[*MT evaluation: basis for future directions*. Proceedings of a workshop...2-3 November 1992, San Diego]

NEW DIRECTIONS: Evaluation of the Machine-Aided Voice Translation (MAVT) System[1]

Christine A. Montgomery, Bonnie Glover Stalls,
Robert E. Stumberger, Naicong Li, Robert S. Belvin, Alfredo Arnaiz
Language Systems, Inc.

1. Introduction

The MAVT project described in the presentation was the first phase of a prototype development to assist Air Force personnel in interacting with speakers of foreign languages. The project, which was begun in 1990 and concluded in 1992, resulted in the development of a speaker-independent continuous speech translation system for English↔Spanish. The MAVT testbed system developed under the project is comprised of three hardware/software subsystems: SSI's DS200 speech recognition system; LSI's DBG natural language processing system, which performs the language understanding and translation functions; and a DECtalk DTC01 speech synthesizer. The system is controlled by a single user interface residing on a Sun workstation, which also hosts the DBG software and the software components of the SSI recognition system. The testbed software is written in Prolog and C and runs under the Unix operating system. In 1992 the MAVT testbed system was installed in the Speech Laboratory/IRAA at Rome Laboratory, where it currently provides a demonstration of the component natural language processing and speech technologies.

2. Functional Description

A brief description of the MAVT testbed components and functions is given below. For a detailed discussion of all components and processes, as well as a complete account of test and evaluation, see Montgomery et al. (1992).

2.1 Speech Recognition

In processing, the speech recognizer takes a spoken utterance as input and produces a list of the n-best text-transcriptions of that utterance based on a speech grammar and phonetic dictionary. Through the user interface, the user selects one of those transcriptions (defaulting to the highest-scoring) or types in a new text input, and the selected text is passed into the language translator. The processing can also be performed autonomously, with the MAVT system taking the initial item in the n-best text-transcription list and passing it to the DBG NLP subsystem for understanding and translation. SSI's DS200 recognition system is equipped with English speaker models and a vocabulary of 30,000 active words. The Spanish speaker model was bootstrapped from the English male model using SSI's adaptation software. The speech grammars used for recognition of English and Spanish are partitioned into biographic (bio) and mission (msn) domains. These grammars contain mainly WH- and yes/no questions for English and responses to these question types in Spanish. Other types of interaction are relegated to "chat" grammars, which are more complex.

[1] The work reported in this paper was supported by AFMC, Rome Laboratory/IRAA, Griffiss Air Force Base, NY, under Contract No. F30602-90-C-0058. Continuing work is supported under Contract No. F30602-93-C-0098.

2.2 Natural Language Understanding and Translation Generation

LSI's Data Base Generation (DBG) system is the NLP component of the MAVT prototype; it is responsible for deriving the semantic representation of the orthographic text output by the speech recognition component and generating the corresponding text in the target language to be sent to the speech synthesizer. The DBG system is a syntax-driven natural language processing system that performs full-scale lexical, syntactic, semantic, and discourse analyses of text to produce a knowledge representation of the text that can serve as input to a downstream system or external data structure. For the MAVT application, a Spanish language lexicon and morphological component were added, as well as modules to produce translation output. The same syntactic and semantic processors are used for both languages.

2.3 Speech Synthesis

The DECtalk speech synthesizer is a hardware-based text-to-speech (TTS) system with a large English vocabulary, several male and female voices, and controls for pitch and speaking rate. It was adapted for Spanish using a set of orthographic representations approximating Latin American Spanish phones.

3. MAVT I Summary

To summarize the capabilities of the MAVT testbed system, the speech recognition subsystem contains the following numbers of lexical items and syntax rules in the listed grammars:

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

The DBG natural language processing subsystem, as configured for the MAVT testbed application, has over 1,000 lexical items for each language.[2]

The DECtalk Spanish speaker model for the MAVT testbed has approximately 400 words in its repertoire.

4. MAVT Testbed System Performance

The presentation reported on detailed test and evaluation of the MAVT testbed system, including three black-box tests of the overall system and 12 glass-box tests of the individual subsystems and components (e.g., Spanish speech recognition, English syntactic parsing, Spanish translation generation, etc.).

[2] For some English text understanding applications (e.g., MUC), the DBG lexicon contains approximately 30,000 items (see Montgomery et al. 1993).

Some examples of English and Spanish utterances contained in the 100-sentence test corpus are the following:

What is your military identification number?
Indicate your rank.
Do you speak Russian at all?
Were they repositioning to the south?
Is your mission offensive or defensive?

Mi rango es comandante en jefe.
Nací el once de abril de mil novecientos sesenta y ocho.
Era defensiva.
Su misión es encontrar unidades americanas.
Mantener la paz.

Detailed analyses of system performance for the total spoken language understanding and translation system, as well as individual components and modules, are contained on Montgomery et al. (1992). In this report, accuracy of performance is assessed from various perspectives, including functional or translation accuracy.

To summarize, for the whole MAVT I testbed, including NLP components, average accuracy overall was 87%. For the NLP components only, average accuracy was 94%. For the speech recognizer only, average recognition accuracy for English was 88%, while Spanish was 81.3%, with overall recognition accuracy being 84.6%.

5. MAVT Advanced Development Model

The second phase of MAVT development began in 1993, aiming at the construction of a prototype suitable for hands-on experimentation by Air Force end users in an office environment. In addition to Spanish, the MAVT ADM will also include translation from English to Arabic and Russian, and from these languages to English. The development objectives include a vocabulary size of 1,000 items per language and coverage of a core set of sentence types for all languages. Our development approach is based on an interlingual knowledge representation, rather than the concept of a direct transfer between languages.

The MAVT ADM speech processing components are licensed from Entropic Research Laboratory. These are the HTK (Hidden Markov Model Software Toolkit) speaker-independent, continuous-speech recognizer, which will be used for all four languages, and the TrueTalk speech synthesizer, which has software for English and Spanish. Synthesizers for Arabic and Russian will be developed or acquired in the course of the project.

In the current MAVT ADM work, the DBG system is being extended by the incorporation of Lexical Conceptual Structures (CS) as defined in Jackendoff (1990) and Dorr (1993). The LCSs will serve as the language-independent portion of the knowledge representation to provide a basis for interlingual translation. As in the testbed, the basic linguistic analysis and generation processes (e.g., lexicalization, syntactic and semantic processing) will be the same for all languages.

The MAVT ADM system is written in C, C+ +, and Prolog and runs on a Sun SPARCStation 5 under Unix.

References

Dorr, B.J. 1993. *Machine Translation: A View from the Lexicon.* Cambridge (Mass.): MIT Press.

Jackendoff, R.S. 1990. *Semantic Structures.* Cambridge (Mass.): MIT Press.

Montgomery, C.A., E.G. Stalls, R.E. Stumberger, N. Li, R.S. Belvin, A. Arnaiz, P.C. Shinn, A.G. De Cesare, and R.D. Farmer. 1992. Machine-Aided Voice Translation (MAVT). Final Technical Report for Contract No. F30602-90-C-0058, Rome Laboratory, Griffiss AFB, N.Y.

Montgomery, C.A., E.G. Stalls, R.E. Stumberger, R.S. Belvin, N. Li, and S. Hirsh. 1993. Description of the DBG System as Used for MUC-5. *Proceedings of the Fifth Message Understanding Conference.* San Francisco: Morgan Kaufmann. pp. 121-135.