

A New Addition To

Now at last there may be a solution for translators scared to death of being engulfed, devoured, and ultimately replaced by the computer. That solution involves a phenomenon loosely known as hypertext.

Of course, the cloud of hype and poor documentation surrounding hypertext is so thick that skepticism is forgivable. But if you can run a wordprocessor, and if you have some notion of what a database is, you can use hypertext to build your own translator's workstation.

Alex Gross took three off-the-shelf hypertext systems to test their usefulness in building tools for translators: the much advertised IZE; the popular Lotus Agenda; and the less touted but equally capable Houdini and its cousin, PC-Hypertext. Gross's conclusion: only the last can really be described as a true example of hypertext.

What I attempted to build on these programs—after much consultation with colleagues—was a tool that does not yet fully exist: a multi-entry glossary allowing not only

for source and target language listings of any length, but also for extended usage notes for both listings, possible historical or etymological comments, plus all the necessary examples of their correct use in context—also regardless of length.

Such extended *glossaires raisonnés* could prove useful to large organizations, both to ensure consistent terminology among several translators and to train new staff. An essential accessory would be a data-sorting capability that listed entries by keywords, using a minimum of keystrokes. Also desirable would be the usual computer capabilities of printing, copying or switching any entry anywhere an electron can go.

CONTRASTS

Since I suspected it might produce some interesting results, I decided to limit my tests to two brief examples illustrating two extreme translation problems.

The first initially seems quite simple: to provide possible translations back and forth between English and French for phrases containing the English word "table" and its nearly identical French cognate "la table,"

together with all the necessary commentaries and examples.

The second test was more sophisticated: recording the cross-linguistic and cross-cultural contrasts between a small subset of medical diagnostic categories in Western and traditional Chinese medicine, providing all the necessary examples and explanations and showing specifically how Chinese and Western names for lung diseases and syndromes might or might not be translated into each other's terms.

For instance, something we'll call Lung Syndrome # 2 (or "re xieyong fei," heat clogging the lungs, in Chinese) can be interpreted in Western terms as pneumonia, tonsillitis, acute or chronic bronchitis, the common cold, or a pulmonary abscess. But how, working the other way, might any of these latter conditions be represented in Chinese diagnostic terms?

Of these two problems, the former, which initially seemed quite simple, proved far less so in practice—though still admitting of a solution—while the latter, which appeared complex at first sight, turned out somewhat easier to resolve. And from these discoveries springs a series of lessons for translators, linguists, and wordworkers. Perhaps most interestingly, all three programs proved quite equal to both tasks, though two excelled in certain details.

CHAOS

To begin with, with the help of dic-

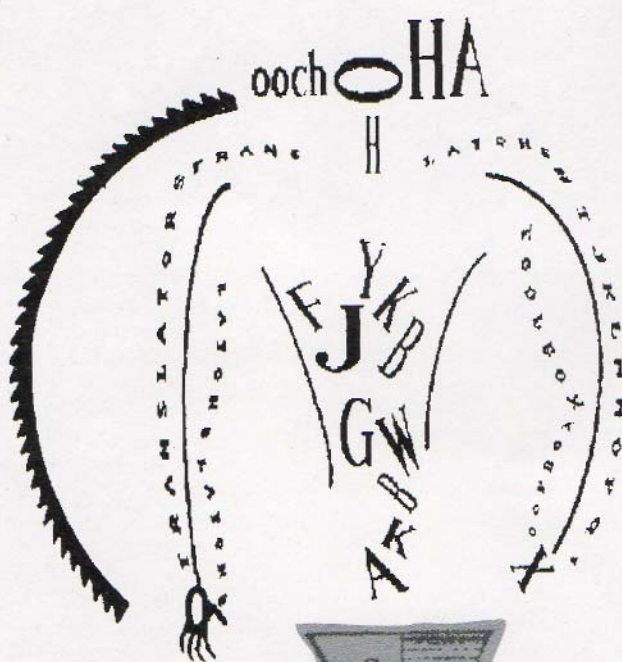


ILLUSTRATION: RYU TAJIRI

The Translator's Toolbox

By Alex Gross

tionaries and colleagues, I found no fewer than 30 different phrases in both languages where direct substitution of the opposite word would not provide a correct translation. I could probably have found close to a hundred if I'd searched all possible technical sources.

On the simplest level, "table salt" is "du sel fin" in French; a "writing table" is un "bureau"; and a "table-land" is "un plateau." Ascending in difficulty, not only would the French never dream of "tabling" a proposal, but even Americans and Britons disagree on what the phrase means — and two separate translations are necessary: "ajourner" for the Yanks and "déposer" for the Brits.

