

Dictating Progress to Machines

REPORT BY
ANDREW JOSCELYNE

Now that speech-recognition technology has become affordable and PC-compatible, automatic dictation is within reach of translators. Language International asked a number of users about their experience with this new productivity tool.

Remember HAL, the talking computer in Stanley Kubrick's "2001"? Although writer Arthur Clarke set his famous story in 1997, real-life progress to date in speech processing has so far proved a little slower than fiction. Computers will not yet answer you back as HAL did, but we can talk to them and get them to take dictation. And as the two main general-purpose dictation software suppliers IBM (with VoiceType) and Dragon Systems (with DragonDictate) have discovered, there is an emerging market for heavy-duty text-workers ready to use an alternative to key-boarding—among them language professionals.

Key initial users were medical staff needing to enter data to formatted documents without using their hands, and lawyers drafting letters and contracts. But 1996 seems to have marked the take-off of the translator market for dictation software. Growth has been partly pulled by freelancers retooling with Pentium chips (133 MHz micro-processors and 24 Mb of RAM seem to be a required minimum for dictation systems) and larger hard disks (compressed sound files can be bulky), and partly pushed by the net price drop for the systems themselves. More accurate product information and a natural decline in primary technophobia in the profession have also played a role in boosting interest. As a result, there is a growing body of useful experience about the dos and don'ts of automatic dictating, together with strong partisanship over products.

SPEECH MODEL

Apart from a few very specific contexts such as listing numbers, computers cannot yet recognize continuous speech uttered at a natural rhythm. But they can manage text spoken as a string of words. For a computer to recognize your speech and take down your dictation, therefore, it first has to create

a model of your way of pronouncing words. It then uses this model as a benchmark against which to match your actual dictation input. This means that when you dictate to a computer, you have to speak with brief pauses—between—your—words, giving the machine time to rapidly check your input against a list of stored forms on the basis of your speech model. (A single workstation equipped with dictation software can, of course, store several different speech models, one for each different user.) In short, automatic dictation systems require a behavioral adaptation that can seem quite unnatural to the novice. Given the inherent strangeness of the technology (and its reputation for making language mistakes) what, then, has driven translators to begin using it?

Reasons vary from one case to another. UK Italian/French-to-English translator Michael Benis was looking for a productivity tool when he started using IBM's VoiceType two years ago. Already an experienced dictator, he noticed that his secretaries "often failed to spell technical terms correctly, so I had to do a lot of careful correcting afterwards." For France-based Shankaran "Vishwa" Viswanathan who translates both into and out of French and English, taking up VoiceType helped him improve his slow typing of French. Roger Fletcher, the doyen of dictation software users in the UK and a long-time Chinese-to-English translator, took to VoiceType as a "natural" method immediately after he tried it, even though he is an 80 wpm typist and didn't expect much gain in input speed. On the other hand, for Tom Barkas, another English translator, the problem was more serious: like many PC-based translators who have never learned to type properly, he found he was suffering from Repetitive Strain Injury (RSI) after 10 years of bashing at the keyboard, and turned to

T
r
a
n
s
l
a
t
i
o
n

dictation software as a serious replacement technology for doing his work.

Dictator-Friendly

As suggested, the initial stage in the learning process is providing the system with your speech model. The first commercial versions of both VoiceType and DragonDictate required a 30-minute start-up phase or longer for the user to speak in a set of sentences that were then "compiled" by the machine into an individual profile. Today, VoiceType 3.0 and DragonDictate 2.5 (both for Windows) have reduced this start-up time to next to nothing.

**To use dictation effectively,
you must master a new
foreign language—
"paused" speech that the
computer can understand.**

But the key stage for novices is mastering the art of paused dictating. Since Vishwa had no experience with dictation, he found the whole process of learning to use VoiceType fairly difficult, especially as he was trying to meet tight deadlines. Michael Benis concurs: "Many translators find they can only think on the written page. They had better be sure they feel comfortable with dictating in general before getting dictation software." Yet the evidence suggests that if you persevere sufficiently, paused speech is just one more masterable skill. Roger Fletcher now feels that his dictation speech is getting "near continuous." He does, however, notice that a lot of beginners tend to "over-exaggerate the word-pause requirement and end up breaking them into syllables." What seems to be required is the capacity to develop an easy, steady rhythm. Until dictation methods form part of the typical translator-training program, teams of translators with typing skills of 80 wpm working on software localization projects might not be natural candidates for dictation-software technology.

