted sentences, existentials, relatives, and reduced relatives) are transformed into simple active assertions. For example, passive assertions are transformed into active assertions. Thus the passive *PA current can be adjusted to 8 volts min.* is transformed into its active counterpart *Someone can adjust to 8 volts min. PA current.* Some elements missing from sentence fragments are filled in. For example, in the fragment *PA current OK when system is keyed in operate mode,* the verb 'be' is filled in, resulting in the completed assertion *PA current be OK when system is keyed in operate mode.* If a sentence does not have subject-verb-object word order, the subject-verb-object order is created. Figure 3 is the regularized version of the parse tree in Fig. 2. The inverted clause (Q-INVERT), *as is input to PA,* in the sentence *Driver current and RF power out are zero, as is input to PA* is transformed into a simple active assertion: *as input to PA is zero.* The order of the verb and subject have been reversed and the object of the sentence has been reconstructed. The syntactic regularization procedure also expands most conjoined structures into conjunctions of complete assertions. For our parsed sentence in Fig. 2, the conjoined modifier + noun constructions are expanded, so that two assertions are conjoined in Fig. 3. For example, *driver current and RF power out are zero* becomes *driver current be zero and RF power out be zero.* Figure 3 only shows the first assertion of the regularized conjunction. The position of the second assertion is indicated by **1** in the tree.

The third stage of processing moves the phrases in the syntactically regularized parse trees into the information format, as discussed in Section 3 above. This procedure involves two steps: (1) stripping off connectives and (2) mapping into the information format. Connective words are those that relate two clauses. They indicate causal, conjunctional, or time relations between the two clauses that they connect. In the formatting procedure, a connective word is extracted from its sentence and

*Had we chosen a semantic parser, we could not have ported a system so readily, since it would have required encoding a large amount of semantic information not previously available.

```
*QREPS  21. 1. 8
IN OPERATE MODES DRIVER CURRENT AND RF POWER OUT ARE ZERO, AS IS INPUT TO PA.


SENTENCE
|
CENTER------ENDMARK
|      |
ASSERTION  '.'
|
SA---SUBJECT---VERB---OBJECT-----SA
|    |         |      |          |
PN   NSTG      VVAR   OBJECTBE   CSSTG
|    |         |      |          |
|    LNR       TV     OBJBE      SUB8
|    |         |      |          |
|    |         ARE    NSTG       CS8---Q-INVERT
|    |                |          |     |
|    |                *1*        AS    VERB1---SUBJECT
|    |                                 |       |
|    LN-----NVAR---ANDSTG              LTVR    NSTG
|    |      |      |                   |       |
|    NPOS   N      'AND'---Q-CONJ      TV      LNR
|    |      |              |           |       |
|    NNN    CURRENT        LN-----NVAR IS      NVAR---RN
|    |                     |      |            |      |
|    N                     NPOS   N            N      RNP
|    |                     |      |            |      |
|    DRIVER                NNN    POWER---OUT  INPUT  PN
|                          |                          |
P----NSTGO                 N                          P----NSTGO
|    |                     |                          |    |
IN   NSTG                  RF                         TO   NSTG
     |                                                     |
     LNR                                                   LNR
     |                                                     |
     LN-----NVAR                                           NVAR
     |      |                                              |
     VPOS   N                                              N
     |      |                                              |
     V      MODES                                          PA
     |
     OPERATE
```

