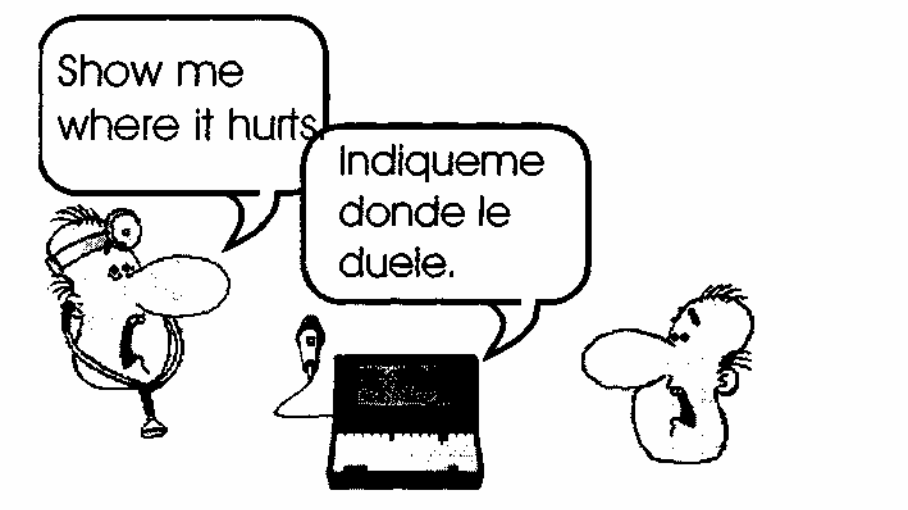
[*Translating and the Computer 22*: Proceedings of the Twenty-second international conference…16-17 November 2000. (London: Aslib, 2000)]

A Voice-Enabled Phrase-Based Translation System

Jay Tucker
Dragon Systems, Inc.
320 Nevada Street
Newton, MA 02460
USA
jayt@dragonsys.com

Ace Sarich
VoxTec, Inc.
1571 St. Margarets Rd.
Annapolis, MD 21401
USA
ace@sarich.com



Introduction

In our ever-shrinking world, the need to communicate with individuals with whom we do not share a common language is an ever-increasing reality. For people in numerous professions such as emergency medical care, law enforcement, and travel and tourism, this need is already present on an almost daily basis. The voice-enabled Phrased-Based Translation System (PTS) developed by Dragon Systems, Inc. in conjunction with VoxTec, Inc. provides a means of bridging the gap between speakers of different languages. Field experience with this system has demonstrated that even relatively simple translation technology can perform a surprisingly useful task.

Why phrased-base translation?

At its most basic level, the PTS can be thought of as a computerized phrase book. The system reads in a data file containing a list of phrases (the "module") relevant to a given topic. There is no real technical limit to the number of phrases in a given module, but real-world modules which have already been developed and deployed tend to range from several hundred to several thousand phrases. The largest module consisted of some 4000 phrases. Since most users of that module found it difficult to deal with that many phrases at once, we recommend that modules be limited to no more than about 1000 phrases.

After the module has been loaded, the user speaks one of these phrases in his own language into a microphone. The speech input is analyzed and interpreted by

Dragon's large-vocabulary continuous speech recognition engine and converted into text. The PTS compares this text to the list of known phrases. If a match is found, it plays a digital recording of that phrase in the selected target language. If no exact match is found, the system can use a "best guess" algorithm to find the known phrase closest in meaning to the user's spoken input phrase and then to prompt the user for confirmation of its guess. If the user confirms, then a recording of the "best guess" phrase is played. The perceived effect of the system is that of a true voice-to-voice translator with natural language understanding.
Like a tourist phrasebook, the PTS is designed with primarily one-way communication in mind. It can translate only in one direction: from the source language to the target language. This may at first appear to be a crippling limitation since spoken communication is normally a two-way process. There are, however, situations where this is all the functionality that is required — such as during a simple medical examination. The PTS can be used to convey information not requiring any sort of a response such as *I am a doctor, You have a fever, I must examine you,* or *This injection will help you feel better.* The PTS can also be used to issue instructions such as *Stand up, Point to where it hurts,* or *Take a deep breath* to which no response (other than compliance with the instructions) is required from the listener.
At other times, one would like to question a listener and get an answer. Even in these cases, the PTS often can be used to good effect as long as the questions are carefully posed. This can be often be accomplished by phrasing questions so that they can be answered with *yes* or *no,* for instance, *Have you had an accident?, Have you lost consciousness?, Have you eaten today?* Listeners can be prompted to give an appropriate response by phrases like *Nod your head like this for yes* and *Shake your head like this for no.* Negative questions of the type *You haven't eaten today, have you?* should be avoided since the meaning of the responses using the words *yes* and *no* differs between languages.
Other questions can be phrased so that they can be answered with a number: for instance, *How old are you?, How many days have you been feeling sick?* As above, listeners can be prompted to give an appropriate response by phrases like *Hold up the number of fingers* or *Write the number here.*
Still other types of information can be acquired by using props along with the appropriate questions such as *Point to your city on this map* or *Point to the picture of any object you've seen in this* area. Even these simple forms of questions and answers can permit the user to conduct surprisingly detailed interviews.