As for the typical "French" phrase "table d'hôte," it turns out to mean something else in French and has to be retranslated into that language as "menu à prix fixe." Worst of all, a game universally famous in Europe but almost unknown stateside is called "table football" in my dictionary, a phrase I've never uttered in my life. But in French, it allegedly becomes something likely to raise the hackles of the anti-Franglais brigade: "le baby-foot."

The point here is not whether any of these translations are adequate or even accurate. Whatever the perfect choices may be, even these two seemingly similar words in French and English land us on the brink of linguistic chaos — and we've only just begun. A similar morass awaits us when we attempt to wade in from the

French side and deal with such phrases as "tenir table," "table de tir," or "table d'harmonie."

We haven't got onto diplomatic, chemical, or commercial nuances yet. But we already need every kind of explanation and example we can devise, including ones not too readily found in our dictionaries. We've slipped unwittingly into the borderlands of lexicography and found them less well patrolled than we supposed.

AGENDA, IZE

Can these computer programs possibly help us? I believe they can, provided we're willing to submit to the structures they impose.

Lotus Agenda is quite competent in this regard. Almost any word or phrase you're likely to need can be treated as an Agenda "item," even if it's as long as 350 characters or 4 1/2 lines of text. If you need still more room to create examples or commentaries, you can turn them into an Agenda "note," which can be as long as 10 Kb, or five pages of double-spaced typing — long enough to satisfy even scholarly purposes.

On top of this, you can categorize your entries in numerous different ways of your own choosing and almost instantaneously create phrase lists according to those categories.

But Agenda has certain drawbacks. Although it's remarkably supple and easy to use — I was in fact able to jump in and create my first glossary in less than half an hour — you'd never know this from its top-

heavy documentation. You'd also have to include both the phrase and its translation in your first 350-character entry, because the spreadsheet-like columns Agenda creates are too narrow for any extended text. The instructions for printing are particularly dense.

Nonetheless, the program is so powerful and versatile that a number of interesting compromises await inventive users.

IZE, Persoft Inc.'s "personal information" tool, has its own virtues and drawbacks. Text entry is extremely easy, and you can easily switch between screens containing translations, or examples or comments. You can even import longer commentaries than Agenda allows from many wordprocessors.

Furthermore, the program can automatically check every instance of a word in either language to make sure translations have been consistent, plus it can make an outline showing every usage of a particular word, organized according to other keywords surrounding it. It can even make allowances for various synonyms representing the same reality, such as "Britain," "England," "the U.K.," and "the British Isles."

It does most of this with blinding speed when it starts, but some of the preparatory keywork is just plain tedious. It also creates vast numbers of extremely small files which are hard to tell apart. Its documentation and tutorial are better than Agenda's but still no bargain. On the whole, Agenda

comes out slightly ahead.

HOUDINI, PC-HYPertext

Which brings us to MaxThink's *Houdini* and its half-sister *PC-Hypertext*. On grounds of price alone, there's good reason to choose these two. Whereas both Agenda and IZE sell for hundreds of dollars, Houdini's official price is \$89. And PC-Hypertext is available, astoundingly, free of charge on almost any bulletin board, or for a nominal charge from the manufacturer.

Such generosity can probably be attributed to the fact that some of its capabilities are best exploited by other MaxThink programs such as Houdini. I would choose Houdini and PC-Hypertext over the other two — and not just for price.

While Agenda and IZE have been shaped primarily as business tools employing some degree of text management, the whole Houdini outlook is explicitly a textual and linguistic one. And this clearly shows in the way it and PC-Hypertext operate.

Not only can either of these two programs do everything I've described so far, but they are specifically oriented towards linking different sizes and shapes of text, even towards delineating the semantic fields between them.

Neil Larson, the brain behind these programs, actually invokes a mixture of Whorfian and Chomskian theory to justify the programs' *raison d'être*. And although they were origi-

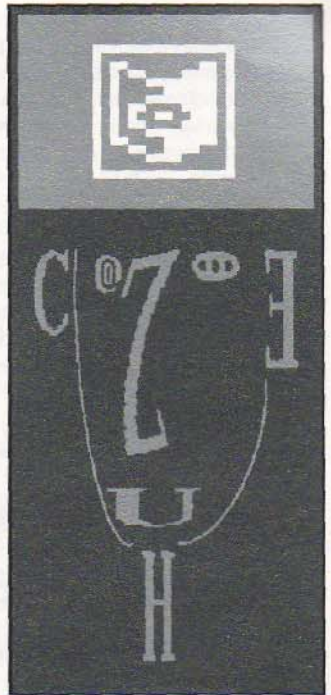


ILLUSTRATION: RYU TAJIRI

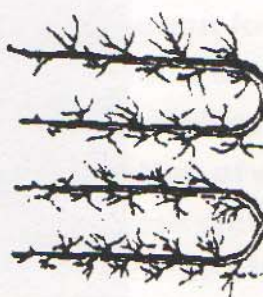
nally planned with a single language in mind, they will also work well for a translator's purposes.

