THE RETURN OF LOW-LINGUISTICS MT

PC-TRANSLATOR: CAN A LOW COST MACHINE TRANSLATOR DO THE JOB?

Linguistic Products, a Texas-based company, has been developing and selling PC-based machine translation systems since 1984. The company's current software, PC-Translator, is offered in seven language pairs covering Spanish, French, English, Swedish, and Danish. Each pair has English as its source or target language. The following evaluation was conducted by our resident MT expert Claude Bédard on both the English-Spanish and French-English packages.

Throughout the fifties, the general approach in machine translation was high on engineering and low on linguistics. Then came Chomsky followed by ALPAC – and the "Engineer's Song" fell into disfavor. The *linguists* picked up the gauntlet, and for thirty years, they've been spending millions teaching natural language to computers – to this very day without coming up with really convincing solutions.

In recent years, though, the engineers have had their second chance. While computational linguistics has been making interesting though slow progress, computer technology has exploded. Within the past decade, inexpensive micros have landed on everyone's desk and been intensively applied in office automation.

This has made low-linguistics MT once more thinkable – this time based on microcomputers. Compared to high-linguistics mainframe systems, these systems are user-controllable, relatively simple to use, less costly to develop, run on inexpensive hardware, and have their import/export operations considerably facilitated by the new office automation technology, which also permits onscreen postediting.

These advantages can now make MT an attractive proposition for hard-pressed document-cranking operations.

SEARCH AND REPLACE

PC-Translator is an eloquent example of a low-linguistics MT package. In fact, its linguistic features can be summarized in a few words.

Some words are shifted during processing; essentially, this is limited to adjectives and nouns. Regular adverbs don't have to be in the dictionary; if a word has a regular adverb ending and if the corresponding adjective is in the dictionary, the system will supply a translation based on the adjective. Inflected forms of nouns and adjectives – if quite regular – don't have to be in the dictionary; the system will find the root form. Adjectives agree with directly adjacent nouns.

The rest is essentially search-and-replace.

Now what exactly are the implications of a search-and-replace MT system from the user's point of view? The answer depends very much on the particular language pair considered.

On the *input* (dictionary coding) side, it means that the system has limited morphological capabilities – morphology is not only cute, it also provides for dictionary *economy*. Main result: user fatigue if the source language is highly inflected.

For instance, you have to enter each inflected form of every verb. This is bearable when *English* is the source language ("play, plays, played, playing"). But it can get out of hand in Spanish or French, with dozens of different forms per verb.

On the other hand, search-and-replace has the advantage that coding each entry has never been easier. Syntactic information is reduced to the word class – like V for verb. Nouns require gender and/or number as applicable.

PC-Translator has three types of dictionary: a "core dictionary" and several modifiable "user dictionaries," both for single-word entries; and a modifiable "phrase dictionary" for two-or-more-word entries. There's no morphology for the latter, so noun compounds, for example, have to be entered in both their singular and plural forms.

The number of entries in the core dictionaries is 21,000 for English and 72,000 for French and Spanish as source languages. This clearly reflects how many more entries have to be made for the more inflected languages. (Note that these figures refer to *forms*, not root words, and can't therefore be compared directly with the number of entries quoted for systems with full morphology.)

RAW OUTPUT

On the *output* (raw translation) side, a search-and-replace MT package such as PC-Translator has no sense of context and therefore no basis for making decisions. The rule is simple: any given word can have only one translation. Main result, you guessed it: crude translation – especially if the target language is highly inflected and if word order is quite different. Here are five aspects of crudity:

The first is word agreement. Since verbs have no morphology, you can easily imagine what happens with conjugation: *I work = Yo trabajar, they work = ellos trabajar*, etc. You do get *trabajamos* for *we work,* provided you write an entry for it in the phrase dictionary.

Though adjectives and nouns are inflected, articles, surprisingly, don't agree with their accompanying nouns. The masculine article is used as the default and is applied in all cases. *El* mujer?!? Yes-sirree, bob . . . unless you code *the woman* as *la mujer* in the phrase dictionary.

