RL-TR-95-265
Final Technical Report
February 1996

MACHINE-AIDED VOICE TRANSLATION (MAVT)

Language Systems, Inc.

Christine A. Montgomery, Bonnie G. Stalls,
Robert Stumberger, Naicong Li, Robert S. Belvin,
Alfredo Arnaiz, Philip C. Shinn, Armand G. DeCesare,
and Robert S. Farmer

APPROVED FOR PUBLIC RELEASE; DISTRIBUTION UNLIMITED.

Rome Laboratory
Air Force Materiel Command
Rome, New York

This report has been reviewed by the Rome Laboratory Public Affairs Office (PA) and is releasable to the National Technical Information Service (NTIS). At NTIS, it will be releasable to the general public, including foreign nations.

RL-TR-95-265 has been reviewed and is approved for publication.

APPROVED:

SHARON M. WALTER
Project Engineer

FOR THE COMMANDER:

JOSEPH CAMERA
Technical Director
Intelligence & Reconnaissance Directorate

REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 0704-0188

1. AGENCY USE ONLY (Leave Blank)	2. REPORT DATE	3. REPORT TYPE AND DATES COVERED
	February 1996	Final Jun 90 - Aug 92

4. TITLE AND SUBTITLE

MACHINE-AIDED VOICE TRANSLATION (MAVT)

5. FUNDING NUMBERS

C - F30602-90-C-0058
PE - 63260F
PR - 3481
TA - 00
WU - 14

6. AUTHOR(S)

Christine A. Montgomery, Bonnie G. Stalls, Robert Stumberger, Naicong Li, Robert S. Belvin, Alfredo Arnaiz, Philip C. Shinn, Armand G. DeCesare, and Robert S. Farmer

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)

Language Systems, Inc.
6269 Variel Ave., Suite 200
Woodland Hills CA 91367

8. PERFORMING ORGANIZATION REPORT NUMBER

N/A

9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)

Rome Laboratory/IRAA
32 Hangar Rd
Rome NY 13441-4114

10. SPONSORING/MONITORING AGENCY REPORT NUMBER

RL-TR-95-265

11. SUPPLEMENTARY NOTES

Rome Laboratory Project Engineer: Sharon M. Walter/IRAA/(315) 330-4025

12a. DISTRIBUTION/AVAILABILITY STATEMENT

Approved for public release; distribution unlimited.

12b. DISTRIBUTION CODE

13. ABSTRACT (Maximum 200 words)

Military field interrogators require both foreign language proficiency and interrogation skills. The Machine-Aided Voice Translation (MAVT) prototype system was developed in response to the shortage of experienced interrogators with this mix of abilities. The MAVT accepts an interrogator's spoken English question and translates it into spoken Spanish. The spoken Spanish response of the potential informant can then be translated into spoken English. Other military applications for spoken language translation technology include foreign language training, multi-national operation assistance, and foreign communications processing. Civilian applications include law enforcement, diplomatic or business briefing aids, and tourist travel aids.

14. SUBJECT TERMS

Speech translation, Language translation, Natural language processing

15. NUMBER OF PAGES 132

16. PRICE CODE

17. SECURITY CLASSIFICATION OF REPORT	18. SECURITY CLASSIFICATION OF THIS PAGE	19. SECURITY CLASSIFICATION OF ABSTRACT	20. LIMITATION OF ABSTRACT
UNCLASSIFIED	UNCLASSIFIED	UNCLASSIFIED	UL

CONTENTS

1. INTRODUCTION AND SUMMARY

1.1 Introduction

This Final Technical Report summarizes work performed under Contract No. F30602-90-C-0058, "Machine-Aided Voice Translation (MAVT)". The project resulted in the development of a speaker-independent,[1] continuous speech, translation system for English => Spanish => English. An overview diagram of the system is presented in Figure 1-1, while the component functions are shown in Figure 1-2.

The MAVT testbed system developed in the course of this project effort and delivered to Rome Laboratory is designed to be language-independent, featuring a multilingual lexicon and a language-independent syntactic parser. These innovations are based on extensions to LSI's DBG natural language processing system, which was originally developed in previous projects for Rome Laboratory and serves as the core component of the MAVT testbed.

This introductory section (1.1)

- explains the functional concept of MAVT (Section 1.1.1);
- summarizes work performed on the development effort (Section 1.1.2);
- outlines the hardware and software components of the MAVT testbed configuration (Section 1.1.3);
- describes test and evaluation procedures (Section 1.1.4);
- and summarizes results of the development (Section 1.1.5).

Section 1.2 provides a summary of Sections 2 - 7 and appendices.

1.1.1 The MAVT Concept

The MAVT project is the first phase of a prototype development to assist Air Force interrogation personnel in interacting with potential informants in an unfamiliar foreign language, as depicted in Figure 1-3. The initial phase has resulted in aiming the MAVT system toward the functions of screening and preliminary evaluation of potential informants.

A novice interrogator (or an experienced interrogator unfamiliar with the required foreign language) will be able to utilize an MAVT device to screen prisoners of war or other potential informants, and evaluate whether further questioning by an experienced interrogator with the requisite language skills would be productive. MAVT can thus be seen as a productivity-multiplier for the skilled interrogator, who is a scarce commodity, expensive to train, and should be concentrated on high-yield informants.

To support the screening function, the work of this initial phase has been focused on the collection of information in the critical domains of biographic and mission-related data, based on interrogation scenarios generated by LSI's IPW (Interrogation of Prisoners of War) consultants [TIR: see Section 8, References]. Background on IPW and rationale for the scenario development is given in Section 2.0.

1. This distinguishes MAVT from AT&T's recently announced English=>Spanish=>English system, which is speaker-dependent (Bindra 1992).

MACHINE - AIDED VOICE TRANSLATION

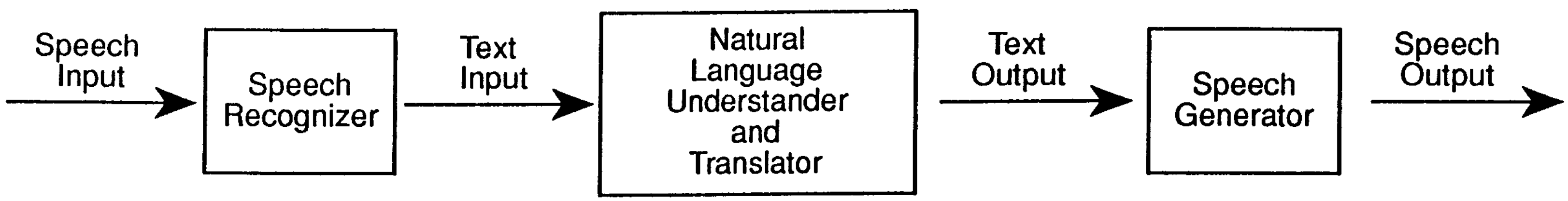


Figure 1-1. MAVT System Diagram

Machine - Aided Voice Translation

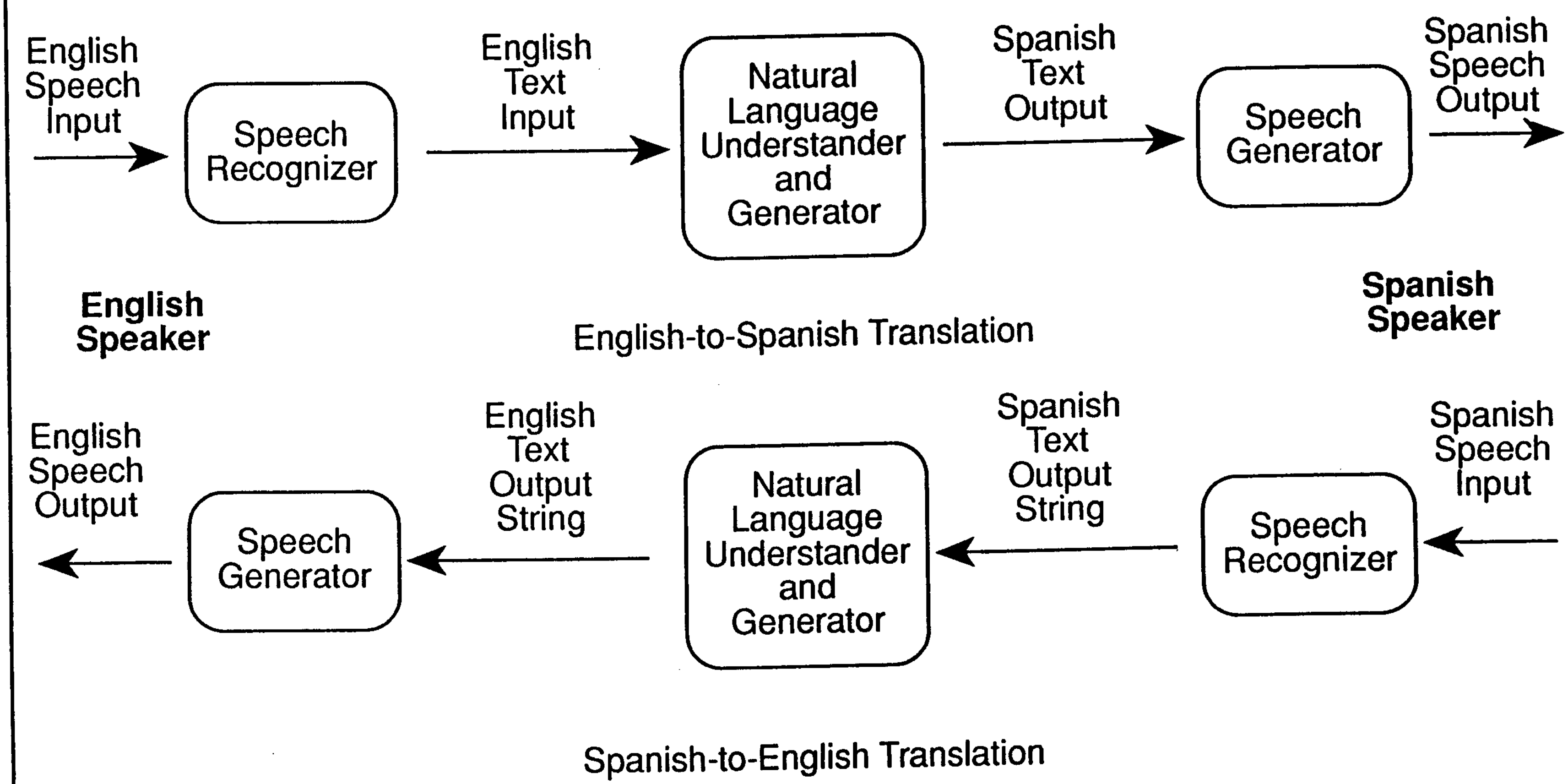


Figure 1-2. MAVT Functions

MACHINE- AIDED VOICE TRANSLATION
MAVT
THE CONCEPT

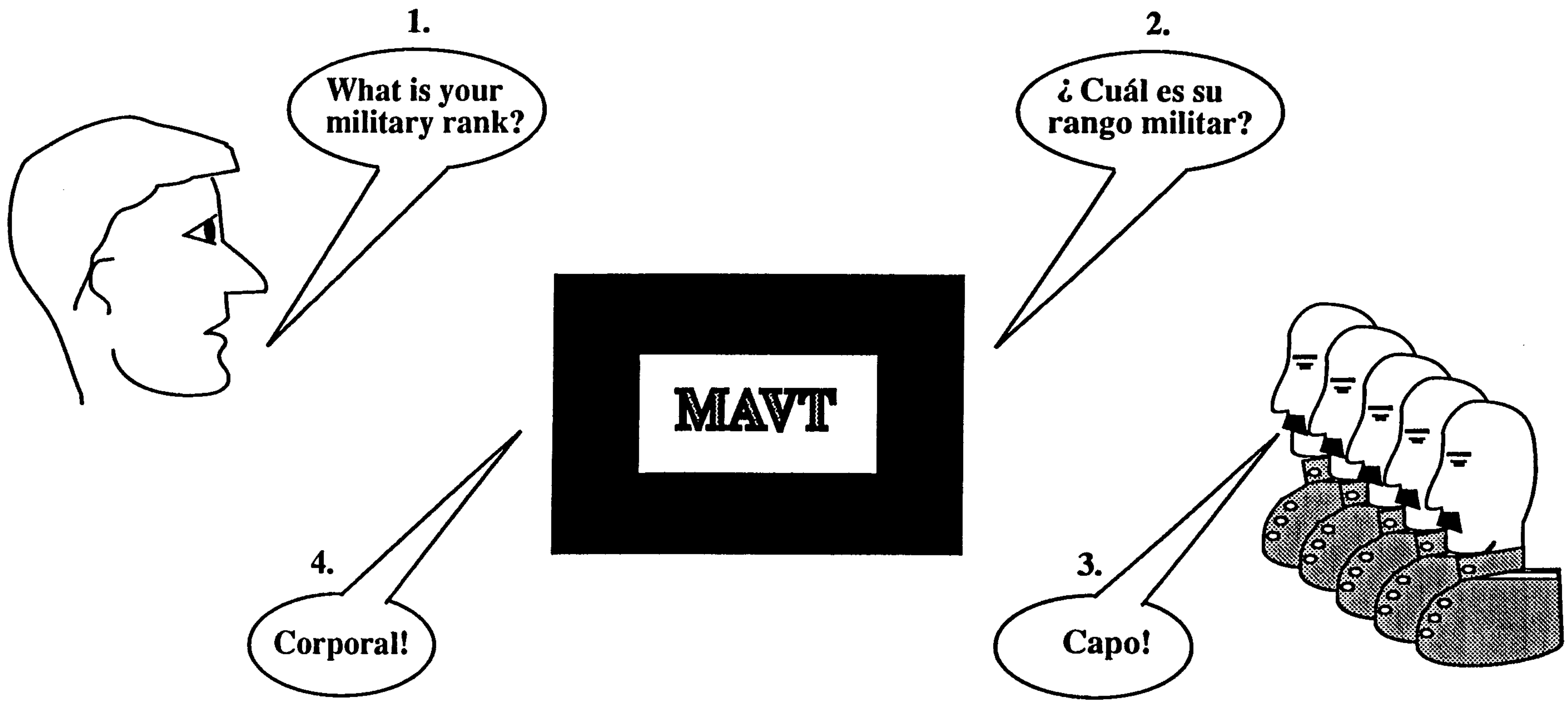


FIgure 1-3. Concept of MAVT Function

1.1.2 Summary of the MAVT Project Effort

Recognizing the limitations of state-of-the-art speaker-independent, continuous speech recognition systems, as well as available funding, the Rome Laboratory Statement of Work specifically provided for a constrained approach to development of the initial MAVT system. It would therefore have been possible to considerably limit the scope of the work carried out under this contract, e.g., by utilizing a small vocabulary speech recognizer, by relying on speaker-dependent technology, by cobbling together existing NLP and translation software, etc. LSI did not follow this approach for the reasons discussed below.

In the first place, such a development would have had very limited legacy for the following phase of MAVT. Lessons learned would have been virtually inapplicable elsewhere, and test and evaluation results could not have been extrapolated to MAVT developments requiring larger vocabularies, more complex sentence structures, additional foreign languages, additional domains or applications. In addition, an endeavor of such limited scope would have made no contribution toward the definition of basic or applied research issues for speech or natural language processing, the hardware and software technologies that support them, or interrogation training or practice.

The approach adopted by LSI was rather to construct a foundation for future development emphasizing multilingual, multidomain capabilities, based on the SOW requirement for extensibility to numerous languages and additional interrogation domains. Thus, we extended the DBG NLP component with a multilingual lexicon, a multilingual morphological component, and a syntactic parser designed to be language independent. (These features of the system are discussed in Section 5.)

In line with this approach of building a solid foundation for MAVT development, we selected a large vocabulary, continuous speech, speaker-independent automated speech recognition (ASR) system, which allowed us to perform detailed evaluations of the effect of vocabulary size and syntactic complexity on robustness of speech recognition in two interrogation domains (biographics and mission) for both English and Spanish. (This work is discussed in Section 4.1 and Section 6.3.)

We also evaluated the advantages and disadvantages of two methods of synthesizing speech, Digital Audio Playback (DAP) and Text-to-Speech (TTS), for the IPW application, as described in Section 4.2.

Technical tasks specified in the RL SOW and performed in the course of the MAVT contract effort include -- but are not limited to -- the following:

1. **Data Collection.** Collected data for interrogation scenarios and user requirements. Prepared three interrogation scenarios with supporting materials. Defined screening function as paramount for MAVT.

2. **Data Analysis.** Analyzed collected data to produce constrained interrogation sequences. Performed word and phrase frequency analyses to assist in developing linguistic and domain knowledge bases for MAVT testbed.

3. **Language Constraint Development.** Developed and tested speech recognition demonstration of constrained interrogation questions. Formulated speech syntax for English biographics and mission domains.

4. **Response Constraint Development.** Developed specifications for constrained responses to interrogation questions. Formulated speech syntax for Spanish biographics and mission domains, collected Spanish speech data and developed limited speaker model for Latino Spanish.

5. **Language Translation Experimentation.** Developed and tested natural language (text) processing for translation of constrained interrogation sequences into Spanish and English. Developed linguistic and domain knowledge bases required to support translation.

6. **Word and Phrase Speech Recognition.** Evaluated speech recognition technology with respect to the constrained interrogation sequences in Spanish and English. Experimented with various branching factors and vocabulary sizes.

7. **Speech Recognition and NLP Model Interaction.** Experimented with possible interactions between speech recognition and language processing components.

8. **Speech Synthesis.** Evaluated Text-to-Speech (TTS) and Digital Audio Playback (DAP) for voice output in English and Spanish.

9. **System Design Development.** Developed design for prototype system architecture.

10. **Test and Evaluation of System.** Tested demonstration prototype with constrained interrogation scenarios in English and Spanish.

In addition to these technical tasks, LSI performed all the major system tasks of MAVT testbed hardware and software selection, acquisition, and integration, as well as preparation of reports, documentation, and other deliverables (see Section 8, References: Contract Technical Documentation).

To perform all these tasks required a non-trivial amount of effort. Although we feel that more work could be done on all of them, we believe that the current MAVT testbed is a substantial achievement within the limitations of time and resources.

1.1.3 The MAVT Testbed

The current MAVT testbed is comprised of three hardware/software components corresponding to the boxes in Figure 1-1:

1. the Phonetic Recognition System of Speech Systems, Inc. (SSI)
2. LSI's DBG natural language processing system, which performs the core functions of language understanding and translation;
3. a DECtalk speech synthesizer.

The speech recognition subsystem of the current MAVT testbed contains the following numbers of lexical items and syntax rules in the listed grammars:

Speech grammar	Number of lexical items	Number of syntax rules
English biographics:	66	28
English mission:	60	51
Spanish biographics:	238	212
Spanish mission:	113	58
Totals	477	349

The DBG natural language processing subsystem, as configured for the MAVT application, currently has over 1,000 lexical items for both languages, and is estimated as being comprised of approximately 15,000 lines of Prolog code.

The DECtalk Spanish speaker model has approximately 400 words in its current repertoire.

Section 7 contains a detailed assessment of the status of the MAVT testbed as delivered to Rome Laboratory, including coverage of the various subsystems and components, and types of items not currently covered.

1.1.4 Test and Evaluation

Regression testing based on a diagnostic set of sentences for both English and Spanish (Appendix A) was carried out on a regular basis during the the most intensive phases of system development. A formal Test Plan was submitted to RL in January, 1992, which specified a rigorous test and evaluation of all subsystems and components of the MAVT system [TP]. The Test Plan defined 15 distinct tests, of which three were "black-box" tests of the total system, and the other 12 were "glass-box" tests of individual subsystems and components, e.g., Spanish speech recognition, English syntactic parsing, Spanish translation generation, etc.

The test corpus of 100 sentences was selected by Lt. Bradford Clifton; an initial acceptance test was carried out during his visit to LSI in June, 1992, and final acceptance testing was performed at LSI prior to delivery of the MAVT testbed software and hardware to Rome Laboratory.

Speech recognition testing included part of three overall system "black-box" tests, as well as two "glass-box" tests of the ASR subsystem only. Results of speech recognition testing were scored both according to current DARPA SLS conventions and in terms of LSI measurements of accuracies along four different dimensions of speech recognition and functional validity for the IPW application. English recognition and generation are performed using standard English speaker models furnished with the SSI ASR and DECtalk equipment, while Spanish recognition and generation are performed using limited speaker models and adaptations developed by the LSI MAVT team for this project.

Since the NLP subsystem forms the core of the MAVT testbed, tests of the NLP subsystem comprised the majority of the test and evaluation activities carried out (part of three overall system "black-box" tests, 8 "glass-box" tests of NLP components only). Detailed test results are given in Section 6, which contains 13 tables listing the individual tests, test data, and test results for all tests.

1.1.5 MAVT Project Results

1.1.5.1 Summary of Test Results

In the final test and evaluation of the most current version of the MAVT experimental system, average accuracy for all sentences in the test corpus for all 15 tests including both English and Spanish is 87%. Based on LSI scoring methods, speech recognition accuracy is 88% for English and 81% for Spanish, averaging results of the biographics and mission domains, and figures for word, utterance, semantic, and translation accuracy. Using DARPA SLS scoring methods, average accuracy is 95% for English and 86% for Spanish.

1.1.5.2 Basic and Applied Research Issues

As discussed in Section 1.1.2, LSI adopted the approach of making the MAVT testbed a foundation for future development rather than a dead end implementation. Based on the results obtained using this approach, a number of issues merit further research. These include, but are not limited to the following:

- degree of integration between the component subsystems: in particular, the degree of integration of the ASR and NLP subsystems can be seen as a continuum ranging from unintegrated to fully integrated. Integration possibilities range from lexical and grammatical compatibility (as in the current MAVT testbed) to use of the NLP to select which of the N-best interpretations made by the ASR is actually the best (based on NLP criteria including lexical, syntactic, semantic, and discourse properties) to direct coupling of the ASR and NLP lexicons and grammars for fully integrated processing.
- degree of unconstrained dialog that can be handled by the ASR subsystem in an interrogation situation (experimentation to determine the optimum point between use of the constrained grammars, such as the biographics and mission grammars, versus the chatter grammar).
- degree of language-independence that is achievable, for each NLP component: i.e., lexicon, syntactic parser, semantic (functional) parser, translation generation, etc.)
- use of discourse phenomena to achieve selection of best candidate utterance interpretation output by ASR system, anaphora resolution, determination of register, substitution of semantically equivalent phrases, etc.
- use of the MAVT testbed for foreign language training and proficiency maintenance.
- applied psychological and sociological research in the discipline of interrogation: interaction with an MAVT device, training via MAVT, methodology for handling unknown words, determining errors, interacting with potential informants using an MAVT device, etc.

1.2 Summary

In the discussions that follow, Section 2 describes the interrogation function, information goals of interrogation, the screening process, and the interrogation scenarios developed by LSI's expert consultants and documented in [TIR].

Section 3 gives an overview of the hardware and software environment of the MAVT testbed, while Section 4 focuses on the speech processing components, with Section 4.1 treating the ASR subsystem and 4.2 the speech synthesis subsystem.

Section 5 describes the natural language processing components of the DBG system, concentrating on the multilingual aspects.

Section 6 gives detailed test and evaluation analyses, with 12 tables enumerating results to the sentence level, while Section 7 presents an in depth description of the current status and capabilities of the MAVT testbed system.

Section 8 lists project documentation and other relevant references and Appendix A contains diagnostic corpora of sentences in both English and Spanish.

2. THE MILITARY APPLICATION ENVIRONMENT

2.1 Military Linguists in the Air Force C^3I Environment

One serious issue facing military planners under the "new world order" is the dwindling resources that will be available to the military for personnel recruitment, training in a given specialty --e.g., interrogator-- and maintaining/improving the skills associated with the given specialty. In the case of interrogators, the primary skills are the cognitive and psychological skills required for interrogation, and the ability to speak and understand a foreign language with sufficient fluency to perform these cognitive and psychological interactions in an interrogation context.

Since military linguists typically are trained at the Defense Language Institute, Monterey, for a year in their designated language, and then must still attend several months of training to be able to apply the language to C^3I activities within their specialty (e.g., Interrogator, Voice Intercept Operator), training is an expensive and lengthy proposition. In particular, skilled interrogators require not only a long training period, but a number of years of experience in order to sharpen their interrogation and linguistic skills.

However, as resources become less, the demands for linguistic competence are on the increase. Many more situations will arise that will require interoperability with the armed forces of other nations, including large military operations such as Desert Storm, small scale Special Operations, Low Intensity Conflicts (LIC) in world trouble spots, peacekeeping force deployments, and other activities. Although interoperability is an important issue, European nations have typically been involved, and a fair number of military linguists are trained in European languages.

On the other hand, operations such as Desert Storm dramatize the need for military linguists trained in languages other than the familiar European ones like French and German. At the outbreak of Desert Storm, there were only a handful of military linguists who were competent to conduct interrogations in Arabic, and very shortly, there were thousands of Iraqi prisoners of war to be interviewed.

The fundamental problem is that the skilled interrogator -- who is part linguist, part tactician, part psychologist, part intelligence analyst, part actor -- is a very scarce commodity, precisely because of those unique requirements of the interrogation specialty.

Clearly, there is an acute need for more skilled interrogators, but two even more critical needs are:

1. to maximize the skilled interrogator's considerable expertise in interviewing potential informants using his fluent foreign language ability through exploitation of an MAVT device to screen out candidate informants who are uncooperative, unimportant, have little or no relevant information to offer, or are otherwise of dubious value;

2. to reduce the time and expense involved in acquiring and maintaining fluency in a foreign language and ability in interrogation -- which can also be achieved through exploitation of an MAVT device for embedded training.

2.2 The Context of Interrogation of Prisoners of War (IPW)

As noted previously, LSI's project team has acquired specific familiarity with the IPW context in connection with the development of an experimental MAVT testbed to assist Air Force interrogators as described in items 1) and 2) above.

To provide realistic dialogs and situations, our Army and Marine IPW consultants generated three interrogation scenarios based on materials supplied by USAICS (Combat Interrogator Course) and AFSAC (Interrogation Guide), as well as their own expertise. (Both consultants served as interrogators in the Vietnamese language during the war in Southeast Asia.) Two scenarios dealt with a single POW in a guerrilla action in the Caribbean, and the third was based on Desert Storm. The Caribbean scenarios were designed to illustrate interrogator verbal behavior in two different circumstances, the first assuming an uncooperative POW who potentially has significant knowledge, and the second assuming a cooperative informant with significant knowledge. Each simulated interrogation dialog was associated with an event history and Order of Battle information, as well as associated annotations explaining what the interrogator was attempting to accomplish with each transaction in the dialog. In the Desert Storm scenario, several POWs were interrogated to illustrate the application of screening a large number of potential informants to isolate the few appearing the most productive for exploitation by a skilled interrogator. The notion underlying this scenario is that the MAVT system could be used by a novice interrogator in the screening application, to reduce a cast of thousands to a tractable number requiring further exploitation by more experienced IPW personnel.

The following discussion briefly describes the context and the goals of an interrogation. While reading through these requirements and procedures, it is important to keep in mind that all this interaction is carried out in a language which is not the interrogator's own.

2.2.1 Interrogation Requirements and Procedures

Interrogation is the art of questioning and examining a source to obtain the maximum amount of usable information. The goal of any interrogation is to obtain usable and reliable information -- in a lawful manner and in the least amount of time -- which meets intelligence requirements of any echelon of command. Sources may be civilian internees, insurgents, EPWs, defectors, refugees, displaced persons, and agents or suspected agents. A successful interrogation produces needed information which is timely, complete, clear, and accurate. An interrogation involves the interaction of two personalities: the source and the interrogator. Each contact between these two differs to some degree because of their individual characteristics and capabilities, and because of variations in the circumstances of each contact and in the physical environment.

The interrogation process involves the screening and selection of sources for interrogation, and use of interrogation techniques and procedures. Both screening and interrogation involve complex interpersonal skills, and many aspects of their performance are extremely subjective. Each screening and interrogation is unique because of the interaction of the interrogator with the source.

There are five interrogation phases: planning and preparation, approach, questioning, termination, and reporting. Every phase is complicated and should be reported separately.

Questioning techniques are extremely important in the context of an interrogation. An interrogator must know when to use different types of questions. With good questioning techniques, the interrogator can extract the maximum amount of information in the minimum amount of time. There are many types of questioning techniques. For example, direct questions are basic interrogatives (who, what, when, where, and how plus qualifier, e.g., "how many, how long," etc.). Questions that a skilled interrogator avoids are leading questions (questions that suggest an answer), compound questions (questions that require two separate answers), and, where possible, questions that can be answered "yes" or "no". An interrogator will always attempt to ask questions that require a narrative response from the source. (The latter two strategies require modification when using an MAVT device, since it is important to stay within the constraints imposed by the state of the art in speech recognition.)

The three scenarios developed by LSI's IPW consultants in the course of the MAVT project use the direct approach, which is appropriate for the screening function, and lends itself the most readily to usage in the context of state-of-the-art automated speech recognition (ASR) technology. In any case, although there are over a dozen approach techniques and many more combinations of such techniques, the direct approach is always tried first. In this approach, the interrogator simply begins to ask questions and the source answers them. The interrogator will continue to ask direct questions as long as the source continues to answer. If the source becomes uncooperative, a new approach may be necessary. The direct approach is the quickest way to extract the most information in the shortest period of time as well as the most effective. According to statistical records, it was 85 percent to 95 percent effective in World War II, 90 to 95 percent effective in Vietnam. The direct approach works best on lower enlisted personnel as they have little or no resistance training and have had minimal security training. Due to its effectiveness, the direct approach is always tried first and used at the tactical echelons where time is limited. The interrogation scenarios presented in LSI's Interim Report provide extensive illustrations of this technique.

3. THE MAVT HARDWARE/SOFTWARE ENVIRONMENT

This section discusses the hardware and software configurations of the MAVT testbed. The testbed is composed of three principal subsystems: the speech recognizer, the language translator, and the speech generator. The hardware and software configurations mirror the higher-level system configuration.

The speech recognition component takes the speech utterance, as an audio signal, as input, and converts it into a list of text transcriptions (sentences). The translation component takes a single text transcription (sentence) and converts it into a sentence (text) in the target language. The speech generation component takes a sentence (text) and converts it into an audio signal (generated speech) and outputs the signal ("speaks").

3.1 Hardware

The hardware portion of the testbed consists of three primary elements: a host-server, the speech recognizer and the speech generator. The host-server is the central compute-engine for the testbed, providing operating system utilities, and supporting the user interface, the language translation subsystem, and a part of the speech recognition subsystem.

3.1.1 Host-Server

The host-server is a Sun Microsystems workstation (a SPARCstation-1), which is an entry-level Sun4 class machine. Any Sun4-class machine may be used for the testbed. The host-server has serial line (RS-232) connections to the speech recognition and speech generation hardware. These are the only hardware elements which must be directly connected to the server. System-level peripherals such as disk drives, tape drives, CDROM drives, etc., may be remotely accessed over a network.

3.1.2 Speech Recognition Hardware

The complete speech recognition system, known as the DS200 Speech Input Development System from Speech Systems, Inc. (SSI), consists of two components: a hardware-based speech recognizer, and a software-based speech decoder. The hardware component will be described in this section, and the software component will be described in Section 3.2. A detailed description of speech recognition processing is presented in Section 4.1.

The speech recognition hardware, known as the PE200 Phonetic Engine, consists of a special-purpose speech recognizer which is used to segment the audio speech signal and generate the feature codes for the segmented signal. This process is referred to as "phonetic encoding"; the speech recognition hardware element is called a "Phonetic Encoder". The feature codes produced by the speech recognition hardware are used by "phonetic decoding" software residing on the host-server, which traverses a grammar to map the phonetic feature codes onto one or more candidate text transcriptions of the initial speech utterance.

The speech recognizer has an attached disk drive (which is the second element of the speech recognizer hardware subsystem). This disk drive stores files required by the speech recognizer. These files are downloaded from the host-server by the speech recognition software.

The speech recognition hardware connects to the host-server through a serial line (RS-232) interface.

3.1.3 Speech Generation Hardware

The speech generation hardware consists of a DECtalk DTC01 speech generator from Digital Equipment Corporation (DEC). It is connected to the host-server through a serial line (RS-232) interface. It converts text which has been downloaded from the server, into an audio transcription of the text, which is played through the generator's speaker, or "spoken". The speech generator also accepts commands from the host-server which are used to alter the spoken voice, pitch, and speaking rate. The host server can also download additions to the generator's pronunciation dictionary, to allow for the addition of new words.

3.2 Software

The software portion of the testbed consists of two main elements (speech recognition and language translation) which support their respective primary subsystems. (Note that speech generation is entirely performed in the DECtalk hardware). Additional software elements consist of the user interface, which controls end-user interactions with the entire system, and interfaces to the speech recognition, translation, and speech generation systems.

3.2.1 Speech Recognition Software

The speech recognition software consists of the SSI phonetic decoding software and speech application development tools which run on the Sun workstation. All of this software is documented in the SSI manual, "Speech Input Development System (Model DS200): Reference Manual" [SSI-DS200]. (Note that DS200 is the combined configuration of the PE200 speech recognition hardware and the speech decoding software, which is currently at release 3.6.)

The speech decoding software consists of system calls which permit an application to interface with the Phonetic Engine. These calls allow an application to initialize the speech recognizer, use a specific grammar and dictionary and retrieve decoded utterances. The speech application development tools provide a higher-level interface to the entire speech development system for application developers, and are used to develop and test the various files, such as dictionaries and grammars, which are used by the decoding software.

3.2.2 Language Translation Software

The language translation software was entirely developed by LSI. It takes input text in a source language, and translates that text into output text in a target language. The processing stages are as follows:

1. Segment input text into individual words
2. Look up words in lexicon, or apply morphological rules to identify lexical category
3. Construct syntactic parse tree from lexicalized words
4. Derive functional parse (predicate-argument structure) from parse tree
5. Determine possible translation mappings for individual words and phrases in functional parse
6. Generate target text by applying sentence generation and target-language morphology rules

The language translation software is written in Quintus Prolog (version 3.1.1).