PC-Hypertext works best when it's linking large multi-paragraph sections of text, while Houdini handles smaller, closely reasoned connections, though both programs can deal with both problems to some extent.

HYPER-STRUCTURE

It should be emphasized that not one of these programs was designed to help translators. Agenda is marketed essentially as a freeform database, whose structure can be determined and changed as you go along. IZE is being promoted as a "personal

TABLELAND
TABLE SALT
TABLE FOOTBALL
WRITING TABLE
TABLING



information manager," whose keywording capabilities make it useful for text retrieval in a single language. PC-Hypertext and Houdini are intended to help in linking and organizing texts and ideas respectively.

This rapid linkage of words and concepts, even images and music, is what hypertext is really supposed to be about, and only PC-Hypertext really approaches this. But there's been such a marketing storm around the hypertext concept that all have laid some claim to the term in their promotional campaigns.

Yet all these programs have some hypertext capabilities, and all four can be easily handled by most translators. It should be emphasized that computer programs don't have to impose a structure on people. Rather, people, after they've learned a little, are free to impose the structures that *they* need on programs.

THE FUN PART

And now to the fun part: how my two test examples turned out, and why I think they worked out that way.

In technical terms, both tests came out quite well on all these programs, though they turned out differently from what I expected. The "table-latable" example — which looked simple — wasn't, but there were no real problems in representing the discontinuities between these words on any of the programs.

On the other hand, the translation between Western and Chinese medical diagnostic categories, which for cross-cultural, scientific, and linguistic reasons sounded almost impossible, turned out to be a snap on all three programs, though this was most obvious on Houdini and Agenda.

IZE lagged behind the others because it's not as good at linking categories on the fly, but those who think out their keywords beforehand can produce comparable results. None is a "bad" program. Rather, each one is quite capable in somewhat different areas.

This doesn't mean that any one of the programs provided a perfect translation of any Chinese diagnostic category into western terms or converted any western diagnostic category into a precise Chinese equivalent.

We are still a long way from the fantasy constantly reinvented by many non-linguists that someone will suddenly devise a Black Box for Translation with a slot on either side.



Illustration: Ryu Tajiri

Rather, they correctly and accurately delineated the differences in a situation where there were no precise equivalents. They showed which Western conditions each Chinese category could be, and also which Chinese conditions each Western category could be.

They accorded both systems complete textual equivalence, devoid of value judgements, and in so doing provided a map of the total semantic field linking them. Thus, they not only showed that Lung Syndrome # 3 (termed "tanshi u fei," or mucus-damp hindering the lungs) could be either chronic bronchitis or bronchial asthma in Western terms, thus reassuring those who choose to be reassured by western terminology. But they also showed that bronchial asthma, for instance, in Chinese terms could be either Lung Syndromes # 3 or # 1.

Houdini produced careful step-by-step frames showing precise linkages in this area, while Agenda automatically wrote entire overview screens summing up the same links. Both these programs could also, at a single keystroke, branch out into any number of specifics concerning treatment or complicating factors or alternative prognoses, dependent only on how richly they had been programmed.

In so doing, these programs provided examples of "semantic fields" so meaningful that even laymen would be able to see why such a seemingly exotic concept was important.

However, it's one thing for linguists to quibble over the meaning of words, as they frequently do. It's quite another to learn that you may be suffering from one illness or condition in Western terms, requiring a certain course of treatment, but simultaneously have a dramatically different condition in

terms of another system, prompting a completely different approach — all supported by intrinsically logical computerized steps.

GLOSSEMATICS

But there are other quite unrelated reasons why I believe these two tests have produced stunning results.

Ever since Descartes, it has been assumed that real knowledge must be mathematical in nature: either mathematics itself or the so-called exact sciences that mathematics supports. Concomitantly, it has also been assumed that so-called verbal or language-based knowledge must be in some way inferior, since language does not easily lend itself to mathematical precision.

But now, inadvertently, unexpectedly, and with unforeseeable consequences, through such concepts as hypertext and its inevitable spinoffs, language may at last be in a position to make a comeback on the knowledge ladder.

While it may never be possible to assign any absolute numerical value to any word or combination of words, this may no longer be necessary. It may now be possible to draw a map showing the relationships of all words relative to one another, and not for just one language but for most major languages together.

This idea was in fact first broached some fifty years ago by two Danish Glossematics school linguists, Louis Hjelmslev and Hans Jorgen Uldall, who postulated that the true study of linguistics was in fact the area covered not only by all existing languages but by all languages that could be imagined.

They also held out the possibility that linguistics might not be merely that branch of science which deals with language, but rather that science might be a mere sub-branch of language purporting to deal with universal and objective phenomena, trapped in its own linguistically determined preconceptions and errors.

The advent of such programming concepts as hypertext provides us for the first time with a tool that might prove capable of proving or disproving this thesis.

