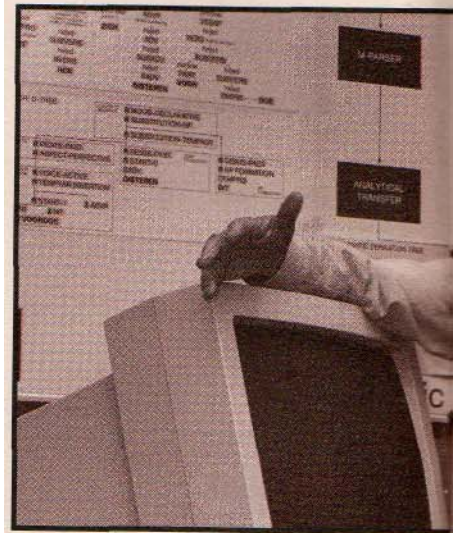The guiding light of Philips' ambitious Rosetta MT project, Jan Landsbergen now divides his time between Eindhoven and the University of Utrecht, where he recently became professor of computational linguistics.

Well, can they?

"If a human translator holds a gun to my head and forces me to answer yes or no, I'll say 'no.' But once out of range, I'll add: 'But they can help you.'

"If a linguist asks, I'll reply: 'Yes, machines can supply all linguistically possible translations.'

"But my favorite answer is: 'Yes, machines offer people the possibility of expressing themselves in a foreign language.'"

This was how Jan Landsbergen answered the question he asked himself aloud in his inaugural lecture at Utrecht university in November 1989.

Rhetoric aside, the Rosetta project, whose approach MT observers regard as the epitome of purist formalism, is gradually moving forward. From solely theoretical research, Rosetta has progressed to the application-oriented phase. The product-to-market stage, however, is still some time off.

Landsbergen has been thinking hard about computers and language for many years now. What originally drew him to natural language processing?

"While studying math at the Technical University of Delft, I became interested in formal language theory. Attracted by its mathematical basis, I went on to write my graduate thesis on transformational-generative grammar, as described in Chomsky's *Aspects of the Theory of Syntax*.

"That was the link to linguistics," says Landsbergen. "But at the time, I wasn't working with computers."

In 1971, Landsbergen felt like a change from the rarified atmosphere of academia. So he signed up with Philips to
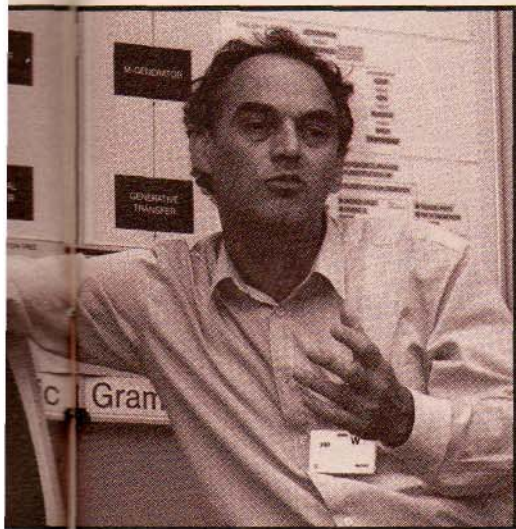
work in the company's Apeldoorn-based Computer Industry Laboratory.

There, he found himself in the same research group as Harry Bunt and Remko Scha (now professors in computational linguistics at Tilburg and Amsterdam respectively), who were researching artificial intelligence. The group, including Landsbergen, later moved to the Nat Lab, Philips renowned research hothouse in Eindhoven, home of the company's headquarters.

"As early as university, I'd realized that it was possible to describe a language *formally* to a certain extent. The next step was to try to describe it *constructively*, so that the computer could do something with it," explains Landsbergen. "It was pretty much a new field."

PHLIQA TO ROSETTA

Landsbergen's first NLP project at Philips, from 1972 to 1979, was PHLIQA, which centered around the development of a natural language interface to databases, also called a "question-answering" system. "The idea was that you could both formulate queries and get direct answers back in natural language," he says.