Fig. 2 — Parse tree

```
*QREPS  21. 1. 8
IN OPERATE MODES DRIVER CURRENT AND RF POWER OUT ARE ZERO, AS IS INPUT TO PA.


SENTENCE
!
CENTER-----------------------------------------ENDMARK
!                                              !
ASSERTION-----------------------------ANDSTG   '.'
!                                     !
SA---SUBJECT---VERB---OBJECT-----SA   'AND'---Q-CONJ
!    !         !      !          !             !
PN   NSTG      VVAR   OBJECTBE   CSSTG        *1*
!    !         !      !          !
!    LNR       V      OBJBE      SUB8
!    !         !      !          !
!    !         BE     NSTG       CS8---ASSERTION
!    !                !          !     !
!    LN-----NVAR      LNR        AS    SUBJECT---VERB---OBJECT
!    !      !         !                !         !      !
!    NPOS   N         LN-----COMMASTG  NSTG      LTVR   OBJECTBE
!    !      !         !      !         !         !      !
!    NNN    CURRENT   QPOS   '.'       LNR       V      OBJBE
!    !                !                !         !      !
!    N                LQR              NVAR---RN BE     NSTG
!    !                !                !      !         !
!    DRIVER           QVAR             N      RNP       LNR
!                     !                !      !         !
P----NSTGO            Q                INPUT  PN        LN-----COMMASTG
!    !                !                       !         !      !
IN   NSTG             ZERO                    P----NSTGOQPOS   '.'
     !                                        !    !    !
     LNR                                      TO   NSTG LQR
     !                                             !    !
     LN-----NVAR                                   LNR  QVAR
     !      !                                      !    !
     VPOS   N------N                               NVAR Q
     !      !      !                               !    !
     V      MODE   PLURAL                          N    ZERO
     !                                             !
     OPERATE                                       PA
```

Fig. 3 — Regularized parse tree

mapped into the CONNective column. In Table 2, the words in the CONNective column are read as connecting the row above to the row below. Our example sentence from Figs. 2 and 3 is shown in lines 20 to 26 of Table 2. The coordinating conjunction *and* and the subordinating conjunction *as* are mapped into the CONN column of the format table. The arguments of a connective are mapped into separate format rows, and their words are mapped into the appropriate format columns. The mapping process is controlled in large part by the sublanguage (semantic) word classes associated with each word (these classes, along with syntactic information about the word, are recorded in each word's dictionary entry). The order of the rows in the table is dependent on the order of the sentences in a text and the connective involved. For example, if *X is the cause of Y*, then the order would be *X cause Y*, with the information in *X* in one row, followed by a row containing the information in *Y*, these connected by the connective *cause*. On the other hand, if *X is due to Y*, the order of the rows containing the arguments is reversed so that *Y* is the first argument of the causal connective *due to*, and *X* is the second argument, i.e., essentially this is handled as *Y cause X*. The regularized version of our example sentence, as shown above in Fig. 3, is:

> *In operate modes driver current be zero as input to PA be zero and in operate modes RF power out be zero as input to PA be zero.*

We read its mapped version in format Table 2 in the order that follows: Line 20 (*Driver current zero*), Line 21 (*as*), Line 22 (*input to PA zero*), Line 23 (*and*), Line 24 (*RF power out zero*), Line 25 (*as*), Line 26 (*input to PA zero*). In our example sentence, the verb *be*, which is present in the parse trees, has not been formatted because it is not informationally important in the text and, thus, does not have a sublanguage semantic category; the sentence adjunct, *in operate modes*, has not been formatted either because it is not associated with any one lexical host. (cf. Section 3).

5. APPLICATIONS

We have implemented prototype knowledge bases for two application areas: dissemination and summary generation. In each area, our current system consists of a set of productions, implemented in a Lisp-based version of the OPS5 production system programming language.* Organizing the knowledge base as a production system makes modification easy and promotes user understanding. Productions operate on an initial data base of working memory elements that includes data from the *pro forma* set and the information formats.

Permanent domain knowledge resides in the initial choice of what fields are available in the format system devised for the domain. Additional domain knowledge and knowledge of the nature of the application are embodied in the production rules of the expert system. Some production rules reflect an understanding of the subject matter of the equipment failure reports, while others are based on general principles of summarization. The end use that will be made of the summaries is also a guiding factor in some of the productions. For example, to guide future equipment specification and procurement, one must know not only what went wrong and how often, but also why. Thus, causality is important to the summaries. Taken together, the productions are attentive to such matters as malfunction, causality, investigative action, uncertainty, and level of generality. Some of these will arise in the example provided below. We will describe the action of the production system as if it worked directly on the information format table in Table 2, although in reality it handles the corresponding working memory elements. Table 2, containing information from the narrative assist amplification and remarks data sets, is part of a more complete data base that incorporates information from both the *pro forma* and narrative portions of the text. It is a neutral representation of the information in the text and is not application-specific. As a result, any number of different applications can be performed on the same data base. The applications of concern here are dissemination and summary generation.