3.2.3 User Interface

The user interface provides access to functions of the system which allow the user to demonstrate the system's capabilities. It provides interfaces to each major component (speech recognizer, language translator, and speech generator). The user interface controls the grammars, dictionaries, and speaker models of the speech recognizer, the source and target language settings of the language translator, and the additional pronunciation dictionaries for the speech generator.

The user interface also controls the processing of text through the system. It receives the text transcriptions from the speech recognition system, and passes the selected transcription into the language translation system. It then receives the translated text from the language translation system and passes that text into the speech generation system.

The user interface is a window/menu driven system, built with the Xview toolkit, which runs on the OpenLook window manager, which runs on top of Xwindows. The implementation language is 'C'.

The user interface is shown in Figure 3.1.

3.2.4 Interface to Speech Recognition Subsystem

The interface between the user interface and the speech recognition system has several aspects. The user interface initializes the Phonetic Engine through Phonetic Decoder Interface (PDI) calls. The user interface manages the speaker model, speech grammar and speech dictionary, also through PDI calls. Finally, once the entire application has been initialized, the user interface starts a loop which monitors encoded speech utterances from the Phonetic Engine. Encoded utterances are passed through the currently active speech grammar, producing the "N-best" speech transcriptions, which are displayed to the user in a menu. In the larger percentage of trials, the correct transcription is at the top of the N-best list, and is the default phrase to be translated and spoken. When the correct transcription is farther down the N-best list, the user can select it for translation and output.

3.2.5 Interface to Language Translation Subsystem

The user interface initializes the language translator on application startup. It also sets the source language in the language translation system, thereby activating language-specific lexicons, frames, and rule sets. During translation, the user interface passes the text selected from the speech recognizer to the language translator, and receives the translated text after translation processing.

3.2.6 Interface to Speech Generation Subsystem

The user interface initializes the speech generator on application startup, and downloads the Spanish-specific pronunciation dictionary to DECtalk. During speech generation, the user interface passes text strings, which have been received from the language translator, to the speech generator.

3.3 Guide to User Interface Functions

This section provides a guide to the controls and functions available in the user interface, and specifies how to use each of them. As mentioned above, the user interface provides access to

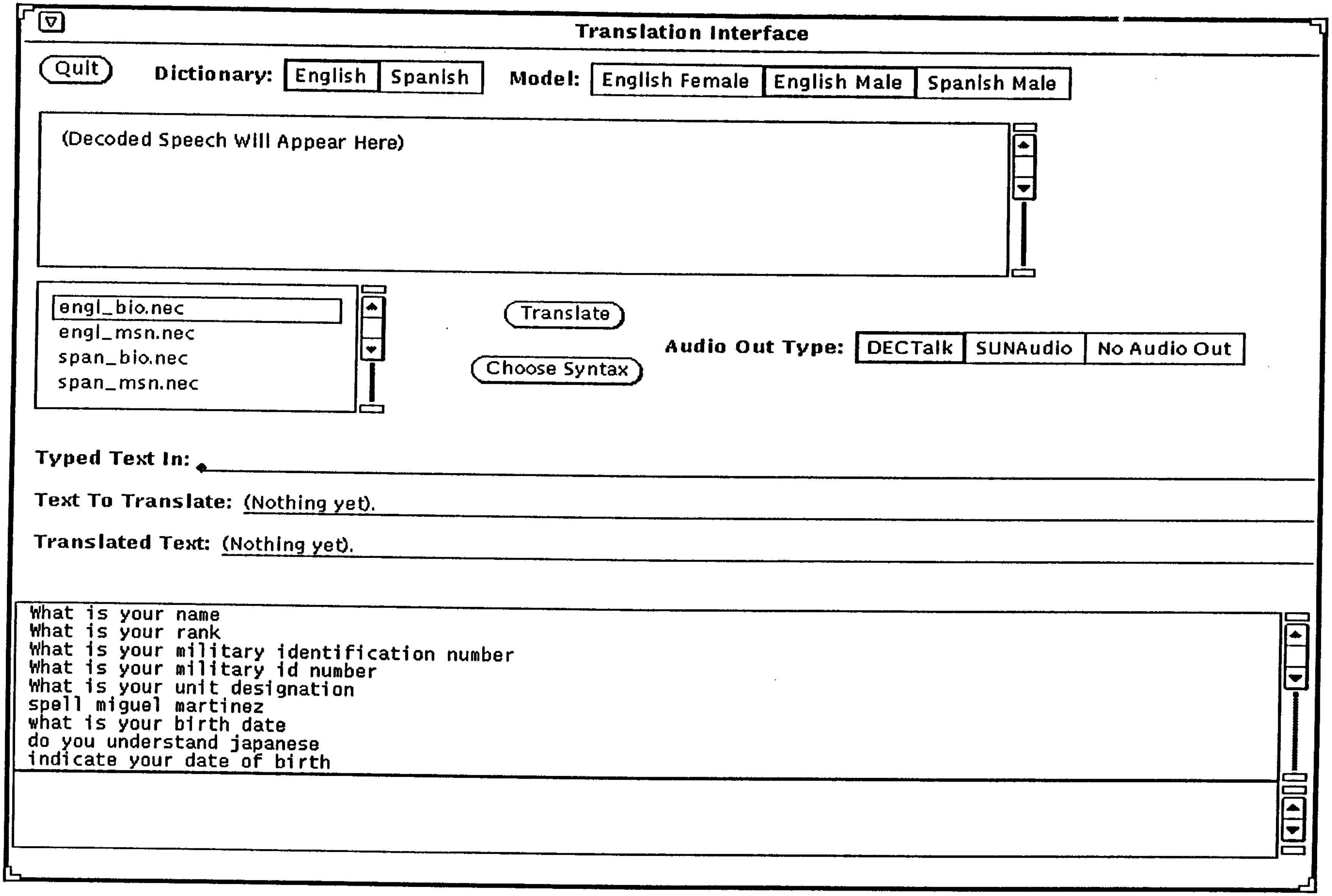


Figure 3.1 MAVT User Interface

the functions needed to demonstrate the MAVT system.

Information on the DS200 speech development system is given in [SSI-DS200]. Information of the DECtalk speech generation system is given in [DECTALK]. Information on OpenWindows is given in [OW-USER].

The three mouse buttons are specified as follows:

SELECT	Left button
ADJUST	Middle button
MENU	Right button

3.3.1 Speech Input

After the MAVT system has been initialized, the speech recognizer is continuously waiting for speech input. As specified in the DS200 Reference Manual from Speech Systems, Inc. [SSI-DS200], the user must press the 'Talk Button' on the Headset Interface Unit during input of a speech utterance.

3.3.2 Changing Languages

The MAVT system accepts speech input in English or Spanish mode. Changing from one input language mode to another is a three-step process:

1. Set speaker model
2. Then, set speech dictionary
3. Finally, set speech grammar

3.3.3 Setting Speaker Model

The speaker models are organized by language (English or Spanish) and by voice-type (male or female). The current version of the user interface provides three speaker models: English Male, English Female, and Spanish Male.

To change the speaker model, position the cursor in the appropriate 'Model:' button (either English Female, English Male, or Spanish Male) and click with the SELECT mouse button.

Whenever a new model is selected, the system will prompt the user to speak: "Testing One Two Three" into the headset. This will recalibrate the microphone gain for the new speaker model.

3.3.4 Setting Speech Dictionary

The speech dictionary may be set to either English or Spanish. To do this, position the cursor over the 'Dictionary:' button for the desired language (either English or Spanish) and click with the SELECT mouse button.

3.3.5 Setting Speech Grammar

There are currently four speech grammars:

engl_bio.nec	English Biographic
engl_msn.nec	English Mission
span_bio.nec	Spanish Biographic
span_msn.nec	Spanish Mission

which are located in the Grammar choice list (to the left of the 'Translate' and 'Choose Syntax' buttons.

To choose a grammar, position the cursor over the name of the grammar in the Grammar choice list, and click with the SELECT mouse button. Then position the cursor over the 'Choose Syntax' button and click with the SELECT mouse button. This will set the new grammar, and will bring up a new list of sample sentences in the 'Sample Text' list at the bottom of the user interface.

3.3.6 The 'Decoded Speech' List

When the speech recognizer processes a speech utterance, it returns the N-best text transcriptions to the user interface, and these are displayed in the 'Decoded Speech' list (which is immediately below the first row of buttons). The most-likely text transcription is at the top of the list, and is highlighted (selected). No further selection is necessary if the user wants to translate the selected text-transcription (see 'Translating Text' below).

3.3.7 Selecting an Alternate Text-Transcription

If the user decides to select an alternate text-transcription (to be translated), that is done by positioning the cursor over the desired transcription in the 'Decoded Speech' list, and clicking the SELECT mouse button. The system will highlight the selected text-transcription.

3.3.8 Typing-in Text for Translation

The user may also directly type-in text to be translated. The desired text is typed into the 'Typed Text In' (Typed-text Input) item.

3.3.9 Translating Text

Once text has been selected for translation, it may be passed into the translation component with the 'Translate' button. To do this, position the cursor in the 'Translate' button, and click with the SELECT mouse button.

After text has been processed by the translation component, the original and translated text are put in the 'Text to Translate' and 'Translated Text' text items.

3.3.10 The 'Sample Text' List

The 'Sample Text' list shows a set of sample sentences which may be used with the current speech grammar. The list can be scrolled to show additional entries with the scrollbar at the right-hand side of the list.

3.3.11 Specifying Audio Output

This button is currently not functioning. In the future, it will be used to specify the type of speech generation, either Text-to-Speech (e.g., DECtalk), Digital Audio Playback (SUNAudio), or no audio output.

3.3.12 Ending the Session

To end the session and exit the program, use the 'Quit' button. Position the cursor in the 'Quit' button and click the SELECT mouse button.

4. SPEECH PROCESSING COMPONENTS

This section is comprised of two subsections, Section 4.1, which describes the operation of the speech recognition system utilized for the MAVT testbed, and Section 4.2, which discusses the speech synthesizer used in the testbed.

4.1 The Speech Recognition Subsystem

In response to a selection and evaluation task specified in the Rome Laboratory Statement of Work, the project team performed an evaluation of state-of-the-art technology in automated speech recognition (ASR). The ASR system selected for the MAVT testbed was SSI'S Phonetic Recognition System. In the selection process, the following dimensions of features/performance of state-of-the-art ASR systems were considered:

- speaker dependence/independence
- vocabulary size
- run-time perplexity
- response time
- hardware unit size
- continuous/isolated word speech
- user adaptability/trainability
- programmability/ease of application development
- accuracy
- channel characteristics
- modularity

Many of these dimensions can take on a range of values, and they also interact. For example, vocabulary size is a discrete variable that can range from very small (< 20 words) up to the very large (> 30,000 words) in currently available systems. Yet with systems that do have large vocabularies, it is rare for all the words to be active at the same time, since response time and accuracy are affected (i.e., the larger the perplexity, the slower the response time and the less accurate the system).

There is no system which is optimal for all applications. All make tradeoffs in the complex feature space. The first step in designing a speech application is to determine which features are absolutely necessary for the application, and which are desirable, but not necessary.

For the MAVT testbed, the selection of the SSI system was based on several key criteria specified in the Rome Laboratory Statement of Work for this effort. In the first place, the requirements of the IPW application imply large vocabulary and continuous speech. Moreover, although the role of the interrogator would allow use of a speaker-dependent system, the SOW requirement for interpreting responses from arbitrary informants dictates a speaker-independent capability. In addition, although some of the DARPA-sponsored ASR developments could satisfy these requirements to some extent at the time this work was begun, only the SSI system was commercially available (another requirement specified in the MAVT solicitation). Another advantageous feature of the SSI ASR was the SSI-provided environment for application development, which was not available for any other system at that level of capability. The software tools in this environment considerably facilitated the MAVT testbed development and made possible the collection of a corpus of spoken Spanish and the development of a preliminary version of a speaker model for Spanish based on that corpus. If the SSI development environment had not been available to us, it is doubtful whether it would have been possible to achieve this goal within the time and resources of the contract.

4.1.1 SSI System Architecture

In the following discussion, an overview of the architecture is first presented, followed by a detailed discussion of the interaction of the system components and subcomponents. Examples of the knowledge sources used by the system are given in Section 4.1.2, which describes the SSI ASR as configured for the MAVT application.

4.1.1.1 Overview The SSI phonetic recognition system is composed of two major components: the Phonetic Engine® (PE), and the Phonetic Decoder™ (PD).

At the front end of the system, the PE translates the speech signal into a sequence of phonetic codes representing the basic sounds or phonemes of a particular language. Figure 4-1 shows a spectrogram of the sound patterns associated with the phonemes representing the utterance "What is your military rank?". In the SSI ASR system, the variant of a given phoneme actually occurring in the utterance is represented by a complex phonetic code (see discussion below). The PE is thus a speech-to-phonetic-code device.

The PD takes as input the output of the PE, and further decodes the phonetic code string into (orthographic) text. The PD is thus a phonetic-code-to-text translator. The whole system is a speech to orthographic text system, as illustrated in Figure 4-2.

The PE consists of specialized hardware and firmware developed and built at SSI. The PD is a C-language software package that resides in a general purpose computer along with application-specific software. The connection between the PE and the PD is a low-speed RS-232 cable, which the computer treats as a terminal line. An application program takes the text output from the PD and responds appropriately.

Figure 4-3 shows a schematic of the information sources used in the recognition system, divided into the on-line recognition system information flow, and off-line knowledge compilation processes. The PE uses a speaker model built off line; given speech, the PE produces phonetic codes.

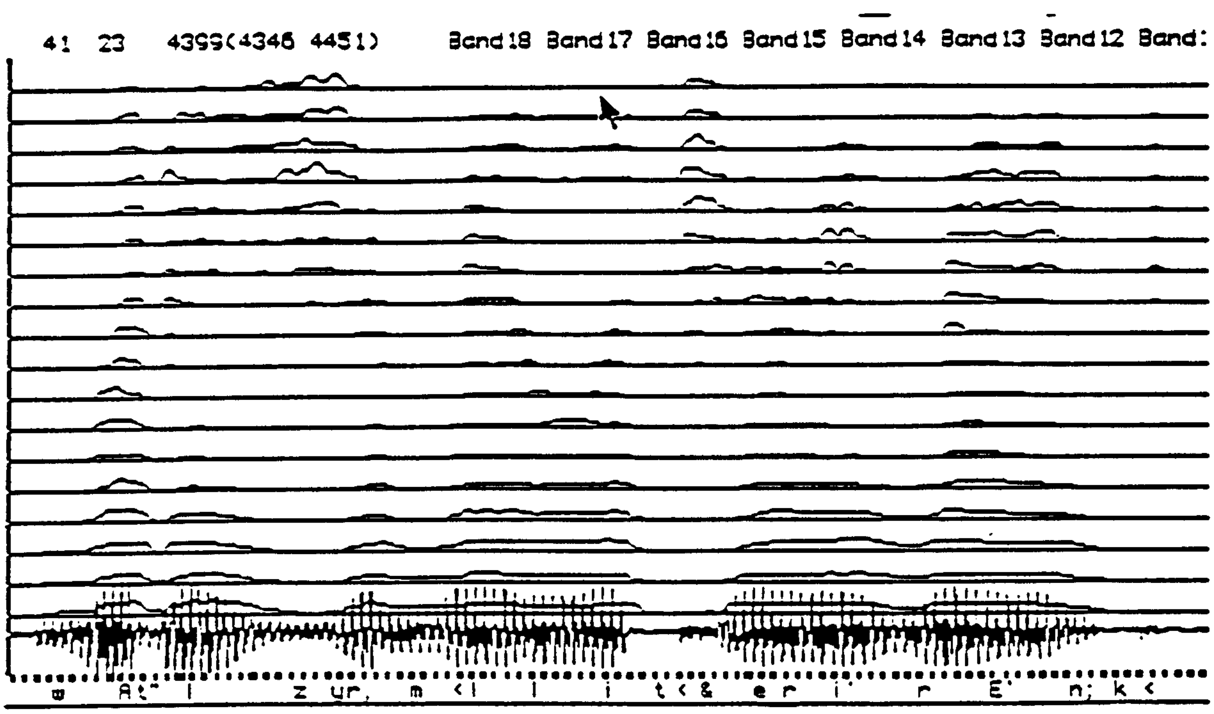
The PD takes the phonetic codes as input in live recognition, and outputs a text string (or a set of the N-best alternative text strings). To do the decoding, the PD uses the following:

- a phonetic codebook, which tells it what the codes mean;
- a syntax, which tells it what word sequences are allowable (see Figure 4-8 in the following section); and
- a phonetic dictionary, which tells it what phoneme sequences correspond to each word (see Figure 4-7 in the following section).

The codebook and the speaker model are matched (however, see the discussion in the following section concerning modification of the speaker model for Spanish). The syntax and dictionary are created off line, and are application-specific.

MACHINE-AIDED VOICE TRANSLATION (MAVT)

ACOUSTIC PATTERN



"WHAT IS YOUR MILITARY RANK?"

Figure 4-1. Sound Spectrogram for Query.

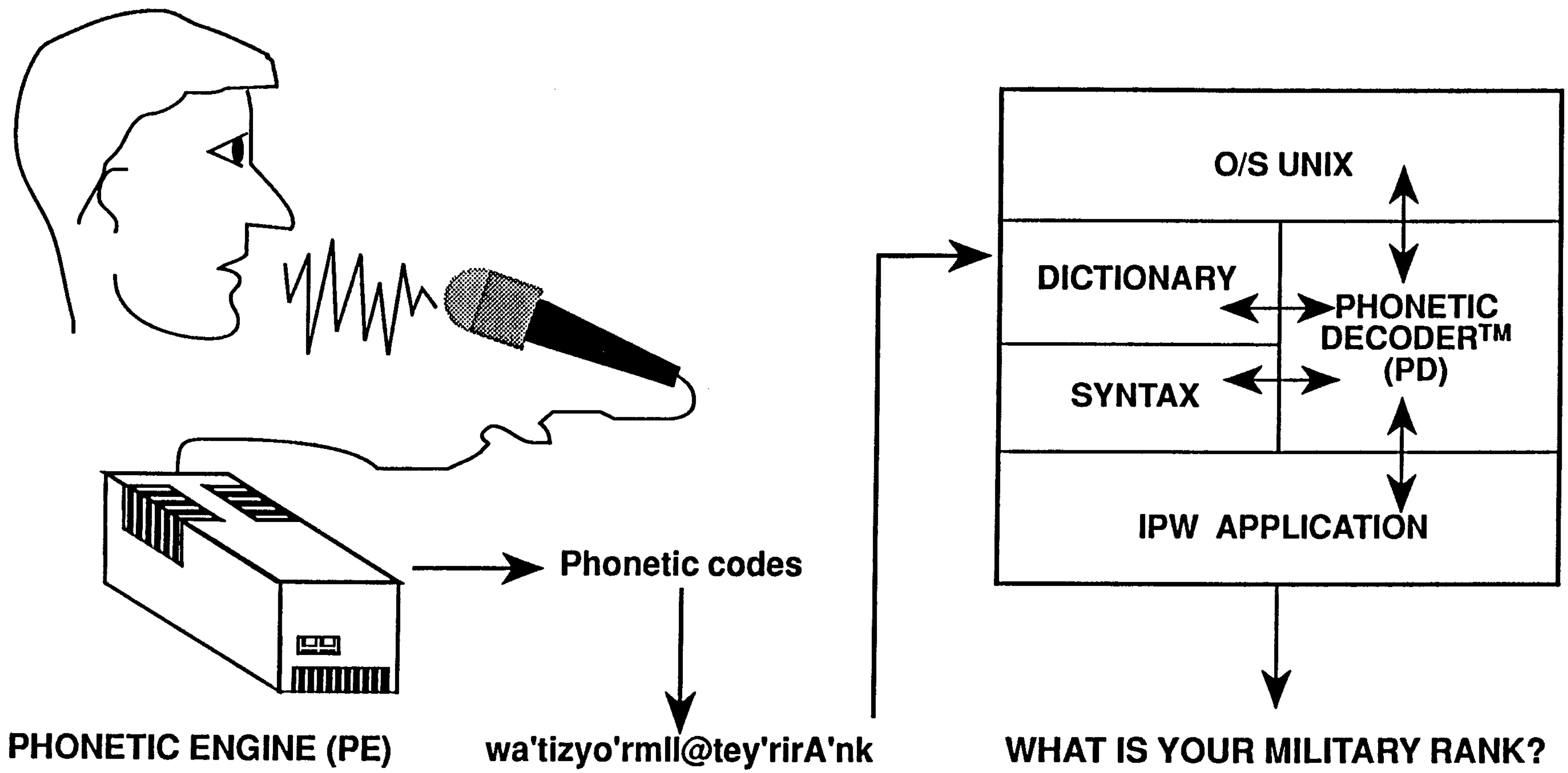


Figure 4-2. Overview of SSI Speech Recognition System

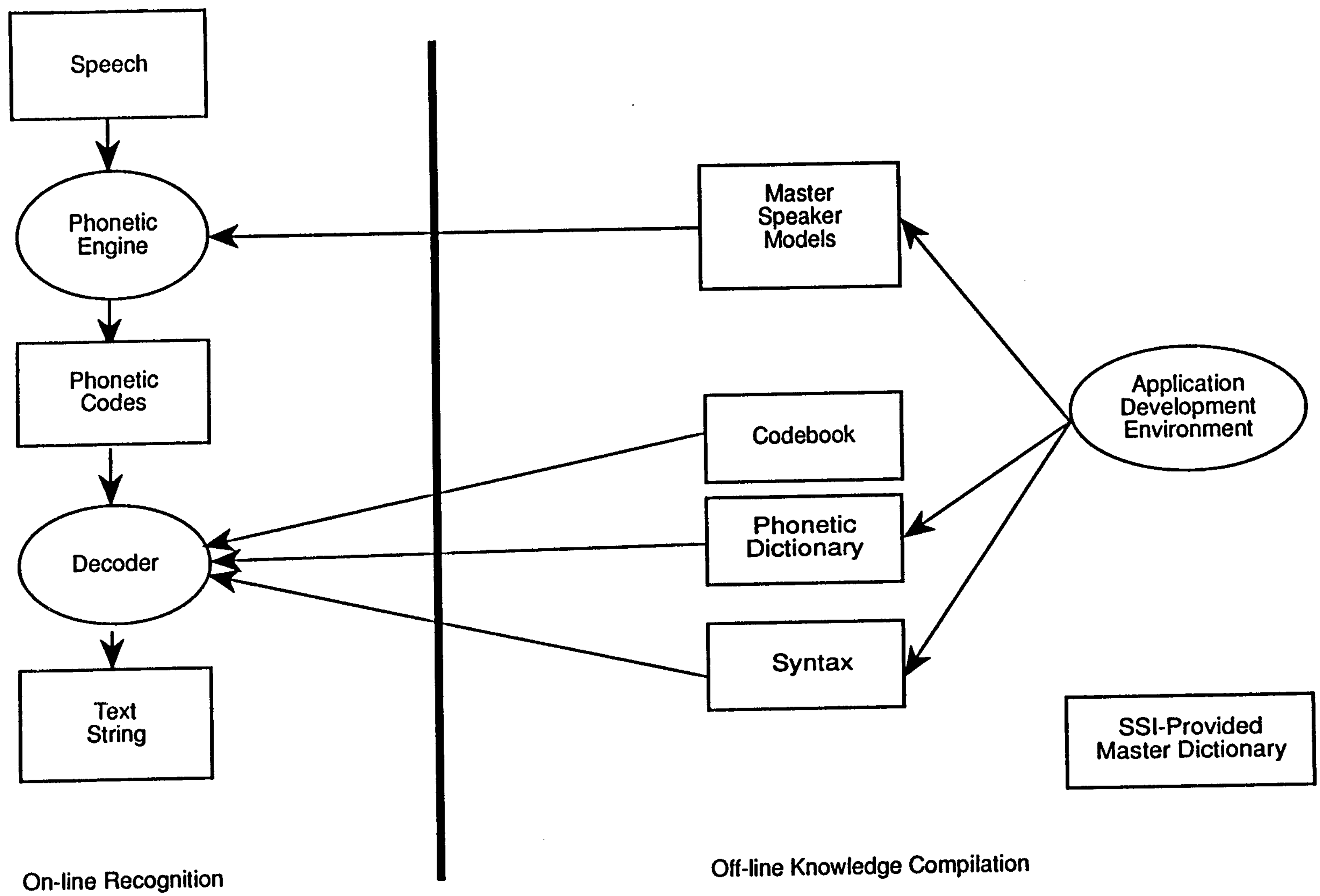


Figure 4-3. Information sources for SSI ASR

4.1.1.2 Recognition Processing

In on-line recognition mode, the PE takes in and digitizes the acoustic data and then segments the data into frames based upon pitch periods. The speaker model then performs an initial classification of each frame into phonetic classes. Runs of frames with the same class are concatenated, forming segments.

After the initial segmentation and classification, the PE applies a series of phonotactic rules to insert and delete segments, based upon limited context. Phonotactic knowledge dictates, for example, that a stop burst follows a stop closure, not a vowel. Similarly, if a very short weak fricative is found between a long vowel and a long strong fricative, it is likely that the weak fricative is actually the low-amplitude start-up portion of the strong fricative, and hence should be collapsed into the fricative. The phonotactic rules, which are loaded into the PE as part of the speaker model, are built by comparing the classification algorithm's segmentation performance to human expert knowledge.

After segmentation, the segments are further categorized into phonetic codes. Each code can be considered a vector of probabilities of phonetic classes. Thus whereas segmentation classes are scalars of one major class out of a small set of classes, phonetic code classes are vectors over a larger set of classes. There must be, in principle, enough phonetic classes to perform all the minimal pair distinctions in the language. Figure 4-4 shows a simplified example of a string of phonetic codes that is passed to the Decoder from the PE. The segments 1 through 4 are segments in time. In practice, each segment is represented by one phonetic code. The phonetic code book tells the Decoder what phonetic class probabilities are associated with what phonetic code, so one can think of the Decoder as being passed a matrix of phonetic class probabilities. In the figure, two paths, one for the word 'purple' and one for the word 'yellow', are shown. In this case, the word 'purple' is more likely.

Figure 4-5 illustrates the stages of phonetic processing in the SSI system, showing the set of processes that are applied along the vertical axis, and the output of each process (which is input to the next process) as a horizontal layer, beginning with the acoustic signal at the top and ending with a phonemic transcription at the bottom.

The back end of the SSI ASR system, or Phonetic Decoder (PD), combines several sources of knowledge to produce English text output. First, the strings of phonetic codes which are the output of the PE inform the PD about the characteristics of the speech signal, and a phonetic codebook helps interpret these codes. A phonetic dictionary provides information about what words and pronunciations are available, and a syntax provides information about the syntactic and semantic constraints of the application. Using these sources of knowledge, (described in the following section), the PD decodes the phonetic code string into English text.

4.1.2 The SSI ASR as Configured for the MAVT Application

The SSI Phonetic Recognition System is provided with two American English speaker models, one for male speakers and one for female speakers. In the initial phase of the MAVT development, only the male model was utilized. As discussed in the preceding section, the model incorporates rules for assigning the frame segments into phonetic classes as well as phonotactic rules for applying contextual knowledge to insert and delete segments.

Since no SSI speaker model existed for Spanish, it was necessary to develop a preliminary model through collection of speech data and application of the speaker profiling software provided in the SSI development environment. As the primary focus of the MAVT effort was

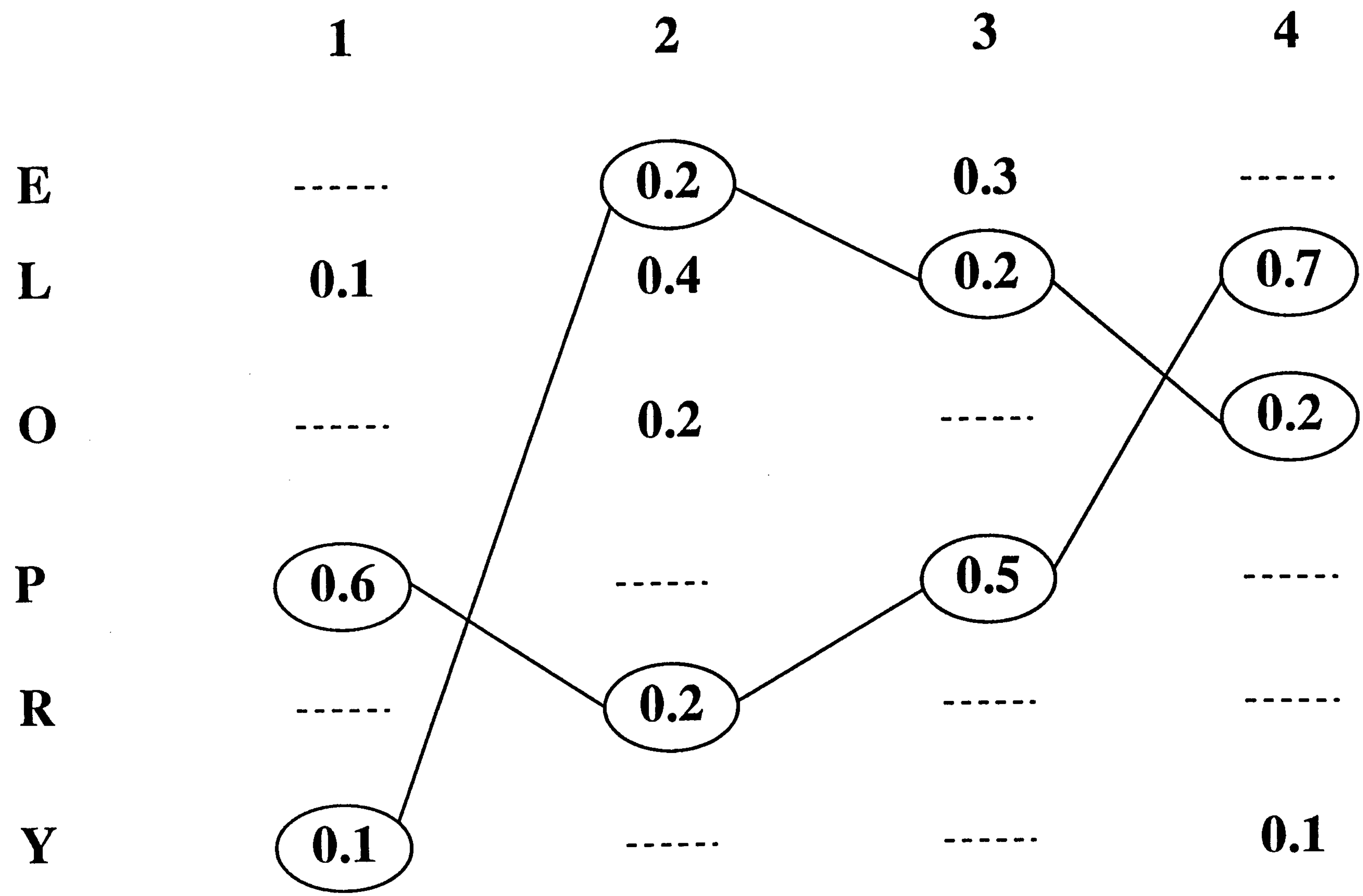


Figure 4-4 Acoustic Matrix Input to Decoder

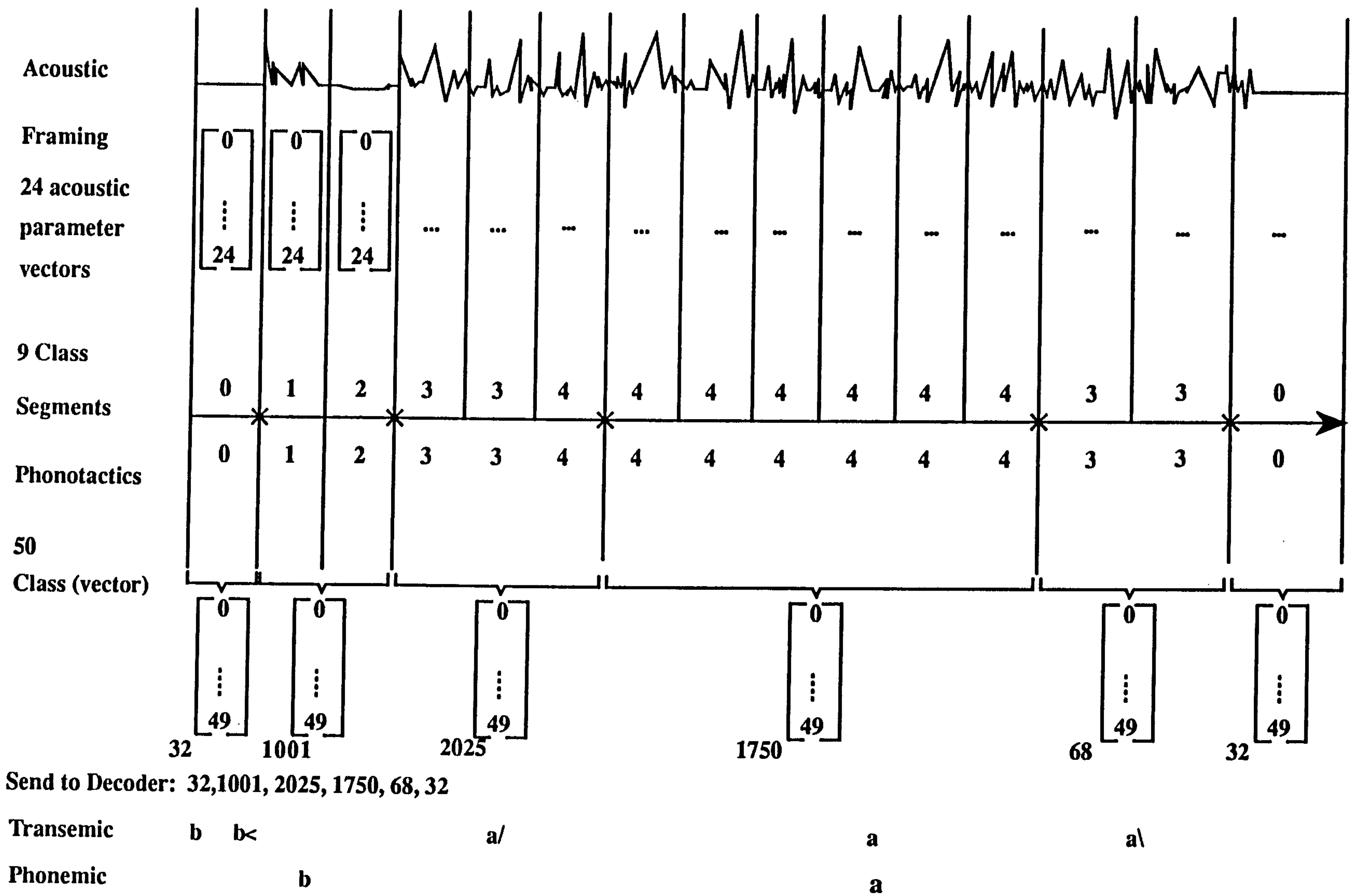


Figure 4-5 Stages of Phonetic Processing in the SSI ASR System

the development of a testbed to evaluate the feasibility of the IPW application -- rather than the construction of a robust speaker model for Spanish -- speech data were collected from only two male speakers from the same dialect area (Lima, Peru) for the construction of the MAVT Spanish speaker models. For the initial model, approximately 640 utterances were collected from speakers 'ara' and 'cmx', while approximately 1000 additional utterances were collected from speaker 'ara' for use in the profiling experiments. Three Spanish speaker models were developed in the course of this experimentation: ara00, the initial model; ara03, which is biased toward the Peruvian dialect; and ara06, the first step toward a more generic (Latin American) model of Spanish.

