

Language personality of the month

Dr Herman Caeyers



Armed with its integrated arsenal of language technology, Belgium-based company LANT s.a. appears set to take the world of multilingual documentation by storm.

Bob Clark interviews LANT CEO Herman Caeyers to discover what makes the company, and the man, tick.

CLARK:

You received your Masters Degree in Theoretical Physics in 1978 and your PhD in Mathematics in 1985, both from the Catholic University of Leuven. Were these the sort of qualifications to which you aspired whilst still at school? Did you ever want to be a footballer or a train driver?

CAEYERS:

When I was young - I come from a big family, four brothers and one sister. I am the second one - I had to do everything the same as my brother, Jan, who was the eldest. Now he is a famous conductor, in music. We always did the same things. When I was 17 and he was 18 he had to choose what he was going to do. The problem was that he was the first to make his choice. I always dreamt of being a musician and then he decided to go into music. So I had to do something else. To go my own way. Because when you are 18 years old and we did the same studies in secondary school, always together, we decided to go our own way. I had to do something else. I decided to go into mathematics and physics. If he had made the choice to go into mathematics, maybe today I would be a musician. But that's life. Now we are both very happy about what we did. Sometimes, when I go to a concert and I see him conducting a big orchestra, I'm still a bit jealous.

CLARK:

Do you play a musical instrument yourself now?

CAEYERS:

Yes, I play violin and piano. I got a degree for violin and piano at the same time. All my brothers are artists and I am the only one doing something different, in business, and mathematics and physics. One of them is a ballet dancer, another is a composer. So it's very strange how life can go.

CLARK:

Your PhD thesis was based on a study about mathematical models for mental processes that are the basis of music perception. Could you tell us how you came to choose this topic?

CAEYERS:

This is linked to the first answer. I was still very interested in music. When I obtained my degree in Theoretical Physics my idea was to combine the mathematics and music. I decided to work on how to make my doctoral thesis on music. What I did was the following: Theoretical musicology is still based on theories from the 18th century, Rameau. He produced a wonderful work on music theory. But what I see is that there are people today that are still using this old theory. Going in the same direction that he did. It's based on scores. In music theory today, they examine music scores. The score itself is a way to communicate music. But in reality it's a human artefact. Something you create in your mind, like language. The other person has to use another system to understand what the other wants to say. The score itself is a kind of means to communicate. To say, OK, how can I put down on paper what I have in my mind so that an interpreter can perform it and have someone else hear it? My idea was to start from the mental process because that is where it's created. To start music theory from scratch. Starting from what a tone is and the relation between two tones. Just two tones. And, starting from that primitive function, to make a system out of it using mathematics. It was very strange that I could follow the music history by doing that. That you have hierarchical structures in music which can grow. Like life can grow. And it was fascinating and I found out that the score itself is very limited. That we are doing much more, like language, than you can see on the scores. When we have the representation of a tone, an A or a B for

example, it can represent different tones, even when you have the same frequency. With an A that frequency can have another meaning in our mind, but it's the way you come into that tone that makes the difference. It's very strange and there are a lot of questions, simple questions that have never been dealt with in music theory. The way melody goes. I've never seen studies about this. They've done work on chords and the relationship between chords. This was the basic theory of Rameau. But simple questions about melody, when you sing and you stop in the middle of that melody you may say it's not finished. Why? It's a very simple question. Sometimes you have the feeling that you have to do something, add 'pahm-pahm'. Yes, now I feel that it's finished. What's the relationship between rhythm and the frequencies? You have a song and you sing it putting the accents in a different way. We say that's not the melody any more. It doesn't sound as good. Why? We don't know. These are very simple questions never addressed. This was part of my work. To start from the mind. To see how this can be organised. In this respect I follow the ideas of Chomsky. I think it was in the '70s that there was a big conference where the ideas of Chomsky and Piaget were debated. A big debate between these two people about art - is it some universal thing in the mind or is it something you learn as a child? That you are building your mind by learning. And Chomsky said, 'No, no. Everything is already there'. The system is there already. What you learn is how to use that system. That human beings are the same all over the world. That there are basic principles. I follow his ideas. I say he's right. And it's the same thing in music. I can prove that Chinese music and African music are the same system, but it's used differently. I'm still very interested in this subject and my dream is to write two books on musical theory beginning in the brain. I refuse to die before I have done so.