To formulate accurate rules for the dissemination system, we first conducted investigations of CASREP distribution and processing in several Navy organizations, focusing on NAVSURFLANT (Navy Surface Forces for the Atlantic) in Norfolk, VA. We found that in general, proper distribution of CASREPs depends on two types of message information: (1) the identity of the equipment that failed, e.g., propulsion systems, combat systems, etc., and (2) identifying data on the ship that filed the report (e.g., ship name and fleet). At NAVSURFLANT, dissemination is also influenced by whether or not the message asks for assistance and, if so, then what type of assistance and from whom. The assistance details, if they are given, usually appear in the amplification narrative following the ASSIST

*A prototype dissemination system for CASREPs about steam turbine systems and gun mounts has previously been implemented as a knowledge base [14] using the KES production system [15].

data set. For example, (4) is the amplification narrative from our sample message (Fig. 1); the amplification lines in (5) come from other messages in our *corpus.*

(4) Request assistance from MOTU twelve Mayport.

(5) a. Request COMSERFORSIXTHFLT assist in locating replacement antenna.
b. Request MOTUVO arrange tech assist.
c. Parts expediting assistance required.

The demonstration system automatically generates a distribution list within NAVSURFLANT for each input message by extracting information from the formatted and narrative portions of the message and then applying the production rules to this information. Fixed format material is sent directly to the data base; this includes ship name, equipment identification, CAT rating, and ASSIST data. If ASSIST is augmented by narrative, then the narrative is mapped into a format entry (cf. lines 1 and 2 of Table 2) and this too becomes part of the data base. Example (6) summarizes a set of OPS5 rules that determine how CASREPs are disseminated to the Combat Systems Group (N442) at NAVSURFLANT in Norfolk.

(6)	If		(1)	ship fleet	= Atlantic
		and	(2)	equipment	= EIC-Q,
		and	(3)	either ship category	= combatant
				or assist	= technical
	then			distribution	= N442.

Briefly, message (6) indicates that the Combat Systems Group N442 should receive all CASREPs that report problems with electronic equipment (i.e., have a Q equipment identification code) and that come from combatant ships in the Atlantic Fleet; they should receive CASREPs from noncombatants only if the message asks for technical assistance.

In our example message, shown in Fig. 1 and formatted in Table 2, the USS XXXXXXXX, an Atlantic Fleet ship, reported problems with a piece of equipment that has the equipment identification code (EIC) QEIN. The Q indicates that the failed equipment was electronics equipment, specifically a communications and data system. The XXXXXXXX is not a combatant ship (this information is incorporated in another production rule), but it is requesting technical assistance from *MOTU* (mobile technical unit) *twelve.* The dissemination list will therefore include desk N442. In these cases, the dissemination system derives the information needed for its decisions both from the *pro forma* data and from the narrative that has been mapped into the information format. *MOTU 12,* present in the narrative AMPN line of the message, is recognized as a technical unit providing technical assistance by the dissemination rules.

The second application is the generation of a summary describing the equipment malfunction reported in the narrative RMKS portion of the message. The summary typically consists of a single clause, extracted from several sentences of text, so that there is a five- to tenfold reduction of material. The summaries rarely contain text that is not present in the narrative of the RMKS section and usually restate a clause that is already in the message. The generation of each summary usually involves reading the entire message and then selecting an appropriate clause as the summary. Such summaries, which up to now have been generated by hand by a NAVSEA contractor, are used to detect patterns of failure for particular types of equipment. This failure information is crucial to decisionmakers who procure equipment for new and existing ships. Clearly, the sharp reduction in reading material can ease the decision-making process, provided that the key information from the report regularly finds its way into the summary.

The data for the summarization system are obtained entirely from the information format, and not from any of the *pro forma* data sets. A set of production rules is used to identify the crucial clause that will be used for the summary. The criteria for the production rules are based on the manual summarization that is currently performed. In our illustration, we will refer to only format rows 7 to 9 of Table 2, although the rules discussed apply in the same way to rows 3 to 5, respectively.

The summarization system proceeds in three stages: (a) inference, (b) scoring the format rows for their importance, and (c) selection of the appropriate format row as the summary.