Figure 4-6 shows an input utterance (labeled "PROMPT TEXT") together with the string of phonetic codes output by the Phonetic Engine (PE) based on a particular speaker model (in this case, the standard English male speaker -- SSI's 3013). To interpret such strings of phonetic codes, the Phonetic Decoder (PD) uses a phonetic dictionary and a syntax, or set of grammar rules.[2] While the phonetic dictionary for English is provided with the SSI system, the phonetic dictionary for recognizing Spanish was developed in the course of the MAVT project. A segment of the Spanish dictionary, currently in its 7th version, is presented in Figure 4-7.

In the course of the project effort, syntaxes used by the PD were defined for the biographics and mission-related information domains for both English and Spanish. A syntax for the PD is specified by a series of replacement, or rewrite rules, as illustrated in Figure 4-8, which shows a segment of the biographics syntax for Spanish. In the syntax, items on the left of the '-->' are replaced by those on the right, starting with the root symbol 'S'. Items on the right separated by ' | ' are alternative choices, and can be delimited by '{}'. Hence '{x | y}' means 'either x or y'. Items delimited by parentheses are optionally deleted. The symbol '==' means the items on the right are all members of the category on the left. The Spanish biographics syntax shown in Figure 4.8 recognizes/generates sentences such as *mi nombre es Carlos Guzma'n, mi apellido es Guzma'n, naci' el once de febrero de mil novecientos cincuenta y dos,*, etc. Figure 4-9 shows the expansion of the rule +RANK_IS in Figure 4-8.

To recognize such sentences, the PD utilizes a heuristic search strategy. The syntax specifies the sequences of allowable words. The PD look-up scores the likelihood of a word by comparing the string of input phonetic codes to the set of allowable pronunciations in the dictionary. Only words allowed by the syntax are considered as candidates. The heuristic search considers only the most likely paths, given the information processed so far in a left-to-right parse. The output of the search is presented in the Decode Log File (*dlf*), as shown in Figure 4-10. A summary of the *dlf* file (i.e., a report of the given set of utterances using a particular speaker model) is presented in Figure 4-11. Finally, branching and other statistics for the given syntax or grammar are presented in the *nec* report, illustrated in Figure 4-12. These last three figures exemplify reports produced by the SSI development software which are utilized to compile the test and evaluation statistics given in Section 6, Tables 6-6 - 6-12.

For a discussion of the current capabilities of the speech recognition component, see Sections 6 (Test and Evaluation), and 7 (System Status).

---

2. In the MAVT testbed, these are distinct from the more powerful rules used in the NLP component for syntactic analysis (see Section 5.3).

```
VERSION 2
SPEAKING TIME   2566
SPEAKER NAME    ara
AGC ATTENUATION VALUE   117
MAXIMUM SIGNAL VALUE    3211
PROMPT TEXT     soy el comandante
EVENT NAME      30
PE ENCODE STATUS        0
NUMBER OF SEGMENTS      20
TRANSEME CODES  240 875 1486 1021 949 1131 502 232 752 1401 405 1473 343 462
                1474 342 9 425 987 246
PCI     mavt_f_s_b_a015 PE200 MPS


Rev 10.215, Copyright (C) 1990 Speech Systems, Inc.
32.10 4/19/91, PDK Version:  35.7 11-April-1991
ara03   a271f2  150792                          395     164050
```

Figure 4-6. Example of Phonetically Encoded Speech (pci) File

PHONETIC DICTIONARY FOR SPANISH LEXICON

ae'rea [a'rI'a']
bateri'a [ba'ta'rI'a']
cabeza [k'a'bE'sa']
cargo [k'A'rgo']
carnet [k'arnE']
carretera [k'ar'atE'ra]
cero [s'E'ro']
cinco [s'I'nko']
clase [k'l'A'se']
comandante [k'o'ma'ndA'nte']
comando [k'o'ma'ndo']
completo [k'o'mpl'Eto']
creo [k'r'E'o']
cua'l [k'w'A'l`]
cuatro [k'w'A'tr'o']
de [d'E']
defensa [d'e'fE'nsa']
defensiva [d'e'fe'nsI'va']
del [d'El`]
derecha [d'e'rE'ca']
desde [d'Esde']
desempena [d'e'sempE'nya']
desplazaba [d'e'spl'a'sA'va'']
desplazan [d'e'spl'a'sA'n`]
desplazando [d'e'spl'a'sA'ndo']
direccio'n [d'i'reksi'O'n`]
dirige [d'i'rI'he']
dos [d'O's`]
es [E's`]
el [El`]
en [En`]
era [E'ra']
esa [E'sa']
estaba [e'stA'va']
estan [e'stA'n`]
fuerzas [f'u'Er`sa's]

Figure 4-7. Spanish Phonetic Dictionary Segment

```
#start Srule

Srule -> ( +NAME_IS ) A_NAME

+NAME_IS -> { (mi nombre) es | me llamo }

Srule -> ( +LAST_NAME_IS ) A_SURNAME

+LAST_NAME_IS ->  { (mi apellido) es | me llamo }

Srule -> ( +RANK_IS ) A_RANK

+RANK_IS -> mi rango (militar) es

Srule -> ( +UNIT_NAME_IS ) A_UNIT

+UNIT_NAME_IS -> { ((el nombre ({completo|entero}) de) mi unidad) es
                 | (mi unidad) se llama }

Srule -> { (mi cargo es) A_FUNCTION
                 | (soy) (el) { A_POSITION | A_FUNCTION }
                 | sirvo como A_FUNCTION }

Srule -> { (1ST_NAME_SPELLED) LAST_NAME_SPELLED (LAST_NAME_SPELLED)
                 | LAST_NAME_SPELLED }

Srule ->  +BIRTH_DATE_IS  el DAYS_OF_MONTH de  MONTHS_ de
                 mil novecientos 30_to_90  (y_sp  SPAN_DIGIT_1to9)

+BIRTH_DATE_IS -> { naci' |  mi fecha de nacimiento es }

Srule ->  +ETHNIC_ORIGIN

Srule ->  +BIRTH_PLACE (la ciudad de)  CITIES_

+BIRTH_PLACE -> naci' en
```

Figure 4.8 Segment of Spanish Biographic Speech Grammar

```
------------------------------------------------------------
|>>>>>>>>>>>>>>>>>>>>>>>>>|general                          |
------------------------------------------------------------|
|mi rango |>>>>>>>>>| es |sargento |>>>>>>                  |
|         |militar  |    |         |mayor                   |
|         |         |    ---------------------------------- |
|         |         |    |teniente |>>>>>>>>>               |
|         |         |    |         |primera                 |
|         |         |    |         |segunda                 |
|         |         |    ---------------------------------- |
|         |         |    |teniente |>>>>>>>>>               |
|         |         |    |         |coronel                 |
|         |         |    |         |general                 |
|         |         |    ---------------------------------- |
|         |         |    |mayor                             |
|         |         |    |coronel                           |
|         |         |    |soldado                           |
|         |         |    |capita'n                          |
|         |         |    |cabo                              |
|         |         |    ---------------------------------- |
|         |         |    |sargento de |primera | clase      |
|         |         |    |            |segunda |            |
|         |         |    |            |tercera |            |
-------------------------------------------------------------
```

Figure 4-9. Expansion of Speech Syntax Rule for Rank.

File Name	mavt_f_s_b_a015.pci
Encode Time	395
Speaking Time	2566
Decode Time	1578
Number of Segments	20
Utterance Score	740
Matching Status	0
Decode Status	1
Tags Found Status	2
Tag Match Status	1
Words Prompted	3
Words Transcribed	2
Words Correct	2
Words Inserted	0
Words Substituted	0
Words Deleted	1

PROMPT		soy	el	comandante	
TRANS	#_	soy		comandante	#_
MATCH	SIL	OK	DEL	OK	SIL
SCORE	834	693		740	881
SEG LEN	1	5		13	1
PROMPT PTAGS					
TRANS PTAGS					
PTAGS MATCH					

* * *

File Name	mavt_f_s_b_a018.pci
Encode Time	129
Speaking Time	2778
Decode Time	2731
Number of Segments	23
Utterance Score	709
Matching Status	1
Decode Status	1
Tags Found Status	2
Tag Match Status	1
Words Prompted	2
Words Transcribed	2
Words Correct	2
Words Inserted	0
Words Substituted	0
Words Deleted	0

PROMPT		jesu's	martinez	
TRANS	#_	jesu's	martinez	#_
MATCH	SIL	OK	OK	SIL
SCORE	834	693	697	881
SEG LEN	1	7	14	1
PROMPT PTAGS				
TRANS PTAGS				
PTAGS MATCH				

Figure 4-10. Segment of the Decode Log File for Spanish Biographics (Speaker Model ara03)

SUMMARY OF DECODE LOG FILE t3_s6.rlf

SYNTAX	span_bio.nec
DICTIONARY	span.phd
SPEAKER MODEL	ara03
MODEL SERIAL	a271f2
DATE/TIME	270892 174141
HOST PLATFORM	SUN-4
SLIDER SETTING	6

UTT ERROR RATE	0.38
WORD ERROR RATE	0.05
PROCESSING TIME RATIO	0.82
UTTS CORRECT	16
UTTS PROMPTED	26
WORDS CORRECT	238
WORDS PROMPTED	250
DECODING TIME	91.70
ENCODING TIME	15.99
SPEAKING TIME	131.89
UNDECODED	0
WORDS TRANSCRIBED	247
WORDS SUBSTITUTED	9
WORDS DELETED	3
WORDS INSERTED	0

Figure 4-11. Summary of Decode Log File for Spanish Biographics (Speaker Model ara03)

Analysis of span_bio.nec.

```
# NODES:  622
# EDGES:  1702
# EDGE LISTS:  294
        AVE # EDGES PER LIST:  2.94
# CATEGORIES:  238
# WORDS:  238

AVERAGE # WORDS IN A LEXICAL CATEGORY:  1.02
MAXIMUM # WORDS IN A CATEGORY:  2

BRANCHING STATISTICS:   Categories         Words
    AVERAGE FOR NEC          3.15             3.19
    SENTENCE INITIAL        94               96
    AVERAGE INTERNAL         2.98             3.01
    MAXIMUM INTERNAL        46               48

GRAMMAR SIZE:     Bytes   Kbytes
    NODES          2492    2.43
    EDGES          2320    2.27
    CATEGORIES      962    0.94
    WORD LIST      1754    1.71
    TOTAL          7528    7.35

# OF TERMINAL NODES:  1
```

LENGTH	# LIN RULES	TOTAL	INCR BR	AVG BR
1	30	30	94.00	94.00
2	288	318	4.49	20.54
3	1392	1710	4.22	12.12
4	1389	3099	1.30	6.94
5	1327	4426	1.38	5.02
6	1266	5692	1.18	3.95
7	37	5729	0.94	3.22
8	59	5788	1.17	2.83
9	2445	8233	2.03	2.73
10	300	8533	1.67	2.60
11	22493	31026	2.93	2.63
12	2006	33032	0.74	2.37
13	15324	48356	1.65	2.30
14	389	48745	1.26	2.20
15	19727	68472	2.24	2.21
16	3951	72423	1.46	2.15
17	2072	74495	1.38	2.10
18	22807	97302	1.27	2.04
19	64108	161410	1.10	1.97
20	84940	246350	0.94	1.90
21	83312	329662	0.98	1.84
22	84714	414376	1.03	1.79
23	65680	480056	0.88	1.74
24	41479	521535	0.80	1.68
25	82946	604481	0.85	1.64
26	103680	708161	0.64	1.58
27	41472	749633	0.29	1.48

Figure 4-12. Statistical Summary for Spanish Biographics Grammar

LENGTH	# SENTENCES	TOTAL	INCR BR	AVG BR
1	30	30	96.00	96.00
2	288	318	4.45	20.66
3	1392	1710	4.25	12.20
4	1416	3126	1.35	7.04
5	1327	4453	1.39	5.09
6	1428	5881	1.25	4.03
7	37	5918	0.98	3.29
8	59	5977	1.18	2.89
9	2931	8908	2.06	2.79
10	300	9208	1.74	2.66
11	31241	40449	3.10	2.70
12	2006	42455	0.77	2.43
13	54690	97145	2.37	2.42
14	414	97559	0.83	2.24
15	19752	117311	2.25	2.24
16	4445	121756	1.53	2.19
17	2566	124322	1.40	2.13
18	31126	155448	1.28	2.07
19	88052	243500	1.10	2.01
20	116734	360234	0.94	1.93
21	114662	474896	0.98	1.87
22	116483	591379	1.03	1.82
23	90093	681472	0.88	1.76
24	57129	738601	0.80	1.71
25	114246	852847	0.85	1.66
26	142805	995652	0.64	1.60
27	57122	1.05e+06	0.29	1.50

Figure 4-12. Statistical Summary for Spanish Biographics Grammar (continued)

4.2 The Speech Synthesis Component

There are two available technologies for speech synthesis: text-to-speech (TTS) and digital audio playback (DAP). Both of these were used in the course of the MAVT experimental development, as discussed below.

4.2.1 Background

TTS starts from ASCII orthographic text. The text is input to a series of routines that generate a phonetic transcription. The phonetic transcription is in turn input to routines which generate synthesizer parameters, which are then sent to a synthesizer for audio output. DAP, on the other hand, starts from the recording of real speech. The signal is digitized and usually compressed and stored on disk. At playback time, the stored data is uncompressed and sent through a digital to audio converter and then output.

The advantage of DAP over TTS is its auditory quality: DAP sounds like the person who donated the speech. TTS synthesizers do not sound like natural human speech. In addition, due to the extra computational power needed for TTS, it is more expensive, in terms of processing power, than DAP. The advantages of TTS over DAP is that in DAP, everything that is to be output must be planned and recorded in advance, and the data storage requirements of TTS are less than that of DAP. It should be noted that the process of digitizing and organizing speech samples for a reasonable application can be an extremely labor-intensive task.

The appropriate technology depends upon the nature of the application. For use of the MAVT device as an automated language trainer, DAP may be superior to TTS, since one important aspect of learning a second language is acquiring the ability to distinguish and reproduce the sounds of that language. The requirement for extensibility to other languages is another possible argument for going with DAP instead of TTS, since the latter requires re-engineering the orthographics, phonology, phonetics and synthesizer parameters of the TTS synthesizer -- i.e., building a new model for each new language added -- an expensive consideration. Moving to another language using DAP only requires recording the outputs in that new language (although this is a labor-intensive process, as noted above). Although TTS devices do exist for languages other than English, there still are few available.

In general, DAP will be superior to TTS for applications where output quality is critical *and* the vocabulary is completely known in advance. For applications where the speech output is not completely determined in advance, TTS is the only solution for speech generation.

4.2.2 Speech Synthesis for the MAVT Testbed

Since there was no overriding criterion for selecting either approach to speech synthesis to the exclusion of the other, both approaches were tried as part of the experimentation performed in the course of the contract.

DAP experimentation was carried out early in the testbed development, using the DAP processor built into the Sun Workstation. Synthesized speech in the MAVT concept videotape (ELIN A005, CLIN 0002) was prepared using the Sun DAP capability. Figure 4-13 shows one of the experiments performed to illustrate the difference between the American English vowel sound in the word "say" (phonetically, a glide [ey], represented in Figure 4-13a) and the Spanish vowel sound in the word "se" (a pure vowel [e], shown in Figure 4-13b).

TTS experimentation was carried out using an available DECtalk synthesizer which duplicated equipment at IRA's Speech Laboratory, allowing contract resources to be devoted to the

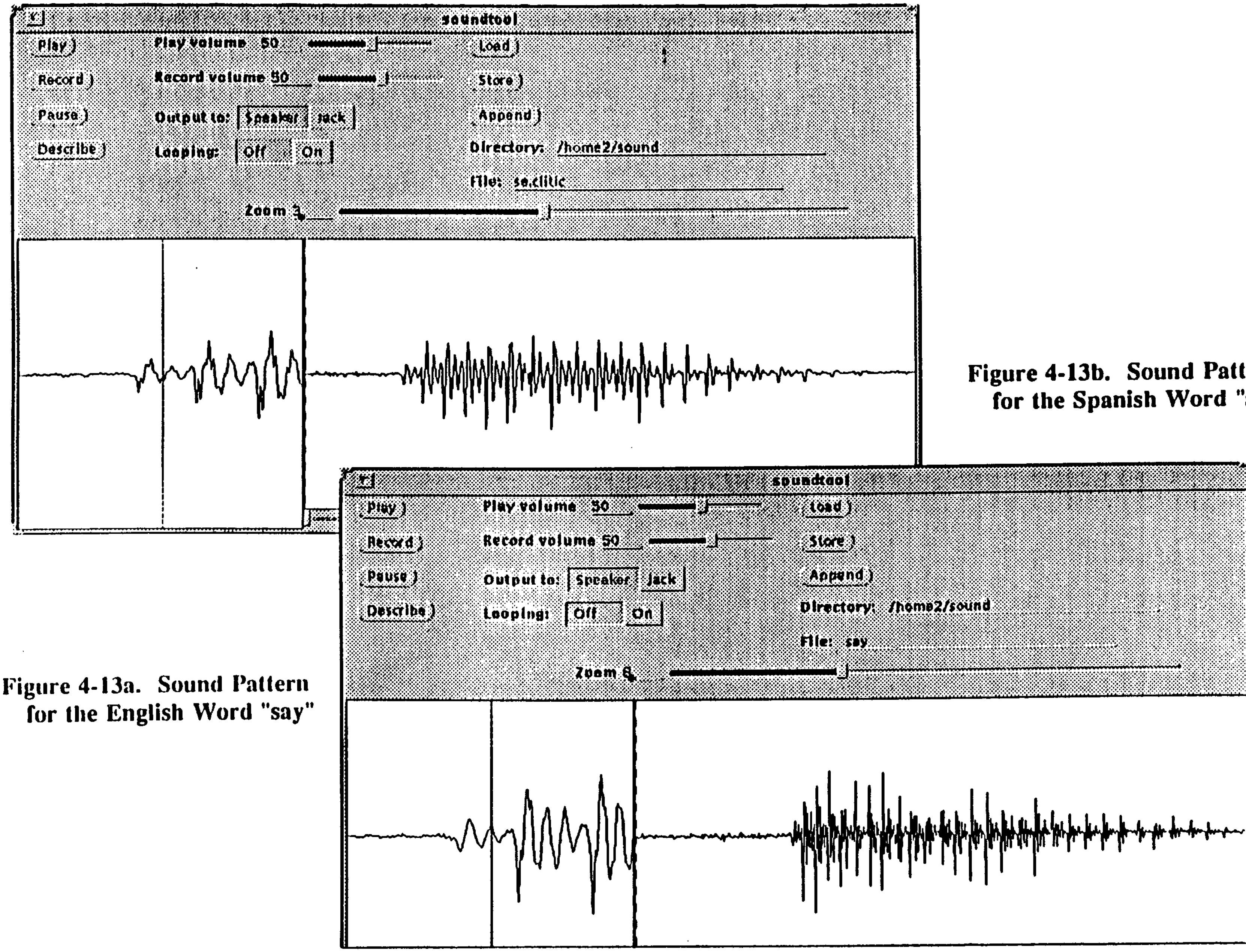


Figure 4-13b. Sound Pattern for the Spanish Word "se"

Figure 4-13a. Sound Pattern for the English Word "say" project effort, rather than the purchase of additional equipment for the testbed version of the MAVT system. As in the case of speech recognition, it was necessary to develop a preliminary speaker model for Spanish via adaptation of the English model. Adaptation rules and symbols are summarized in Table 4-1, while Figure 4-14 shows a segment of the Spanish DECtalk speaker model based on the rules in Table 4-1.

In the context of the MAVT application, DAP is clearly preferable for language learning functions, since the use of DAP would produce an overall utterance quality most closely approximating that of natural speech. Verbatim prerecording of all utterances to be used in a foreign language tutorial would require a substantial amount of effort for a tutorial of any reasonable size. Prerecording phrases which can be combined with other phrases to form viable utterances is a more realistic endeavor, but demands considerable effort in terms of composing suitable intonation contours for generated utterances comprised of two or more phrases (partial utterances).

Even if the DAP component constructed for language tutorials is large enough to cover most of the anticipated interrogation dialog within the MAVT context, the problem of handling previously unencountered items entered via the keyboard still exists, necessitating some TTS capability.

Description	Symbol	Sample Word
intervocalic "s"	ss	cabeza
intervocalic "b" or "v" before unstressed vow.	v	avanzaba
intervocalic "b/v" before stressed vowel.	b	cabeza
word-initial "b/v"	b	bien, vehi'culos
"n" between vowel and "s", "t" or "d"	nnn	central, avanzaba
"r"-flap	dx	era
"r"-flap before consonant	dxix	cuarto
"r"-flap after consonant	ixdx	cuatro
"rr" trill	dxixdx	carretera
intervocalic "t" before unstressed vowel (prevents flapping)	tt	completo
"t" after n & before unstressed vowel (prevents "nt" reduction to "n")	tt	identidad
"d" after n & before unstressed vowel	dd	reubicando
"ch" always (sounds better than [ch])	tsh	derecha
"a" in word final position	ax	ha'cia
"g" before "i" & "e"; "j" always	hxhx	dirige, jefe
"i" before "l" (prevents vowel reduction)	iyyx	militar
"e" in syllables which do not have primary or secondary stress	eh	carretera
"e" before consonants that cause lowering (mostly voiced alveolar consonants)	eh	el, era tienen

Table 4-1. Rules for Representing Spanish Sounds in DECTALK Synthesizer

```
a'rabe   'aadxaavey
a_       'aa
abril    aavixdx'iyyxl
ae'rea   aa'ehdxiyax
agosto   aag'owsstow
al       aal
alema'n  `aalehm'aan
alemana  `aalehm'aanax
alemanes         `aalehm'aaneys
amador   `aamaadh'owdx
americanas       aam`eydxiyyxk'aanaas
americanos       aam`eydxiyyxk'aanows
aniquilar        aann`iyyxkiyyxl'aadx
apellido         `aapehyx'iydhow
atravesar        aattixdx`aaveyss'aadx
avanzaba         `aavannnss'aavax
avanzaban        `aavannnss'aavaan
avanzada         `aavaannnss'aadhax
avanzando        `aavaannnss'aannnddow
b_       b'ey
bateri'a         b`aatteydx'iyyxax
batista  baat'iysstax
bien     byx'ehn
buscar   buwssg'aadx
c_       sey
ca'rdenas        k'aadxixdh`eynnaas
cabeza   kaab'eyssax
camagu:ey        kaam'aagwey
camino   kaam'iynow
campos   k'aammmpows
capita'n         k`aapiyt'aan
capturar         k`aaptuwdx'aadx
capturarlo       k`aaptuwdx'aadxixlow
capturarlos      k`aaptuwdx'aadxixlows
cargo    k'aadxixgow
carlos   k'aadxixlows
carnet   k`aadxixn'eytt
carretera        k`aadxixdxeht'eydxax
```

Figure 4-14. Lexicon Segment for DECtalk Spanish Speaker Model

5. NATURAL LANGUAGE PROCESSING COMPONENTS

In the MAVT testbed, the output from the speech recognition component is a sentence or phrase of written text in the source language. This text serves as input to the natural language processing (NLP) component, which translates the written text into the appropriate written text in the target language. Alternatively, the user may bypass the speech recognition system and type in a sentence for the NLP component to translate. In either case, the text output of the NLP component is then passed on to the speech synthesizer, which produces a spoken utterance in the target language. The flow of processing among the speech recognition, NLP, and speech synthesis components of the MAVT testbed is illustrated in Figures 1-1 and 1-2. In this section, we discuss the major internal modules of the NLP component.

5.1 Multilingual Lexicon

LSI's NLP system consists of a series of modules that process message text in stages. Each major level of analysis is contained in a separate module, as shown in Figure 5.1. Processing is performed sequentially: the output of each module is a temporary data structure that serves as input to the succeeding module and is then available to all later modules. Each individual module contains a processing mechanism and a knowledge base (rule set). The knowledge bases allow the incorporation of general linguistic knowledge as well as domain-sensitive (in the case of MAVT, containing information specific to the military domain or to the interrogation scenario) and language-sensitive (e.g., Spanish or English) features. The lexicon is one of the core knowledge bases.

During processing, for each sentence the words and multi-word phrases are matched with the lexical definitions of items in the lexicon, yielding a lexicalization for the entire sentence. If the input has been derived from the speech recognition component, then the input is assumed to be well-formed, because all possible output is specified by the speech recognition grammar (see Section 4.1). In the case of sentences typed into the system, however, a typographical error may occur, or part of the input may be unknown to the system. In these cases, the item may be derived by means of Unexpected Input processing by either the Lexical Unexpected input module (LUX), which corrects errors by allowing partial matches between words in the text relating to typographical errors, or by the on-line Word Acquisition Module (WAM1) which allows preliminary classification of new or unidentified material by the user by means of menu selection. Alternatively, the WAM system can operate in an autonomous mode, wherein a word class is assigned based on the system's morphological analysis of the word. The new words can also be stored for later incorporation into the system by means of a second, more extensive mode of the Word Acquisition Module (WAM2), which operates off-line to allow periodic lexicon update by the System Administrator. An example of a lexicalization for the sentence "What is your unit designation?" is shown in Figure 5.2.

Each entry in the lexicon contains morphological information concerning any irregularities in form, morphosyntactic features pertaining to reference and agreement, subcategorization features, selectional restrictions, and links into the frame subsystem (the semantic hierarchy), as described further below. The output of this stage of processing is the lexicalization data structure, which is then passed to the syntactic parser.

For the MAVT testbed, we developed individual lexicons for both English and Spanish in a generic form useful to language processing and translation/generation, and applicable to other languages as well.

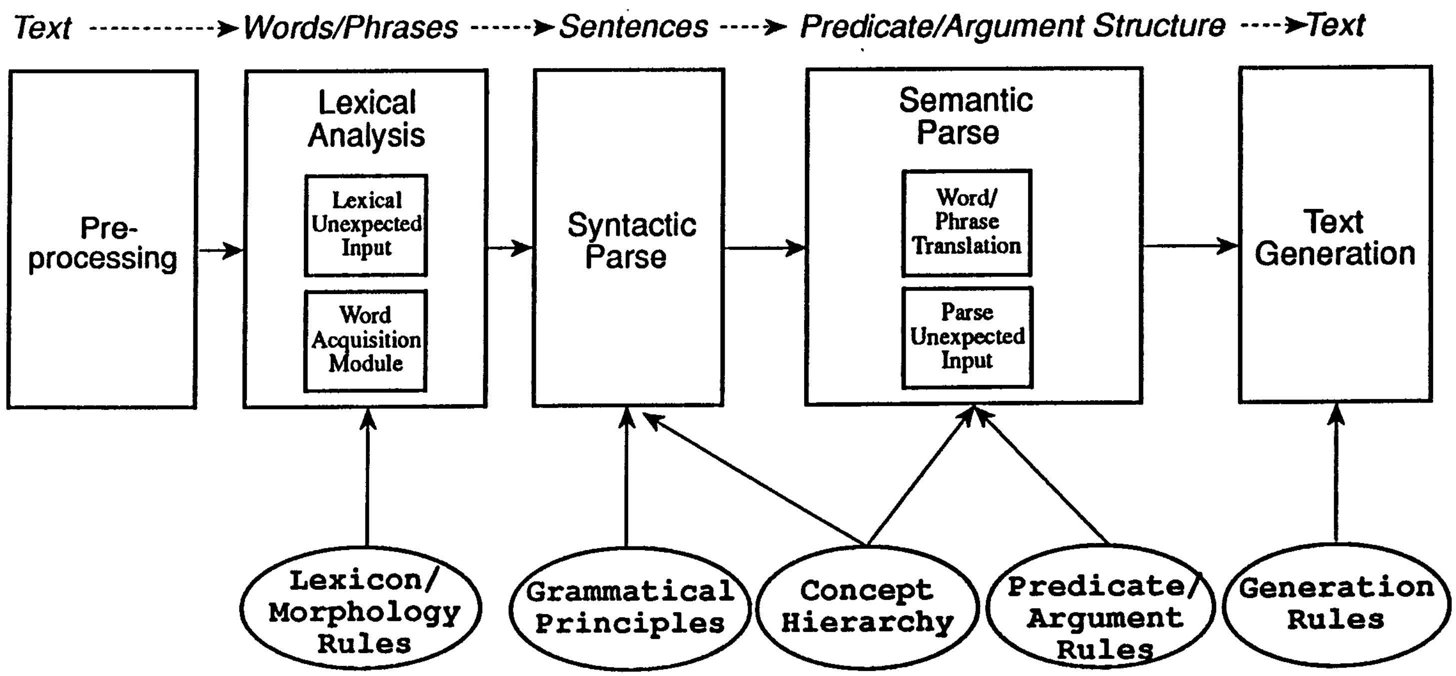


Figure 5.1. MAVT Natural Language Translation

```
what is your unit designation

There are no 'Header' instantiations at this time.

   Transmission 0   Paragraph 1

 1     what,is,your,unit,designation

   Transmission 0   Paragraph 1   Sentence 1

 1     lxi(det,what,what,[wh],[],[],[],[opt(qp,ap,np)],[],[],[what])
       lxi(pronoun,what,what,[wh],[],[],[],[none],[],[],[what])
 2     lxi(aux,is,is,[cont],[],[],[],[xp('-agr','-past')],['+agr','-past'],[],
           [is])
       lxi(aux,is,is,[passive],[],[],[],[xp('-agr','+past')],['+agr','-past'],
           [],[is])
       lxi(third_pres,is,be,[],[],[],[],[strict(pred(adj,np),pp)],
           ['+agr','-past'],[],[be])
 3     lxi(det,your,your,[],[],[],[],[strict(qp,ap,np)],[],[],[pospro])
 4     lxi(noun,unit,unit,[],[],[],[],[],[],[],[military_unit])
 5     lxi(noun,designation,designation,[],[],[],[],[],[],[],[abstract_object])


   Transmission 0   Paragraph 1   Sentence 1

'Cmax1':
Cmax(Dmax+3(Dbar(D([what]:pronoun))),
     Cbar(C+1+2([is]:third_pres),
          Imax(Dmax(Dbar(D([your]:det),
                         Nmax(Nbar(N([unit]:noun),
                                   Nmax(Nbar(N([designation]:noun))))))),
               Ibar(I+2(*empty*),
                    Vmax(Vbar(V+1(*empty*),
                              Dmax+3(Dbar(D(*empty*))))))))).
```

Figure 5.2 Lexicalization and Parse Output for English Sentence

5.1.1 Current Form and Content of the Lexicons

LSI's MAVT English and Spanish lexicons contain a number of different kinds of information. First, each individual entry contains the spelling of the item and any alternative spellings and alternate names or symbols, such as acronyms.

The lexicons also contain all of the morphological information necessary to derive the appropriate morphological inflectional features from an analysis of the actual word. (For a more complete description of the morphological analysis, see Section 5.2.) The lexicons contain two main types of morphological features: referential features and agreement features. Referential features are those features that are properly or inherently part of the meaning of the item, whereas agreement features indicate the kinds of modifications the item must undergo to be used correctly in a sentence. For example, 'nuestro', "our" has the referential features first person plural. At the same time, the word 'nuestro' formally agrees in gender and number with the noun it modifies. The forms are masculine singular, 'nuestro'; feminine singular, 'nuestra'; masculine plural, 'nuestros'; and feminine plural, 'nuestras'. A sample of the kinds of morphological features that are specified in the lexicon or can be derived from lexical information combined with morphological processing is given below. These features may be referential features or agreement features, depending on what the feature applies to. For example, gender is referential for nouns in Spanish, but is an agreement feature for modifiers, such as adjectives and demonstratives. These features are applicable to a wide variety of the world's languages.

tense:	present, past, future, conditional, imperfect, preterite
mood:	indicative, subjunctive, imperative, jussive
person:	1, 2, 3, 4
number:	singular, plural, dual, trial, inclusive_plural, exclusive_plural
gender:	masculine, feminine, neuter
case:	nominative, genitive, dative, accusative, ablative, vocative (Latin: nouns, adjectives, pronouns, demonstratives)
	nominative, genitive, dative, accusative, instrumental, prepositional (Russian: nouns, adjectives, pronouns)
	nominative, accusative, genitive (Modern Standard Arabic: nouns, adjectives, pronouns)
	nominative, possessive, objective (English: pronouns)
	nominative, possessive, prepositional (Spanish: pronouns)
	dative, accusative (Spanish: clitics)
class:	human, nonhuman; other semantically-based classes (e.g., long objects)
level:	formal, nonformal, semiformal

As described in detail in Section 5.2, the lexicon also contains information as to semi-regular patterns or unpredictable morphological information, including the spellings of any irregular stems.