Building the Word Knowledge Base

Although the user's voice profile is established once and for all at the beginning, the machine is continually learning from a translator's usage, and this offers a key advantage to the technology. The systems come with built-in dictionaries (some 130,000 for DragonDictate and 40,000 words for IBM) plus various facilities for developing additional personal and domain dictionaries. Words which tend to be statistically rare in the language but very frequent in a certain corpus of texts—esoteric technical terms, product names, in-house preferences, etc.—

can be entered into the system just once and recognized for ever after. Rather like memory technology for automatic translation, automatic dictation systems seem better adapted for repetitive work on the same subject matter or lexical domain, than on unrelated texts from many different domains. As with any computer-based utility, it is necessary to build up that knowledge base by sheer repetition. Vishwa, for example, uses dictation software for translating a series of press releases from the same firm: "I can get real productivity gains here since the machine always recognizes the technical words, whereas it can easily make mistakes over ambiguous simple common words in the language."

For Michael Benis, however, the idea that dictation software is better suited to some types of texts than others is "all nonsense." He started off with ideas about what dictation software was good at, but has reached the conclusion that "the best way to use it is for increasing

(continued on page 16) ▼



Shankara
Viswanathan



Roger Fletcher

DragonDictate Successfully Road-Tested at European Commission

During the first months of 1996, a team of 15 translators at the European Commission's Translation Department (SDT) in Brussels put DragonDictate version 1 through rigorous testing over several months. And the software came through with flying colors. Said Tim Cooper of the IT Support and Training section in charge of evaluating the system, "We are convinced that automatic dictation is the way forward." Testers included a mix of computer literates and those without keyboarding skills, and ranged over five of the EC's 11 official languages. The SDT hopes to upgrade to Windows NT in the near future, and plans to implement automatic dictation support for any of the 600 translators that request it.

The SDT did encounter a problem worth underlining for international organizations where multiple languages are processed. To avoid maintaining an 11-language helpdesk and product support for 11 versions of its in-house word-processing package (WordPerfect), the SDT uses the UK version as the single backbone. Naturally, DragonDictate comes in a different version for each language and automatically assumes that you are using the national (and latest) version of the word-processing software. For example, the conflict caused by using French DragonDictate with UK English WordPerfect meant that the dictation system would interpret the English speech macro for bold (Control + B or "Bold") as the wrong command (it would normally be Control + G or "Gras" in French).

productivity in any text at all, provided it is seen as a first draft." Although over time the rate at which the system recognizes the right words seems to improve (Roger Fletcher reckons he gets 98-99% for a large number of texts), a translator always has

Contrary to what nonusers might think, dictation systems adapt easily to highly formatted text such as tables, and translators can dictate into existing word-processing source files that arrive by network or on diskette.

to check for misrecognized words in the dictated text, now made easier in VoiceType by the playback feature, which allows users to listen to what they have input while reading the text. Erroneously recognized words stand out distinctly. Benis uses this stage "to check that I haven't left anything out." He then word-processes the resulting draft to finalize it for delivery.

ON-SCREEN CORRECTION

In normal mode, dictators great or otherwise have to look at the screen as they work, since the machine will display homophones ("brake" or "break") and ask for confirmation of what you want, and will also need to have words not found in its native dictionaries typed in. Users of IBM's VoiceType however, claim that they do not have to look at the computer screen while dictating—the machine just records everything, wrong or not. This runs the risk of missing certain errors and therefore reducing the machine's ability to continually learn from its mistakes. However, Michael Benis sometimes prefers to "sit back and rattle on merrily, even gazing around the garden while finishing off a sentence." Roger Fletcher, since he spends a lot of time translating from faxed pages of Chinese characters, finds he can concentrate on the often poorly printed source text while translating, without having to check on the screen output and lose his place. Obviously the choice between correcting as you go, and working fast but blind and editing the translated version extensively later is an individual productivity issue.