First, inferences are drawn by a set of production rules. For example, words such as *inhibit, impair, prevent, etc.* are grouped into a single category, here the *impair* category. The presence of one of the words in this category triggers an inferencing rule. The inferencing rule is sensitive to the ordering of rows in the information format table. If part1 impairs part2, we can infer that part1 causes part2 to be bad, and we can also infer that part1 is bad. A set of production rules, summarized as rules (7) and (8) below, operate on the format lines to draw such inferences; these rules are sensitive to the order of the arguments of the connectives and, therefore, sensitive to the order of the rows in the information format table. The production rule in (7) infers that the second argument (part2) of CONN is bad.

(7) If	both	(1)	CONN contains an 'impair' word
	and	(2)	the STATUS column of the 2nd argument of CONN is empty
then	both	(3)	fill the STATUS column of the 2nd argument with 'bad'
	and	(4)	assign the word in CONN the attribute 'cause.'

For illustration, we refer to Table 2, the information table derived for our sample message (Fig. 1). *Inhibit*, in row 8, has been mapped by the formatting procedure into the CONN column, connecting the two format rows 7 and 9. The first argument of CONN is row 7, and its second argument is row 9. Rows 7 and 9 both have the PARTs column filled: row 7 with *APC-PPC circuit* and row 9 with *PA driver.* By a previous production rule, the verb stem "inhibit" has been categorized to the class of impairment verbs. Rule (7) replaces impairment by a format version of "cause to be bad." Specifically, the verb *inhibit* in the CONN column gets assigned the attribute "cause." Since the STATUS column in row 9 is empty, *bad* is inserted into the STATUS column of row 9. Thus, it is inferred that the PA driver is *bad* because it has been impaired. Another production rule, summarized as (8), infers that the STATUS column of the first argument (part1) of CONN is also 'bad' and inserts *bad* into the STATUS column since it has caused something else to have a bad status.

(8) If	both	(1)	the head of CONN has the attribute 'cause'
	and	(2)	the STATUS of the first argument of CONN is empty
	and	(3)	the STATUS of the second argument of CONN is 'bad'
then		(4)	insert 'bad' into the empty STATUS column.

Since 'inhibit' in the CONN of row 8 now has the attribute 'cause', by rule (7), and the STATUS of *APC-PPC circuit* of row 7 is empty while the STATUS of row 9 contains 'bad' by rule (7), 'bad' is inserted into the STATUS column of row 7 by rule (8).

The second stage of the summarization system rates the format rows for their importance to the summary. When it comes time to score the various formats to determine which is the most appropriate for the summary, the fact that "bad" is a member of the class of words signifying malfunction will cause format rows 7 and 9 to be promoted in importance. An additional scoring increment will accrue to 7 but not the others because it is a cause rather than an effect. Another rule increments a format row referring to an assembly, a midlevel component, since such a format is more revealing than a format containing a statement about a whole unit or an individual part, such as a transistor. For example, *circuit*, the head of the PART phrase in row 7, is identified as belonging to a class of components at the assembly level. As a result, the score of row 7 is incremented again.

In addition, the system has rules excluding format rows containing very general statements from summaries. For instance, universal quantification and mention of the top level in a part of tree betray a clause that is too general to be useful.

The third and final stage of summarization is to select the format row or rows with the highest rating. As a result of the various production rule actions, the winning format is "PART: APC-PPC circuit; STATUS: bad." From the format table that has been modified by the production rules, this selects rows 3 and 7. Both rows have been selected because they received identical scores. This arose because they were both part of expanded conjoined sentences having the CONN *and*, row 6 of Table 2. Other format rows also have positive scores. Another causal CONN occurs, namely, *allow*, in row 35, and there are other bad STATUSes, *zero units* in rows 23, 24, 27, and 29, and *not adjustable* in row 31. However, these format rows are not selected because their scores are not the highest.

We view these two application systems as members of a family of systems that will perform not only dissemination and summarization, but also such tasks as data base update, message creation, and question answering for a variety of operational reports. The common features of each family member are (a) a linguistically motivated message analyzer that generates computer interpretations of message content [16], and (b) an application system that defines those aspects of interpretation that are needed to perform a specific task.