Each lexical entry must also include at least one link into the semantic hierarchy. This hierarchy, which we call the frame system, is a set of linked concepts in the form of a hierarchical tree structure with the more general categories at the higher levels. At the top of the tree is the most general node, '*thing*'. The point at which a lexical entry is linked into the hierarchy tells how the meaning of the entry is related to the meanings of other concepts to which are associated other lexical entries. Furthermore, an important property of LSI's frame system is the ability to inherit semantic properties from higher up in the hierarchy. For example, the word "gun" has the lexical feature "isa('*weapon*')", which means that the concept that this word represents has all of the properties associated with the concept of 'weapon' (it is an instrument that can injure or kill a person, and so on). Also, '*weapon*' has the feature "isa('*equipment*')" and '*equipment*' in turn has the feature "isa(constructed_object)," so the properties associated with both of these nodes (*equipment* and constructed_object) are inherited by the concept "gun" as well.

The property of inheritance is an important and useful one because it can provide additional information about a word that may not be explicit in the message. The frame hierarchy also provides a single, unified repository for all semantic information and relations that are used during processing. This makes it possible not only to characterize general concepts in detail, but also to add domain-specific information (marked as such) to the frame system. Additional information, such as the organization of a particular country's military, can easily be incorporated into the frame system. A lexical item may have more than one link into the frame system, that is, the item may represent more than one concept. For example, "post" can refer to a physical object, a location, or a position to which a person is appointed. Each of these meanings is represented by a separate link into the frame system.

Syntactic properties of lexical items are also indicated in the lexicon. One of these is subcategorization. This specifies the syntactic categories of the items that either optionally or obligatorily occur with the lexical item. For example, some verbs like 'find', must have a noun phrase object (one can say "he found it" but not simply "he found"), whereas verbs like 'attack' (as in "he attacked") can occur with or without a direct object. Some other verbs like 'go' never have a direct object. This information is crucial for the syntactic parser in helping it to assign correct verb-argument relations in the parse tree for the noun phrases of the sentence. Knowing these relations is also an important key to interpreting the meaning of the sentence and translating it. In another example, a few adjectives like 'previous' can occur only followed by a noun, and not predicatively, like 'new' (e.g., "The show was new" but not *"The show was previous"). The adjective 'previous', then, strictly (that is, obligatorily) subcategorizes for a noun, whereas for most adjectives, having a following noun in optional.

Selectional restrictions are another type of property specified in the lexicon. Selectional restrictions state the semantic category limits on the items for which an entry subcategorizes. For example, the verb 'kidnap' strictly subcategorizes for a direct object noun phrase. That noun phrase can only be a person, or possibly an animal. In terms of the frame hierarchy, the patient of 'kidnap' can only be something that is linked to the hierarchy to a node that has as a direct ancestor the node 'animate_object.' Certain prepositions, too, can be restricted as to the category of the noun phrase following (e.g., 'aboard' can only be followed by a vehicle). This kind of information is also extremely useful in building the parse tree for a sentence because it limits possibilities and provides a means for checking whether a given relation is correct (i.e., does the argument fall into the proper semantic category to qualify?). Selectional restrictions also provide a further means for distinguishing one meaning of a word from another. For example, the verb 'kill' has a different meaning when applied to 'time' than it does when applied to an animate object. These distinctions can be very helpful, particularly in translation.

In addition to these properties, there are others that we are currently in the process of including in the lexicons. One of these is aspectual type, referring to states, processes, or actions. These are now distinguished by means of general categories within the frame hierarchy.

However, more detailed information about the implications of temporal meaning for particular words has yet to be incorporated. We are also investigating means to include other kinds of information that make up the meaning of particular words, such as causality and intentionality for verbs. All of these kinds of information will help in the translation and generation processes in future versions of the system.

5.1.2 Lexicon Expansion

In this project, expanding the number of entries in the MAVT lexicons has not been a focus of development (see Section 7 for further discussion of the current status of the testbed). Rather, we have concentrated our efforts on constraining the interrogation scenario as specified in the SOW, so that we could focus on the system integration effort required to insure lexical and grammatical compatibility between the NLP subsystem, the ASR subsystem, and the speech generation subsystem. Because the ability of the recognizer to produce accurate output decreases markedly as the number of lexical items in the speech grammar goes up, the definition of the set of lexical items that ultimately should be incorporated into the speech recognizer grammar requires experimentation and careful consideration.

In order to help maintain the speech recognizer output parameters within usable ranges and yet still have available to it an adequate set of lexical items, we have divided the speech grammars and lexicons for English and Spanish into biographics and mission subgrammars and sublexicons, as discussed in Section 4.1. This kind of division corresponds well with topic-oriented aspects of the interrogation framework that interrogators are trained to use. We have also written a more general "chatter" grammar and lexicon, but the speech recognizer output for these, as expected, is not as good, and the chatter grammar has not been integrated into the current testbed. During the past year, in conjunction with another project, LSI has greatly expanded its general English lexicon, to about 30,000 items. We expect to be able to draw from this larger lexicon as we expand the MAVT lexicons. We also expect that feedback from potential users will be crucial in deciding what range of vocabulary is actually workable and necessary for them to perform their tasks adequately.

5.2 Multilingual Morphology

Morphological analysis is performed during the lexicalization stage of processing, as described in the previous section, by using sets of stem, affix, and clitic tables, which are described in detail in this section.

The morphological component of the DBG system for MAVT is truly multilingual. It incorporates the capability for handling all of the morphology of English and Spanish, and could easily be extended to analyze the morphology of many other languages.

English and Spanish are alike in that both have primarily suffixal verbal morphologies. However, English has very little inflectional morphology, whereas Spanish has a rich morphological system. Most of the inflectional morphology of English is quite regular, although there are some irregular past tense verbs and past participles and a few irregular noun plurals. Spanish, on the other hand, has a complex verbal morphology as well as a system of pronominal clitics.

Spanish, like other Romance languages (e.g., French, Italian, Portuguese), is highly inflected, particularly in the verbal system. There are a number of unpredictable irregularities in Spanish morphology, as well as several different systematic types of irregularity (what we have called "semi-regular" forms). All of these need to be encoded into the system. A single verb in

Spanish has over sixty different possible forms, taking into account variations for tense, person, and number. The system needs to be able to recognize all of these forms as belonging to the appropriate verb and to identify the correct inflectional features (that is, tense, person, and number) of the form. The morphological processing should furthermore be flexible enough to be extensible to languages with even more complex morphology.

The irregularities that are found among Spanish verbal stems are of three main types, and are similar to the types of irregularity found in English and in other languages. They can be categorized as: 1) basic irregularity, where the form or stem is totally unpredictable and must be given in the lexicon; 2) patterned irregularity, where internal morphological change is predictable by feature or by the shape of the stem; and 3) orthographic morphotactic variation that is predictable based on the shape of the form and is conditioned by certain stem-affix combinations. These three types of irregularity are found in English also. To illustrate, an example of 1) is the verb "go" in English, the past tense of which is "went" and the past participle of which is "gone." These verbal stems are entirely unpredictable and must be given in the lexicon. The lexical representation of morphological features is described in the previous section, 5.1, on the lexicon. In contrast, verbs like "ring, rang, rung" and "sing, sang, sung" can be said to vary according to a pattern of ablaut, or vowel variation. However, because such patterns affect only a few verbs in English, they are usually treated simply as irregular stems. Patterned irregularity is much more common in Spanish than in English and each pattern generally affects a greater number of verbs. The third type of irregularity can be shown by the English changes in spelling when /-ing/ is added to a verb ending in /-Ce/, where C is a consonant. In those cases the /-e/ drops out at the morpheme boundary, as in "moving", from /move/ + /-ing/.

Affixes are also more complex in Spanish and the other Romance languages than they are in English. In English there is one main person/number inflectional ending, which is the /-s/ in third person singular present tense verbs (e.g., "I run" but "he runs"). In Spanish, on the other hand, there are six different person/number combinations (1st, 2nd, and 3rd persons, singular and plural), each with its own ending. Furthermore, these suffixes vary according to which of three conjugations the verb belongs to and which of eight tenses is being conjugated. In addition, in Spanish the infinitive and present participle, although they are not inflected verbs, can have one or two clitic pronouns attached.

In a translation system such as MAVT, it is necessary both to analyze and to generate stem and affix combinations. To list, in order to do a simple match, all of the inflected forms and the forms with clitics attached individually in the lexicon would be extremely inefficient and produce an enormous and unwieldy lexicon. Therefore, we designed the morphological processing mechanism to take advantage of the productive nature of the morphology.

The way that we analyze a verbal form is to match as much of the actual form in the text as closely as possible with the set of all possible verbal stems in a stem table. After the best stem match has been made, the remaining part of the form is assumed to be an ending and is matched with the entries in the affix table. If a match is found, then the stem entry is checked to verify whether that particular suffix is appropriate for the tense and conjugation of the stem. If so, a further check is then performed to ensure that the hypothesized stem/affix combination is an actual allowable form and that there is no blocking index indicating that there is an irregular form that supersedes it. Processing of the Spanish verbal form 'dara'', "he will give" as found in the following sentence is illustrated below:

El sargento me dara' la carta. 'The sergeant will give me the map.' (1)

dara':

STEM TABLE
stem: dar-
ending: -a'
verb class: -ar
conj. index: reg
blocking index: [irrpres, irrsjvpres],
ending type: r

ENDING TABLE
conj.features: ind, fut, 3, s

CONJUGATION INDEX TABLE

BLOCKING INDEX TABLE
conj. features ok.

These results are derived from the lexical entry and the stem and affix tables shown below.

SAMPLE VERB ENTRY IN LEXICON

```
verb(dar, eng(give),
    mstem(present(doy, das, da, damos, dais, dan),
    preterite(di, diste, dio, dimos, disteis, dieron),
    pres_subjunctive('de''', des, 'de''', demos, deis, den))).
```

STEM TABLE

```
[. . .]

dar

[---]
stem_table([100,97,114|_7625],_7625,dar,dar,ar,#,[],c)
stem_table([100,97,110,100,111|_7716],_7716,dando,dar,ar,#,[],c)
stem_table([100,97,114|_7765],_7765,dar,dar,ar,inf,[],r)
stem_table([100|_15724],_15724,d,dar,ar,reg,[irrpres,irrsjvpres],r)
[. . .]
```

AFFIX TABLE

```
[. . .]
end_table(r, "e'", ar, ind, fut, 1, sg, '#', '#').
end_table(r, "e'", er, ind, fut, 1, sg, '#', '#').
end_table(r, "e'", ir, ind, fut, 1, sg, '#', '#').
end_table(r, "a's", ar, ind, fut, 2, sg, '#', '#').
end_table(r, "a's", er, ind, fut, 2, sg, '#', '#').
end_table(r, "a's", ir, ind, fut, 2, sg, '#', '#').
end_table(r, "a'", ar, ind, fut, 3, sg, '#', '#').
end_table(r, "a'", er, ind, fut, 3, sg, '#', '#').
end_table(r, "a'", ir, ind, fut, 3, sg, '#', '#').
end_table(r, "emos", ar, ind, fut, 1, pl, '#', '#').
end_table(r, "emos", er, ind, fut, 1, pl, '#', '#').
[. . .]
```

Finally, the information about the tense, person, and number of the form that is derived from the tables is passed back with the analyzed form as part of the lexicalization, where subsequent processing modules can have access to it.

```
Transmission 0    Paragraph 1    Sentence 1

4    lxi(tensed_verb,'dara''',dar,[],[m(ind),t(fut),p(3),n(s),g(#)],[],[],['+agr','-past'],[],[dar])
```

In generation for a given verb, the stem of the appropriate tense is linked to the ending having the appropriate person and number features allowable for the tense and conjugation of the verb.

In addition to inflectional endings, Spanish has a system of "clitic pronouns," that is, pronouns that can be suffixed to infinitives and present participles (for further discussion of the functions of Spanish clitics and how they compare to English pronouns, see Section 5.3.1.4). Clitic attachment to infinitives and present participles is handled in much the same way as inflection, although it is more straightforward. Once the stem match is made, the possible clitic material is then matched with the first clitic table. If a match is found, then any leftover material is matched with the entries in the second clitic table. The stem and clitics are then represented as independent items in the lexicalization, with the features derived from the tables available for subsequent stages of processing.

An example using the present participle of the verb 'dar' "to give" is given below:

El sargento esta' da'ndomela. "The sergeant is giving it to me." (3)

da'ndomela:

STEM TABLE
stem: dando-
ending: -mela'
verb class: -ar
ending type: c

CLITIC TABLES
first clitic: me
1, s, i, ...
second clitic la
3, s, f, ...

The stem table is the same as that given for 'dar' in the previous example. The clitic tables, from which the clitic analysis is derived, are given below.

CLITIC TABLES

A. first_clitic(Clitics, Clitic1, Clitic2Char, Person, Number, Gender, Case, Level, Humanness)

Clitics - char list containing clitic endings of a verb.
Clitic1 - atom, first clitic in Clitics.
Clitic2Char - char list, second clitic in Clitics.
Person - person feature of Clitic1.
Number - number feature of Clitic1.
Gender - gender feature of Clitic1.
Case - case feature of Clitic1.
Level - level feature of Clitic1.
Humanness - humanness feature of Clitic1.

```
first_clitic([109, 101 | Clitic2Char], me, Clitic2Char, 1, s, i, ad, i, human).
first_clitic([116, 101 | Clitic2Char], te, Clitic2Char, 2, s, i, ad, informal, human).
first_clitic([108, 111 | Clitic2Char], lo, Clitic2Char, 3, s, m, a, i, i).
first_clitic([108, 111 | Clitic2Char], lo, Clitic2Char, 2, s, m, a, formal, human).
first_clitic([108, 97 | Clitic2Char], la, Clitic2Char, 3, s, f, a, i, i).
first_clitic([108, 97 | Clitic2Char], la, Clitic2Char, 2, s, f, a, formal, human).
first_clitic([108, 101 | Clitic2Char], le, Clitic2Char, 3, s, i, d, i, i).
first_clitic([108, 101 | Clitic2Char], le, Clitic2Char, 2, s, i, d, formal, human).

first_clitic([115,101 | Clitic2Char], se, Clitic2Char, 3, i, i, ad, i, i).
first_clitic([115, 101 | Clitic2Char], se, Clitic2Char, 2, i, i, ad, formal, human).

first_clitic([110, 111, 115 | Clitic2Char], nos, Clitic2Char, 1, p, i, ad, i, human).
first_clitic([111, 115 | Clitic2Char], os, Clitic2Char, 2, p, i, ad, informal, human).
first_clitic([108, 111, 115 | Clitic2Char], los, Clitic2Char, 3, p, m, a, i, i).
first_clitic([108, 111, 115 | Clitic2Char], los, Clitic2Char, 2, p, m, a, formal, human).
first_clitic([108, 97, 115 | Clitic2Char], las, Clitic2Char, 3, p, f, a, i, i).
first_clitic([108, 97, 115 | Clitic2Char], las, Clitic2Char, 2, p, f, a, formal, human).
first_clitic([108, 101, 115 | Clitic2Char], les, Clitic2Char, 3, p, i, d, i, i).
first_clitic([108, 101, 115 | Clitic2Char], les, Clitic2Char, 2, p, i, d, formal, human).
```

B. second_clitic(Clitics2, Person, Number, Gender, Case, Level, Humanness)

```
second_clitic(me, 1, s, i, a, i, human).
second_clitic(te, 2, s, i, a, informal, human).
second_clitic(lo, 3, s, m, a, i, i).
second_clitic(lo, 2, s, m, a, formal, human).
second_clitic(la, 3, s, f, a, i, i).
second_clitic(la, 2, s, f, a, formal, human).
second_clitic(nos, 1, p, i, a, i, human).
second_clitic(os, 2, p, i, a, informal, human).
second_clitic(los, 3, p, m, a, i, i).
second_clitic(los, 2, p, m, a, formal, human).
second_clitic(las, 3, p, f, a, i, i).
second_clitic(las, 2, p, f, a, formal, human).
```

A possible lexicalization for 'da'ndomela' is given below. Note that the clitics 'me' and 'la' are separated from the present participle, and that each clitic has two possible interpretations, corresponding to distinct entries in the lexicon, to be resolved during later stages of processing.

```
Transmission 0   Paragraph 1   Sentence 1

4   lxi(prespart,'dando',dar,[],[m(prt),t(pres),p(#),n(#),g(#)],[],[],['-agr','-past'],[],[dar])
5   lxi(clitic, me, me, [psv], [], [p(1),n(s),g(i),c(h),l(#)], [], [], [], [me])
    lxi(clitic, me, me, [psdrflx], [], [p(1),n(s),g(i),c(h),l(#)], [], [], [], [me])
6   lxi(clitic, la, la, [], [], [p(3),n(s),g(f),c(i),l(#)], [], [], [], [la])
    lxi(clitic, la, la, [], [], [p(2),n(s),g(f),c(i),l(f)], [], [], [], [la])
```

The above mechanism can be easily extended to handle prefixes, infixes, and even circumfixes, which are prefix/suffix combinations. In Spanish, we have handled mainly suffixes and clitics attached to the ends of words. The main difference in processing prefixes is that recognition of the lexical stem is delayed until the prefix is processed.

The MAVT testbed has the capability to handle the two major morphological processes necessary for language processing--inflection (morphological variation) and clitic attachment (merging of morphemes)--as demonstrated in the processing of verbal inflection and clitic attachment in Spanish. The mechanism that we have developed can also be applied to the inflection of other parts of speech, such as nominal or adjectival inflection for languages like Russian, German, and Classical Arabic, with the effort lying mainly in the analysis and implementation of the particular morphemes from the language in question. Similarly, the mechanism can be implemented to handle clitics attached to other parts of speech, such as to nouns, pronouns, and inflected verbs in Arabic, by incorporating those particular morphemes into the system.

5.3 Principle-Based Parsing for Multilingual MAVT

The NLP parser is a principle-based parser that uses grammatical principles from Government-Binding Theory to construct a parse tree for each sentence being processed. The parser combines a bottom-up, data-driven approach to attaching incoming words into the parse tree, with a top-down expectation that a complete tree will be built around a verbal projection (shown in Figure 5.3 as C''-I''-V''). The parser mechanism works by projecting incoming words to maximal X-bar projections (three-level node-graphs), and then attempting to attach

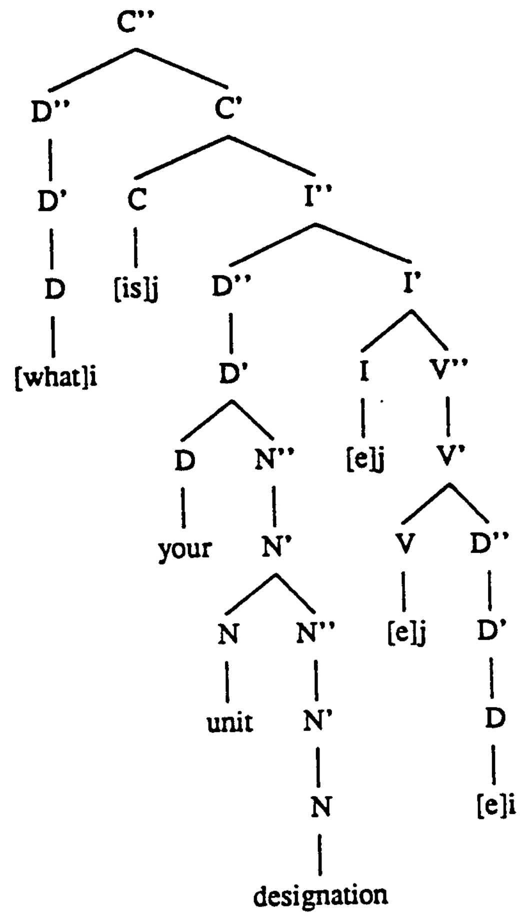


Figure 5.3. Graphical Representation of Parse Tree for English Sentence

the projections into currently available "docking locations" on the existing tree, using syntactic and semantic checks to validate the attachment. The parse structure which is built up through these attachments is represented as an acyclic, directed graph. The mechanism itself can be thought of as a "window" which moves through the emerging parse-graph of the sentence, examining/attaching a pair of nodes at a time. The parser places theta-role information (similar to case frames) in properly attached verb-argument nodes. The syntactic parse for the sentence "what is your unit designation?" is shown as a Prolog term in Figure 5.2, and as a tree in Figure 5.3.

After the syntactic parse is completed, the parse structure/graph for a sentence is passed to the semantic parse module. This module traverses the graph to extract semantic elements and their relations, based on the local graph structure, theta-role assignment, and semantic labels derived from the underlying concept hierarchy, described in Section 5.1.1. This semantic (or 'functional' parse) for the example sentence of Figures 5.2 and 5.3 is shown in Figure 5.4.

There are generally two different approaches taken in designing a parser for use in NLP. The two types of parsers are commonly referred to as principle-based and rule-based parsers. Principle-based parsers are built on universal grammatical principles that can be parameterized for particular languages. Rule-based parsers, on the other hand, are built on a complete set of explicit grammatical rules, with a number of rules to handle every construction in each language.

One of the advantages of a principle-based parser is that it can be used to parse many different languages. Once the parser has been designed for a particular language, the transition to another language requires only some adjustments of the parameter settings that govern the interaction of the principles of grammar. In the case of a rule-based parser, for every new language a new set of rules must be developed or, in the best situation, some existing rules will have to be modified.

Here is a brief set the syntactic categories and their abbreviations that will be used in this section, together with examples of how they can be realized:

CP	complementizer phrase	(Comp + S)
Comp	complementizer	
S	sentence	(NP + VP)
NP	noun phrase	(N)
N	noun	
VP	verb phrase	(V + NP)
V	verb	
PP	prepositional phrase	(P + NP)
P	preposition	

5.3.1 Advantages of Principle-Based Parsing

What follows are four examples of the advantages of principle-based parsers over rule-based ones, namely, the use of the head principle, word order, null subjects, and clitic pronouns.

5.3.1.1 The Head Principle

The position of the head in a phrase is not fixed across languages. Some languages, such as English, Spanish and French, are "head-initial", where the head of the phrase precedes its complement. Other languages, like Japanese and Turkish, are "head-final", where the head of the phrase follows its complement.

In the non-English examples that follow, the first line represents sentence as it is said in the original language. The second line is the gloss of the sentence, that is, a morpheme-by-morpheme translation of the sentence, and the third line is the translation of the sentence into English. In the glosses, some morphemes, such as case and tense/aspect markers, have functional labels rather than literal translations into English. The abbreviations for the functional labels of these morphemes and what the abbreviations stand for are given below.

NOM	nominative case
ACC	accusative case
COMP	complementizer
GEN	genitive case
CAUS	causative
FUT	future
POSS	possessive
PROG	progressive

Verb Phrase

1. English:
   Mary saw John (Verb(Head) Object)

2. Japanese:
   Mary-ga John-o mita. (Object Verb(Head))
   Mary-NOM John-ACC saw
   "Mary saw John"

Adpositional Phrase

3. English:
   Mary gave that book to John. (Preposition(Head) NP)

4. Japanese:
   Mary-ga John-ni sono hono-o watasita (NP Postposition(Head))
   Mary-NOM John-to that book-ACC handed
   "Mary handed that book to John" (Saito 85:41)

Complement Sentences

5. English:
   John thinks that [Mary saw him] (Comp(Head) S)

6. Japanese:
   John-ga [Mary-ga kare-o mita] to omotte iru (S Comp(Head))
   John-NOM Mary-NOM he-ACC saw COMP think

"John thinks that Mary saw him" (Saito 85:75)

7. Turkish:
   Hasan [Fatma-nin on-u ol-dur-eceg-in]-i dusun-uyor
   Hassan Fatima-GEN he-ACC die-CAUS-FUT-POSS-ACC think-PROG
   "Hassan thinks that Fatima will kill him" (Lehmann 84:52)

The parse trees for verb phrases, adpositional phrases, and complement sentences, respectively, in a head-initial language such as English, shown above in examples 1, 3, and 5, would look like:

```
      V''          P''        C''
      |            |          |
      V'           P'         C'
     / \          / \        / \
    V   NP       P   NP   COMP  S
```

Parse trees for analogous phrases in a head-final language (examples 2, 4, 6, and 7 above) would look like:

```
      VP           PP         CP
      |            |          |
      V'           P'         C'
     / \          / \        / \
   NP   V       NP   P      S   COMP
```

To parse the example sentences given above from a head-initial language, a rule-based parser would require a list of Context-Free Phrase-Structure Rules that would look like:

VP -> V NP
PP -> P NP
CP -> COMP S

Furthermore, to parse a head-final language the parser would require the additional Context-Free Rules:

VP -> NP V
PP -> NP P
CP -> S COMP

A principle-based parser, on the other hand, using X-bar theory as its principle, would use a single template as a basis for all phrase types

```
            X''
           / \
  Specifier   X'
             / \
            X   Complement
        (Head)
```

The linear order of the head with respect to the complement, and of the specifier with respect to the X' is derived by direction of semantic-role assignment. In head-initial languages this assignment is rightward, while in head-final languages it is leftward.

The X-bar trees, derived from the template, for these phrases in a head-initial language would now look like:

```
     VP           PP         CP
    /  \         /  \       /  \
   V    NP      P    NP   COMP  S
```

X-bar trees for analogous phrases in a head-final language would look like:

```
      VP          PP        CP
     /  \        /  \      /  \
   NP    V     NP    P    S    COMP
```

By comparison, then, a rule-based parser will need separate sets of rules for every type of language, while a principle-based parser will only need the one X-bar template which is used together with principles of semantic-role assignment.

5.3.1.2 Word Order

In English the basic word order is Subject-Verb-Object:
8. John solved the problem (SVO)

A language like Spanish admits more possibilities:

9. Juan resolvio' el problema (SVO)
John resolved the problem
"John resolved the problem"

10. Resolvio' el problema Juan (VOS)
Resolved the problem John
"John resolved the problem"

11. Resolvio' Juan el problema (VSO)
Resolved John the problem
"John resolved the problem"

To deal with these simple cases, a rule-based parser would need one rule for English and three rules for Spanish. A principle-based parser, on the other hand, will only need the one X-bar template described above, but will use principles of Case Theory to account for the possible structures.

Case is a grammatical feature which is assigned by a class of grammatical categories (the [-N] class, including V's, finite INFL's and P's). It is assigned only to an immediately adjacent category. Case is required on all syntactic arguments (which are usually nominal expressions (ie, some variety of NP)); this principle is expressed as follows (Chomsky 1981:49):

12. *NP if NP has phonetic content and has no Case

English and Spanish are both head-initial languages. This fact takes care of the absence of OVS and OSV order. The fact that there is only SVO order in English follows from the mechanisms available for nominative Case assignment: Spec-Head Agreement. For a subject to receive Case in English it must be in a special configuration, a configuration that happens to be SV (subject preceding verb). Meanwhile, in Spanish there are two mechanisms available for nominative Case assignment: Spec-Head Agreement and Government. Spec-Head Agreement allows the SVO order, just like in English. In other words, Government relies on a structural relation that requires a configuration in which the verb is higher than the subject, in other words, preceding the subject. Case assignment by government allows the VOS and VSO orders.

Notice that the word-order variation between the two languages can be accounted for by only one difference. In Spanish there are two mechanisms of Case assignment and that in English there is only one. See the following chart:

	Case Assignment Spec-Head Agreement	Government
English	+	-
Spanish	+	+

5.3.1.3 Null Subjects

Another interesting difference between English and Spanish is related to the possibility of having "null" subjects. In Spanish, it is possible to "omit" the subject of a sentence. Consider the following example:

13. Llegamos al cine temprano
(We) arrived to the movie theater early
"(We) arrived early to the movie theater"

In sentence (13) there is no overt (or phonetically realized) subject. The English gloss shows the pronoun "we" as subject, this means that even if the subject is absent phonetically, it is active syntactically and semantically.

In English, however, the subject must always be present:

14.* Arrived early to the movie theater

In the case of a rule-based parser, a specific set of rules would be needed to allow for both these cases. In a principle-based parser, on the other hand, the Spanish case would be covered easily simply by switching on the "null subject" parameter.

The reason for this discrepancy is based on the fact that Spanish is "morphologically uniform". This means that it is uniform with respect to its verbal inflectional morphology. In other words, it has different morphological markers (for person and number) for each of the forms of the conjugation paradigms. All null-subject languages are morphologically uniform. Consider the present-indicative paradigms for English and Spanish:

15.	Person	English	Spanish
	1 sing.	arrive	llegO
	2 sing.	arrive	llegAS
	3 sing.	arriveS	llegA
	1 pl.	arrive	llegAMOS
	2 pl.	arrive	llegAIS
	3 pl.	arrive	llegAN

Notice that Spanish is uniform in that it has a particular marker for each case in the paradigm. English, on the other hand, has one marker that distinguishes only third-person singular, all the other cases are marked in the same way.

Once it has been determined that a language presents null subjects (it shows morphological uniformity) the null-subject parameter is switched on. In such a language, it is the interaction of other principles and mechanisms, available also for English, that interpret these sentences with null subjects. The Extended Projection Principle (EPP) requires all sentence to have a subject. The mechanism of Spec-Head Agreement establishes a relation, in this case, between the subject and the verb. In the case of Spanish, the EPP requires a subject (it does not matter if it is phonetically realized or not) and the relation of Spec-Head Agreement reconstructs the subject. In example in (13), this relation tells us that the subject has to be first person plural ("we"), recoverable from the morphology on the verb.

5.3.1.4 Clitic Pronouns

Another interesting difference between English and Spanish concerns "clitic pronouns". Consider the following examples:

16 a. John saw Mary
b. John saw HER

17 a. Juan vio a Mari'a
John saw Mary
"John saw Mary"
b. Juan LA vio
John her saw
"John saw her"

In the case of English, the object NP "Mary" in (16a) is replaced by a pronoun ("her") in (16b). In the case of Spanish, something similar appears to happen; the object NP "Mari'a" in (17a) has been replaced by a pronoun ("la") in (17b). At first sight, there seems to be no major difference between the two languages. Apparently, the only difference is related to the position these pronouns occupy with respect to the verb. Now consider the following cases:

18 a. John writes letters to Mary

b. John writes letters to HER
c. *John writes letters to HER to Mary

19 a. Juan escribe cartas a Mari'a
John writes letters to Mary
"John writes letters to Mary:
b. Juan LE escribe cartas
John her writes letters
"John writes letters to her"
c. Juan LE escribe cartas a Mari'a
John her writes letters to Mary
"John writes letters to Mary"

These examples are similar to the ones in (16) and (17), but in examples (18) and (19) it is the indirect object which is replaced by a pronoun. The interesting cases are the ones in (18c) and (19c). While in English an (indirect) object pronoun cannot co-occur with the full NP, in Spanish this is possible. In English, object pronouns stand instead of the full NP. In Spanish, this is not the case. First, they do not occur in the same position; second, they may co-occur. These pronouns in both languages cannot have the same properties. The difference is that these pronouns in Spanish, but not in English, are "clitics".

Clitics are special pronouns that cannot occur by themselves, they usually attach to another element (in this case to the verb), and no element (except another clitic pronoun) may intervene between them and the verb. In a rule-based parser, specific rules are necessary to treat these pronouns, rules concerning their placement and interpretation. Meanwhile, in a principle-based parser, they may be treated as agreement markers (they are like object-agreement markers, i.e. affixes to the verb that establish a relation with objects). Their placement and interpretation follows from other components of grammar.

5.3.2 Additional Advantages

Aside from these basic advantages of choosing a principle-based parser, there are some other advantages tied into the parsing of ungrammatical input and support for Foreign Language Training (FLT).

5.3.2.1 Parsing Ungrammatical Input

It is not sufficient for a parser to be able to process grammatical input, but rather, it must also be able to handle ungrammatical input. Given the ungrammatical sentences

20. *John to go alone is dangerous
21. *John hit probably Bill.
22. *We need a 50 gallon drum oil.

a rule-based parser would need a set of special weights to account for their ungrammaticality:

20'. Add weighted rule which makes "for" optional.
21'. Add weighted rule which allows an adverb to intervene between verb and Direct Object.
22'. Add weighted rule which allows one NP immediately after another.

A principle-based parser would rely simply on Case Theory to rule out the ungrammatical cases.

The problem with (20), then, is that there is a non-finite INFL (inflectional node - "to" functions as a non-finite INFL) and only finite INFL can assign Case. This "knowledge" is in essence part of the lexical specification of "to", and as such is available during parsing. (21) is ungrammatical because there is an adverb intervening between the Case-assigner ("hit") and the nominal argument ("Bill"); this violates the adjacency requirement on Case-assignment.

Comparing (20) and (21), both violate the Case-filter. In the first instance the violation arises because, even though the structural configuration is correct, the quality of the would-be Case-assigner is incorrect, that is, INFL is non-finite, rather than finite. In the second instance the violation arises because, even though there is a Case assigner present ("hit"), the structural configuration is incorrect, that is, the Case assigner is not adjacent to the Case assignee.

The problem with sentence (22) is that there is no Case assigner for "oil" at all, since the NP [a 50 gallon drum] gets the Case feature assigned by the verb "need", and the subject "we" gets the Case feature of the finite INFL. Thus, "oil" is left stranded. "of" functions as a Case assigner (just as other prepositions), so when it is present, the sentence is grammatical. In order to make all of these sentences parsable, then, the Case-filter needs to be relaxed, and a grammatical structure can be assigned. Moreover, since the system would know that there is a violation, not only could it assign a structure, but it could also rate the sentence on a grammaticality scale.

Comparing the principle-based approach to parsing the sentences in (20-22) to the rule-based approach, the principle-based requires only one coherent principle compared to several apparently unrelated rules to cope with the ungrammaticality. Moreover, there are myriad other instances of Case-filter violations which can be adduced to exhibit the usefulness of the Case-filter, each of which would require their own separate rule to be parsable in a rule-based system. A sampling of Case-filter violations are given below. Notice that any of these examples are close enough to grammatical sentences that they might arise in texts via typos or even speech (especially certain dialects), and some way of parsing them despite their ungrammaticality is needed:

23. *That's the man you looking for.
24. *Which street did you go?
25. *Who do he think he is?
26. *Which sentry does it appear to see those snipers?
27. *The soldier gun jammed.

The Case-filter demonstrates that sentences contain a violation of a grammatical principle, but a well-constructed parser can still assign a syntactic and semantic representation to a sentence containing a Case violation.

