

UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE

PORT DOCUMENTATION PAGE

AD-A205 499

FILE COPY

2b. DECLASSIFICATION/DOWNGRADING		1b. RESTRICTIVE MARKINGS	
4. PERFORMING ORGANIZATION REPORT NUMBER(S)		3. DISTRIBUTION/AVAILABILITY OF REPORT	
6a. NAME OF PERFORMING ORGANIZATION		7a. NAME OF MONITORING ORGANIZATION	
6b. OFFICE SYMBOL (if applicable)		7b. ADDRESS (City, State and ZIP Code)	
6c. ADDRESS (City, State and ZIP Code)		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	
11. TITLE (include Security Classification)		PROGRAM ELEMENT NO.	
12. PERSONAL AUTHOR(S)		PROJECT NO.	
13a. TYPE OF REPORT		TASK NO.	
13b. TIME COVERED		AGENCY ACCESSION NO.	
14. DATE OF REPORT (Year, Month, Day)		15. PAGE COUNT	
16. SUPPLEMENTARY NOTATION		17. COSATI CODES	
18. SUBJECT TERMS (Continue on reverse if necessary and identify by block number)		19. ABSTRACT (Continue on reverse if necessary and identify by block number)	
FIELD		GROUP	
SUB-GROUP		Accession For	
		NTIS GRA&I	
		DTIC TAB	
		Unannounced	
		Justification	
		By	
		Distribution/	
		Availability Codes	
		Dist Avail and/or Special	
		A-1	
Viewgraph presentation.		DTIC ELECTE	
QUALITY INSPECTED 4		16 MAR 1989	
Presented at the Natural Language Systems Evaluation Workshop, 8 - 9 December 1988, Philadelphia, PA.		S D	
20. DISTRIBUTION/AVAILABILITY OF ABSTRACT		21. ABSTRACT SECURITY CLASSIFICATION	
UNCLASSIFIED/UNLIMITED ☒ SAME AS RPT ☐ DTIC USERS		UNCLASSIFIED	
22a. NAME OF RESPONSIBLE PERSON		22b. TELEPHONE (include Area Code)	
B. M. Sundheim		(619) 553-4145	
		22c. OFFICE SYMBOL	
		Code 444	