6. EXPERIMENTAL RESULTS

The purpose of this experiment was to test the feasibility of automatically summarizing narrative text in Navy equipment failure messages using techniques of computational linguistics and artificial intelligence. Computer-generated results were compared to those obtained by manual summarization procedures to evaluate the performance of the system.

Both the natural language processing components and the applications programs were under development while this experiment was being carried out. The implementation just described has been under development for about 10 months. The messages were preselected only to the extent that they all contain some REMARKS narrative. Our initial test corpus consisted of a set of 26 electronic equipment CASREPs from a batch received from SPCC (the Ships Parts Control Center). The assistance amplification and remarks sections of these messages together contain 109 sentences. Our parsing procedure has successfully analyzed 92 of these sentences (84.4%), and in particular, has been successful in analyzing all the sentences in 12 of the messages. These 12 casualty reports were used for debugging the programs and have been successfully processed through the dissemination and summarization components. Subsequently, 12 other reports were used for the computer-human comparison.

For an appropriate summary line to be generated, it is necessary that 100% of the sentences in a text be processed correctly by the natural language procedures. The natural language analysis procedures processed 100% of the sentences contained in the second set of documents; this percentage includes 9 sentences (25%) that were paraphrased and rerun because they were not correctly processed on their first run. Paraphrasing these sentences brought the total number of sentences from 36 to 38.

The sentences were paraphrased to expedite processing since the major purpose of running the second set of messages was to test the performance of the summarization system, not the performance of the natural language processing system. Seventy format lines were generated from 38 sentences in 12 messages.

The computer-generated results of the summarization program compare favorably to those obtained manually. Table 3 shows a comparison of the two sets of results for the 12 test documents. The discrepancies between the computer-generated results and the manual results are summarized in Table 4.

Table 3 — Comparison of Machine and Manual Summary Results

Doc.	Machine *# format rows*	Manual *# sentences*	Agreement Machine/Manual
1.	1	1	1/1
2.	1	1	1/1
3.	1	1	1/1
4.	1	1	0/1
5.	1	2	1/2
6.	2	1	1/1
7.	1	1	1/1
8.	2	1	1/1
9.	1	1	0/1
10.	1	2	1/2
11.	1	1	1/1
12.	1	2	1/2
	14	15	10/15

Table 4 — Analysis of Machine and Manual Summary Results

#	Discrepancy	Doc.
1	word not included in category list	4
1	second manual summary not about bad-status	5
1	second manual summary not contained in text narrative	10
2	different summaries generated	9,12

Agreement between machine and manual summaries is obtained when the text contained in the format row selected by the automatic procedure agrees with the text in the sentences manually generated sentences. The discrepancies in the *Agreement* column of Table 3, as specified in Table 4, are illustrated in Table 5.

In our tally, we considered the manual- and machine-generated summaries as matching. This is illustrated for message 1 in Table 5. One (message 4) is the result of a failure to enter a word on a category list in the production rule system. As a result, the word was not categorized as a BAD-STATUS, and the score of its format row was not correspondingly boosted. Two errors (messages 5 and 10) were due to the program selecting one format line, although manual generation produced two sentences. In the first case (message 5), the additional text in the manual summary did not concern a description of a bad status. Rather it was a description of a good function status (i.e., *Drive shaft was found to rotate freely.*). In message 10, the extra manual summary consisted entirely of text (*loss of*

Table 5 — Examples of Machine and Manual Summaries

Doc.	Machine	Manual
1.	starting air regulating valve fail	starting air regulating valve failed
4.	unable (to maintain) lube oil pressure to SAC	inspection of LO filter revealed metal particles
5.	splines extensively worn	drive shaft rotates freely; splines were extensively worn
6.	NR 4 SAC oil pressure dropped start air pressure dropped	SAC oil pressure dropped below alarm point
9.	clog strainers; (due to)wiped bearing	loss of pressure when SAC engaged
10.	faulty high speed rotating assembly	loss of SAC faulty high speed rotating assy.

SAC) that was not contained in the message narrative. Our system does not automatically generate text, nor could it have made the inferences necessary to do so. In both these cases, however, the line that the program selected agreed with one of the manual summaries.