The Case-filter and X-bar theory are only two principles of many which can serve to aid in parsing and to cope with ungrammatical input. There are other important principles. The Theta-Criterion and the Projection Principle provide well-formedness conditions on semantic role assignment. The Empty Category Principle presents a well-formedness condition stating the context in which phonologically null elements may occur in a sentence. The Binding Principles provide well-formedness conditions on antecedent-anaphor/pronoun relations. The Bounding Principles present well-formedness principles on movement of syntactic categories. Control Theory provides principles of interpretation of certain phonologically null elements.

These principles are not all of equal importance in assessing grammaticality; for example, a violation of the Theta-Criterion usually leads to a much worse sentence than a violation of the Case-filter. This knowledge can be exploited in parsing and evaluation of parsed sentences.

5.3.2.2 Foreign Language Training

There are tremendous possibilities for using a principle-based parser for language training. The reason for this is related to the reasons just discussed above in connection with parsing ungrammatical input. Specifically, since a principle-based parser is not only in a position to parse an ungrammatical sentence, but also to assess the problem and its severity, it is also in a position to guide a language learner in acquiring a foreign language.

Clitic placement in the Romance languages, for example, is closely tied to semantic role and Case factors. Clitics are frequently difficult for non-Romance speakers to acquire, but with the guidance of a principle-based parser, the user would be in a position to experiment with clitic placement and be evaluated by the parser.

A principle-based parser would also be especially useful in other areas of grammar. Basic sentence structure problems could be assessed using X-bar Theory, Theta-Theory and Case Theory. Question formation and variant word orders could be explained through the Empty Category Principle, Bounding Theory and Case Theory. Theta-Theory, Binding Theory and Control Theory will aid in Complex-sentence formation (e.g. complement sentences and relative clauses). Antecedent-anaphor(reflexives)/pronoun relations could be described through Binding Theory and Case Theory.

5.4 The Text Generation Process

The text generation phase of NLP processing is, in effect, the translation step from the source language to the target language. The input to text generation is the semantically-analyzed parse (the Functional Parse, or FP) of the sentence in the source language; the output is the written text of the sentence translated into the target language. The stages in between are described below.

Text generation from the FP is handled by two components: a Target Language Functional Parse (TLFP) module, and a Target Language Text Generation (TLGEN) module. The two new modules operate in sequence after the FP module. In the TLFP module, target language equivalents of significant sentential elements identified in the FP are found by means of the lexical links and features in the lexicon that indicate appropriate usage. The sentential predicate/argument structure of the English input is thus mapped into a corresponding

English to Spanish

```
fp1:
  'MAINPRED'('1.0')                  = 'INDEX'('1.1')
  'UTTERANCE TYPE'('1.1')            = 'WH QUESTION'
  'PREDICATE NOUN PHRASE
        /PREDICATE'('1.1')           = 'INDEX'('1.2')
  'WH EXPRESSION'('1.2')             = what            ==   [cua'l,que']
  'TENSE'('1.1')                     = 'PRESENT'
  'SUBJECT/AGENT'('1.1')             = 'INDEX'('1.3')
  'POSSESSOR'('1.3')                 = 'INDEX'('1.4')
  'POSSESSIVE PRONOUN'('1.4')        = your            ==   [su,vuestro,tu]
  'NOUN QUALIFIER
        /MILITARY_UNIT'('1.3')       = unit            ==   [unidad]
  'NOUN/ABSTRACT_OBJECT'('1.3')      = designation     ==   [nombre]
  'VOICE'('1.1')                     = 'ACTIVE'
  'PREDICATE'('1.1')                 = be              ==   [ser,estar]
```

Figure 5.4 Functional Parse Output for English Sentence with Spanish Translations

GENERATION

Processes:

1. Constituent generation

   Constituent selection
   Internal ordering
   Internal agreement

2. Word choice

3. Sentential constituent ordering

4. Sentential agreement
   (e.g., subject/verb, indirect object/verbal clitic)

Figure 5.5 Stages in Target Language Sentence Generation

Generation Output

```
yes
| ?- generate(spanish, S).

S = "ser nombre el de unidad su cua'l"

gen1 + gen2 + gen3 + gen4 + gen5 + gen6 + gen7


[1]
| ?- show_gennode(gen1).
GENnode:
  gen_node_id:    gen1
  gen_flabel:     pred
  gen_tlabel:     pred
  gen_word:       ser    --> es
  gen_tense:      pres
  gen_prev:       gen7
  gen_next:       gen6
  gen_lexcat:     pred
  gen_inflected:  inflected
  gen_dectalk:    es


yes
[1]
| ?- show_gennode(gen2).
GENnode:
  gen_node_id:  gen2
  gen_flabel:   subj
  gen_tlabel:   noun
  gen_word:     nombre
  gen_person:   3
  gen_number:   s
  gen_gender:   m
  gen_prev:     gen6
  gen_next:     gen5
  gen_lexcat:   noun
  gen_dectalk:  nombre


yes
[1]
| ?- show_gennode(gen3).
GENnode:
  gen_node_id:  gen6
  gen_flabel:   subj
  gen_tlabel:   posPron
  gen_word:     su
  gen_prev:     gen5
  gen_next:     gen4
  gen_lexcat:   det
  gen_dectalk:  su
```

```
yes
[1]
| ?- show_gennode(gen4).
GENnode:
  gen_node_id:  gen4
  gen_flabel:   subj
  gen_tlabel:   noun
  gen_word:     unidad
  gen_prev:     gen3
  gen_next:     '*none*'
  gen_lexcat:   noun
  gen_dectalk:  unidad


yes
[1]
| ?- show_gennode(gen5).
GENnode:
  gen_node_id:  gen5
  gen_flabel:   subj
  gen_tlabel:   prep
  gen_word:     de
  gen_prev:     gen2
  gen_next:     gen3
  gen_lexcat:   prep
  gen_dectalk:  de


yes
[1]
| ?- show_gennode(gen6).
GENnode:
  gen_node_id:  gen7
  gen_flabel:   subj
  gen_tlabel:   det
  gen_word:     el
  gen_prev:     gen1
  gen_next:     gen2
  gen_lexcat:   det
  gen_dectalk:  el


yes
[1]
| ?- show_gennode(gen7).
GENnode:
  gen_node_id:  gen8
  gen_flabel:   wh_predNP
  gen_tlabel:   wh_exp
  gen_word:     'cua''l'
  gen_next:     gen1
  gen_lexcat:   det
  gen_dectalk:  'cua''l'

S = "cua'l es el nombre de su unidad"
```

Figure 5.6 Generation Steps for Spanish Sentence predicate/argument structure in the target language.

The next step is text generation. Once the appropriate target language elements have been specified, they must be organized into an acceptable sentence. This is done by the TLGEN module. In this process, the translated elements are arranged by a series of generative rules into a linear word order with a hierarchical syntactic structure, including word order. Then morphological rules of government and agreement are applied, as well as certain lexical constraints, and other syntactic characteristics of the target language that are necessary to produce an acceptable sentence.

The steps in text generation processing are outlined in Figure 5.5. First the individual constituents are generated, i.e., the noun phrases, verb phrase, prepositional phrases, and so on. The constituents are first selected or identified based on the labeling and indexing of the structure of the sentence in the Functional Parse (Figure 5.4). Then the internal ordering of the constituents (e.g., whether adjectives should precede or follow the noun) is done based on language-specific information (e.g., in Spanish adjectives generally follow the noun whereas in English they usually precede the noun). Then internal constituent agreement is done (e.g., in Spanish, adjectives agree in number and gender with the noun that they modify). In the functional parse, sometimes more that one word is given as a possible translation for the source language word, and in the next step the appropriate word is selected. The reason that this step is not done first is that information about the constituents of the sentence is important in making the appropriate word selection. For example, in Spanish, 'cua'l' is used only when the interrogative word is independent, and 'que'' is used when it modifies a noun. The third step is the overall ordering of the constituents within the sentence. In English, for example, the normal word order is subject-verb-object(s); in the case of questions, the interrogative word is found at the beginning of the sentence; and so on. Finally, sentential agreement across constituents is accomplished; this includes subject-verb agreement, verb-object agreement, and so on. These text generation steps are basically the same for any language, although the rules themselves are language-specific. The specific steps involved in the text generation of the Spanish sentence illustrated in Figure 5.6 are discussed below.

In the text generation phase, we first generate all the sentence constituents that are present in the functional parse of the source sentence and produce a generation node or "gennode" for the translation of each word in the source sentence. Take for example the translation of "What is your unit designation?" from English to Spanish. Figure 5.6 shows the the Spanish constituents produced on the basis of the functional parse (Figure 5.4): a predicate (gen1), a subject (gen2 - gen6), and a predicate noun phrase (gen7). The subject noun phrase consists of a head noun (gen2), a determiner (gen6), a genitive phrase (gen3 - gen5), which consists of a genitive marker (gen5), a possessive pronoun (gen3), and a head noun (gen4). The gennodes at this stage record the base form (uninflected form) of these words, as well as the morphological, syntactic, and other relevant information about these words in the target language. (Note that the actual order of gennode generation is irrelevant, because generation of the correct surface word order for the target sentence will be done as part of the normal processing at this stage.) Notice that the noun qualifier (e.g., "unit") in the source English sentence is turned into a genitive phrase in Spanish, since noun qualifiers in English usually are expressed by genitive structures in Spanish.

In the next stage of text generation processing, we check to see if we need to generate any extra words which are not in the source sentence but which are required in the target sentence, such as the auxiliary "do" for certain target interrogative and negative sentences in

English; the reflexive pronoun as the object when the verb in the target language strictly subcategorizes for (in other words, must take) an object (e.g., "desplazar" in Spanish), but there isn't one in this particular source sentence; the subject when the subject in the source language is not expressed by a separate lexical item (such as pronominal subjects in Spanish), but is required in the target language, and so on. For our present case, nothing needs to be done at this stage. Next, we order the sentence constituents produced so far according to the word ordering rules of the target language. Thus, the gennodes mentioned above are arranged according to the word order rules of Spanish (i.e., interrogative predicate noun phrase, predicate verb "ser", subject). Next we cliticize the object pronouns, if any, for languages such as Spanish. Nothing needs to be done for the present case. Next, agreement is done depending on the language: the verb (or the modal or the auxiliary verb if present) is conjugated according to the semantic features of the subject (thus the infinitive form "ser" becomes "es" in agreement with the third person singular features of the head noun of the subject "nombre"); the predicate adjective phrase and predicate noun phrase also agree with the features of the subject in Spanish. The agreement within a noun phrase (agreement between the head noun and the adjective modifiers, the possessive pronoun, and the determiners) also needs to be done according to language specific rules. Finally we check to see if we have to delete certain constituents which exist in the source sentence but should be omitted in the target sentence, such as the pronominal subject in Spanish. This does not apply in our present case.

The output of text generation -- the sentence generated in written form in the target language -- is then passed on to the speech synthesizer, which converts the written text into speech as described in Section 4.2.

6. MAVT TESTING AND EVALUATION

This section describes the test corpora and procedures used for speech and natural language testing, and the test results which were obtained. In general, the description follows the evaluation format presented in the MAVT Test Plan (CLIN A002, ELIN A003). The set of tests given therein was aimed at evaluating the integrated speech/NLP testbed, as well as individual processing components of the testbed. Testing thus included "black-box" tests of the integrated system as a whole (tests #1 - 3), as well as "glass-box" tests of individual speech and NLP processing components (tests #4 - 15). Summary descriptions of each of the tests are presented in Table 6-1.

For convenience of presentation of the test and evaluation material, this section is comprised of the five subsections listed below. The initial section presents test data and briefly describes the test procedure. The remaining four sections describe sets of tests of the functionality of the overall MAVT system, its subsystems, and their components, corresponding to the displays presented in Figures 1-1 and 1-2 in Section 1. Thus,

1. **Section 6.1** lists the test corpora utilized for speech and natural language processing, and summarizes the overall test procedure as well as formal and informal testing that has been carried out in the course of the MAVT testbed development;
2. **Section 6.2** treats black-box tests of the total system (tests #1 - 3);
3. **Section 6.3** details glass-box tests of the speech recognition subsystem only (tests #4 and 5);
4. **Section 6.4** describes glass-box tests of the natural language processing subsystem components (tests #6 - 13);
5. **Section 6.5** deals with the speech synthesis subsystem only (tests # 14 and 15).

It should be noted that the procedure used in MAVT speech recognition testing and described throughout Section 6 -- as well as the associated scoring methodology -- is not limited to the DARPA Spoken Language Systems (SLS) scoring approach defined in Pallett [1987], but includes several other measures of accuracy as well. In most respects, the testing and scoring methodology used in MAVT testing is more rigorous than the DARPA procedure, since several types of scores are computed for each test sentence individually, and for each processing stage, from spoken input through natural language understanding and translation to spoken output.

In order to provide some basis for comparison with DARPA-supported speech recognition systems, performance statistics for each speech syntax and speaker model for both English and Spanish are presented in Table 6-6, Section 6.3. In addition, three columns of DARPA statistics have been added to each of the six tables providing detailed test results for individual sentences in the speech recognition testing described in that section. A comparative analysis of DARPA vs. LSI scoring is also given, to facilitate comparison of the test scores.

In general, LSI average accuracy measures are more conservative than the DARPA "Words Correct (Corr)" measures, since the LSI average accuracy scores are lower than or equal to the Corr scores in all cases. A detailed explanation of LSI and DARPA measures is given in Section 6.3.

Test # 1	Dialog Pairs Spoken English input => Spoken Spanish output and Spoken Spanish input => Spoken English output
Test # 2	English-to-Spanish sentences Spoken English input => Spoken Spanish output
Test # 3	Spanish-to-English sentences Spoken Spanish input => Spoken English output
Test # 4	English speech recognition Spoken English input => English text output
Test # 5	Spanish speech recognition Spoken Spanish input => Spanish text output
Test # 6	English-to-Spanish text translation English text input => Spanish text output
Test # 7	Spanish-to-English text translation Spanish text input => English text output
Test # 8	English text understanding English text input => English syntactic parse
Test # 9	English text understanding English text input => English functional (semantic) parse
Test #10	Spanish text understanding Spanish text input => Spanish syntactic parse
Test #11	Spanish text understanding Spanish text input => Spanish functional (semantic) parse
Test #12	Spanish text translation/generation English/Spanish functional parse input => Spanish text output
Test #13	English text translation/generation Spanish/English functional parse input => English text output
Test #14	Spanish speech generation Spanish text input => Spoken Spanish output
Test #15	English speech generation English text input => Spoken English output

Table 6-1. Summary of MAVT Tests.

6.1 Test Data

An initial acceptance test was performed on June 4, 1992, in the presence of the Rome Laboratory Project Engineer, Lt. Bradford Clifton. Lt. Clifton selected 100 sentences from the following lists:

a. sentences generated by the speech syntax[1] for the biographics domain (English)

b. sentences generated by the speech syntax for the mission domain (English)

c. sentences generated by the speech syntax for the biographics domain (Spanish)

d. sentences generated by the speech syntax for the mission domain (Spanish)

e. sentences used in the interrogation training course at the Defense Language Institute (DLI).

The list of sentences from the DLI course contained a number of words as well as syntactic and semantic structures which are not handled in the current versions of the speech syntaxes, and thus could not be processed through the entire system. Since these sentences are essentially invalid as test items for the speech recognizer, they are excluded from test and evaluation of the speech recognition subsystem; thus, they are not utilized in Tests #4 and #5, nor could they be utilized in tests of the overall system (Tests #1 - 3). They are, however, included in the sentences processed by the DBG NLP component in Tests #6 - 13, although the syntactic and semantic structures of some of these (e.g., Sentences 89, 95, 96) are also beyond the current capabilities of the NLP components of the MAVT testbed system (see Section 7 for a detailed discussion of present capabilities and plans for future development).

The set of sentences selected by Lt. Clifton for use as a test corpus is presented in Table 6-2.

6.1.1 Test Corpus for Speech Testing

Table 6-3 lists the subset of sentences given in Table 6-2 which were actually used for speech testing. All of these sentences were uttered twice in the tests, yielding two unique utterances per sentence (in a few cases, more than two). Thus the number of test items in the speech test corpora is actually more than double the number of sentences listed in Table 6-3, as described in Section 6.3.

1. See Section 4.1.2 for an explanation of the generative speech syntax and how it is used in recognition.

Table 6-2. Test Corpus.

Table 6-2a. Biographics Test Sentences (English)

1. State your name.
2. State your full name.
3. Tell me your full name.
4. Spell your name.
5. What is your military identification number?
6. Indicate your unit designation.
7. Tell me your unit.
8. What is your rank?
9. Indicate your rank.
10. What is your duty position?
11. What is your birth date?
12. Can you read english?
13. Do you speak russian at all?

Table 6-2b. Mission Test Sentences (English)

14. What is your mission?
15. What was your mission?
16. Was your mission offensive?
17. Is his mission offensive or defensive?
18. Why was your unit moving out to the south?
19. Is the main force heading in that direction?
20. Can the forward element see our tanks from the road?
21. Are they repositioning to the right of your unit?
22. What kind of vehicles do they have?
23. Why is she heading to the north?
24. Can she hear its tanks from the road?
25. Why is the main force heading to the east?
26. Are you heading to the east?
27. Why are you heading to the east?
28. Were they repositioning to the south?
29. Was the main force heading to the right of your unit?
30. Why was the main force heading to the right of your unit?
31. Can he hear our tanks from the road?
32. Can they observe his vehicles from the command post?

Table 6-2. Test Corpus (continued)

Table 6-2c. Biographics Test Sentences (Spanish)

33. Jesu's Martinez.
34. Sargento de segunda clase.
35. Naci' en Santa Clara.
36. Mi nombre es Oscar Batista.
37. Mi rango es comandante en jefe de tercera clase.
38. Mi unidad es bateri'a de defensa ae'rea de'cimo regimiento de infanteri'a mecanizada de'cima divisio'n de infanteri'a mecanizada.
39. El nombre completo de mi unidad es primera seccio'n se'ptimo peloto'n bateri'a de defensa ae'rea octavo regimiento de infanteri'a mecanizada primera divisio'n de infanteri'a mecanizada.
40. Naci' el once de abril de mil novecientos sesenta y ocho.
41. Es He'ctor Herna'ndez.
42. Mi madre es mulata y un poco italiana.
43. Naci' el catorce de junio de mil novecientos treinta.
44. Mi padre es portugue's y un poco italiano y mi madre era norteamericana y china.
45. La de'cima seccio'n primer peloto'n bateri'a de defensa ae'rea tercer regimiento de infanteri'a mecanizada cuarta divisio'n de infanteri'a mecanizada.
46. Mi rango es teniente general.
47. Mi unidad es tercer peloto'n bateri'a de defensa ae'rea se'ptimo regimiento de infanteri'a mecanizada cuarta divisio'n de infanteri'a mecanizada.
48. Soy el comandante.
49. Cua'l es su nombre?

Table 6-2d. Mission Test Sentences (Spanish)

50. Mi misio'n es proteger tanques del regimiento.
51. Su misio'n es encontrar unidades americanas.
52. Era defensiva.
53. Proteger el puesto de comando del regimiento.
54. Porque el puesto de comando se desplazaba en esa direccio'n.
55. Si'.
56. Tanques.
57. Por que' su unidad se desplazaba hacia el sur.
58. Porque el comando evacuaba al norte.
59. Pueden observar del camino.
60. Mantener la paz.
61. Mi misio'n era aniquilar.
62. Porque unidades estaban desplazando al este.
63. Ofensiva.
64. No pero puede ser.
65. Si' mis unidades podri'an.

Table 6-2. Test Corpus (continued)

Table 6-2e. Test Sentences from the DLI Interrogator Training Courses

66. When was your unit so designated?
67. Where are the subordinate units located?
68. What is the name of your unit commander?
69. What are your alternate bases of operation?
70. Where are your alternate bases of operation located?
71. Give a sketch of the installations in your home base.
72. How many officers do you have?
73. How many tanks do you have?
74. How many enlisted men are there in your unit?
75. How many persons were killed?
76. How many officers were killed?
77. How many persons were wounded?
78. How many persons deserted your unit?
79. Which persons deserted your unit?
80. How many individuals in your unit are relatively new?
81. Volunteers only.
82. Where do you get the weapons to arm replacements?
83. Specify the types of weapons available to your unit.
84. Who makes the actual plan of attack?
85. Can you hear your vehicles from the road?
86. What are the communications requirements?
87. What intelligence-gathering means are available to your unit?
88. How do you obtain the support of local population?
89. What method is employed to plant one of your personnel or a sympathizer in a government installation?
90. What is the current mission of your unit?
91. What is your unit's mission?
92. What is the mission of your unit?
93. What is your personal mission?
94. What happened to these people?
95. Man-trap and boobytrap setting.
96. When was the first time you were exposed to political propaganda indoctrination?
97. When was the first time that you were exposed to political propaganda indoctrination?

Table 6-3. Test Corpus for Speech Processing.

Table 6-3a. Biographics Test Sentences (English)

1. State your name.
2. State your full name.
3. Tell me your full name.
6. Indicate your unit designation.
7. Tell me your unit.
8. What is your rank?
9. Indicate your rank.
10. What is your duty position?
11. What is your birth date?
12. Can you read English?
13. Do you speak Russian at all?

Table 6-3b. Mission Test Sentences (English)

14. What is your mission?
15. What was your mission?
16. Was your mission offensive?
17. Is his mission offensive or defensive?
18. Why was your unit moving out to the south?
19. Is the main force heading in that direction?
20. Can the forward element see our tanks from the road?
21. Are they repositioning to the right of your unit?
22. What kind of vehicles do they have?
23. Why is he heading to the north?
24. Can he hear its tanks from the road?
25. Why is the main force heading to the east?
26. Are you heading to the east?
27. Why are you heading to the east?
28. Were they repositioning to the south?
29. Was the main force heading to the right of your unit?
30. Why was the main force heading to the right of your unit?
31. Can he hear our tanks from the road?
32. Can they observe his vehicles from the command post?
    Can you hear your vehicles from the road?

Table 6-3. Test Corpus for Speech Processing (continued)

Table 6-3c. Biographics Test Sentences (Spanish)

33. Jesu's Martinez.
34. Sargento de segunda clase.
35. Naci' en Santa Clara.
36. Mi nombre es Oscar Batista.
38. Mi unidad es bateri'a de defensa ae'rea de'cimo regimiento de infanteri'a mecanizada de'cima divisio'n de infanteri'a mecanizada.
39. El nombre completo de mi unidad es primera seccio'n se'ptimo peloto'n bateri'a de defensa ae'rea octavo regimiento de infanteri'a mecanizada primera divisio'n de infanteri'a mecanizada.
40. Naci' el once de abril de mil novecientos sesenta y_sp ocho.
41. Es He'ctor Herna'ndez.
43. Naci' el catorce de junio de mil novecientos treinta.
45. La de'cima seccio'n primer peloto'n bateri'a de defensa ae'rea tercer regimiento de infanteri'a mecanizada cuarta divisio'n de infanteri'a mecanizada.
46. Mi rango es teniente general.
47. Mi unidad es tercer peloto'n bateri'a de defensa ae'rea se'ptimo regimiento de infanteri'a mecanizada cuarta divisio'n de infanteri'a mecanizada.
48. Soy el comandante.

Table 6-3d. Mission Test Sentences (Spanish)

50. Mi misio'n es proteger tanques del regimiento.
51. Su misio'n es encontrar unidades americanas.
53. Proteger el puesto de comando del regimiento.
54. Porque el puesto de comando se desplazaba en esa direccio'n.
59. Pueden observar del camino.
60. Mantener la paz.
61. Mi misio'n era aniquilar.
62. Porque las unidades estaban desplazando al este.
63. Ofensiva.
64. No pero puede ser.
65. Si' mis unidades podri'an.

6.1.2 Test Data for Language Translation Testing

The test set for language translation testing is the complete test set as given in Table 6-2.

6.1.3 Tests, Test Results, and Evaluation

Throughout the course of the project, informal regression testing was carried out routinely as changes were implemented to determine whether the addition of new capabilities negatively impacted any components of the existing system. A set of sentences for regression testing covering basic syntactic structures of English and Spanish was developed to test and gradually expand the functionality of the system. These are described in Section 7 (System Status), and included in the report as Appendix A.

Formal testing involved two tests:

1. The initial acceptance test performed on June 4, 1992, in the presence of the Rome Laboratory Project Engineer, Lt. Bradford Clifton;
2. Final acceptance testing performed at LSI prior to delivery of the MAVT hardware and software to Rome Laboratory.

The initial test was flawed in several ways through inadvertent procedural errors, including the utilization of test sentences with words outside the vocabulary of the speech recognizer, as pointed out above. In addition, since it turned out that use of the collection tools provided in the SSI development environment conflicted with the formal procedure defined in the MAVT Test Plan document, these could not be utilized, making it difficult to provide valid test statistics for the speech recognizer. A third difficulty with the initial test was loss of test and evaluation information normally provided by the tracing facility, which was not used in order to speed up throughput.

For these reasons, the test and evaluation discussion below is based on the final acceptance test carried out at LSI, which utilized the test corpus selected by Lt. Clifton.

The following criteria for scoring (full and partial credit) are listed in the Test Plan:

- Correct text in speech recognizer output
- Correct lexicalization in NLP subsystem[2]
- Correct syntactic parse
- Correct functional (semantic) parse in NLP subsystem (both source and destination)
- Correct text generated by translator
- Correctly spoken text
- Correct total interaction

Test results for each of these stages of processing are presented in the tables displayed throughout this section, where each sentence of the test corpus can be found using the given index number. Table 6-4 gives a summary of all tests conducted, by test number and sentence number. Scoring reflects above criteria, using the following scale:

2. This and the following three processes are internal to the DBG natural language processing subsystem, described most recently in Montgomery et al [1992].

Final Test Results

n.t. = not tested n.a. = not applicable

Sentence #	Lang. E/S	T-1	T-2	T-3	T-4	T-5	T-6	T-7	T-8	T-9	T-10	T-11	T-12	T-13	T-14	T-15	Result
1	E	n.t.	n.t.	n.a.	1.00	n.a.	1.00	n.a.	1.00	1.00	n.a.	n.a.	1.00	n.a.	1.00	n.a.	1.00
2	E	n.t.	n.t.	n.a.	1.00	n.a.	1.00	n.a.	1.00	1.00	n.a.	n.a.	1.00	n.a.	1.00	n.a.	1.00
3	E	n.t.	n.t.	n.a.	1.00	n.a.	1.00	n.a.	1.00	1.00	n.a.	n.a.	1.00	n.a.	1.00	n.a.	1.00
4	E	n.t.	n.t.	n.a.	n.t.	n.a.	1.00	n.a.	1.00	1.00	n.a.	n.a.	1.00	n.a.	1.00	n.a.	1.00
5	E	n.t.	n.t.	n.a.	n.t.	n.a.	1.00	n.a.	1.00	1.00	n.a.	n.a.	1.00	n.a.	1.00	n.a.	1.00
6	E	n.t.	n.t.	n.a.	1.00	n.a.	1.00	n.a.	1.00	1.00	n.a.	n.a.	1.00	n.a.	1.00	n.a.	1.00
7	E	n.t.	n.t.	n.a.	1.00	n.a.	1.00	n.a.	1.00	1.00	n.a.	n.a.	1.00	n.a.	1.00	n.a.	1.00
8	E	n.t.	n.t.	n.a.	1.00	n.a.	1.00	n.a.	1.00	1.00	n.a.	n.a.	1.00	n.a.	1.00	n.a.	1.00
9	E	n.t.	n.t.	n.a.	1.00	n.a.	1.00	n.a.	1.00	1.00	n.a.	n.a.	1.00	n.a.	1.00	n.a.	1.00
10	E	n.t.	n.t.	n.a.	1.00	n.a.	1.00	n.a.	1.00	1.00	n.a.	n.a.	1.00	n.a.	1.00	n.a.	1.00
11	E	n.t.	n.t.	n.a.	1.00	n.a.	1.00	n.a.	1.00	1.00	n.a.	n.a.	1.00	n.a.	1.00	n.a.	1.00
12	E	n.t.	n.t.	n.a.	1.00	n.a.	1.00	n.a.	1.00	1.00	n.a.	n.a.	1.00	n.a.	1.00	n.a.	1.00
13	E	n.t.	n.t.	n.a.	1.00	n.a.	1.00	n.a.	1.00	1.00	n.a.	n.a.	1.00	n.a.	1.00	n.a.	1.00
14	E	n.t.	n.t.	n.a.	0.56	n.a.	1.00	n.a.	1.00	1.00	n.a.	n.a.	1.00	n.a.	1.00	n.a.	0.93
15	E	n.t.	n.t.	n.a.	0.56	n.a.	1.00	n.a.	1.00	1.00	n.a.	n.a.	1.00	n.a.	1.00	n.a.	0.93
16	E	n.t.	n.t.	n.a.	1.00	n.a.	1.00	n.a.	1.00	1.00	n.a.	n.a.	1.00	n.a.	1.00	n.a.	1.00
17	E	n.t.	n.t.	n.a.	1.00	n.a.	0.75	n.a.	0.50	0.50	n.a.	n.a.	0.75	n.a.	0.75	n.a.	0.85
18	E	n.t.	n.t.	n.a.	1.00	n.a.	1.00	n.a.	1.00	1.00	n.a.	n.a.	1.00	n.a.	1.00	n.a.	1.00
19	E	n.t.	n.t.	n.a.	1.00	n.a.	1.00	n.a.	1.00	1.00	n.a.	n.a.	1.00	n.a.	1.00	n.a.	1.00
20	E	n.t.	n.t.	n.a.	0.43	n.a.	1.00	n.a.	1.00	1.00	n.a.	n.a.	1.00	n.a.	1.00	n.a.	0.91
21	E	n.t.	n.t.	n.a.	0.67	n.a.	1.00	n.a.	1.00	1.00	n.a.	n.a.	1.00	n.a.	1.00	n.a.	0.95
22	E	n.t.	n.t.	n.a.	1.00	n.a.	1.00	n.a.	1.00	1.00	n.a.	n.a.	1.00	n.a.	1.00	n.a.	1.00
23	E	n.t.	n.t.	n.a.	1.00	n.a.	1.00	n.a.	1.00	1.00	n.a.	n.a.	1.00	n.a.	1.00	n.a.	1.00
24	E	n.t.	n.t.	n.a.	1.00	n.a.	1.00	n.a.	1.00	1.00	n.a.	n.a.	1.00	n.a.	1.00	n.a.	1.00
25	E	n.t.	n.t.	n.a.	1.00	n.a.	1.00	n.a.	1.00	1.00	n.a.	n.a.	1.00	n.a.	1.00	n.a.	1.00
26	E	n.t.	n.t.	n.a.	1.00	n.a.	1.00	n.a.	1.00	1.00	n.a.	n.a.	1.00	n.a.	1.00	n.a.	1.00
27	E	n.t.	n.t.	n.a.	1.00	n.a.	1.00	n.a.	1.00	1.00	n.a.	n.a.	1.00	n.a.	1.00	n.a.	1.00
28	E	n.t.	n.t.	n.a.	1.00	n.a.	1.00	n.a.	1.00	1.00	n.a.	n.a.	1.00	n.a.	1.00	n.a.	1.00
29	E	n.t.	n.t.	n.a.	1.00	n.a.	1.00	n.a.	1.00	1.00	n.a.	n.a.	1.00	n.a.	1.00	n.a.	1.00
30	E	n.t.	n.t.	n.a.	0.69	n.a.	1.00	n.a.	1.00	1.00	n.a.	n.a.	1.00	n.a.	1.00	n.a.	0.95
31	E	n.t.	n.t.	n.a.	0.66	n.a.	1.00	n.a.	1.00	1.00	n.a.	n.a.	1.00	n.a.	1.00	n.a.	0.94