MACROS

Recent versions of both VoiceType and DragonDictate offer full editing control over documents as well as the capacity to command the com-

puter to carry out various nonediting commands such as opening files, switching to other applications and so on, showing the way towards full speech control over the interface. In terms of text management (and contrary to what nonusers might think), dictation systems adapt easily to highly formatted text such as tables, and translators can dictate into existing word-processing source files that arrive by network or on diskette. To handle a number of repetitive editing tasks, voice macros can be created to automate such commands as "bold previous paragraph" or "title 3," or insert certain lexical items (e.g., saying "em-em" and having the software automatically convert your speech input into the whole word "millimeter"). Interestingly, users distinguish commands from text input by speaking in continuous rather than paused speech. Although macros can theoretically manipulate the cursor for complex formatting, users such as Michael Benis find that jumping around between cells or labels in drawings is probably marginally quicker using a mouse.

The next logical step in translator-technology convergence would be to integrate dictation software into other productivity components in the standard tool set. In a recent

(continued on page 17) ▶

Talk About Speech Recognition

Information about speech-recognition products in general and dictation systems in particular is available at a wide variety of Web sites (use your favorite search engine). For example, it is worth finding out the latest language versions for these fast-evolving products (DragonDictate and VoiceType both currently offer five or six European languages). Besides IBM and Dragon, Philips and Kurzweil both ship speech-recognition products, but do not target the translator market. The Eldorado for all these companies is, of course, multispeaker continuous speech recognition, allowing users of all kinds to interact with information technology by voice alone.

As for the specific case of dictation software for text and language professionals, a large body of user experience is currently being gathered and will allow associations, service companies, and information circles to properly assess dictation software's capabilities. In the UK, Trevor Holloway is said to be compiling a comprehensive comparison between rival systems for the ITI. Information may well be posted on the ITI site in due course. Useful forums on CompuServe include FLEFO and VOICETYPE, where set-up and technical questions are treated at length. *Language International* editors would naturally be interested to receive any feedback on the evolution of the dictation-software user community.

Dictation (continued from page 16)

test, users at the European Commission Translation Department (see sidebar) have successfully integrated DragonDictate with translation workbench systems, using voice commands to switch to dictionary look-up functionalities. Ironically, technological progress in automatic dictation is becoming a victim of its own success. Melvin Hunt of Dragon Systems UK himself points out that the distinction between continuous-speech macros and paused-speech text input is gradually blurring as the recognition capacity for near-continuous speech increases almost by the month. Agreed “stop” words will no doubt be used to separate dictation input from machine commands.

GET A GOOD MIKE

Many nonusers wonder how dictation software systems handle the problem of ambient noise—can one actually work properly in an office full of busy dictating translators, phone calls, and the shrill music of electronic gadgetry? In fact the problem has been largely solved by the use of so-called noise-canceling microphones that receive only the input speech of the dictator and nothing else. But microphone quality and control is absolutely vital. Vishwa found the microphone bundled with his version of VoiceType was totally inadequate, since he could not position it correctly at the corner of his mouth to avoid breath interference. Michael Benis expects that the new (somewhat flimsy) Andrea microphone will allow you to dictate effectively even with background music on. He notes, however, that “careful positioning can significantly increase recognition accuracy.”

GREAT PRODUCT.

WHERE'S THE SUPPORT?

Like any complex engineering product, dictation systems need after-sales support. When it comes to product support, at least for VoiceType, there is massive agreement on how bad it has been. The general feeling is that although IBM has put tens of person-years into the design and testing of its successful system, the company has underestimated the need for serious helpdesk facilities. Yet product support it certainly needs, as Tom Barkas found out at his own expense. He notes that users should take with a large grain of salt the third word in the small print that says “works with most Pentiums.” After installing the hardware, each time he managed to dictate over 200 words into a file, his version of VoiceType would freeze, and it has taken him several months of calls, discussions, re-installations, and fruitless conversations with help staff to learn that the program may not work with all systems. For a working translator suffering from RSI seeking to quadruple his output, this is a truly painful way to waste time.

Michael Benis would agree that dealer support (at least in the UK) has been “hopeless,” but that overall, installation problems are on average no worse than for other information technology. Most users find that the dedicated speech-recognition product forums on CompuServe or the Internet in general are by far the best source for advice on system bugs. The growth of dedicated user groups and perhaps serious evaluations of intensive usage will further encourage uptake. So remember folks, big brother HAL is now prepared to listen to you again.

*Andrew Joscelyne is a General Editor
of Language International.*