Such a tool may even come to challenge more ancient preconceptions than Cartesian ones. A basic principle of Western logic is the "syllogism," the most famous formulation of which was "All Men Are Mortal. Socrates Is a Man. Therefore Socrates Is Mortal." This "A = B, B = C, therefore A = C" structure also lies at the heart of computer matches and sorts.

But in both real life and in language, this structure does not always apply. A will often have some aspects of B and C, not to mention D, E, and F. B may have some elements of A but not as many of C. And C will have differing proportions of both, plus some of P, Q, and R.

Until now, there has been no simple notation to record such complex relationships.

VERBAL NO-MAN'S-LAND

Hypertext may provide such a simple notation. This means it may also provide a way of escaping the crude and rigid tree structure imposed on reality by most other computer approaches. And this again makes it ideal for dealing with language, where a "leaf" with one value in Language A may have a quite different value in Language B, and may even turn out not to be growing on the same tree or in the same jungle.

The Chinese-Western medical example came out as a perfect set of connections, joining neatly in the middle with each side having its complements on the other.

But the English-French and French-English examples ended up largely as two unjoined systems, like two ends of a seesaw with nothing in the middle. Though such phrases as "table of the elements" or "table of the laws" balance out in both languages, many "idioms," "semi-idioms," and "non-corresponding phrases" fail to do so (no adequate vocabulary may yet exist for this verbal no-man's-land).

Thus, although "propos de table" became "table talk" and "to lay the table" became "mettre le couvert," nothing joined the words "talk," "lay," "mettre" or "couvert" to their opposite sides. The reason for this is simple and provides, I believe, yet further evidence that hypertext can function as a tool in delineating linguistic relationships.

The medical example join in the middle because it deals with a limited, specialized field, where ultimately comparable realities are being addressed, though in radically different terms. The "table-la-table" example fail to join because it's part of an entirely unspecialized field, namely the total complexity of relationships between the entire vocabularies of two rich languages.

Yet, by using hypertext-related methods, the middle could be constructed in detail, though not simply or soon, and certainly not on a micro-computer, since this would entail mapping out — in something resembling hypertext form — all the relationships between all words in both languages.

As unwieldy as this task may seem, it may furnish the first example in several decades of a linguistic theory that can actually be tested by documented experimentation. Many problems

might be encountered during such testing, but thanks to hypertext it may now become at least possible to conceive of the space separating two languages in a dimensional manner and devise ways of delineating it.

NO BLACK BOX

This is scarcely to say that the modest tests I've run on these programs will change the universe tomorrow or soon prove capable of solving all translation problems. Hypertext will find its place most readily in large offices or among translators dealing with repetitive but highly technical specialties. Typing a glossary into a computer takes time, and no one is going to do it unless the work pays for itself in some way.

Hence, there is still plenty of room in the profession for custom work by "old-fashioned" translators using typewriters and "hard-copy" dictionaries. Those who do high quality work have nothing to fear from computers but also no reason not to experiment with them.

In many cases, Hypertext will serve best by showing where discontinuities of sense may exist between two languages and suggesting alternative ways of dealing with them, leaving the final decision up to human translators.

This means that we are still a long way from the fantasy constantly reinvented by many non-linguists that someone will suddenly devise a Black Box for Translation with a slot on either side — you feed any kind or form of written or spoken "Language A" text into one slot, and out it comes a perfectly polished piece of "Language B" from the other one.

This fantasy probably belongs in the same category as claims of progress in inventing a Perpetual Motion Machine. There are very real reasons why it cannot be done, but there are also very good reasons why human minds will not stop playing with the possibility.

Although such programs as these hold out the possibility of telling us much about human knowledge and supplying many clues for machine translation, so far they will only assist with one aspect of language: vocabulary.

They will not help us with parsing, with evaluating context, or with determining tone or level of discourse. Their use also assumes that the primary goal of language is communication, totally ignoring the larger reality that language is used at least as often to express emotion, confirm class or group status, or obscure the truth as it is to communicate.

However, the possibilities opened by such programs may prove so stunning that rather than blame them for shortcomings, we ought to welcome them for what they can do.

Alex Gross is chair of the Machine Translation Committee of the New York Circle of Translators.

Houdini (US\$89) and **PC-Hypertext** (US\$6) are available from MaxThink Corp., 44 Rincon Road, Kensington, CA 94707, U.S.A. Tel: +1 (415) 428-0104. PC-Hypertext is also available free of charge on many bulletin boards.

IZE costs US\$445, from Persoft, Inc, 465 Science Drive, Madison, WI 53711, U.S.A. Tel: +1(608) 273-6000.

Agenda goes for US\$395 and can be ordered from Lotus Development Corp., 55 Cambridge Parkway, Cambridge, MA 02142, U.S.A. Tel: +1 (617) 577 8500.