Software Components

Speech Recognition

Speech input to the PTS is analyzed and transcribed into text by Dragon's large-vocabulary continuous speech recognition engine and converted into text. This speech engine is the same one used in Dragon's commercial product, NaturallySpeaking. Although the PTS was originally developed for users whose primary language is American English, it could in principle be easily reconfigured to accept input in any of the other languages in which NaturallySpeaking is available; currently, those include British English, French, Italian, German, Spanish, Dutch,

Japanese, and Mandarin Chinese. In practice, due to the limited length and number of the phrases the system can translate, speakers of British English as well as other national variants of English can often achieve satisfactory performance using a system set up for American English.

Translation

In reality, the system does not perform any actual translation. The well-known shortcomings of machine translation — primarily its inaccuracy and unreliability — have been circumvented by simply not using machine translation. All phrases are translated in advance by a native speaker of each desired target language. Back translation is used during the module development process as an additional step to ensure accuracy.

Although seemingly overly limited, this approach offers several advantages. Not only does the PTS eliminate the uncertainties of machine translation, but it can also be more reliable and consistent in its performance than a human translator, who may inject his or her own prejudices or agendas into a translation or otherwise alienate interviewees. This has proven to be a particular problem with military translators. For instance, ethnic Bosnian interviewees are often unwilling to speak either to or through an ethnic Serb translator (and vice versa). Additionally, the PTS can be taken into areas of military conflict where civilian translators cannot or would not want to go.

Another advantage stems from the fact that the PTS is often used to translate into less common languages such as Albanian or Hmong for which machine translation is unlikely to exist in the near future. Such languages present no difficulty to the PTS since the translation approach used eliminates any need for either machine translation into or synthesized speech for target languages.

There is no limit to the supported target languages other than the need to find a native speaker of each desired target language to do the translations and recordings. Similarly, since the PTS is entirely data-driven, the speed with which a new set of phrases or new target languages can be incorporated into the system is limited only by the speed with which the new set of phrases and/or recordings are made available.

Voice Output

No speech synthesis is used in the PTS. All the translated phrases are prerecorded by a native speakers and stored as digital audio files in the standard WAV format. Once the PTS has identified the spoken input phrase, it plays the recording of the same phrase in the desired output language. Foreign language speakers listening to the PTS thus hear the familiar sound of actual native speech as opposed to the often alienating and difficult to understand sound of synthesized speech.

Moreover, this approach offers the advantage for uncommon languages of eliminating the need for available speech synthesis in those languages. As noted in the section above, even languages such as Albanian or Hmong present no difficulty for the PTS.

Hardware Platform

The PTS currently runs on Windows 95/98/NT-based computers. The hardware platforms of choice have been mini-laptops and tablet computers due to their small size. The current version of the application program is written in Java. We expect to take advantage of Java's platform independence and to port the system to even smaller, handheld platforms in 2001.

Operational History

The origins of the PTS go back to the Gulf War. A large, unexpected problem encountered by the Allied Forces towards the end of the conflict was the need to communicate with the very large numbers of Iraqi prisoners of war. After returning from the Gulf, Dr. Lee Morin, a reserve member of the U.S. Navy medical corps, created the Medical Translator, a simple program with a point-and-click interface that allowed its user to conduct medical interviews.

Anticipating a similar situation in Bosnia, the Defense Advanced Research Projects Agency (DARPA) asked Dragon Systems to add a voice interface to the Medical Translator. Dragon complied and then rewrote the Medical Translator, adding many new features in the process. The result was the first version of the PTS, which first saw operational use in Serbo-Croatian with the US forces assigned to the UN in Bosnia, where it was used primarily to help both UN forces and the local population to locate and clear minefields. One of the advantages of the PTS surprisingly did not involve communication in the conventional sense at all. Rather, the soldiers' use of the system was a source of fascination to local civilians, especially the young, who were drawn into better relations with the soldiers because of it and thus became more valuable sources of information.

A more recent version of the PTS was used in the Persian Gulf aboard US Navy ships enforcing UN sanctions against Iraq. This version contained approximately 350 phrases related to ship-boarding and searching in the most common languages used aboard merchant vessels in the Persian Gulf, namely, Arabic, Farsi (Persian), Hindi, and Urdu. In response to the troubles in Kosovo, an Albanian version has been developed. The PTS has also been used in several humanitarian relief exercises, most recently during the Strong Angel exercise, which took place in June 2000 in Hawaii (http://www.strongangel.org). More information on the PTS can be found at http://www.sarich.com/translator.

Acknowledgements

Development of the Voice-Enabled Phrased-Based Translation System was funded primarily by the Defense Advanced Research Projects Agency (DARPA).