**A PROPOSED PLAN FOR ESTABLISHING
TASK-ORIENTED BENCHMARKS
FOR TEXT UNDERSTANDING SYSTEMS**

**BETH SUNDHEIM
NAVAL OCEAN SYSTEMS CENTER**

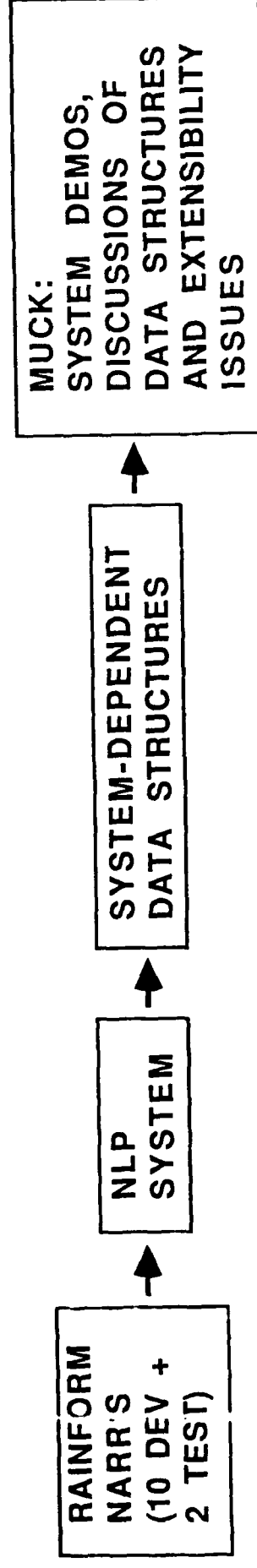
**NL SYSTEMS EVALUATION WORKSHOP
PHILADELPHIA, PA
8-9 DECEMBER 1988**

The DESIGN OBJECTIVES

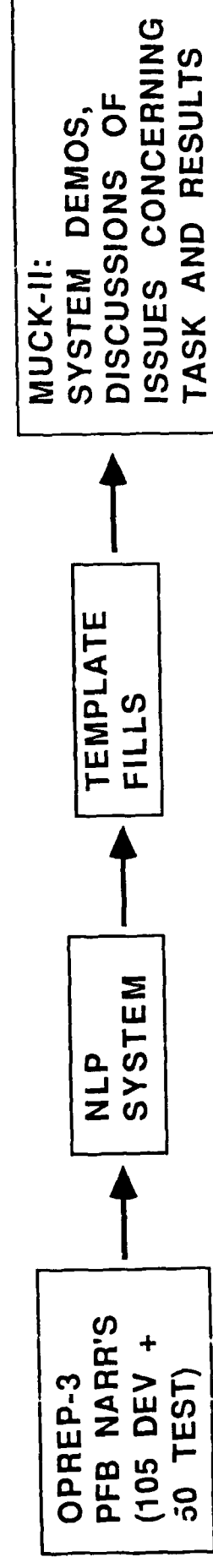
- 1-● RESTRICTED, HIGH-INTEREST DOMAIN;
- 2-● SIZEABLE CORPUS IN ONE MESSAGE FORMAT;
- 3-● REASONABLE TASK WITH WELL-DEFINED OUTPUT REQUIREMENTS
- 4-● TASK FOCUSED ON TEXT UNDERSTANDING

PREVIOUS AND CURRENT EFFORTS

1987 -- EXTENSIBILITY OF TEXT UNDERSTANDING SYSTEMS



1988-89 -- BENCHMARKING NL UNDERSTANDING SYSTEMS



**SIMULATED OPREP-3 PFB MESSAGE
(BASED ON DATA FROM NOSC TD 1060
AND OPNAVINST 3100.6D)**

**Z 222157Z DEC 00
FM USS PETERSON**

OPREP-3 PINNACLE/FRONT BURNER/N03361/001

- 1. INCIDENT. ATTACK ON HOSTILE SUBSURFACE**
- 2. CDR'S ESTIMATE. CONTINUING MISSION**
- 3. REFERENCE. NONE**
- 4. DETAILS.**

A. TIME. 222118Z DEC 00

B. LOCATION. 1000N1 - 10000W1

**C. NARRATIVE. VISUAL SIGHTING OF PERISCOPE AND RADAR CONTACT,
SONAR ACTIVE CONTACT. OPENED FIRE WITH 51 54 GUNS - 16 ROUNDS
EXPENDED CONDUCTED URGENT ATTACK WITH BLOODHOUND.**

5. LOSS/DAMAGE. UNKNOWN

**6. REMARKS. DARKNESS PRECLUDES SIGHTING DEBRIS. AMPLIFYING INFO TO
FOLLOW.**

INPUT VERSIONS OF MESSAGE NARRATIVES

1. BASIC FORM (VERBATIM NARRATIVE LINE)

#2-40

VISUAL SIGHTING OF PERISCOPE AND RADAR CONTACT, SONAR ACTIVE CONTACT. OPENED FIRE WITH 5IN 54 GUNS - 16 ROUNDS EXPENDED CONDUCTED URGENT ATTACK WITH BLOODHOUND.

2. MODIFIED FORM (BASIC ELIDED ELEMENTS SUPPLIED; SPELLING AND PUNCTUATION CORRECTED; EQUIPMENT TERMS REGULARIZED)

#2-40

A VISUAL SIGHTING OF A PERISCOPE WAS MADE BY USS PETERSON, AND RADAR CONTACT AND SONAR ACTIVE CONTACT WERE GAINED.

**PETERSON OPENED FIRE WITH 5/54 GUNS.
16 ROUNDS WERE EXPENDED.**

PETERSON CONDUCTED AN URGENT ATTACK WITH A BLOODHOUND.