Table 6-4. Summary of All Tests

Sentence #	Lang. E/S	T-1	T-2	T-3	T-4	T-5	T-6	T-7	T-8	T-9	T-10	T-11	T-12	T-13	T-14	T-15	Result
32	E	n.t.	n.t.	n.a.	1.00	n.a.	1.00	n.a.	1.00	1.00	n.a.	n.a.	1.00	n.a.	1.00	n.a.	1.00
33	S	n.t.	n.a.	n.t.	n.a.	1.00	n.a.	1.00	n.a.	n.a.	1.00	1.00	n.a.	1.00	n.a.	1.00	1.00
34	S	n.t.	n.a.	n.t.	n.a.	1.00	n.a.	1.00	n.a.	n.a.	1.00	1.00	n.a.	1.00	n.a.	1.00	1.00
35	S	n.t.	n.a.	n.t.	n.a.	1.00	n.a.	1.00	n.a.	n.a.	1.00	1.00	n.a.	1.00	n.a.	1.00	1.00
36	S	n.t.	n.a.	n.t.	n.a.	1.00	n.a.	1.00	n.a.	n.a.	1.00	1.00	n.a.	1.00	n.a.	1.00	1.00
37	S	n.t.	n.a.	n.t.	n.a.	n.t.	n.a.	1.00	n.a.	n.a.	1.00	1.00	n.a.	1.00	n.a.	1.00	1.00
38	S	n.t.	n.a.	n.t.	n.a.	0.87	n.a.	1.00	n.a.	n.a.	1.00	1.00	n.a.	1.00	n.a.	1.00	0.98
39	S	n.t.	n.a.	n.t.	n.a.	1.00	n.a.	0.00	n.a.	n.a.	0.00	0.00	n.a.	0.00	n.a.	0.00	0.17
40	S	n.t.	n.a.	n.t.	n.a.	1.00	n.a.	0.75	n.a.	n.a.	0.25	0.25	n.a.	0.75	n.a.	0.75	0.63
41	S	n.t.	n.a.	n.t.	n.a.	1.00	n.a.	1.00	n.a.	n.a.	1.00	1.00	n.a.	1.00	n.a.	1.00	1.00
42	S	n.t.	n.a.	n.t.	n.a.	n.t.	n.a.	1.00	n.a.	n.a.	1.00	1.00	n.a.	1.00	n.a.	1.00	1.00
43	S	n.t.	n.a.	n.t.	n.a.	1.00	n.a.	0.75	n.a.	n.a.	0.25	0.25	n.a.	0.75	n.a.	0.75	0.63
44	S	n.t.	n.a.	n.t.	n.a.	n.t.	n.a.	1.00	n.a.	n.a.	1.00	1.00	n.a.	1.00	n.a.	1.00	1.00
45	S	n.t.	n.a.	n.t.	n.a.	0.96	n.a.	1.00	n.a.	n.a.	1.00	1.00	n.a.	1.00	n.a.	1.00	0.99
46	S	n.t.	n.a.	n.t.	n.a.	1.00	n.a.	1.00	n.a.	n.a.	1.00	1.00	n.a.	1.00	n.a.	1.00	1.00
47	S	n.t.	n.a.	n.t.	n.a.	1.00	n.a.	1.00	n.a.	n.a.	1.00	1.00	n.a.	1.00	n.a.	1.00	1.00
48	S	n.t.	n.a.	n.t.	n.a.	1.00	n.a.	1.00	n.a.	n.a.	1.00	1.00	n.a.	1.00	n.a.	1.00	1.00
49	S	n.t.	n.a.	n.t.	n.a.	n.t.	n.a.	1.00	n.a.	n.a.	1.00	1.00	n.a.	1.00	n.a.	1.00	1.00
50	S	n.t.	n.a.	n.t.	n.a.	0.77	n.a.	1.00	n.a.	n.a.	1.00	1.00	n.a.	1.00	n.a.	1.00	0.96
51	S	n.t.	n.a.	n.t.	n.a.	0.75	n.a.	1.00	n.a.	n.a.	1.00	1.00	n.a.	1.00	n.a.	1.00	0.96
52	S	n.t.	n.a.	n.t.	n.a.	n.t.	n.a.	1.00	n.a.	n.a.	1.00	1.00	n.a.	1.00	n.a.	1.00	1.00
53	S	n.t.	n.a.	n.t.	n.a.	1.00	n.a.	1.00	n.a.	n.a.	1.00	1.00	n.a.	1.00	n.a.	1.00	1.00
54	S	n.t.	n.a.	n.t.	n.a.	1.00	n.a.	1.00	n.a.	n.a.	1.00	1.00	n.a.	1.00	n.a.	1.00	1.00
55	S	n.t.	n.a.	n.t.	n.a.	n.t.	n.a.	1.00	n.a.	n.a.	1.00	1.00	n.a.	1.00	n.a.	1.00	1.00
56	S	n.t.	n.a.	n.t.	n.a.	n.t.	n.a.	1.00	n.a.	n.a.	1.00	1.00	n.a.	1.00	n.a.	1.00	1.00
57	S	n.t.	n.a.	n.t.	n.a.	n.t.	n.a.	1.00	n.a.	n.a.	1.00	1.00	n.a.	1.00	n.a.	1.00	1.00
58	S	n.t.	n.a.	n.t.	n.a.	n.t.	n.a.	1.00	n.a.	n.a.	1.00	1.00	n.a.	1.00	n.a.	1.00	1.00
59	S	n.t.	n.a.	n.t.	n.a.	0.63	n.a.	0.00	n.a.	n.a.	0.00	0.00	n.a.	0.00	n.a.	0.00	0.11
60	S	n.t.	n.a.	n.t.	n.a.	1.00	n.a.	1.00	n.a.	n.a.	1.00	1.00	n.a.	1.00	n.a.	1.00	1.00
61	S	n.t.	n.a.	n.t.	n.a.	0.31	n.a.	1.00	n.a.	n.a.	1.00	1.00	n.a.	1.00	n.a.	1.00	0.89
62	S	n.t.	n.a.	n.t.	n.a.	0.13	n.a.	1.00	n.a.	n.a.	1.00	1.00	n.a.	1.00	n.a.	1.00	0.86
63	S	n.t.	n.a.	n.t.	n.a.	1.00	n.a.	1.00	n.a.	n.a.	1.00	1.00	n.a.	1.00	n.a.	1.00	1.00
64	S	n.t.	n.a.	n.t.	n.a.	1.00	n.a.	1.00	n.a.	n.a.	1.00	1.00	n.a.	1.00	n.a.	1.00	1.00
65	S	n.t.	n.a.	n.t.	n.a.	1.00	n.a.	0.00	n.a.	n.a.	0.00	0.00	n.a.	0.00	n.a.	0.00	0.17
66	E	n.t.	n.t.	n.a.	n.t.	n.a.	0.00	n.a.	0.00	0.00	n.a.	n.a.	0.00	n.a.	0.00	n.a.	0.00
67	E	n.t.	n.t.	n.a.	n.t.	n.a.	1.00	n.a.	1.00	1.00	n.a.	n.a.	1.00	n.a.	1.00	n.a.	1.00
68	E	n.t.	n.t.	n.a.	n.t.	n.a.	1.00	n.a.	1.00	1.00	n.a.	n.a.	1.00	n.a.	1.00	n.a.	1.00

Sentence #	Lang. E/S	T-1	T-2	T-3	T-4	T-5	T-6	T-7	T-8	T-9	T-10	T-11	T-12	T-13	T-14	T-15	Result
69	E	n.t.	n.t.	n.a.	n.t.	n.a.	1.00	n.a.	1.00	1.00	n.a.	n.a.	1.00	n.a.	1.00	n.a.	1.00
70	E	n.t.	n.t.	n.a.	n.t.	n.a.	1.00	n.a.	1.00	1.00	n.a.	n.a.	1.00	n.a.	1.00	n.a.	1.00
71	E	n.t.	n.t.	n.a.	n.t.	n.a.	1.00	n.a.	1.00	1.00	n.a.	n.a.	1.00	n.a.	1.00	n.a.	1.00
72	E	n.t.	n.t.	n.a.	n.t.	n.a.	1.00	n.a.	1.00	1.00	n.a.	n.a.	1.00	n.a.	1.00	n.a.	1.00
73	E	n.t.	n.t.	n.a.	n.t.	n.a.	1.00	n.a.	1.00	1.00	n.a.	n.a.	1.00	n.a.	1.00	n.a.	1.00
74	E	n.t.	n.t.	n.a.	n.t.	n.a.	0.00	n.a.	1.00	1.00	n.a.	n.a.	0.00	n.a.	0.00	n.a.	0.40
75	E	n.t.	n.t.	n.a.	n.t.	n.a.	0.75	n.a.	1.00	0.75	n.a.	n.a.	0.75	n.a.	0.75	n.a.	0.80
76	E	n.t.	n.t.	n.a.	n.t.	n.a.	1.00	n.a.	1.00	1.00	n.a.	n.a.	1.00	n.a.	1.00	n.a.	1.00
77	E	n.t.	n.t.	n.a.	n.t.	n.a.	0.75	n.a.	1.00	0.75	n.a.	n.a.	0.75	n.a.	0.75	n.a.	0.80
78	E	n.t.	n.t.	n.a.	n.t.	n.a.	0.75	n.a.	1.00	0.75	n.a.	n.a.	0.75	n.a.	0.75	n.a.	0.80
79	E	n.t.	n.t.	n.a.	n.t.	n.a.	1.00	n.a.	1.00	1.00	n.a.	n.a.	1.00	n.a.	1.00	n.a.	1.00
80	E	n.t.	n.t.	n.a.	n.t.	n.a.	0.00	n.a.	0.00	0.00	n.a.	n.a.	0.00	n.a.	0.00	n.a.	0.00
81	E	n.t.	n.t.	n.a.	n.t.	n.a.	1.00	n.a.	1.00	1.00	n.a.	n.a.	1.00	n.a.	1.00	n.a.	1.00
82	E	n.t.	n.t.	n.a.	n.t.	n.a.	0.50	n.a.	0.50	0.50	n.a.	n.a.	0.50	n.a.	0.50	n.a.	0.50
83	E	n.t.	n.t.	n.a.	n.t.	n.a.	0.00	n.a.	0.00	0.00	n.a.	n.a.	0.00	n.a.	0.00	n.a.	0.00
84	E	n.t.	n.t.	n.a.	n.t.	n.a.	0.50	n.a.	1.00	0.50	n.a.	n.a.	0.50	n.a.	0.50	n.a.	0.60
85	E	n.t.	n.t.	n.a.	n.t.	n.a.	1.00	n.a.	1.00	1.00	n.a.	n.a.	1.00	n.a.	1.00	n.a.	1.00
86	E	n.t.	n.t.	n.a.	n.t.	n.a.	0.75	n.a.	1.00	0.75	n.a.	n.a.	0.75	n.a.	0.75	n.a.	0.80
87	E	n.t.	n.t.	n.a.	n.t.	n.a.	0.00	n.a.	1.00	1.00	n.a.	n.a.	0.00	n.a.	0.00	n.a.	0.40
88	E	n.t.	n.t.	n.a.	n.t.	n.a.	0.50	n.a.	1.00	1.00	n.a.	n.a.	0.50	n.a.	0.50	n.a.	0.70
89	E	n.t.	n.t.	n.a.	n.t.	n.a.	0.00	n.a.	0.00	0.00	n.a.	n.a.	0.00	n.a.	0.00	n.a.	0.00
90	E	n.t.	n.t.	n.a.	n.t.	n.a.	1.00	n.a.	1.00	1.00	n.a.	n.a.	1.00	n.a.	1.00	n.a.	1.00
91	E	n.t.	n.t.	n.a.	n.t.	n.a.	1.00	n.a.	1.00	1.00	n.a.	n.a.	1.00	n.a.	1.00	n.a.	1.00
92	E	n.t.	n.t.	n.a.	n.t.	n.a.	1.00	n.a.	1.00	1.00	n.a.	n.a.	1.00	n.a.	1.00	n.a.	1.00
93	E	n.t.	n.t.	n.a.	n.t.	n.a.	1.00	n.a.	1.00	1.00	n.a.	n.a.	1.00	n.a.	1.00	n.a.	1.00
94	E	n.t.	n.t.	n.a.	n.t.	n.a.	1.00	n.a.	1.00	1.00	n.a.	n.a.	1.00	n.a.	1.00	n.a.	1.00
95	E	n.t.	n.t.	n.a.	n.t.	n.a.	0.50	n.a.	0.50	0.50	n.a.	n.a.	0.50	n.a.	0.50	n.a.	0.50
96	E	n.t.	n.t.	n.a.	n.t.	n.a.	0.50	n.a.	0.50	0.50	n.a.	n.a.	0.50	n.a.	0.50	n.a.	0.50
97	E	n.t.	n.t.	n.a.	n.t.	n.a.	0.00	n.a.	1.00	1.00	n.a.	n.a.	0.00	n.a.	0.00	n.a.	0.40
Average					0.92	0.89	0.83	0.89	0.91	0.88	0.86	0.86	0.83	0.89	0.83	0.89	0.87

1. 1.0 for correct results
2. 0.75 for minor errors
3. 0.5 for moderate errors
4. 0.25 for severe errors
5. 0.0 for unusable output

Each test was treated as a black-box. Thus, the test score for a particular test reflects the quality of the output without regard to the results of internal processing. Since more than one repetition of a given sentence was scored in speech recognition testing, the highest average accuracy score for the sentence was utilized in the table. The column labeled "Result" is an average across all applicable tests.

The remaining tables of test results are described in the appropriate section below.

6.2 Tests of the Overall MAVT System

(Black-box Tests #1 - 3)

6.2.1 Test #1

(Black-box #1, Dialog Pairs)
Spoken English input => Spoken Spanish output
and Spoken Spanish input => Spoken English output

As specified in the Test Plan, the purpose of this test was to determine the overall effectiveness of the system in a dialog-translation setting. Success was to be determined by correctly translating the "meaning" of the English input and the "meaning" of the Spanish input so as to ultimately receive a semantically valid response to the original statement. Since it was not possible to conduct a spontaneous test dialog as planned, (because of incompatibility with the sentence selection procedure described above), a set of dialogs were composed from the corpus of test sentences to show the dialog capability of the testbed system. These dialogs, which show a query or a query set and a response or response set, are presented in Table 6-5. Dialog sentences are numbered to allow comparison with the evaluation tables presented throughout this section.

Table 6-5. Dialog Sequences.

Table 6-5a. Biographics Dialogs.

1. State your name.
2. State your full name.
3. Tell me your full name.

***

33. Jesu's Martinez.
36. Mi nombre es Oscar Batista.
41. Es He'ctor Herna'ndez.

---

8. What is your rank?
9. Indicate your rank.

***

34. Sargento de segunda clase.
37. Mi rango es comandante en jefe de tercera clase.
46. Mi rango es teniente general.

---

11. What is your birth date?

***

40. Naci' el once de abril mil novecientos sesenta y_sp ocho.
43. Naci' el catorce de junio mil novecientos treinta.

---

10. What is your duty position?

***

48. Soy el comandante.

Table 6-5a. Biographics Dialogs (continued).

6. Indicate your unit designation.
7. Tell me your unit.

***

38. Mi unidad es bateri'a de defensa ae'rea de'cimo regimiento de infanteri'a mecanizada de'cima divisio'n de infanteri'a mecanizada.

39. El nombre completo de mi unidad es primera seccio'n se'ptimo peloto'n bateri'a de defensa ae'rea octavo regimiento de infanteri'a mecanizada primera divisio'n de infanteri'a mecanizada.

45. La de'cima seccio'n primer peloto'n bateri'a de defensa ae'rea tercer regimiento de infanteri'a mecanizada cuarta divisio'n de infanteri'a mecanizada.

47. Mi unidad es tercer peloto'n bateri'a de defensa ae'rea se'ptimo regimiento de infanteri'a mecanizada cuarta divisio'n de infanteri'a mecanizada.

Table 6-5b. Mission Dialogs.

---

16. Was your mission offensive?

***

50. Mi misio'n es proteger tanques del regimiento.
52. Era defensiva.
55. Si'.
61. Mi misio'n era aniquilar.
63. Ofensiva.

---

18. Why was your unit moving out to the south?

***

53. Proteger el puesto de comando del regimiento.
54. Porque el puesto de comando se desplazaba en esa direccio'n.
62. Porque unidades estaban desplazando al este.

6.2.2 Test #2

(Black-box #2, English-to-Spanish sentences)
Spoken English input => Spoken Spanish output

This test is similar to the first half of test #1. The purpose of the test is to determine the correctness of translation from spoken English to spoken Spanish. Success is determined by correctly translating the "meaning" of the English input into a Spanish output with the same meaning. As an example, "What is your name?" could result in: "?'Cua'l es su nombre?" or "?'Co'mo se llama?".

Scoring: Same as for test #1.

6.2.3 Test #3

(Black-box #3, Spanish-to-English sentences)
Spoken Spanish input => Spoken English output.

This test is similar to the second half of test #1. The purpose of the test is to determine the correctness of translation from spoken Spanish to spoken English. Success is determined by correctly translating the "meaning" of the Spanish input into an English output with the same meaning. As an example, "Soy el comandante" could yield: "I am the commander" or just "Commander".

Scoring: Same as for test #1.

6.3 Test and Evaluation of the Speech Recognition Subsystem

As discussed previously, the scoring method used in test and evaluation of the speech recognition subsystem was not limited to the DARPA SLS evaluation approach, since our goal was to derive more detailed information on a variety of aspects of system performance, including adequacy for the IPW application. In order to facilitate comparison with DARPA SLS benchmarks, Table 6-6 contains performance statistics compiled by SSI's evaluation software, which closely follows the DARPA scoring convention described in Pallett [1987].[3]

Thus in Table 6-6, the word error rate (usually abbreviated to *Err*) is computed as

$$\frac{100*(\text{insertions} + \text{deletions} + \text{substitutions})}{\text{total number of word tokens in the test}}$$

while the percentage of words correct (usually abbreviated to *Corr*) is computed as 100 - Err. The sentence error rate (popularly *Sent Err*) is computed as :

$$\frac{100*(\text{number of sentences with errors})}{\text{total number of sentences in the test}}$$

3. SSI's scoring software excludes "silence words", and apparently weights all types of errors equally. Although the 1.5 negative weight on substitutions is discussed in the context of the Pallett article [SSI 90], and the intent of the discussion is that this scoring weight is used, we found no evidence of this in the statistics compiled by the SSI software. Indeed, from the recent literature on DARPA SLS benchmark testing, it does not appear that the 1.5 penalty is still being used in scoring. For the most part, it appears that substituted words are counted in the same way as other errors. This was the approach we adopted in the MAVT scoring after investigating the literature on current scoring practices of the DARPA SLS community.

grammar	# edges	# cat	#word	Avg. branch. factor		Total # Sentences	Words Correct	Word Error	Sentence Error	model	# utt.
				Cat.	Word						
Eng-Bio-nec	149	58	66	2.01	2.99	829	0.96	0.04	0.04	3013	24
Eng-Msn-nec	285	56	60	2.18	2.36	8982	0.94	0.06	0.43	3013	42
Span-Bio-nec	1702	238	238	3.15	3.19	1.05e+06	0.94	0.06	0.42	ara06	26
Span-Bio-nec	1702	238	238	3.15	3.19	1.05e+06	0.95	0.05	0.38	ara03	26
Span-Msn-nec	222	76	113	2.36	3.87	70198	0.78	0.22	0.45	ara06	22
Span-Msn-nec	222	76	113	2.36	3.87	70198	0.74	0.26	0.55	ara03	22

Notes:

Grammar = the speech grammar or syntax
Bio = the speech syntax for collection of biographic data
Msn = the speech syntax for collection of mission related data
Scen = the interrogation scenario on which the syntax is based
# edges = the number of atomic rules in the given speech syntax
# cat. = the number of atomic categories in the speech syntax
# word = the number of words in the given speech syntax
Avg. Branch. Factor = the average branching factor of the given syntax (average number of possible word choices per rule element)
% Accuracy = Accuracy percentage in tests of the syntax (by utterance and by word)
Model = the speech model used by the phonetic decoding software
(0313 = English male
 ara00 = Profiled Spanish male with span_mavt dictionary
 ara03 = Profiled Spanish male with span3_mavt dictionary
 ara06 = Profiled Spanish male with span6_mavt dictionary)
Utterance = spoken sentence
#Utt. = the number of utterances in the text

Table 6-6. MAVT Statistics: Speech Grammars

In the remaining tables of this section (Tables 6-7 through 6-12), test results are given based on individual utterances. The initial column in each table gives the sentence index number, linking the item analysis to the test sentences in Tables 6-2 and 6-3, as well as other test results presented elsewhere in this section. The second column is a unique utterance number distinguishing the particular instance of speech behavior from other instances of speech behavior for the same sentence (i.e., other utterances of the same sentence).

Since the DARPA scoring procedure was not adequate to represent all the dimensions of the test data that were of interest to us, we formulated a set of accuracy measurements that are better suited for defining success/failure in terms of the project goals. First, for our purposes, the impact of insertions, deletions, and substitutions is not limited to the word level, but has a substantial impact at the sentence or utterance level as well. Although substitutions at the word level provide some information for phonetic fine tuning, these are largely phonologically predictable (e.g., substitution of *our* for *your*, and vice versa).

Much more interesting is the effect of word insertions, deletions, and substitutions at the utterance level. For example, the reference word string for Utterance 4 of Table 6-12 was the response:

"*Porque las unidades se estaban desplazando al este.*"
("Because the units were repositioning to the east.")

In the recognition process, the verb "*trasladando*" ("moving") was substituted for "*desplazando*", requiring deletion of the reflexive clitic "*se*". On the whole, this is an acceptable transformation which preserves the meaning of the original sentence. Similarly, of the 10 utterances containing errors in Table 6-12, 8 preserve sufficient meaning to be acceptable in the context of the MAVT application.

Since there are varying degrees of degradation/preservation of meaning of utterances depending upon the number and type of errors, we felt that it made more sense in the MAVT context to incorporate the error penalties into the utterance scores as a measure of utterance accuracy. The DARPA approach of scoring an utterance as wrong if it contains any errors does not appear to provide insightful results for our purposes.[4] This is especially true since our test analysis is based on individual utterances rather than on test corpora or sets of utterances (except for Table 6-6, which was compiled to facilitate comparison with DARPA SLS results).

Our fundamental objective is to assess MAVT testbed performance along several different dimensions which appear significant in the context of the IPW application. We therefore defined four types of accuracy measures, which are computed as follows:

4. Scientists at the Naval Research Laboratory appear to have encountered a similar problem in their evaluation of a spoken language interface developed with the SSI ASR system. They found that "raw accuracy" scores (presumably those computed by the SSI system based on the DARPA scoring method) were less indicative of the usability of the spoken language interface than a "functional accuracy" score based on the ability of an utterance to elicit the desired system action, regardless of recognition errors it might contain (Everett, Wauchope and Perzanowski 1992).

Word Accuracy

$$\frac{\text{no. correct words}}{\text{no. words in utterance}}$$

Utterance Accuracy

$$\frac{\text{no. correct words - (deletions + insertions + substitutions)}}{\text{no. words in utterance}}$$

Semantic accuracy is correctness of the meaning of the utterance as a whole, while **translation accuracy** is a functional measure which trades off explicit distinctions in one language of a language pair versus implicit distinctions in the other. For example, Sentence 14 of the English mission test corpus "What is your mission?" was recognized as "What is their mission?" The semantic accuracy score is zero, since the substituted word "their" causes the meaning to be incorrect in English. However, in either case in Spanish, the pronoun *su* is used, therefore, the translation accuracy is 100%.

Finally, an average accuracy score is computed based on the preceding four scores. As no attempt was made to normalize the four scores, the average is somewhat sensitive to the distribution of these scores for individual items. However, the resulting score in most cases seems to parallel intuitive judgments about the degree of success/failure of recognition for the particular item.

The columns to the right of the LSI accuracy measures in each table are the DARPA statistics; in this case, they are also given on a sentence by sentence basis, with the totals customarily utilized shown at the bottom. In all cases, the DARPA "Corr" scores are greater than or equal to the LSI average accuracy scores.

A final note deals with the N-best capability of the MAVT testbed. For any spoken input, the speech recognition subsystem of the MAVT testbed produces the N-best matches from all the sentences generated by the given speech syntax for the input utterance. The N-best choices are displayed on the user interface, allowing the user to select whichever one is correct for further processing:

What is their mission?
What is your mission?
What is our mission?
What is her mission?

In the test procedure, only the first option was selected, and since the MAVT testbed software currently does not compile statistics on the set of N-best options, it was not possible to evaluate the number of times the correct choice was in the N-best set, but not listed first (having the highest likelihood of being correct, according to the system's best estimate).

The following paragraphs and associated tables present a detailed analysis of the speech recognition testing for English and Spanish.

6.3.1 Test #4
(Glass-box #1, English speech recognition)
Spoken English input => English text output

This test measures the quality of the speech recognizer output for English utterances based on the English male speaker model (3013) supplied with the SSI recognizer (see Section 4.1 for a discussion of speaker models). Tables 6-7 and 6-8 present detailed analyses of the speech testing for the biographics and mission domains respectively.

6.3.2 Test #5
(Glass-box #2, Spanish speech recognition)
Spoken Spanish input => Spanish text output

This test is similar to test #4, except that it tests the conversion of Spanish utterances into Spanish text. The test is scored in the same manner as the previous one, except that two different Spanish speaker models, ara03 and ara06 (described in Section 4.1.2), were tested, each being reported in a separate table. Spanish biographics utterances are presented in Tables 6-9 and 6-10, while Spanish mission utterances are given in Tables 6-11 and 6-12.

6.4 Test and Evaluation of NLP Subsystem Components

The tests covered in this section include all the glass-box tests of natural language processing subsystem components (Tests # 6 - 13). The test sentences were scored according to where errors occurred in the flow of natural language processing and on how well each component processed the sentences as evidenced by the corresponding output data structures. The results of the scoring are shown in Table 6-13.

As shown in Table 6-13, errors were scored at 0.75 for minor errors, 0.5 for more serious errors, and 0.0 for major errors, incomplete processing, or fatal errors. The set of data structures for each sentence scored in the table are the Source Lexicon, the Parse (PRS), the Functional Parse (FP), the Target Lexicon, and the Text Generation output. Because processing is sequential, once an error is found, it is assumed to affect all of the remaining processing structures for that sentence. A plus indicates that an output data structure was produced but that that particular data structure was not scored because an error was already found and scored in a previous column. The final "Result" column represents an average of the scores for all of the previous columns for which there is a score (the plus is not counted in computing the results). This means that the overall result for a sentence containing a parse error, for example, is lower than a sentence containing a generation error of the same magnitude. This correlates with our perception that the earlier on in processing that an error occurs, the greater that error's effect on the output.

As Table 6-13 shows, the individual natural language components performed relatively well. In the 97 test sentences, there was a total of 21 errors. Nearly three-quarters of these errors (16 out of 21, or 74 %) occurred in the Defense Language Institute (DLI) sentences (##66-97) chosen by Lt. Clifton, which constituted less than one-third of the sentences tested. The DLI sentences, though relevant to an interrogation scenario, represent a somewhat broader domain than than the more constrained domain of the scenarios we used as the development corpus

Syntax: English Biographics

Speaker Model 3013														
							Percent Accuracy					DARPA Statistics		
Sentence No.	Utterance No.	Words Correct	No. of Words	Words deleted	Words inserted	Words substituted	Utterance	Word	Semantic	Translation	Average	Sent Err	Err	Corr
8	1	4	4	0	0	0	1.00	1.00	1.00	1.00	1.00	0.00	0.00	1.00
10	2	5	5	0	0	0	1.00	1.00	1.00	1.00	1.00	0.00	0.00	1.00
11	3	5	5	0	0	0	1.00	1.00	1.00	1.00	1.00	0.00	0.00	1.00
13	4	6	6	0	0	0	1.00	1.00	1.00	1.00	1.00	0.00	0.00	1.00
7	5	4	4	0	0	0	1.00	1.00	1.00	1.00	1.00	0.00	0.00	1.00
9	6	3	3	0	0	0	1.00	1.00	1.00	1.00	1.00	0.00	0.00	1.00
12	7	4	4	0	0	0	1.00	1.00	1.00	1.00	1.00	0.00	0.00	1.00
1	8	3	3	0	0	0	1.00	1.00	1.00	1.00	1.00	0.00	0.00	1.00
2	9	4	4	0	0	0	1.00	1.00	1.00	1.00	1.00	0.00	0.00	1.00
3	10	5	5	0	0	0	1.00	1.00	1.00	1.00	1.00	0.00	0.00	1.00
10	11	5	5	0	0	0	1.00	1.00	0.75	1.00	0.94	0.00	0.00	1.00
6	12	4	4	0	0	0	1.00	1.00	0.00	1.00	0.75	0.00	0.00	0.00
8	13	4	4	0	0	0	1.00	1.00	1.00	1.00	1.00	0.00	0.00	1.00
10	14	5	5	0	0	0	1.00	1.00	1.00	1.00	1.00	0.00	0.00	1.00
11	15	5	5	0	0	0	1.00	1.00	1.00	1.00	1.00	0.00	0.00	1.00
13	16	6	6	0	0	0	1.00	1.00	1.00	1.00	1.00	0.00	0.00	1.00
7	17	0	4	0	0	4	0.00	0.00	0.00	0.00	0.00	1.00	1.00	0.00
9	18	3	3	0	0	0	1.00	1.00	1.00	1.00	1.00	0.00	0.00	1.00
12	19	4	4	0	0	0	1.00	1.00	1.00	1.00	1.00	0.00	0.00	1.00
1	20	3	3	0	0	0	1.00	1.00	1.00	1.00	1.00	0.00	0.00	1.00
2	21	4	4	0	0	0	1.00	1.00	1.00	1.00	1.00	0.00	0.00	1.00
3	22	5	5	0	0	0	1.00	1.00	1.00	1.00	1.00	0.00	0.00	1.00
10	23	5	5	0	0	0	1.00	1.00	1.00	1.00	1.00	0.00	0.00	1.00
6	24	4	4	0	0	0	1.00	1.00	1.00	1.00	1.00	0.00	0.00	1.00
Avg(item)		4.17	4.33	0.00	0.00	0.17	0.96	0.96	0.91	0.96	0.95	0.04	0.04	0.92
Sum	24	100	104	0	0	4	N/A	N/A	N/A	N/A	N/A	1	N/A	N/A
Avg(all)	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	0.04	0.96

Table 6-7. Test Results for English Biographics Utterances (Speaker Model 3013)

Syntax: English Mission

Speaker Model 3013														
							Percent Accuracy					DARPA Statistics		
Sentence No.	Utterance No.	Words Correct	No. of Words	Words deleted	Words inserted	Words substituted	Utterance	Word	Semantic	Translation	Average	Sent Err	Err	Corr
0	1	7	8	0	0	1	0.75	0.88	0.00	0.00	0.41	1.00	0.13	0.88
14	2	3	4	0	0	1	0.50	0.75	0.00	1.00	0.56	1.00	0.25	0.75
16	3	3	4	0	0	1	0.50	0.75	0.00	1.00	0.56	1.00	0.25	0.75
15	4	3	4	0	0	1	0.50	0.75	0.00	1.00	0.56	1.00	0.25	0.75
18	5	9	9	0	0	0	1.00	1.00	1.00	1.00	1.00	0.00	0.00	1.00
19	6	8	8	0	0	0	1.00	1.00	1.00	1.00	1.00	0.00	0.00	1.00
20	7	9	10	0	0	1	0.80	0.90	0.00	0.00	0.43	1.00	0.10	0.90
21	8	8	9	0	0	1	0.78	0.89	0.00	0.00	0.42	1.00	0.11	0.89
22	9	7	7	0	0	0	1.00	1.00	1.00	1.00	1.00	0.00	0.00	1.00
17	10	6	6	0	0	0	1.00	1.00	1.00	1.00	1.00	0.00	0.00	1.00
23	11	6	7	0	0	1	0.71	0.86	0.75	1.00	0.83	1.00	0.14	0.86
14	12	3	4	0	0	1	0.50	0.75	0.00	1.00	0.56	1.00	0.25	0.00
24	13	8	8	0	0	0	1.00	1.00	1.00	1.00	1.00	0.00	0.00	1.00
25	14	9	9	0	0	0	1.00	1.00	1.00	1.00	1.00	0.00	0.00	1.00
26	15	6	6	0	0	0	1.00	1.00	1.00	1.00	1.00	0.00	0.00	1.00
27	16	7	7	0	0	0	1.00	1.00	1.00	1.00	1.00	0.00	0.00	1.00
28	17	6	6	0	0	0	1.00	1.00	1.00	1.00	1.00	0.00	0.00	1.00
29	18	10	11	0	0	1	0.82	0.91	0.00	1.00	0.68	1.00	0.09	0.91
30	19	11	12	0	0	1	0.83	0.92	0.00	1.00	0.69	1.00	0.08	0.92
31	20	7	8	0	0	1	0.75	0.88	0.00	0.00	0.41	1.00	0.13	0.88
32	21	9	9	0	0	0	1.00	1.00	1.00	1.00	1.00	0.00	0.00	1.00
0	22	8	8	0	0	0	1.00	1.00	1.00	1.00	1.00	0.00	0.00	1.00
14	23	3	4	0	0	1	0.50	0.75	0.00	1.00	0.56	1.00	0.25	0.75
16	24	4	4	0	0	0	1.00	1.00	1.00	1.00	1.00	0.00	0.00	1.00
15	25	3	4	0	0	1	0.50	0.75	0.00	1.00	0.56	1.00	0.25	0.75
18	26	9	9	0	0	0	1.00	1.00	1.00	1.00	1.00	0.00	0.00	1.00
19	27	8	8	0	0	0	1.00	1.00	1.00	1.00	1.00	0.00	0.00	1.00
20	28	9	10	0	0	1	0.80	0.90	0.00	0.00	0.43	1.00	0.10	0.90
21	29	8	9	0	0	1	0.78	0.89	0.00	1.00	0.67	1.00	0.11	0.89
22	30	7	7	0	0	0	1.00	1.00	1.00	1.00	1.00	0.00	0.00	1.00
17	31	6	6	0	0	0	1.00	1.00	1.00	1.00	1.00	0.00	0.00	1.00
23	32	7	7	0	0	0	1.00	1.00	1.00	1.00	1.00	0.00	0.00	1.00
14	33	3	4	0	0	1	0.50	0.75	0.00	1.00	0.56	1.00	0.25	0.75
24	34	8	8	0	0	0	1.00	1.00	1.00	1.00	1.00	0.00	0.00	1.00
25	35	9	9	0	0	0	1.00	1.00	1.00	1.00	1.00	0.00	0.00	1.00
26	36	6	6	0	0	0	1.00	1.00	1.00	1.00	1.00	0.00	0.00	1.00
27	37	7	7	0	0	0	1.00	1.00	1.00	1.00	1.00	0.00	0.00	1.00
28	38	6	6	0	0	0	1.00	1.00	1.00	1.00	1.00	0.00	0.00	1.00
29	39	11	11	0	0	0	1.00	1.00	1.00	1.00	1.00	0.00	0.00	1.00
30	40	11	12	0	0	1	0.83	0.92	0.00	1.00	0.69	1.00	0.08	0.92
31	41	7	8	0	0	1	0.75	0.88	0.00	1.00	0.66	1.00	0.13	0.88
32	42	9	9	0	0	0	1.00	1.00	1.00	1.00	1.00	0.00	0.00	1.00
Avg(item)		7.00	7.43	0.00	0.00	0.43	0.86	0.93	0.59	0.88	0.81	0.43	0.07	0.91
Sum	42	294	312	0	0	18	N/A	N/A	N/A	N/A	N/A	18	N/A	N/A
Avg(all)	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	0.06	0.94

Table 6-8. Test Results for English Mission Utterances (Speaker Model 3013)