"At the time, it was pretty ambitious. Now, such interfaces are on the market. Symantec's Q & A, for example. Of course, commercial systems are still very limited syntactically, but they *are* useful – especially for people who don't formulate database queries on a daily basis."

PHLIQA never saw the light of day as a product, but the ideas it generated proved useful in other research projects, such as the Siemens-Philips joint venture SPICOS, a natural language question-answering system combined with speech recognition and synthesis (see EW#17).

The Rosetta project is also descended from PHLIQA. "For the natural language interface," Landsbergen explains, "we developed a grammar parser based on the theories of the logician Richard Montague, extended with certain aspects of Chomsky's transformational-generative grammar.

"Later it occurred to me that 'M-grammar,' as we nicknamed it, might well be suited to translating. So we proposed an MT project to Philips in 1981." That project became Rosetta.

"We basically set out to study whether M-grammar could describe a symbolic relationship between two languages, one which might be of use to a translation system. That was the question we posed, and not 'how soon can we make a translation system for Philips?'

"Of course," he adds, "Rosetta has grown since then. And now it's important for us to have applications to show for our work too."

DICTIONARY STANDARDS

One subject close to Landsbergen's heart is the importance of collaboration among different MT projects in the building of dictionaries.

"Dictionary-building is essential, but it's not the most exciting part of MT research. That's why it's idiotic for everyone to be doing it on their own. I'm all for autonomous research projects, but for dictionaries I think it's vitally important to collaborate.

"The problem is that there are lots of linguistic theories to deal with. One dictionary format is based on one theory, another on another. For broad cooperation, you'd have to agree on how you're going to structure the data you put into your dictionaries."

Philips is currently negotiating with Dutch dictionary publisher Van Dale and project leaders at Eurotra (the EC's MT project) to base a Spanish-Dutch lexical database on Van Dale's Spanish-Dutch (paper) dictionary. It would then be made available for use by both Rosetta and Eurotra. EC funding may be forthcoming.

"In practice, grammars come first, and then come dictionaries," says Landsbergen, "because only after developing the grammars do you know what should go into the dictionaries. But you shouldn't wait too long to plan them. Otherwise, your grammar will make wrong assumptions about the word types available. You'll always make mistakes – you might need more word types than you'd anticipated. So the construction of a large dictionary is an interesting test for any linguistic theory."

THE MT HORIZON

In his new capacity as professor of computational linguistics, Landsbergen advises researchers. One of them, Louis des Tombe, is developing of a theory of translation. "This is necessary," Landsbergen explains, "because there are a number of different aspects of translation.

"In the first place, equivalence of meaning between source and target texts is required. But there are other aspects, like stylistic and cultural considerations. Sometimes these considerations are *more important* than strict equivalence of meaning.

"Human translators, of course, follow their instincts. But when you try to automate this instinctive process, you find there are still no rules, no formal theories."

This absence of theory doesn't necessarily prevent the progress of MT. Landsbergen favors a comparison between MT and aerodynamics. First came the airplane, *then* the theories. And the theory didn't copy the natural process.

"Just as you shouldn't try to make a plane by constructing an imitation bird, neither should you try to make an MT system by imitating a human translator. The aim is the same, but the means of achieving it is of necessity different, because of the difference between the human brain and the computer."

Surveying the broad landscape of current MT research, Landsbergen clearly has more faith in the logical mapping of natural language syntax and semantics than in knowledge-based systems.

"What you're doing with computers is processing symbols. Now, you can represent *some* aspects of human language with symbols. But it's very difficult to represent certain kinds of *world knowledge* – for example, images – in such a way. And that's precisely the sort of knowledge often needed in translation – especially for the solution of ambiguities."

A solution may be interactive MT. "An interactive system lets you call on the user's knowledge. In cases of ambiguity, the user is interrogated in his own language. But it's important that an interactive system be dependable – the user may be monolingual and unable to check the result. That's what makes interactive MT interesting."