OUTPUT TEMPLATE AND FILL

0
1
2
3
4
5
6
7
8
9
10

MESSAGE ID
EVENT: HIGHEST LEVEL OF ACTION
FORCE INITIATING EVENT
CATEGORY(S) OF EVENT AGENT(S)
CATEGORY(S) OF EVENT OBJECT(S)
ID(S) OF 0-TH LEVEL AGENT(S)
ID(S) OF 0-TH LEVEL OBJECT(S)
INSTRUMENT(S) OF 0-TH AGENT(S)
LOC OF OBJECT(S) AT EVENT TIME
TIME(S) OF EVENT
RESULT(S) OF EVENT

#2-40
ATTACK
FRIENDLY
SURF
SUB
[USS PETERSON]
NO DATA
5/54 GUN, BLOODHOUND TORPEDO
NO DATA
NO DATA
NO DATA

EXPECTED RESULTS

<u>MEASURE</u>	<u>MEANS OF CALCULATION</u>	<u>INSIGHTS GAINED</u>
"RECALL"	PERCENT TEMPLATES AND SLOTS FOR TEST SET FILLED THAT SHOULD BE FILLED	1) COVERAGE OF ENGLISH; 2) DEPTH OF UNDERSTANDING; 3) ROBUSTNESS
"PRECISION"	OF THOSE SLOTS CORRECTLY RECALLED FOR TEST SET, PERCENT CORRECTLY FILLED	
"GENERALITY/ POTENTIAL"	COMPARISON OF RECALL AND PRECISION MEASURES FOR DEVELOPMENT VS TEST SET	1) GENERALITY OF SOFTWARE; 2) POTENTIAL FOR HANDLING NOVEL TEXT
"UTILITY"	RECALL AND PRECISION MEASUREMENTS USING ORIG NARR'S AS INPUT (VS MODIFIED NARR'S)	ABILITY TO HANDLE TELEGRAPHIC, ILL-FORMED TEXT AND SUBLANGUAGE CONSTRUCTIONS
"PROGRESS"	RESULTS USING FIRST TEST SET VS RESULTS USING SECOND TEST SET	PACE AT WHICH STATE OF THE ART IS ADVANCING

MATERIALS IN PREPARATION

- NARRATIVES:
 - SANITIZED ORIGINALS, WITH AND WITHOUT MODIFIED VERSIONS
- REPORT ON CORPUS:
 - COMPLETE, SANITIZED MESSAGES
 - GLOSSES
 - GLOSSARY OF TERMS
 - CATEGORIZATION OF AGENTS AND INSTRUMENTS
 - TERMINOLOGY FOR CATEGORIES
 - CHARACTERISTICS OF AGENTS AND INSTRUMENTS
 - BACKGROUND INFORMATION ON DOMAIN AND NAVAL MESSAGES
- REPORT ON TEMPLATES:
 - TEMPLATE FILLS FOR DEVELOPMENT SET
 - "COREFERENCE" TABLE OF NOMENCLATURE
 - BACKGROUND INFORMATION
- DEVELOPMENT SET AIDS:
 - KWIC CONCORDANCE
 - WORD FREQUENCY LIST

PROPOSED SCHEDULE

1 MAR 89

DISTRIBUTION OF DEVELOPMENT PACKAGE TO
MUCK-II INVITEES

1 MAR - 30 APR 89

NOSC PERSONNEL AVAILABLE FOR CONSULTATION
ON OPREP-3 PFB AND TEMPLATE FILLS FOR
DEVELOPMENT SET

1 MAY 89

DISTRIBUTION OF FIRST TEST PACKAGE TO MUCK-II
INVITEES

6-7 JUN 89

MUCK-II

TWO GENERAL OPEN QUESTIONS

- IS TASK DOABLE ENOUGH IN THE TIME ALLOTTED THAT SIGNIFICANT RESULTS CAN BE OBTAINED?
- IS TASK DESIGNED WELL ENOUGH THAT RESULTS WILL TRULY MEASURE SYSTEMS WITH RESPECT TO THE DESIRED EVALUATION CRITERIA?