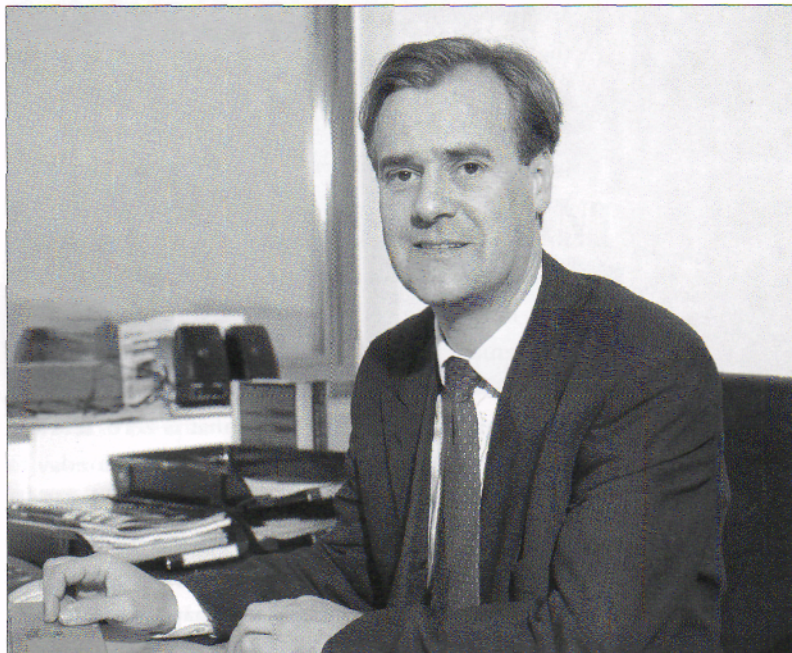
CLARK:

How did a doctoral thesis about mathematical models for mental processes that are the basis of music perception lead to your becoming head of the METAL machine translation project at Siemens, Belgium?

CAEYERS:

When I did my PhD at the university I had to study linguistics. Chomsky did wonderful work in this field. When I started analysing music in a different way than was usual I studied linguistics to see what

kind of system, the mechanisms these people were using to analyse languages; to see if there were some common aspects between music and language. Because music is also some kind of human artefact. Language is a way of communication. Music is also a way to communicate. It's another system. The formal system of Chomsky is that you have a grammar on one side and you can create something using that grammar. And on the other side you have the listener. The listener is using the same system but it's what they call an automata, analysing the stream of language. One line of characters comes in and you have to use your system to make some representation to make some meaning of it. The theory of Chomsky is that there is a kind of isomorphism between the system the listener is using to analyse and the one creating the language, the stream. It's the same in music. It's the acceptor on one side and the creator on the other side. This system has to be the same in a mathematical way. It's another representation of the



HERMAN CAEYERS

same system. You can prove that they are using the same basic system. This was fascinating for me and I could use the same system to apply to music. Then I had to present my thesis and there were people from Siemens sitting in the audience. That evening I received a telephone call. They were thinking about setting up a big project in machine translation here in Belgium. Was I willing to join them and be the project leader? They saw that, on the one side, I was very motivated and that I specialised in formal theory and linguistics as well. I said that I would think about it and in three days I said yes. So I joined Siemens in 1985 as project leader of the METAL project in Belgium. This was the way it went.



CLARK:

After leading the Metal project for five years, you went on to manage the Siemens Research and Development department of Artificial Intelligence, which was quickly followed by your becoming head of European projects in the domain of computational linguistics. Could you tell us about this period?

CAEYERS:

They asked me to go to Munich to manage all these projects within Siemens in computational linguistics. I had the feeling that I had to do this. To develop all these international contacts. To talk to people all over the world about computational linguistics. When I was here in Belgium I had to do a lot of technical stuff. To develop together with my team, to brainstorm about linguistic theories, which theories can be used, and discuss Artificial Intelligence. So this was an interesting step for me. To meet a lot of people all over Europe. When you are responsible for European projects it gives you the opportunity to talk with a lot of people in the field and to see what they are doing. At that time I already had the idea of doing something on my own. Sometimes you have to follow your feelings, your heart. You say, 'I have to do this!. Let's go over to Munich and make this step.' I already knew that this was an intermediate step that I had to make. It was not easy because I have four children. As a mathematician I don't believe in statistics any more, by the way, because I have two sets of twin daughters. The probability for that is zero. It was a hard time. I had to go to Munich from Monday to Friday for over a year. I'm very thankful to my family that it went very well. I had the feeling that I had to do this for a year and a half, even though it was not that easy. Now I am very happy that I did it.

CLARK:

After spending seven years with Siemens, in 1992 you join Sonovision Itep Technologies in France as Deputy Director of the Eurolang department. Were you still commuting?