The most significant discrepancies (messages 9 and 12) were caused by the system selecting more specific causal information than was indicated in the manual summary. In message 9 of Table 5, which contains the sentence *Loss of lube oil pressure when start air compressor engaged for operation is due to wiped bearing*, the manual summary line generated was *Loss of LO pressure*, while the system selected the more specific information that indicated the cause of the casualty, i.e., *wiped bearing*. However, the manually generated line's score was the second highest for that message. This suggests that it may be more appropriate to select all the summary lines in some kind of score window rather than only those lines that have the highest score.

In two cases (messages 6 and 8), illustrated for message 6 in Table 5, the system generated two summary texts, although the manual summary consisted of only one sentence. Two summary lines were selected because both had equally high scores. Nonetheless, one of the two summaries was also the manual summary.

In conclusion, the summarization system was able to identify the same summary line as the manual summary 10/15 times (66.6%). For 10 out of 12 messages (83.3%), the summarization system selected at least one of the same summary lines as the manual generation produced. For two messages, the system was not able to match the manual summary; in one case, because the crucial status word was not in the appropriate list in the production rule system and, in a second case, because the automatic procedure identified the more specific causal agent.

7. DISCUSSION

We believe that the work just described represents a successful first step towards demonstrating the feasibility of automatically processing Navy messages based on their narrative content. At the same time, we recognize that much remains to be done before we have an operational system. Among the areas that require further development are:

Refinement of the format. Our current information format has been developed from a limited *corpus*—the initial set of 26 messages. Even within this *corpus*, not all types of information have been captured—for example, modes of operation, relations between parts and signals, and relations and actions involving more than one part. It is clear that enrichment of the information format is a high priority.

Intersentential processing. Our current implementation does almost no intersentential processing. This has proved marginally adequate for our current applications, but clearly needs to be remedied in the long run. One aspect of this processing is the insertion in the format of information that is implicit in the text. This includes missing arguments (subject and objects of verbs) and anaphors (e.g., pronouns) which can be reconstructed from prior discourse (earlier format entries); such processing is part of the information formatting procedure for medical records [8]. It should also include reconstruction of some of the implicit causal connections between sentences. To a greater degree than the other stages of formatting, the reconstruction of the connections will require substantial domain knowledge, of equipment-part and equipment-function relations, as well as "scriptal" knowledge of typical event sequences (e.g., failure — diagnosis — repair).

Robustness. Perhaps the most crucial issue separating current prototype systems from operational systems is that of *robustness.* By *robustness*, we mean the ability to deal effectively with input that violates some constraint of the analysis procedure or contains some unresolvable ambiguity. Through better analysis procedures and richer domain knowledge, we can expect to gradually reduce the volume of such input, but this is a slow process. It seems that the most fruitful avenue at present is to perform the message analysis at the time of data capture, so that, when problems arise, the system can ask the user for clarification.* Such on-line analysis as part of a message entry system should also be able to detect omissions of crucial information and prompt the user for this information; in this way it may be possible to improve on the present rather uneven quality of the message narratives.

Future applications of natural language processing at the Navy Center for Applied Research in Artificial Intelligence include an on-line message entry system, based on user interaction, that will process messages at the transmission end, rather than at the reception end long after the user has sent the message, and also the development of a natural language interface for querying the Navy's 3M (Maintenance and Material Management) data base.

8. IMPLEMENTATION

The LSP parser and Restriction Language programming language run on a DEC VAX 11/780 under the UNIX and VMS operating systems. Both are implemented in FORTRAN 77. The parser consists of about 15,000 lines of code. It requires 2 megabytes of virtual space when executing; of this, about 2/3 is list space for holding the grammar, dictionary entries, etc. The English grammar, regularization component, and information formatting component are written in Restriction Language, a special language developed for writing natural language grammars [19]. The dissemination and summary generation applications programs are written using the OPS5 production system. In total, there are 63 production rules in the applications programs.

9. ACKNOWLEDGMENTS

This research was supported by the Office of Naval Research and the Office of Naval Technology PE-62721N. The authors thank Dr. Jude Franklin and Dr. James Slagle for their useful comments on earlier versions of this report.

*This is the approach being taken at the University of California—Irvine for a different class of Navy messages [17,18].