Not surprisingly, PC-Translator's designers insist that their system works better *into English*, because of its relative lack of inflection.

Second, the system does not attempt to insert any new words (such as articles or basic prepositions).

Thirdly – and conversely – the system has no rules for merging words. So *will + verb* is translated in Spanish as *va a + verb*, which conveniently avoids the future tense. But then, *would + verb* translates less successfully into *quisiera + verb*. Again, you can get a true future or conditional — if you code the phrase as such.

Fourth, the system does not deal with homographs, or words that belong to more than one word class. For a word such as *light*, for instance, you have to choose the verb, the adjective, or the noun translation – and kiss the rest goodbye.

This can get disturbing for frequent function words, especially in French:

J'ai reçu des fleurs des champs. = I have received of the flowers of the fields.
Le marchand le lui a dit. = The merchant the him has said.

Of course, the same goes for multiple-meaning words. The cure, says PC-Translator's manual, is – as for homographs – to code your own choices in the user dictionary. These new entries then override corresponding ones in the core dictionary.

Fifthly and lastly, word shifts are strictly limited to nouns and adjectives – though this does not apply to noun strings, which are left untouched. For example:

emergency exit door = urgence sortie porte
porte de sortie d'urgence = door of exit of emergency

If you want these two translated right, welcome back to the phrase dictionary. By this time it may be dawning on you that the main asset of a search-and-replace MT system is indeed its phrase dictionary. The designers make no secret of this, encouraging the user to code as many phrases as s/he need.

LET'S GET PRAGMATIC

What can you expect from a search-and-replace MT system? Repetitiveness is the key word here. If all those phrases you pour into your dictionary occur many times in your text – which must be severely restricted in style and vocabulary – then they will be a good investment.

And if your quality requirement is moderate to low, coding can be less exhaustive. Also, despite my comments above on verb forms, for some technical documentation you may only need verbs in the third persons (singular and plural) of the present tense or the infinitive. Even so, in all cases, brace yourself for some hefty postediting.

(continued on page 57)

DESQVIEW (FROM 51)

One minor irritant is that DesqView defaults to its own color scheme for the window – which is dependent on how many windows are displayed – rather than the application's own colors. Most programs today have their own colors, and I feel this should be the default.

DesqView 386 can even run VGA and EGA graphics programs in a small window, and they continue to run in the background. You can have PC Paintbrush merrily running away alongside a WordPerfect window.

Quarterdeck has solved one sticky problem regarding the mouse. If an application uses the mouse, it belongs to that application while it is active. The user can then only use the Alt key to change applications. This is unlike Microsoft Windows/386 which steadfastly refuses to surrender the mouse while an application is running in a window.

DesqView 386 really consists of two programs: a special version of DesqView 2.2 and the Quarterdeck expanded memory manager QEMM.

QEMM turns 386 extended memory into LIM-EMS 4.0 memory and also manages protected mode functions. QEMM is a very useful program in its own right. It allows BIOS ROMs to be copied into RAM for faster execution (although most high-quality 386 machines do this anyway), and can grab unused gaps in the space between 640 Kb and 1 Mb for mapping expanded memory. This space can also be used for running TSR programs with a program called loadhi, so that they don't take up any of the precious 640Kb.

To get the most out of DesqView 386, you do need 2 Mb of RAM or – preferably – more. DesqView will swap programs to disk when you run out of RAM, but the swapping process is much slower, and a program cannot continue to run while it is swapped out to disk.

With the addition of virtual 8086 mode support including memory protection and the ability to run graphics in a window, together with support for true 80386 programs using 15 Mb or more, Quarterdeck has delivered a lot of what IBM and Microsoft have been promising for OS/2-386 – which might be available at the end of next year and will probably need more than 4 Mb of RAM just to load.

And, unlike OS/2, I haven't managed to crash DesqView yet. Now that a Programmer's Toolkit has also been released, DesqView is likely to expand an already strong market position.

It is true that OS/2 and UNIX offer more mainframe-like features, but they cannot deliver what most users expect from multitasking: the ability to use all those good old DOS programs at the same time with plenty of speed (after all that's what they bought a 386 for) and with a minimum of headaches.