Syntax: Spanish Biographics

Speaker Model ARA03														
							Percent Accuracy					DARPA Statistics		
Sentence No.	Utterance No.	Words Correct	No. of Words	Words deleted	Words inserted	Words substituted	Utterance	Word	Semantic	Translation	Average	Sent Err	Err	Corr
33	1	2	2	0	0	0	1.00	1.00	1.00	1.00	1.00	0.00	0.00	1.00
34	2	4	4	0	0	0	1.00	1.00	1.00	1.00	1.00	0.00	0.00	1.00
35	3	3	3	0	0	0	1.00	1.00	1.00	1.00	1.00	0.00	0.00	1.00
36	4	5	5	0	0	0	1.00	1.00	1.00	1.00	1.00	0.00	0.00	1.00
38	5	14	17	1	0	2	0.65	0.82	1.00	1.00	0.87	1.00	0.18	0.82
39	6	25	25	0	0	0	1.00	1.00	1.00	1.00	1.00	0.00	0.00	1.00
40	7	11	11	0	0	0	1.00	1.00	1.00	1.00	1.00	0.00	0.00	1.00
41	8	2	3	0	0	1	0.33	0.67	0.50	0.50	0.50	1.00	0.33	0.67
43	9	9	9	0	0	0	1.00	1.00	1.00	1.00	1.00	0.00	0.00	1.00
45	10	18	19	1	0	0	0.89	0.95	1.00	1.00	0.96	1.00	0.05	0.95
46	11	5	5	0	0	0	1.00	1.00	1.00	1.00	1.00	0.00	0.00	1.00
47	12	19	19	0	0	0	1.00	1.00	1.00	1.00	1.00	0.00	0.00	0.00
48	13	2	3	1	0	0	0.33	0.67	1.00	1.00	0.75	1.00	0.33	0.67
33	14	2	2	0	0	0	1.00	1.00	1.00	1.00	1.00	0.00	0.00	1.00
34	15	4	4	0	0	0	1.00	1.00	1.00	1.00	1.00	0.00	0.00	1.00
35	16	3	3	0	0	0	1.00	1.00	1.00	1.00	1.00	0.00	0.00	1.00
36	17	5	5	0	0	0	1.00	1.00	1.00	1.00	1.00	0.00	0.00	1.00
38	18	16	17	0	0	1	0.88	0.94	0.75	0.75	0.83	1.00	0.06	0.94
39	19	24	25	0	0	1	0.92	0.96	0.50	0.50	0.72	1.00	0.04	0.96
40	20	10	11	0	0	1	0.82	0.91	0.75	0.75	0.81	1.00	0.09	0.91
41	21	3	3	0	0	0	1.00	1.00	1.00	1.00	1.00	0.00	0.00	1.00
43	22	8	9	0	0	1	0.78	0.89	0.75	0.75	0.79	1.00	0.11	0.89
45	23	18	19	0	0	1	0.89	0.95	0.50	0.50	0.71	1.00	0.05	0.95
46	24	5	5	0	0	0	1.00	1.00	1.00	1.00	1.00	0.00	0.00	1.00
47	25	18	19	0	0	1	0.89	0.95	0.50	0.50	0.71	1.00	0.05	0.95
48	26	3	3	0	0	0	1.00	1.00	1.00	1.00	1.00	0.00	0.00	1.00
Avg(item)		9	10	0.12	0.00	0.35	0.90	0.95	0.89	0.89	0.91	0.38	0.05	0.91
Sum	26	238	250	3	0	9	N/A	N/A	N/A	N/A	N/A	10	N/A	N/A
Avg(all)	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	0.05	0.95

Table 6-9. Test Results for Spanish Biographics Utterances (Speaker Model ARA03)

Syntax: Spanish Biographics

Speaker Model ARA06															
							Percent Accuracy					DARPA Statistics		
Sentence No.	Utterance No.	Words Correct	No. of Words	Words deleted	Words inserted	Words substituted	Utterance	Word	Semantic	Translation	Average	Sent Err	Err	Corr
33	1	1	2	0	0	1	0.00	0.50	0.50	0.50	0.38	1.00	0.50	0.50
34	2	4	4	0	0	0	1.00	1.00	1.00	1.00	1.00	0.00	0.00	1.00
35	3	3	3	0	0	0	1.00	1.00	1.00	1.00	1.00	0.00	0.00	1.00
36	4	4	5	0	0	1	0.60	0.80	0.25	0.25	0.48	1.00	0.20	0.80
38	5	17	17	0	0	0	1.00	1.00	1.00	1.00	1.00	0.00	0.00	1.00
39	6	24	25	0	0	1	0.92	0.96	0.50	0.50	0.72	1.00	0.04	0.96
40	7	6	11	2	0	3	0.09	0.55	0.00	0.00	0.16	1.00	0.45	0.55
41	8	3	3	0	0	0	1.00	1.00	1.00	1.00	1.00	0.00	0.00	1.00
43	9	9	9	0	0	0	1.00	1.00	1.00	1.00	1.00	0.00	0.00	1.00
45	10	19	19	0	1	0	0.95	1.00	1.00	1.00	0.99	0.00	0.05	0.95
46	11	5	5	0	0	0	1.00	1.00	1.00	1.00	1.00	0.00	0.00	1.00
47	12	19	19	0	0	0	1.00	1.00	1.00	1.00	1.00	0.00	0.00	0.00
48	13	3	3	0	0	0	1.00	1.00	1.00	1.00	1.00	0.00	0.00	1.00
33	14	2	2	0	0	0	1.00	1.00	1.00	1.00	1.00	0.00	0.00	1.00
34	15	3	4	0	0	1	0.50	0.75	0.75	0.75	0.69	1.00	0.25	0.75
35	16	2	3	0	0	1	0.33	0.67	0.00	0.00	0.25	1.00	0.33	0.67
36	17	5	5	0	0	0	1.00	1.00	1.00	1.00	1.00	0.00	0.00	1.00
38	18	17	17	0	0	0	1.00	1.00	1.00	1.00	1.00	0.00	0.00	1.00
39	19	25	25	0	0	0	1.00	1.00	1.00	1.00	1.00	0.00	0.00	1.00
40	20	10	11	0	0	1	0.82	1.00	0.90	0.90	0.90	1.00	0.09	0.91
41	21	3	3	0	0	0	1.00	1.00	1.00	1.00	1.00	0.00	0.00	1.00
43	22	9	9	0	0	0	1.00	1.00	1.00	1.00	1.00	0.00	0.00	1.00
45	23	17	19	1	0	1	0.79	0.89	0.75	0.75	0.79	1.00	0.11	0.89
46	24	5	5	0	0	0	1.00	1.00	1.00	1.00	1.00	0.00	0.00	1.00
47	25	18	19	0	0	1	0.89	0.95	0.50	0.50	0.71	1.00	0.05	0.95
48	26	2	3	1	0	0	0.33	0.67	1.00	1.00	0.75	1.00	0.33	0.67
Avg(item)		9	10	0.15	0.04	0.42	0.82	0.91	0.81	0.81	0.84	0.38	0.09	0.87
Sum	26	235	250	4	1	11	N/A	N/A	N/A	N/A	N/A	10	N/A	N/A
Avg(all)	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	0.06	0.94

Table 6-10. Test Results for Spanish Biographics Utterances (Speaker Model ARA06)

Syntax: Spanish Mission

Speaker Model ARA03															
							Percent Accuracy					DARPA Statistics		
Sentence No.	Utterance No.	Words Correct	No. of Words	Words deleted	Words inserted	Words substituted	Utterance	Word	Semantic	Translation	Average	Sent Err	Err	Corr
59	1	3	4	0	1	1	0.25	0.75	0.75	0.75	0.63	1.00	0.50	0.50
60	2	3	3	0	0	0	1.00	1.00	1.00	1.00	1.00	0.00	0.00	1.00
61	3	3	4	0	0	1	0.50	0.75	0.00	0.00	0.31	1.00	0.25	0.75
62	4	4	8	1	0	3	0.00	0.50	0.00	0.00	0.13	1.00	0.50	0.50
51	5	5	6	0	1	1	0.50	0.83	0.75	0.75	0.71	1.00	0.33	0.67
63	6	1	1	0	0	0	1.00	1.00	1.00	1.00	1.00	0.00	0.00	1.00
50	7	6	7	0	0	1	0.71	0.86	0.75	0.75	0.77	1.00	0.14	0.86
64	8	4	4	0	0	0	1.00	1.00	1.00	1.00	1.00	0.00	0.00	1.00
65	9	3	4	0	0	1	0.50	0.75	0.75	0.75	0.69	1.00	0.25	0.75
53	10	7	7	0	0	0	1.00	1.00	1.00	1.00	1.00	0.00	0.00	1.00
54	11	10	10	0	0	0	1.00	1.00	1.00	1.00	1.00	0.00	0.00	1.00
59	12	1	4	3	2	0	0.00	0.25	0.00	0.00	0.06	1.00	1.25	0.00
60	13	3	3	0	0	0	1.00	1.00	1.00	1.00	1.00	0.00	0.00	1.00
61	14	3	4	0	0	1	0.50	0.75	0.00	0.00	0.31	1.00	0.25	0.75
62	15	3	8	2	0	3	0.00	0.38	0.00	0.00	0.09	1.00	0.63	0.38
51	16	5	6	0	0	1	0.67	0.83	0.75	0.75	0.75	1.00	0.17	0.83
63	17	1	1	0	0	0	1.00	1.00	1.00	1.00	1.00	0.00	0.00	1.00
50	18	2	7	1	0	4	0.00	0.29	0.00	0.00	0.07	1.00	0.71	0.29
64	19	4	4	0	0	0	1.00	1.00	1.00	1.00	1.00	0.00	0.00	1.00
65	20	4	4	0	0	0	1.00	1.00	1.00	1.00	1.00	0.00	0.00	1.00
53	21	5	7	2	0	0	0.43	0.71	0.75	0.75	0.66	1.00	0.29	0.71
54	22	10	10	0	0	0	1.00	1.00	1.00	1.00	1.00	0.00	0.00	1.00
Avg(item)		4	5	0.41	0.18	0.77	0.64	0.80	0.66	0.66	0.69	0.55	0.24	0.77
Sum	22	90	116	9	4	17	N/A	N/A	N/A	N/A	N/A	12	N/A	N/A
Avg(all)	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	0.26	0.74

Table 6-11. Test Results for Spanish Mission Utterances (Speaker Model ARA03)

Syntax: Spanish Mission

Speaker Model ARA06														
Sentence No.	Utterance No.	Wors Correct	No. of Words	Words deleted	Words Inserted	Words substituted	Percent Accuracy					DARPA Statistics		
							Utterance	Word	Semantic	Translation	Average	Sent Err	Err	Corr
59	1	1	4	0	1	3	0.00	0.25	0.50	0.50	0.31	1.00	1.00	0.00
60	2	3	3	0	0	0	1.00	1.00	1.00	1.00	1.00	0.00	0.00	1.00
61	3	4	4	0	0	0	1.00	1.00	1.00	1.00	1.00	0.00	0.00	1.00
62	4	6	8	1	0	1	0.50	0.75	0.90	0.90	0.76	1.00	0.25	0.75
51	5	6	6	0	0	0	1.00	1.00	1.00	1.00	1.00	0.00	0.00	1.00
63	6	1	1	0	0	0	1.00	1.00	1.00	1.00	1.00	0.00	0.00	1.00
50	7	6	7	0	0	1	0.71	0.86	0.75	0.75	0.77	1.00	0.14	0.86
64	8	4	4	0	0	0	1.00	1.00	1.00	1.00	1.00	0.00	0.00	1.00
65	9	4	4	0	0	0	1.00	1.00	1.00	1.00	1.00	0.00	0.00	1.00
53	10	1	7	0	0	6	0.00	0.14	0.00	0.00	0.00	1.00	0.86	0.14
54	11	7	10	2	0	1	0.40	0.70	0.75	0.75	0.65	1.00	0.30	0.70
59	12	3	4	0	0	1	0.50	0.75	0.75	0.50	0.63	1.00	0.25	0.00
60	13	3	3	0	0	0	1.00	1.00	1.00	1.00	1.00	0.00	0.00	1.00
61	14	4	4	0	0	0	1.00	1.00	1.00	1.00	1.00	0.00	0.00	1.00
62	15	4	8	2	0	2	0.00	0.50	0.75	0.75	0.50	1.00	0.50	0.50
51	16	6	6	0	0	0	1.00	1.00	1.00	1.00	1.00	0.00	0.00	1.00
63	17	1	1	0	0	0	1.00	1.00	1.00	1.00	1.00	0.00	0.00	1.00
50	18	5	7	0	0	2	0.43	0.71	0.75	0.75	0.66	1.00	0.29	0.71
64	19	4	4	0	0	0	1.00	1.00	1.00	1.00	1.00	0.00	0.00	1.00
65	20	3	4	0	0	1	0.50	0.75	0.75	0.75	0.69	1.00	0.25	0.75
53	21	7	7	0	0	0	1.00	1.00	1.00	1.00	1.00	0.00	0.00	1.00
54	22	9	10	1	0	0	0.80	0.90	1.00	1.00	0.93	1.00	0.10	0.90
Avg(item)		4	5	0.27	0.05	0.82	0.72	0.83	0.86	0.85	0.81	0.45	0.18	0.79
Sum	22	95	117	6	0	16	N/A	N/A	N/A	N/A	N/A	10	N/A	N/A
Avg(all)	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	0.19	0.81

Table 6-12. Test Results for Spanish Mission Utterances (Speaker Model ARA06)

for the MAVT project.

In the test set of sentences, there was only one Source Lexicon error, 2 Functional Parse errors, and only 4 minor Target Lexicon errors (3 of which were the same error). The Parse, with 10 errors (7 major), and the Text Generation output, with 5 errors (3 major), had the most problems. (The Parse and Text Generation components were also the only components with errors in translating from Spanish to English as well as English to Spanish). These errors are summarized and classified in Table 6-14 and described in discussions of the appropriate tests below. The outcome is not surprising since the Parse and Text Generation components are the most complex and attempt to perform the deepest analyses of the input and output sentences respectively. The component averages for performance on well-formed input of each component shown in the last row of Table 6-13 also reflect this. The Parser has an average score of 0.91, the Text Generator has a average of 0.96, and all of the other components have averages of 0.99. Future MAVT development efforts in the area of natural language processing will concentrate on these two components. The result average, that is, the average score for each sentence, is 0.94. The average component performance (not shown in Table 6-13) is 0.97.

In addition to averaging the error rates of individual components for scored performance on well-formed input only, we also computed an overall average (shown in the second to the last row of Table 6-13) based on the total number of sentences. This average, which in effect counts the plus as a zero rather than discounting it, demonstrates the overall level of performance of the testbed at each point in processing and the degradation in the overall output. At the Source Lexicon stage, the average is 0.99; at the Parse stage it is 0.90; at the Functional Parse stage the average is 0.88; at the Target Lexicon stage 0.86; and finally at the Text Generation stage, the overall output average is 0.79.

The following paragraphs discuss the individual component tests in terms of the test and evaluation objectives.

6.4.1 Test #6
(Glass-box #3, English-to-Spanish text translation)
English text input => Spanish text output

This test is similar to test #2, with the exclusion of speech input and output. The purpose of this test is to measure the accuracy of translating English text into Spanish text. Success is determined by correctly translating the meaning of the English text into Spanish text.

Scoring: Full and Partial credit

- Correct lexicalization in NLP component
- Correct syntactic parse in NLP component
- Correct functional (semantic) parse in NLP component (source and destination)
- Correct text generated by translator

Final Test Results

Error Analysis for Natural Language Understanding and Generation

Sentence #	Lang. E/S	Source Lex	Prs	FP	Target Lex	Text Gen	Result
1	E	1.00	1.00	1.00	1.00	1.00	1.00
2	E	1.00	1.00	1.00	1.00	1.00	1.00
3	E	1.00	1.00	1.00	1.00	1.00	1.00
4	E	1.00	1.00	1.00	1.00	1.00	1.00
5	E	1.00	1.00	1.00	1.00	1.00	1.00
6	E	1.00	1.00	1.00	1.00	1.00	1.00
7	E	1.00	1.00	1.00	1.00	1.00	1.00
8	E	1.00	1.00	1.00	1.00	1.00	1.00
9	E	1.00	1.00	1.00	1.00	1.00	1.00
10	E	1.00	1.00	1.00	1.00	1.00	1.00
11	E	1.00	1.00	1.00	1.00	1.00	1.00
12	E	1.00	1.00	1.00	1.00	1.00	1.00
13	E	1.00	1.00	1.00	1.00	1.00	1.00
14	E	1.00	1.00	1.00	1.00	1.00	1.00
15	E	1.00	1.00	1.00	1.00	1.00	1.00
16	E	1.00	1.00	1.00	1.00	1.00	1.00
17	E	1.00	0.50	+	+	+	0.75
18	E	1.00	1.00	1.00	1.00	1.00	1.00
19	E	1.00	1.00	1.00	1.00	1.00	1.00
20	E	1.00	1.00	1.00	1.00	1.00	1.00
21	E	1.00	1.00	1.00	1.00	1.00	1.00
22	E	1.00	1.00	1.00	1.00	1.00	1.00
23	E	1.00	1.00	1.00	1.00	1.00	1.00
24	E	1.00	1.00	1.00	1.00	1.00	1.00
25	E	1.00	1.00	1.00	1.00	1.00	1.00
26	E	1.00	1.00	1.00	1.00	1.00	1.00
27	E	1.00	1.00	1.00	1.00	1.00	1.00
28	E	1.00	1.00	1.00	1.00	1.00	1.00
29	E	1.00	1.00	1.00	1.00	1.00	1.00
30	E	1.00	1.00	1.00	1.00	1.00	1.00
31	E	1.00	1.00	1.00	1.00	1.00	1.00
32	E	1.00	1.00	1.00	1.00	1.00	1.00
33	S	1.00	1.00	1.00	1.00	1.00	1.00
34	S	1.00	1.00	1.00	1.00	1.00	1.00
35	S	1.00	1.00	1.00	1.00	1.00	1.00
36	S	1.00	1.00	1.00	1.00	1.00	1.00
37	S	1.00	1.00	1.00	1.00	1.00	1.00
38	S	1.00	1.00	1.00	1.00	1.00	1.00
39	S	1.00	0.00	+	+	+	0.50
40	S	1.00	1.00	1.00	1.00	0.75	0.95
41	S	1.00	1.00	1.00	1.00	1.00	1.00
42	S	1.00	1.00	1.00	1.00	1.00	1.00
43	S	1.00	1.00	1.00	1.00	0.75	0.95
44	S	1.00	1.00	1.00	1.00	1.00	1.00
45	S	1.00	1.00	1.00	1.00	1.00	1.00
46	S	1.00	1.00	1.00	1.00	1.00	1.00
47	S	1.00	1.00	1.00	1.00	1.00	1.00

Table 6-13. Error Analysis for NLP Component

Sentence #	Lang. E/S	Source Lex	Prs	FP	Target Lex	Gen	Result
48	S	1.00	1.00	1.00	1.00	1.00	1.00
49	S	1.00	1.00	1.00	1.00	1.00	1.00
50	S	1.00	1.00	1.00	1.00	1.00	1.00
51	S	1.00	1.00	1.00	1.00	1.00	1.00
52	S	1.00	1.00	1.00	1.00	1.00	1.00
53	S	1.00	1.00	1.00	1.00	1.00	1.00
54	S	1.00	1.00	1.00	1.00	1.00	1.00
55	S	1.00	1.00	1.00	1.00	1.00	1.00
56	S	1.00	1.00	1.00	1.00	1.00	1.00
57	S	1.00	1.00	1.00	1.00	1.00	1.00
58	S	1.00	1.00	1.00	1.00	1.00	1.00
59	S	1.00	0.00	+	+	+	0.50
60	S	1.00	1.00	1.00	1.00	1.00	1.00
61	S	1.00	1.00	1.00	1.00	1.00	1.00
62	S	1.00	1.00	1.00	1.00	1.00	1.00
63	S	1.00	1.00	1.00	1.00	1.00	1.00
64	S	1.00	1.00	1.00	1.00	1.00	1.00
65	S	1.00	0.00	+	+	+	0.50
66	E	1.00	0.00	+	+	+	0.50
67	E	1.00	1.00	1.00	1.00	1.00	1.00
68	E	1.00	1.00	1.00	1.00	1.00	1.00
69	E	1.00	1.00	1.00	1.00	1.00	1.00
70	E	1.00	1.00	1.00	1.00	1.00	1.00
71	E	1.00	1.00	1.00	1.00	1.00	1.00
72	E	1.00	1.00	1.00	1.00	1.00	1.00
73	E	1.00	1.00	1.00	1.00	1.00	1.00
74	E	1.00	1.00	1.00	1.00	0.00	0.80
75	E	1.00	1.00	1.00	0.75	+	0.94
76	E	1.00	1.00	1.00	1.00	1.00	1.00
77	E	1.00	1.00	1.00	0.75	+	0.94
78	E	1.00	1.00	1.00	0.75	+	0.94
79	E	1.00	1.00	1.00	1.00	1.00	1.00
80	E	1.00	0.00	+	+	+	0.50
81	E	1.00	1.00	1.00	1.00	1.00	1.00
82	E	1.00	0.50	+	+	+	0.75
83	E	1.00	0.00	+	+	+	0.50
84	E	1.00	1.00	0.50	+	+	0.83
85	E	1.00	1.00	1.00	1.00	1.00	1.00
86	E	1.00	1.00	1.00	0.75	+	0.94
87	E	1.00	1.00	1.00	1.00	0.00	0.80
88	E	1.00	1.00	0.50	+	+	0.83
89	E	1.00	0.00	+	+	+	0.50
90	E	1.00	1.00	1.00	1.00	1.00	1.00
91	E	1.00	1.00	1.00	1.00	1.00	1.00
92	E	1.00	1.00	1.00	1.00	1.00	1.00
93	E	1.00	1.00	1.00	1.00	1.00	1.00
94	E	1.00	1.00	1.00	1.00	1.00	1.00
95	E	0.50	+	+	+	+	0.50
96	E	1.00	0.50	+	+	+	0.75
97	E	1.00	1.00	1.00	1.00	0.00	0.80
Overall Average		0.99	0.90	0.88	0.86	0.79	
Component Average		0.99	0.91	0.99	0.99	0.96	0.94

Table 6-13. Error Analysis for NLP Component

Sentence#	Lang E/S	Component	Score	Error Description
95	E	Prs	0.50	Lexical item missing
17	E	Prs	0.50	Conjoined Phrase
39	E	Prs	0.00	Predicate not parsed
59	S	Prs	0.00	'de' is preposition instead of genitive
65	S	Prs	0.50	Interjection
66	E	Prs	0.00	'so' is conjunction instead of adverb
80	E	Prs	0.00	Predicate not parsed
82	E	Prs	0.50	Purpose clause
83	E	Prs	0.00	Reduced relative
89	E	Prs	0.00	Purpose clause
96	E	Prs	0.50	Complement clause
84	E	FP	0.50	Subject label on interrogative
88	E	FP	0.50	Adv. phrase label on interrogative
75	E	Target Lex	0.75	Quant-noun gender agreement
77	E	Target Lex	0.75	Quant-noun gender agreement
78	E	Target Lex	0.75	Quant-noun gender agreement
86	E	Target Lex	0.75	Article-noun gender agreement
40	E	Text Gen	0.75	Date formulation
43	S	Text Gen	0.75	Date formulation
74	E	Text Gen	0.00	Existential "there"
87	E	Text Gen	0.00	No output
97	E	Text Gen	0.00	Complement clause

Table 6-14. Classification of NLP Errors

6.4.2 Test #7

(Glass-box #4, Spanish-to-English text translation)
Spanish text input => English text output

This test is like test #6 where we are testing text-to-text translation, excluding speech input and output. The purpose of this test is to measure the translation accuracy going from Spanish text input into English text output. Success is determined by correctly translating the meaning of the Spanish text input into English text output.

Scoring: Same as for test #6.

6.4.3 Test #8

(Glass-box #5, English text understanding)
English text input => English syntactic parse

In this test, the accuracy of the parse tree output by the syntactic analyzer for the English text is evaluated, including all linguistic data affecting accuracy of the parse tree, such as syntactic features in the lexicon. Accuracy is measured in terms of correctness of the syntactic bracketing of constituents and relations among constituents as represented in the parse tree.

Scoring: Full and partial credit

- Correct lexicalization in NLP component
- Correct syntactic bracketing of constituents
- Correct identification of syntactic relations among constituents

Source Lexicon Errors (Total: 1, English to Spanish)

All of the source lexicon information, for both English and Spanish depending on the language of the input sentence, that was necessary to process the test sentences was present in the lexicon and was correct, but one entry, "man-trap" (Sentence 95), was absent from the English lexicon.

Parser Errors (Total: 10; 7 English to Spanish)

Parser errors were scored in two different ways. If the parse output was complete, that is, missing no items of the sentence, and all structures were identified correctly, but the parse was made up of two or more partial parses instead of a single parse (that is, linking of the node structures was incomplete) then the score assigned is 0.5. However, if (1) there were partial parses AND an incorrect structure is chosen within the parse, or (2) the parse output was incomplete such that one or more items were missing from the sentence, then it is scored as 0.0.

Of the seven English to Spanish errors, four contained incorrect or incomplete parses (Sentences 66, 80, 83 and 89) and scored 0.0. Four others (Sentences 17, 82, and 96) contained partial parses but the node structure assigned was otherwise correct, and so these were assigned scores of 0.5. The latter four sentences contained constructions that were not intended to be fully covered in this version of the testbed: Sentences 17 and 95 have conjoined phrases, Sentence 82 has an infinitival purpose clause, and Sentence 96 contains a complement clause without the complement marker "that". These structures will be covered in future versions of the testbed (cf. the discussion of current and future coverage of the MAVT system in Section 7.0). Nevertheless, in the test version of the testbed, these structures were identified correctly or nearly correctly at the lower levels. Because the top-level tree relations were not yet in place, the parser was unable to link these structures into the whole tree. The text generation component, in fact, was able to work with the structures to produce a translation, with some syntactic irregularities, of the whole sentence.

Of the four parse errors scored as 0.0, in three cases (Sentences 66, 80 and 83), the output parse structure was missing elements of the predicate of the input sentence. In one case (Sentences 89), several structures, including an infinitival purpose clause and a quantifier phrase were parsed incorrectly.

6.4.4 Test #9
(Glass-box #6, English text understanding)
English text input => English functional (semantic) parse

The purpose of this test is to measure the accuracy of text understanding for English text. Success is determined by correctly mapping the input text into an appropriate semantic representation (in the DBG NLP component, the functional parse output, which represents predicate/argument relations as well as other semantic functions required to understand the given NL text).

Scoring: Full and Partial credit

- Correct lexicalization in NLP component
- Correct syntactic parse in NLP component
- Correct identification of utterance type, predicate/argument relations
- Correct assignment of predicate/argument indexes

Functional Parse Errors (Total: 2 English to Spanish)

Because the functional parse reflects the parse output quite closely, the functional parse is less likely to be a primary source of error. The majority of functional parse errors that do occur are found in the mislabeling or incomplete functional labeling of a word or phrase. The two functional parse errors in the test set are of this type. In Sentence 84, the subject "who" is labeled in the functional parse as a wh-expression but is not identified as the subject of the sentence. The lack of identification of a functional role makes it impossible to fit into the translated sentence, so the translation of 'who' ('quie'n'), though present in the functional parse, does not appear in the Spanish output because of the incomplete functional parse label. Similarly, in Sentence 88, the functional role (adverbial phrase) of the wh-expression 'how' (Spanish 'co'mo') is not labeled in the functional parse, so it does not occur in the translated output. Both the of these errors are the result of incomplete labeling, rather than mislabeling, so they have been scored as 0.5.

6.4.5 Test #10
(Glass-box #7, Spanish text understanding)
Spanish text input => Spanish syntactic parse

In this test, the accuracy of the parse tree output by the syntactic analyzer for the Spanish text is evaluated. Accuracy is measured in terms of correctness of the syntactic bracketing of constituents and relations among constituents as represented in the parse tree. Accuracy of all inputs (e.g., from lexicon) is also measured.

Scoring: Same as for test #8.

Parse Errors (Total: 10; 3 Spanish to English)

All of the three Spanish to English errors represented incomplete or incorrect/partial parses and so were scored as 0.0. In two instances (Sentences 39 and 65), the parse was incomplete. In the third case, a genitive structure rather than a prepositional structure was chosen to translate a Spanish prepositional phrase (the Spanish word 'de' can be translated both as "of" and "from" in English). and there were partial parses in the output.

6.4.6 Test #11
(Glass-box #8, Spanish text understanding)
Spanish text input => Spanish functional (semantic) parse)

This test is similar to test #9, except that it evaluates the semantic accuracy of Spanish text understanding. Success is determined by correctly mapping the input text into an appropriate semantic representation, as described in test #9.

Scoring: Same as for test #9.

6.4.7 Test #12
(Glass-box #9, Spanish text translation/generation)
English/Spanish functional parse input => Spanish text output)

This test measures the accuracy of the English-to-Spanish semantic translation and Spanish NL text generation. Test input will consist of an English functional parse, associated with Spanish translations of component words and phrases. The output will be the generated Spanish text.

The test has two parts.

1. word accuracy: i.e., have correct translations been provided for all of the English words and phrases in the utterance?
2. utterance accuracy: i.e., has the correct Spanish text been generated for the entire utterance?

Scoring: Full and Partial credit

- Correct functional (semantic) parse in NLP component (source and destination)
- Correct text generated by translator

Target Lexicon Errors (Total: 4 English to Spanish)

Of the four target lexicon errors, three (Sentences 75, 77, and 78) represent exactly the same error: improper gender agreement in the wh-phrase 'cua'ntas personas' ('how many persons'), which comes out in the test output as *'cua'ntos personas', in which 'cua'ntos' is masculine instead of feminine. (In a sentence, such as Sentence 76, where the noun 'oficiales' is masculine, the result is correct since 'cua'ntos' is masculine, i.e. 'cua'ntos oficiales.') The problem in these cases lies in the target language (Spanish) lexicon, which does not contain enough morphological information for the entry for 'cua'ntos', to allow a feminine form to be derived from the plural form 'cua'ntos.' If the proper feminine form had been derived, it would have been selected for the translation. This error is easily remedied by adding the correct morphological features to the lexicon. Because this error only affected agreement on one item in the sentence, it was scored as a minor error, 0.75 in each case. This error is found in Sentence 80 as well, but that sentence also contains a major error in the parse output and so is scored as a parser error instead.

The remaining error is similar. In Sentence 86, *'los comunicaciones' should have been 'las comunicaciones', taking feminine rather than masculine plural agreement. In this case, the definite article did not agree in gender with the noun it modified. The noun 'comunicaciones' ('communications') was not properly identified as feminine plural in the lexicon. In fact, this error was corrected in a later version of the testbed in which the output was correct. However, in the version used for the test, the correction had not been made. This error was also scored as 0.75.

Text Generation Errors (Total: 5; 3 English to Spanish)

The three English to Spanish text generation errors are major errors, two of which include structures that are not covered in the current testbed (cf. Section 7.0). Sentence 74 contains the existential expression "there are" in interrogative form ("How many enlisted men are there in your unit?"), which is translated literally as 'alli'', the locative "there" in Spanish. The Spanish existential 'hay' is not yet in the Spanish lexicon, because existential sentences represent a more specialized sentence type which will be covered in later versions of the testbed. In addition to this major error, which alone is enough to warrant a score of 0.0, there is a phrasal noun in the sentence which is incorrectly pluralized. The Spanish translation for "enlisted man" is 'soldado de tropa'. The plural morpheme -s in Spanish is added to the last word in the phrase, rather than the nominal head, 'soldado'. This is another error that is not included in this version of the MAVT Testbed.

Sentence 87 is the only sentence that did not produce any output at all; instead it hung up at the generation stage, although the analysis was correct up to that point. Sentence 97 contains a complement clause marked by "that". Complement clauses, which characterize more complex sentences, are not covered in this version of the testbed, although the parser and functional parse are able to analyze them correctly when marked by "that" (note that the previous sentence, in which the "that" is omitted, is not correctly parsed). Both of these sentences are scored as 0.0.

6.4.8 Test #13
(Glass-box #10, English text translation/generation)
Spanish/English functional parse input => English text output

This test is similar to test #12, except that it involves Spanish-to-English translation.

Scoring: Same as for test #12.

Text Generation Errors (Total: 5; 2 Spanish to English)

The two Spanish to English text generation errors, in Sentences 40 and 43, involve date formulations. In these sentences, the Spanish dates are translated literally, rather than idiomatically, into English. The result is that the year in English is preceded by the word "of" (a translation of the Spanish 'de' in the same position), whereas in English the year is normally stated with no preposition (i.e. "I was born the eleventh of April (*of) nineteen sixty-eight."). Because this error is very specific to date expressions, and because the translation is otherwise correct, these two errors have been scored as minor errors, i.e. 0.75.

6.5 Test and Evaluation of Speech Generation Subsystem

6.5.1 Test #14
(Glass-box #11, Spanish speech generation)
Spanish text input => Spoken Spanish output

This test provides a very rough measure of the quality of the computer-generated Spanish speech. A segment of text will be sent to the speech generator, and spoken. A Spanish-speaking listener will decide if the speech is intelligible, and provide comments (positive and negative) about the speech quality.

Scoring: Positive or negative evaluation (with comments) by native speaker.

Spanish Speech Evaluation

Overall comment by a native Spanish speaker: "Better than an American tourist." One crucial problem is that words that are not specifically spelled out for Spanish are given an English pronunciation (for further discussion of this see Section 7.3). This problem affected 24 words in the test sentences. In addition, the specific Spanish spellings which were intended to adapt the English speaker model to Spanish were only partially successful. Specific observations by the native Spanish speaker include the following:

- stressed vowels and stressed syllables in general sound too long
- diphthongs (e.g., eu, ay) sound too much like two distinct syllables
- intervocalic /d/ is over-fricativized; it should sound more like a stop
- word-initial /r/ is too "heavy," sounding more like a fricative than a tap
- word-final /r/ is sometimes not heard at all (as in "por")
- medial /rr/ is not sufficiently trilled
- word-initial /h/ is too glottal (this has been corrected for some words)
- postvocalic /l/ is retracted, as in English, rather than the Spanish front /l/
- word-initial /cu/ as in "cual" should sound more like a single phoneme /kw/ rather than a sequence of /k/ followed by /u/.

Using an English speaker model gives the speech generation component a kind of English (i.e., American) accent even when a diligent attempt is made to compensate by specifying Spanish pronunciations.

6.5.2 Test #15
(Glass-box #12, English speech generation)
English text input => Spoken English output

This test is similar to test #14, except that we are testing English speech generation. A segment of text will be sent to the speech generator, and spoken. An English-speaking listener will decide if the speech is intelligible, and provide comments (positive and negative) about the speech quality.

Scoring: Positive or negative evaluation (with comments) by native speaker.

English Speech Evaluation

No significant problems.

7. SYSTEM STATUS

The goal of the MAVT project as expressed in the SOW was to determine the feasibility of integrating state-of-the-art speech recognition, natural language processing, and speech generation capabilities to construct a multilingual voice-to-voice translation system, using a constrained interrogation scenario as the domain. In the course of this project, LSI developed a working prototype system--the MAVT testbed, as described in previous sections of this report and presently installed at Rome Laboratory. The MAVT testbed is a voice-to-voice translation system translating from English to Spanish and Spanish to English, with important additional features that make it a platform for future development, as well as a system that is potentially fieldable in the future.