CAEYERS:

Yes, I had to go to Paris. The only difference was that I flew to Munich and I drove to Paris. The reason that I joined SITE was that, as everybody knows, there was this common EUREKA project called EUROLANG, between SNI and SITE. SITE concentrated on the translation memory and Siemens more on the automatic translation side. I

was asked to move to France as a link between the two companies. To create that synergy in a better way than just two isolated companies. That I could bring my experience from SNI to the other company and to forge this link between the two companies. This didn't work out the way the two companies hoped. The two companies had their own ideas and strong individuals on both sides. It was very difficult to make that link. We never reached a good collaboration between the two companies. I still regret this. I can see that what happened was good for me. But at that time my decision to go on my own was based on the failure of that co-operation. It was part of my job to make that link and I couldn't do it the way I wanted. After some time I decided to say, 'OK, I'll start my own company.' To start from scratch based on my experience of all these years. And now I am very happy that I did it.

CLARK:

This is the point where your life seems to go into 'warp-drive'. You set up the LANT company and, within three years, you have a company that is worth over \$US 15 million. At the same time you acquire the METAL technology lock, stock and barrel, as well as all the rights of the Eurolang Optimiser. How did you manage to accomplish all this?

CAEYERS:

It's because I have no enemies. I am still on very good terms with Siemens and SITE. When I heard that there were problems within Siemens, well I always tell the same story. A company is like a human being. A human being starts as an embryo. And after a while it gets arms, legs, a nose, ears etc. But in the beginning the proportions of these arms and legs are not normal, as with an adult. The arms are longer than the legs and so forth and it's the same with a company. The human being can become very old. Everything is OK, the nose, the ears but the human being is getting old. And then it gets illnesses and diseases and goes to the doctor. The doctors in a company are the Arthur Andersens, the McKenzies, and they have to do something. These big organisations like Siemens, they go to the doctor and get a bypass. They say, 'You have to do this and that'. OK, then it's better for a while. They restructure, 're-engineering' they call it. The doctors are working on this old guy. You can do everything you want but the company will remain old. What I have now with this small company it's nice to see that this embryo gets a

nose, then ears and so on. When I went to Siemens the decision was based on advice from the doctor. They said, 'OK, you have to concentrate on this and this. You have to get rid of all these developments because you are too big to do these kinds of things. There is too much administration. You shouldn't do this kind of development in-house. It's better to put it in a garage with young people doing this kind of thing rather than an old guy doing this kind of development. So they decided that they were going to stop and give it to a small company that can develop that technology further. When I heard that this was going on I got on a plane and went to see them. I said that I know the people with the know-how to do this and that they were ready to join me to work on further development. So they said, 'Let's do it'. The same happened with EuroLang Optimiser. I heard the same thing. That they had to concentrate on their core business, documentation. They had to stop developing technology. When I heard this I drove to Paris. We had the know-how. I knew the system. We were ready to take it over. They said, 'OK, you are a good guy, you're our friend and it's better that you do it'. That's the way it went. It wasn't that simple because it had to do with a lot of investment that they had made. There were a lot of negotiations. In the case of Siemens it took over four days of negotiating, day and night. The result is that we have rights for further development, sales rights and, in a few years, we will have total ownership of the technology. With EuroLang Optimiser we have exclusive rights and with METAL we have non-exclusive rights, but after a period we will have full ownership. For the past year we have done a lot of further development of the METAL technology and it will be marketed under the name LANTMARK. After a while it will be a new system and you will not be able to recognise the METAL system in it anymore. It takes a lot of man-years to make a machine translation system and we are very grateful to Siemens that we have been able to use their basic module for further development.

CLARK:

Do you have problems convincing companies to adopt your approach?

CAEYERS:

Customers in our field are very strange. We have to think a lot about how technology can be used within an organisation. What we see is that organisations are not prepared today. So you have to

combine two things. You have to listen to the customer because they can tell you a lot about how they work, how it's organised. On the other side you have to tell them what to do. You have to balance these two aspects. There are domains where all you have to do is listen to find out how they work. In our field it has to do with the fact that documentation, translation in most companies is a secondary thing. It's not their core business, it's something they have to do. All these people work in another building. They make their product and when they need documentation they send it to that building and they close the door. They don't realise that it is part of the whole process. It's part of the car! It's part of the plane! The way I see it today