10. REFERENCES

1. J. Froscher, R. Grishman, J. Bachenko, and E. Marsh. "A Linguistically Motivated Approach to Automated Analysis of Military Messages," Proc. 1983 Conf. on Artificial Intelligence, Apr. 1983, Oakland Univ., Rochester, MI.

2. J. Bachenko and J. Forscher, "A Classification of Navy Message Types," NRL Technical Memorandum 7590-312:JB:JF, Dec. 1982.

3. J. Froscher, "CASREP Justification," NRL Technical Memorandum 7590-324:JF:jf, Dec. 1982.

4. R. Schank and R. Abelson, *Scripts, Plans, Goals, and Understanding* (Lawrence Erlbaum Associates, Hillsdale, NJ, 1977).

5. C. Montgomery, "Distinguishing Fact from Opinion and Events from Meta-Events," Proc. Conf. Applied Natural Language Processing, Feb. 1983, Assn. for Computational Linguistics, Santa Monica, CA, pp. 55-61.

6. N. Sager, "Syntactic Formatting of Scientific Information," Proc. 1972 Fall Joint Computer Conf., AFIPS Conf. Proc. 41 (AFIPS Press, Montvale, NJ), pp. 791-800.

7. N. Sager, "Natural Language Information Formatting: The Automatic Conversion of Texts to a Structured Data Base," in *Advances in Computers* **17**, M.C. Yovits, ed. (Academic Press, New York, 1978), pp. 89-162.

8. L. Hirschman and N. Sager, "Automatic Information Formatting of a Medical Sublanguage," in *Sublanguage: Studies of Language in Restricted Domains*, R. Kittredge and J. Lehrberger, eds. (Walter de Gruyter, Berlin, 1982), pp. 27-80.

9. R. Grishman, L. Hirschman, and C. Friedman, "Natural Language Interfaces Using Limited Semantic Information," Proc. COLING 82, (Int'l Conf. on Computational Linguistics, 9th, 1981, Prague), North-Holland, Amsterdam, pp. 89-94.

10. R. Grishman, L. Hirschman, and C. Friedman, "Isolating Domain Dependencies in Natural Language Interfaces," Proc. Conf. Applied Natural Language Processing, 1983, Assn. For Computational Linguistics, Santa Monica, CA, pp. 46-53.

11. N. Sager, *Natural Language Information Processing* (Addison-Wesley, Reading, MA, 1981).

12. E. Marsh and N. Sager, "Analysis and Processing of Compact Text," Proc. COLING 82, North-Holland, Amsterdam, pp. 201-206.

13. E. Marsh, "Utilizing Domain-Specific Information for Processing Compact Text," Proc. Conf. Applied Natural Language Processing, Feb. 1983, Assn. for Computational Linguistics, Santa Monica, CA, pp. 99-103.

14. J. Froscher, A. Werkheiser, and J. Bachenko, "Decision Processing in a Dissemination System for Military Messages," Proc. 5th MIT/ONR Workshop on C3 Systems, Dec. 1982, Doc. #557829, MIT, Cambridge, MA, pp. 156-159.

15. J. Reggia, "Knowledge-Based Decision Support Systems: Development Through KMS," University of Maryland Department of Computer Science Technical Report TR-1121, 1981.

16. J. Froscher, K. Hayes, and J. Bachenko, "A Prototype Message Analyzer for the CHARLOTTE System," NRL Technical Memorandum 7590-229:JK:KH:JB, Aug. 1982.

17. R. Granger, C. Staros, G. Taylor, and R. Yoshii, "Scruffy Text Understanding: Design and Implementation of the NOMAD System," Proc. Conf. Applied Natural Language Processing, Feb. 1983, Assn. for Computational Linguistics, Santa Monica, CA, pp. 104-106.

18. R. Granger, "The NOMAD System: Expectation-Based Detection and Correction of Errors During Understanding of Syntactically and Semantically Ill-Formed Text," *Am. J. Comp. Ling.* **9** (3-4), 188-196 (1984).

19. N. Sager and R. Grishman, "The Restriction Language for Computer Grammars of Natural Language," *Comm. ACM* **18**, 390 (1975).