Windows/386 is a great context switcher and wonderful if you use Windows programs, but otherwise . . . forget it. So vote with your wallet and get DesqView 386 instead.

Prices: DesqView 2.2 (XT/AT) $129.95; DesqView 386 $189.90 (including QEMM); QEMM-386 $59.95; DesqView Companions (Notepad, Datebook, Calculator, Link communications program) $99.95; API Reference $59.95.

— Charles Hugo

Quarterdeck Office Systems, 150 Pico Boulevard, Santa Monica, CA 90405, USA. Tel. +1 (213) 392 9851

ILLUSTRATION BY MAX KISMAN

COREL DRAW: EXCELLENT

Don't be fooled by CorelDraw's cheap-looking packaging and advertising. Behind the gaudy design is a truly excellent drawing program. And as a Mac fanatic, I can't pay a much higher compliment than that to a PC-based piece of software.

For those of you who work with the Mac, it will suffice to say that CorelDraw combines the power of Adobe Illustrator, Aldus Freehand and LetraSet LetraStudio. If you've tried any of these programs before, it will be easy for you to learn to work with CorelDraw. If you've never drawn a Bézier curve before, you're in for a little trouble. But the manual is as good as the software itself, and the freehand drawing tool will make things easier for you.

Once you've mastered the basics, CorelDraw gives you powerful options like running text along a path, masking, blending, individual optical letterspacing, radial and linear fills, use of PMS or CMYK (cyan/magenta/yellow/black) colors . . . and on and on.

The software is remarkably fast, even when not run on a 386. It requires MS Windows, a harddisk, an AT and a mouse to run.

I had to think really hard to come up with anything I didn't like. I don't like the fontnames, that's one. I find it somehow unprofessional to call "Helvetica" anything but Helvetica. But by using its own typefaces, CorelDraw is able to convert the lettershapes back to Bézier curves to allow for individual manipulation - impossible with Adobe fonts.

The only other thing I disliked was having to admit that finally I'd found a PC drawing package which – to say the least – equals my beloved Mac software. *(Another – glaring – shortcoming remains the lack of cut and paste to other Windows programs. – Ed.)*

Price: US$495.

— Yonit Swart

Corel Systems Corporation, 1600 Carling Ave., Ottawa, Ontario K1Z 8R7, Canada. Tel.: +1 (613) 728 8200

PC TRANSLATE (FROM 53)

Hardcore pragmatists may view low-end MT as a "logical extension of wordprocessing" – as Ralph Dessau, PC-Translator's chief designer, puts it. Dessau is no linguist; nor does he boast overly about his system's performance, insisting that it's for simple, repetitive texts such as parts lists and technical proposals.

His clients include mainly engineers with no particular resources for translation. Some of them are happy with it and feel that other higher-priced systems are too complicated, at least for people not trained in linguistics.

I think that we'll have to get used to the idea – and reality – of low-linguistics MT. Computational linguists will be horrified by systems such as PC-Translator, but maybe it's time they started wondering whether they themselves are not part of the problem. Most current MT systems are still big and costly, and aim at linguistic sophistication. Meanwhile, as demand grows, there are more and more clients desperate for some kind of MT, however crude, provided it's not too expensive.

Heretical as it may sound to many refined ears, I believe that the market for such systems is bound to keep growing. For more obscure language pairs, they are currently the only solution available. This is probably why Danish and Swedish are included in PC-Translator's line.

PC-Translator has three likable points: the user can import wordprocessing lists of single words or phrases into his dictionary; the system accepts WordStar 2000 formatting codes (except local ones such as underline or italics); and at less than US$1000 per package, the price is right.

All the same, I still think that in 1989, even at such a low price, we ought to be entitled to an MT system less heavily dependent on straight phrases. But tinkering with MT is not a prerogative of professional translators and linguists.

And Linguistic Products does not practice a sell-and-run policy: the company refunds any package returned in 30 days, no questions asked. In other words, you be the judge.

Linguistic Products, P.O. Box 8263, The Woodlands, TX 77387, USA. Tel: +1 (713) 363 9154