One important feature is the fact that the speech recognition capability of the testbed is speaker-independent. This is extremely useful for interrogation in the field, where speakers previously unknown to the system would be interrogated using the system. This feature is especially valuable for the screening process, in which a large number of speakers would be interrogated in a short period of time in order to assess their usefulness for more in-depth interviews. With a speaker-independent system, it would not be necessary to train the system on the voices of each of the persons being interrogated.

Another key aspect of the MAVT testbed is that it has been designed not only as a working system, but also as a platform for future expansion of the system. Such future expansion is anticipated to be in the areas of more robust language coverage, extension to include additional languages and domains, and increased use of the system as a foreign language tutor. All of the components of the system have been assessed for, and where possible, built to accommodate such development. For example, as discussed in Section 5, the NLP processing components, including the lexicon, morphological processor, parser, and text generator, have been designed to be used for languages with a variety of lexical, morphological, and syntactic features.

In this section we discuss the present status of the testbed, in particular the current coverage of the speech recognition, natural language processing (NLP), and speech generation components. Also, the projected future coverage for the NLP processor is outlined. Examples of testbed performance on a basic set of sentences including some structures that are not covered at present can be found in Section 1.1.3, in the discussion of testing and evaluation.

7.1 Speech Recognition

In general, the speech recognition portion of the system is the most constrained part of the MAVT testbed. This is because, as described in Section 4.1, the perplexity of the grammar, which is determined in part by the number of rules and lexical items included in the speech syntax or grammar, has a direct impact on speech recognition performance. From the set of sentences generated by the speech grammar (in some cases, extremely large, as shown in Figure 6-6), the speech recognizer tries to select the sentence which most closely matches the spoken, input utterance. The larger the number of possibilities (the "branching factor") of the grammar, the less accurate the outcome; conversely, the fewer and more distinct the choices are, the faster and more accurate the performance.

To improve the performance of the speech grammar, we have partitioned it into separate biographics and mission-related grammars. These subdomains correlate with subcategories of interrogation as defined by our interrogation experts (described in Section 2). This incremental approach, i.e., using separate speech grammars, can be used in adding further interrogation material to the system, or in incorporating additional domains. In addition, we have developed a third "chatter" grammar, which contains material outside of the biographics and missions subdomains, such as might occur in miscellaneous comments made in the course of interrogation. As we expected, the wider range of the "chatter" grammar makes it much less accurate than the biographics and mission speech grammars. At present, the "chatter" grammar is experimental and has not been incorporated into the MAVT testbed. Such a grammar might be used as a default grammar, to be invoked at certain predetermined points in the discourse or when the other grammars do not produce a feasible output, as determined by the sense of the discourse and other factors.

To further ensure accuracy as mentioned above, the speech grammar has at present been limited to a relatively small set of lexical items. The English speech recognition grammar for the MAVT testbed contains a total of only 113 lexical items (66 in the biographics portion and 60 in the mission portion, with 13 items occurring in both portions), and the Spanish speech recognition grammar has 336 lexical items (238 in the biographics section and 113 in the mission section, with 15 items appearing in both). The reason that the Spanish grammar contains more items is that, unlike the English side of the interrogation scenario which is mainly questions, the Spanish side has dates and military unit designations, which contain sets of ordinal and cardinal numbers and other military unit terms. Also, Spanish nouns, verbs, and other modifiers (including ordinals) have greater morphological variation, and each different form must be represented separately in the Spanish speech grammar (e.g., both the masculine and feminine forms of modifiers).

For similar reasons, that is, in part because morphological variation cannot be derived productively within the simple rewrite grammars of the speech recognition component but must be represented by separate rules (for example, there are separate rules for masculine and for feminine forms), the Spanish grammar contain a total of 270 rules (212 biographics, 58 mission-related) whereas the English speech recognition grammar contains only 79 rules (28 biographics and 51 mission-related). We have attempted to keep the perplexity as low as possible while still adequately covering the basic interrogation scenarios defined by our consultants.

The current system is equipped with the basic English male and female speaker models that are supplied with the SSI recognition system, as well as a preliminary version of a Spanish male speaker model (see the discussion in Section 4.1.2). A goal for future development is to extend the Spanish speaker model to more accurately recognize a wider variety of Spanish voices and accents, and in addition to recognize utterances by Spanish female voices (which the present system does only to a limited extent).

7.2 Natural Language Processing

In contrast to the speech recognition component, where constraints on the grammar(s) are crucial, we developed the NLP portion of the system with the goal of being language-independent. The NLP covers the basic linguistic structures of both Spanish and English and has a capability to accommodate other languages as well. This means that the same set of of NLP components can be used for each language that is added to the system by means of adjusting certain parameters (such as head position, as discussed in Section 5.3). The NLP grammar also needs to be able to process sentences that the interrogator enters via keyboard, so not all of the input sentences are constrained by the speech recognition component. This means that the NLP must be more robust than the speech recognition grammars. Greater robustness also makes the system more useful for language training, because then a greater range of structures can be covered.

Although we have 30,000 lexical items available to our English NLP lexicon, the MAVT testbed contains a much smaller lexicon that has been constrained to relate to the interrogation domain. The current MAVT English and Spanish NLP testbed lexicons consist of just over 500 words each (508 for the English lexicon and 507 for the Spanish lexicon). As described in Section 5.1, the structure of the lexicon is generic in that it will apply to other languages, as well as English and Spanish. The morphology for both English and Spanish, discussed in Section 5.2, is essentially complete, and the system of stem and affix tables will handle other languages as well. In fact, all of the possible irregularities of Spanish morphology are captured in the morphological tables, even though not all of these are yet represented by actual lexical items in the Spanish NLP lexicon. (Although Spanish has a reputation as an "easy" language to learn, there are a great many exceptions to the basic morphological rules, and we took the time to incorporate all of the necessary morphological machinery to process these exceptions into the current testbed in order to facilitate future development.) The lexicon and morphology have now been developed to the point where the process of adding more lexical items for any language is a simple, straightforward process. For what we envision as a fully functional MAVT system, all aspects of the lexical structure and morphological processing are now 90 to 100% complete.

The NLP syntax (which is distinct from the speech recognition syntax described above and in Section 4), discussed greater detail in Section 5.3, includes the basic sentential structures of both English and Spanish. Not to attempt to include these basic structures into the natural language processor is a dangerous course, because it is difficult to constrain or prescribe the syntactic structures employed by a speaker. In addition, as mentioned above, although the speech recognition component acts as a kind of filter, that filter is not in operation when the user simply types a sentence into the system.

In the development of the NLP syntax, we started with a basic set of tree construction rules according to Government Binding principles for both Spanish and English (described in Section 5.3). Then we incorporated into the system those structures that were present in our corpus of interrogation scenarios. In order to make the system more robust and testable on previously unseen sentences, we developed sets of diagnostic sentences displaying core syntactic structures for English and Spanish. In part because we already had implemented many of the basic structures for English, the English coverage of the testbed is more comprehensive than the Spanish coverage. And because the syntactic structure of Spanish differs in certain ways from that of English--for example, adjectives follow rather than precede nouns, basic sentential word order of the subject verb and object is much freer, and Spanish has verbal clitics whereas English does not-- that are reflected in the structure of the parse tree, the language-specific parameters for construction of the Spanish parse tree needed to be ascertained and added to the multilingual NLP grammar. However, once implemented, we can build upon these parameters for Spanish in future efforts as we add certain particular Spanish syntactic structures that have not yet been covered, as discussed in the following sections. The next two subsections describe major structures that are covered in the NLP grammar of the current testbed, and structures that are not covered yet but are targeted for future efforts. This level of coverage also applies to the translation/text generation components, which cover approximately the same structures as the parser at any given stage of development. We estimate that the coverage of the parse, functional parse, and text generation components in the current testbed for English and Spanish is approximately 50 to 60% of what we would expect to see as the functionally effective coverage in a fielded system.

7.2.1 Structures Covered in the NLP Grammar

Major structures that are completely covered in the NLP grammar of the current MAVT testbed are listed below, together with sample input sentences in English and Spanish that the testbed translates correctly. Many of the sample sentences are taken from the list of English and Spanish NLP diagnostic sentences given in Appendix A of this report. These sentences were used as models for adding structures to the MAVT system and in regression testing of the system.

1. Basic subject-verb-object (SVO) sentences

    They speak English.
    The soldiers attacked the tank.
    You saw us.

    El comandante del primer batallo'n habla ingle's.
    No atacaron el puesto de comando.

2. Equational (copular) sentences (having the main verb "to be")

    I am the commander.
    Our mission is defensive.

    Esta es la segunda unidad.
    Es un voluntario.
    Nuestra misio'n es atacar el puesto de comando.

3. Genitive phrases (possessives; expressions with "of")

    They are the sergeants of the second regiment.
    What is the mission of your unit?
    What is your unit's mission?

    Cua'l es el nombre de su unidad?
    Nuestra misio'n es atacar el puesto de comando.

4. Indirect objects

The soldier told the commander his name.
I have told you my name.

El soldado le dije su nombre al comandante.
Le dije mi nombre al comandante.

5. Prepositional phrases

The soldiers near the road were attacked.
They are heading towards the command post from the road.

Varios soldados sin armas fueron capturados.
?Por que' su unidad se desplazaba hacia el sur?

6. Subject/object/possessive pronominalization

He is the commander.
I saw it.
I told them my name.
We are his parents.

Es italiano.
?Cua'l es su rango?

7. Noun modifiers (articles, demonstratives, simple quantifiers, adjectives)

I am the commander.
He was a soldier.
These three American units attacked.

Soy el comandante.
Una unidad se desplazo' al sur.
Estas tres unidades americanas se desplazaron al sur.

8. Tense/aspect (present, past, future/simple, progressive, perfect)

They speak English.	% present
They attacked.	% past
They will attack.	% future
They are attacking.	% present progressive
They were attacking.	% past progressive
They will be attacking.	% future progressive
They have attacked.	% present perfect
They had attacked.	% past perfect
They will have attacked.	% future perfect

Hablan ingle's.	% present
Se desplazaron.	% past
Se desplazara'n.	% future
Se esta'n desplazando.	% present progressive
Se estaban desplazando.	% past progressive
Se estara'n desplazando.	% future progressive

9. Passive voice

They are attacked.	% passive: present
They were attacked.	% passive: past
They will be attacked.	% passive: future
They are being attacked.	% passive: present progressive
They were being attacked.	% passive: past progressive
They have been attacked.	% passive: present perfect
They had been attacked.	% passive: past perfect
They will have been attacked.	% passive: future perfect.

Son atacados.	% passive: present
Fueron atacados.	% passive: past
Sera'n atacados.	% passive: future
Esta'n siendo atacados.	% passive: present progressive
Estaban siendo atacados.	% passive: past progressive
Estara'n siendo atacados.	% passive: future progressive
Han sido atacados.	% passive: present perfect
Habi'an sido atacados.	% passive: past perfect
Habra'n sido atacados.	% passive: future perfect

10. Imperatives (commands)

Tell me your name.
Indicate precisely your unit designation.

11. Negatives

I am not a sergeant.
I did not tell the sergeant my name.
We never attacked the American command post.

No soy un sargento.
No le dije mi nombre al sargento.
Nunca nos desplazamos.

12. Interrogatives (yes/no questions and wh-questions)

Do you speak English?
Did you attack the command post?
What is your personal mission?
Where were you born?

?Se estan reubicando a la derecha de su unidad?
?Co'mo esta'?
?Cua'l era su misio'n?
?A quie'n ataco'?

13. Number agreement

Those are the american soldiers.
The soldiers near the road were attacked.

Nosotros somos italianos.
Esos son los soldados americanos.

14. Gender agreement

She is my mother.
She wounded herself.

Nuestras misio'nes son defensivas.
Ellas son italianas.

15. Reflexives/clitics

I wounded myself.
They wounded themselves.

Se desplazaron.
?Se estan reubicando a la derecha de su unidad?

16. Cardinals and ordinals

The first regiment attacked.
The three units near the command post attacked yesterday.

Son los sargentos del segundo regimiento.
Tres unidades se deplazaron al sur.

17. Military unit names

The sergeant of the first battalion of the second regiment speaks English.

Mi unidad es bateri'a de defensa ae'rea de'cimo regimiento de
infanteri'a mecanizada de'cima divisio'n de infanteri'a mecanizada.

18. Dates

Naci' el once de abril de mil novecientos sesenta y ocho

7.2.2 Structures Not Currently Covered

A number of structures are not covered or only partially covered in the present version of the testbed. (Some of these structures appear in Table 6-4, the Error Classification Table for the test of the system described in Section 6). These include:

Modals. Modals are frozen verbal modifiers, such as can, should, etc., present in some languages. Unlike English, Spanish does not have modals as such. In Spanish, the same meanings are translated by main verbs (such as 'poder', "to be able") with the modified verb embedded in the matrix clause (main sentences) or by tenses, such as the conditional. At present, the system can translate a few very simple modal expressions from English to Spanish, but cannot pick the appropriate English modal based on the Spanish verb or tense. The complicated rules for this, which require matching English words to Spanish tenses or more complex constructions, will be worked out for a future effort.

Complex sentences (complement clauses, relative clauses, reduced relatives, other embedded clauses). The present MAVT testbed covers mainly simple sentences. A few complex sentences, primarily complement clauses in English introduced by "that", are covered. The coverage of complex sentences is in the process of being expanded for future versions of the testbed.

Existentials (English there is/are, Spanish 'hay'). These constructions have not yet been implemented.

Numeric figures (i.e., 1, 2, 3, etc.). Verbal cardinal and ordinal numbers have been incorporated, but the numerals have not yet been implemented. This is partly because the speech recognition component produces a verbal, rather than numeric, output. However, a numeral could be typed into the system, and so should be included in the NLP component.

Plurals of phrases. Plural assignment is currently done by adding the plural morpheme to the end of a noun. Sometimes when the noun is a phrase, such as 'soldado de tropa' "enlisted man", the head of the phrase is not at the end of the phrase. The NLP system does not yet "know" this, so the system attaches the plural morpheme inappropriately, e.g. *'soldado de tropas' instead of the correct 'soldados de tropa'. This problem will be remedied in later testbed versions.

Quantifiers/quantifier phrases. Phrases such as 'all of the X' and other quantifier expressions have not yet been included. Quantifier scope (e.g., whether a quantifier applies only to a noun phrase or to an entire sentence) also needs to be defined for both languages.

Conjunctions (and, or, but). Sentential and phrasal conjuncts can now be identified at lower levels, but how to adjoin them correctly at the highest levels of the tree is a thorny problem for the parser. We have already designed a look-ahead mechanism to accomplish this, however it remains to be implement and tested.

Spanish-specific structures. Among others, these include: the subjunctive mood, which occurs in complex sentences in Spanish; the 'a' marker for human objects; the three-way contrast among demonstratives (a two-way contrast is presently covered); the usage in Spanish of 'de' meaning both the genitival "of" and the prepositional "from" (this can be very confusing for translation into English); and the conditional tense, which corresponds to modals in English (cf. the discussion of modals above).

Discourse. Discourse comprises a set of phenomena, including register level and pronominal reference, which are discussed below. These phenomena need to be incorporated into a fully-functioning translation system for two reasons. One is that a complete and accurate translation is not possible without making use of features that are manifested only elsewhere in the discourse (e.g., pronominal reference) or that characterize the discourse as a whole (e.g., register level). Another reason is that discourse provides a powerful tool in selecting the most plausible sentence from among the N-best possible sentences that are passed to it from the speech recognition component.

1. Register Level. The interrogation scenarios that we have been using employ a formal register. When translating from English to a language where different forms of the language are used by the speaker to indicate the degree of formality of the speech, the MAVT system needs to pick the form appropriate for the speech register. An example can be seen in the formal vs. informal second person pronominal forms in Spanish. Thus, in an informal conversation (between friends, for example), the pronoun for the addressee is "tu" (related possessive adjective "tu", etc.), whereas in a more formal speech situation the pronoun for the addressee is "usted" (related possessive adjective "su", etc.). The NLP system should be able to make the appropriate choice of words according to the degree of formality of the current speech.

In some languages, the degree of formality of the speech is displayed on a larger scale, reflected not only in the choice of certain lexical items, but in the choice of the entire language variety, as in Arabic (Fergusen, 1963). Thus in a less formal speech situation, the speakers may choose to use a certain regional dialect, whereas in a more formal speech situation, standard Arabic will be used. Again, the MAVT system should have the ability to choose the language variety appropriate for the current speech situation.

2. Pronominals. Another discourse problem is recovering pronominal features from the context. In many cases, producing the correct translated text for the current sentence depends on the information contained in previously processed sentences. In particular, we need to keep track of the entities mentioned in the previous discourse in order to produce the correct translation.

For example, the interpretation of a pronoun in the source language and thus the correct choice of a corresponding form in the target language depends on the morphological features of the entities mentioned in the previous discourse. Consider the following example:

[Previous discourse:
- How many soldiers does your unit have?
- Ciento. ("One hundred.")]

-Can we see them from the road?
Translated text: "podamos verlos desde la carretera?"

When translating "them" into a target language which distinguishes the masculine form from the feminine form of "them", we have to know that "them" refers to the soldiers who were mentioned in the previous discourse, and since "soldiers" in Spanish is a masculine noun, the corresponding clitic form should be "los" (instead of "las") in the translation. Without such discourse information, the translation/generation NLP components would have no basis for choosing among the pronominal forms.

Conversely, Spanish doesn't distinguish the various forms of third person possessive pronouns as English does ("his, her, its, their" are all "su" in Spanish). Thus when translating "su" from Spanish to English in the following sentence, the NLP system has to have the number and gender information about the referent for the pronoun (which was mentioned in the previous context) in order to choose the correct pronoun for the output.

[Previous discourse:
- What is their mission?]
- Su mission es ...

Translated text: "Their mission is to ..."

A related problem is that in Spanish, inanimate objects have gender features, thus the forms corresponding to "he/him" and "she/her" (which refer only to human/animate entities in English) are also used in Spanish to refer to inanimate objects:

[Previous context:
- Why is the second unit moving to the South?]
- Porque sus soldados estan atacandola.

Translated text: "Because your soldiers are attacking it."

In the above Spanish sentence, the clitic "la" is feminine in gender, because "unidad" ("unit") is a feminine noun. For English, however, we need to generate "it" instead of "her" for this object pronoun, since English does not make gender distinctions among inanimate objects. This decision, again, depends on the system ability to access the discourse information about the referent of "la" ("the second unit") in the Spanish sentence.

7.3 Speech Generation

As discussed in Section 4.2, the main limitation on the speech generator used in this study is the lack of a speaker model for Spanish for the text-to-speech generation (TTS) technology. For the purposes of the MAVT testbed, we adapted the English model that was developed for the DECtalk synthesizer. This meant that, for Spanish, each word, including all inflectional variants of a given stem, needed to be entered separately using the rules for representing Spanish sounds shown in Table 4-1. Because the capability in the DECtalk speech synthesizer to include extra words was intended as a supplement only (for English), there is a built-in size limitation on the number of entries (a little over 300). Although adequate for demonstration purposes, this number is inadequate for any further development of the system, especially for a language like Spanish that has a large number of morphological variants for each word. The Spanish TTS entries currently in the testbed include for the most part only the limited number of words and forms of words used in our scenarios and in the test sentences (see Section 6). This is especially true of verbs and verbal auxiliaries, which are represented by just a few forms each, whereas each verb in Spanish has nearly sixty separate morphological variants. For future work, then, the availability of language-specific speaker models that "read" words according to the rules of the given language is crucial to being able to use the TTS (rather than the DAP) technology effectively.

8. REFERENCES

8.1 Contract Technical Documentation

The following documents were developed under this project:

[TIR] Language Systems, Inc., "Interrogation Scenarios and Background Material," *Machine Aided Voice Translation: Technical Information Report* (ELIN A002, CLIN 0002, Contract No. F30602-90-C-0058, Rome Laboratory), Language Systems, Inc., October, 1991.

[TP] Language Systems, Inc., *Machine Aided Voice Translation: Test Plan* (ELIN A003, CLIN 0002, Contract No. F30602-90-C-0058, Rome Laboratory), Language Systems, Inc., January, 1992.

[DP] Language Systems, Inc., *Machine Aided Voice Translation: Design Plan* (ELIN A004, CLIN 0002, Contract No. F30602-90-C-0058, Rome Laboratory), Language Systems, Inc., September, 1991.

[VTP] Language Systems, Inc., *Machine Aided Voice Translation: Video Tape Presentation* (ELIN A005, CLIN 0002, Contract No. F30602-90-C-0058, Rome Laboratory), Language Systems, Inc., May, 1991.

[VS] Shinn, P. C., *Speech Processing Vendor Survey*, LSI Technical Note 91-03, Language Systems, Inc., June, 1991.

8.2 Other Technical Documentation

[SSI-DS200] Speech Systems, Inc., *Speech Input Development System (Model DS200): Reference Manual*, version 3.6, Speech Systems, Inc., Tarzana CA, 1991.

[DECTALK] Digital Equipment Corporation, *DECtalk DTC01: Owner's Manual*, (EK-DTC01-OM-003), Digital Equipment Corporation, 1985.

[OW-USER] Sun Microsystems, Inc., *OpenWindows Version 2 User's Guide*, (part# 800-4930-10, revision A, 6/11/90), Sun Microsystems, Inc., 1990.

Additional References

Berwick, R. Principle Based Parsing. MIT AI Laboratory, Technical Report 972, 1987.

Brassey's Multilingual Military Dictionary. Brassey's Defense Publishers, 1989.

Bindra, A. English/Spanish translator works in real time. Electronic Engineering Times, April 27, 1992, p.25.

Chomsky, N. Lectures on Government and Binding. Foris, Dordrecht, 1981.

De la Cadena, M. V., Gray, E., Iribas, J. L. New Revised Velázquez Spanish and English Dictionary. New Century Publishers, Inc., Piscataway, NJ, 1985.

Everett, S. S., Wauchope, K. and Perzanowski, D. Talking to a Natural Language Interface: Lessons Learned. Proceedings of AVIOS 92: Voice I/0 Systems Applications Conference, pp.63-69. Minneapolis, Minnesota, September 22-24, 1992.

Fisher, W.M. et al., "The DARPA Speech Recognition Research Database: Specification and Status," Proceedings of the DARPA Speech Recognition Workshop, 1986.

Green, G, M, Pragmatics and Natural Language Understanding. Lawrence Erlbaum Associates, Hillsdale, NJ, 1989.

Grosz, B. J. and Sidner, C. L. "Attention, Intentions, and the Structure of Discourse". Computational Linguistics 12 (1986), 175-94.

Jackendoff, R. "Parts and Boundaries". Cognition, 41 (1991) 9-45.

Lehmann, C. Der Relitivsatz. Gunter Narr Verlag, Tubingen, 1984.

Levinson, S. C. Pragmatics. Cambridge University Press, 1983.

Montgomery, C. A., Stalls, B. G., Stumberger, R. E., Li, N., Belvin, R.S., Arnaiz, A., Hirsh, S. B. "Description of the DBG System as used for MUC-4" in Proceedings of the Fourth Message Understanding Conference (MUC-4), Morgan Kaufmann Publishers, Inc., San Mateo, CA, 1992.

Pallett, D. S., "Selected Test Material, Test and Scoring Procedures," Proceedings of the DARPA Speech Recognition Workshop, Cambridge, MA, October, 1987.

Price, P.J. et al., "The DARPA 1000-Word Resource Management Database for Continuous Speech Recognition," Proceedings of ICASSP, 1988.

Pustejovsky, J. "The Syntax of Event Structure". Cognition, 41 (1991) 47-81.

Sidner, C. L. "Focusing in the Comprehension of Definite Anaphora" in M. Brady and R. C. Berwick, eds. Computational Models of Discourse, Cambridge, MA: MIT Press, 1983.

Saito, M. Some Asymmetries in Japanese and Their Theoretical Implications. Unpublished Ph.D. dissertation, MIT, 1985.

Simon and Schuster's International Dictionary (English-Spanish/Spanish-English). Prentice Hall, New York, NY.

Webber, B. L. "Discourse Model Synthesis" in M. Brady and R. C. Berwick. eds. Computational Models of Discourse, Cambridge, NA: MIT Press, 1983.

Wheatley, B., et al., "Robust Automatic Time Alignment of Orthographic Transcriptions with Unconstrained Speech," Proceedings of ICASSP 1992.

A. DIAGNOSTIC SENTENCES

English Diagnostic Sentences for the MAVT Testbed

"I am the commander."
"They are soldiers."
"He was a soldier."
"He will be a soldier."
"They speak English."
"They attacked."
"They will attack."
"They are attacking."
"They were attacking."
"They will be attacking."
"They have attacked."
"They had attacked."
"They will have attacked."
"They are attacked."
"They were attacked."
"They will be attacked."
"They are being attacked."
"They were being attacked."
"They have been attacked."
"They had been attacked."
"They will have been attacked."
"I can move."
"He can move."
"A unit attacked."
"The unit attacked."
"This unit attacked."
"Our unit attacked."
"One unit attacked."
"Three units attacked."
"Several units attacked."
"The American unit attacked."
"The American units attacked."
"The second unit attacked."
"The first regiment attacked."
"This is my unit designation."
"This is my unit id."
"The soldiers near the road were attacked."
"The commander of our unit was attacked."
"These American units attacked."
"These three American units attacked."
"The three units near the command post attacked yesterday."
"The soldiers in our unit attacked yesterday."
"The commander of the first battalion speaks english."

"The sergeant of the first battalion of the second regiment speaks English."
"I went to the command post."
"You went to the command post."
"He went to the command post."
"She went to the command post."
"The soldier went to the command post."
"We went to the command post."
"They went to the command post."
"The soldiers went to the command post."
"The soldiers attacked the tank."
"The soldiers will attack the tank."
"They moved to the south."
"They will move to the south."
"I told the commander my name."
"The soldier told the commander his name."
"I saw it."
"You saw us."
"She saw them."
"We saw her."
"They saw you."
"He told me his name."
"I told you my name."
"I told him my name."
"He told us his name."
"I told them my name."
"I wounded myself."
"You wounded yourself."
"He wounded himself."
"She wounded herself."
"We wounded ourselves."
"They wounded themselves."
"The soldier wounded himself."
"The soldiers wounded themselves."
"I can tell you my name."
"I have told you my name."
"They were attacking him."
"I can tell you my name."
"I am a commander."
"You are a commander."
"He is the commander."
"She is my mother."
"We are his parents."
"You are commanders."
"They are American soldiers."
"They are the sergeants of the second regiment."
"They are my units."
"It is the second unit."
"This is our tank."

"These are our tanks."
"Tanks."
"Those are the American soldiers."
"I am American."
"You are American."
"He is Italian."
"She is Italian."
"We are Italian."
"They are Italian."
"Our mission is defensive."
"Our missions are defensive."
"This is different."
"Our unit is moving in that direction."
"Our vehicles are heading to the south."
"I was born in Cuba."
"They can see your tanks from the road."
"They are heading towards the command post from the road."
"We went to the command post."
"We attacked the tank for our commander."
"I can speak English besides Italian."
"We attacked the tank yesterday."
"We attacked the tank last month."
"We attacked the tank there."
"He speaks English clearly."
"We never moved."
"Speak clearly."
"Tell me your name."
"Tell the commander your name."
"Indicate precisely your unit designation."
"I am not a sergeant."
"I don't speak English."
"I did not tell the sergeant my name."
"They did not attack the command post."
"We never attacked the American command post."
"Our unit did not attack the command post at all."
"They will not attack the command post."
"He hasn't moved."
"Our unit is not moving."
"They will not be moving."
"They are not attacked."
"You were not being attacked."
"She wasn't being attacked."
"I did not tell him my name."
"We didn't attack them."
"We will not be attacking them."
"I don't think so."
"I do not think so"
"I don't know."

"I do not know."
"Do you speak English?"
"Did you attack the command post?"
"Will you attack the command post?"
"Is your unit attacking the command post?"
"Were you attacking the command post?"
"Will you be attacking the command post?"
"Who speaks English?"
"Which unit moved to the south?"
"Which units were attacked?"
"What is your unit designation?"
"What was your mission?"
"What were your missions?"
"What is the mission of your unit?"
"What is your unit's mission?"
"What is the current mission of your unit?"
"What is your personal mission?"
"What is the name of your unit commander?"
"What kind of vehicles do you have?"
"What other languages do you speak?"
"How many tanks do you have?"
"Why did the second unit move to the south?"
"Why did you attack the third unit?"
"When did your unit move to the south?"
"When were you attacked?"
"Where were you born?"
"Where is your unit located?"
"How are you?"
"How is she?"
"My father is Italian and my mother is Spanish."
"I can speak English and Spanish."
"My name is Miguel Martinez."
"My rank is Sergeant Second Class."
"One four seven four zero two five."
"Because the command post is moving in that direction."
"Because the command post was moving in that direction."
"Protect the regimental command post."
"Defensive."
"It was defensive."
"Very different."
"No."
"Yes."
"Where are the subordinate units located?"
"What are your alternate bases of operation?"
"Where are your alternate bases of operation located?"
"Give a sketch of the installations in your home base?"
"How many officers do you have?"
"Volunteers only."

"How many persons were killed?"
"How many officers were killed?"
"How many persons were wounded?"
"How many persons deserted your unit?"
"Which persons deserted your unit?"
"What happened to these people?"

Spanish Diagnostic Sentences for the MAVT Testbed

"Soy el comandante."
"Son los soldados."
"Era un soldado."
"Sera' un soldado."
"Hablan ingle's."
"Se desplazaron."
"Se desplazara'n."
"Se esta'n desplazando."
"Se estaban desplazando ."
"Se estara'n desplazando."
"Son atacados."
"Fueron atacados."
"Sera'n atacados."
"Esta'n siendo atacados."
"Estaban siendo atacados."
"Estara'n siendo atacados."
"Han sido atacados."
"Habi'an sido atacados."
"Habra'n sido atacados."
"Me puedo desplazar."
"Se puede desplazar."
"Una unidad se desplazo' al sur."
"La unidad se desplazo' al sur."
"Esta unidad se desplazo' al sur."
"Nuestra unidad se desplazo' al sur."
"Tres unidades se desplazaron al sur."
"La unidad americana se desplazo' al sur."
"Las unidades americanas se desplazaron al sur."
"La segunda unidad se desplazo' al sur."
"Esta es la identificacio'n de mi unidad."
"Varios soldados sin armas fueron capturados."
"Estas unidades americanas se desplazaron al sur."
"Estas tres unidades americanas se desplazaron al sur."
"El comandante del primer batallo'n habla ingle's."
"El sargento del primer batallo'n del segundo regimiento habla ingle's."
"Mi unidad es bateri'a de defensa ae'rea de'cimo regimiento
de infanteri'a mecanizada de'cima divisio'n de infanteri'a mecanizada."
"Le dije mi nombre al comandante."

"Le dije al comandante mi nombre."
"El soldado le dijo su nombre al comandante."
"El soldado le dijo al comandante su nombre."
"Eres el comandante."
"Es mi madre."
"Somos sus padres."
"Son comandantes."
"Son soldados americanos."
"Son los sargentos del segundo regimiento."
"Es la segunda unidad."
"Esta es la segunda unidad."
"Estos son nuestros tanques."
"Esos son los soldados americanos."
"Es un voluntario."
"Es una voluntaria."
"Son voluntarios."
"Son voluntarias."
"Soy americano."
"Soy americana."
"Eres americano."
"Eres americana."
"Es italiano."
"Es italiana."
"Nosotros somos italianos."
"Nosotras somos italianas."
"Ellos son italianos."
"Ellas son italianas."
"Nuestra misio'n es defensiva."
"Nuestra misio'n es muy distinta."
"Nuestras misiones son defensivas."
"Su misio'n era ofensiva."
"Nuestros vehi'culos se dirigen al sur."
"Naci' en Cuba."
"Nunca nos desplazamos."
"Nuestra misio'n es atacar el puesto de comando."
"No somos americanos."
"No soy un sargento."
"No hablo ingle's."
"No le dije mi nombre al sargento."
"No atacaron el puesto de comando."
"No creo."
"No se'."
"?Cua'l es su nombre?"
"?Cua'l es su rango?"
"?Cua'l es su nu'mero de identificacio'n militar?"
"?Cua'l es su carnet de identidad militar?"
"Uno cuatro siete cuatro cero dos cinco."
"?Cua'l es su cargo?"

"?Cua'l es el nombre de su unidad?"
"?Cua'l era su misio'n?"
"?Cua'les eran sus misiones?"
"?A quie'n ataco'?"
"?Que' tipo de vehi'culos tiene?"
"?Que' tipo de vehi'culos tienen?"
"?Cua'ntos tanques tiene?"
"?Cua'ntos tanques tienen?"
"?Por que' la segunda unidad se desplazo' al sur?"
"?Por que' ataco' la tercera unidad?"
"?Por que' atacaron la tercera unidad?"
"?Co'mo esta'?"
"Mi padre es italiano y mi madre es espan˜ola."
"Puedo hablar ingle's y espan˜ol."
"Tanques."
"Era defensiva."
"Defensiva."
"Defensivo."
"?Por que' su unidad se desplazaba hacia el sur?"
"?Se esta'n reubicando a la derecha de su unidad?"

☆U.S. GOVERNMENT PRINTING OFFICE: 1996-710-126-47001

MISSION

OF

ROME LABORATORY

Mission. The mission of Rome Laboratory is to advance the science and technologies of command, control, communications and intelligence and to transition them into systems to meet customer needs. To achieve this, Rome Lab:

a. Conducts vigorous research, development and test programs in all applicable technologies;

b. Transitions technology to current and future systems to improve operational capability, readiness, and supportability;

c. Provides a full range of technical support to Air Force Materiel Command product centers and other Air Force organizations;

d. Promotes transfer of technology to the private sector;

e. Maintains leading edge technological expertise in the areas of surveillance, communications, command and control, intelligence, reliability science, electro-magnetic technology, photonics, signal processing, and computational science.

The thrust areas of technical competence include: Surveillance, Communications, Command and Control, Intelligence, Signal Processing, Computer Science and Technology, Electromagnetic Technology, Photonics and Reliability Sciences.