with most companies is that it's an isolated process. It's not integrated in the production line. And it's that important. They have all these matrix structures so what you see is that the technical writers are here, the translators are somewhere else in this matrix and they don't talk to each other. They say, 'We have nothing to do with translation, that's another business'. So when you tell them, 'Yes, but if you invest a little bit here on the authoring side they can gain a lot on the translation side'. And they answer, 'Yes, but that's not our budget. Are we supposed to invest so that they can make savings on their budget?'. It's a strange situation and the bigger the organisation the worse it is. We have to do missionary work at a high level. We tell them that there a lot of things that they can do but they have to look at the whole process. Right from the start, bringing this know-how together, authoring, translation, distribution, everything. It's one line and you have to streamline it. What we do is go into the business and say that we want to



control the whole process. Our products are the basic modules of solutions. We have an authoring tool, our LANTMASTER system, we have our translation memory, we have automatic translation, we are also working on an object-based documentation solution. We try to integrate all these modules into different solutions. We are not selling products as such. We use our products in our solutions. We also sell EuroLang Optimiser as a product, for smaller organisations. For bigger organisations we work together with the customer and say, 'How can we solve your problem?' More and more I get the feeling that we are not selling products, we are not selling solutions, we are selling savings. In using our modules, integrated in a good way, with a bit of reorganisation of the company we can cut the budget by twenty percent - using translation memory, using controlled language and so on. This is what we are doing for General Motors. We still have to do a lot of missionary work. When you go to a translation department and show them the solutions they say, 'That's a lot of money!'. Because in these departments, they don't have the experience of talking about money and budgets. It's strange because when you have to make an investment in a translation department of a company, they have never done it. They are not used to talking about investment in a translation department. Just a PC! Translators are not used to asking management for money. They think it's expensive but when you calculate the savings, the investment is peanuts. For instance, if you have a translation department of 20 people, 20 percent savings is four people. It's the money related to four people but, in practice, they keep the people and they do more, and faster. It's the speed that's more important than the volume. The time-to-market aspect is very, very important. But it's very hard to calculate. What does time-to-market really mean? It's hard to sell. You can't make an exact calculation of what you are going to gain by the time-to-market principle. But it's changing a lot. People are realising more and more that it's important. After more than 12 years in the language technology business I see that these days it's totally different than five or ten years ago. And for the first time I believe in the market prospects. Ten years ago I would have thought that it was not possible. All these figures and marketing studies that said that it's going to happen next year. No, I didn't believe it then. But for the past two years I have been sure that now it will be true. There is an awareness now with companies. Language technology is important.

CLARK:

You mentioned an object-based documentation solution. Could you explain that?

CAEYERS:

In an automotive production process, if they produced a car the same way that documentation is produced, the management would be fired. You assemble the car, you put the radio in, all the electronics, the seats, the trim, everything. But it still has to be painted. You have to send it to another building where they do that sort of thing. You have nothing to do with painting cars. That's not your business. So you send the car to the paint shop and they have to take everything off the car again to paint it. And, after it's painted, they have to put everything back again. That's the way a multilingual document is built. The author is very proud of his work. Then he sends it to the translation department. Then the translation department has to take all the pictures out of the document. All the tables. They translate the document then they have to put everything back into the document. What I see is that the translator is only translating 20 percent of the time. The rest of the time is spent manipulating the document. This I would like to change. The idea is to use object-oriented databases for the documentation. Where the authors keep all these blocks of data separated. What he writes is treated as a text block. The drawings come from another department will be another object. You have a 'document builder' who brings together all these objects and puts it on paper. This object here, that object there. You keep all these things separate and the language department only has to deal with the text parts. Then you can use technology. It comes in, you translate it, creating new text blocks for the database and, by changing certain labels in the database, you will have a new multilingual document automatically. There again, it's the time-to-market principle. Once you have all these blocks translated by machine with revision and you change the language label of the document, all the text blocks will be replaced automatically by the appropriate translation. This is the kind of approach that we are doing for General Motors. It's a big project in providing all the linguistic aspects, automatic translation, translation memory for their document database.

CLARK:

Could you discuss some of your other activities in greater detail, like the controlled language module?

CAEYERS:

Automatic translation in the technical domain can only have benefits if you start from a controlled language. Automatic translation can also be used for communication purposes, just to have an idea of the content of the text, but not in the documentation business as such. If you start from controlled language you can achieve very high quality in automatic translation. The idea was to come up with a checker for controlled language in combination with our translation engine. There are companies in the world offering checkers for controlled language. And there are other companies with machine translation systems. But the ideal situation, and this is the situation that LANT can offer, is a combination of both, where you can control the output from the controlled language side with the input of the automatic translation. It's the same idea again, to control the whole process. To optimise the whole process so that you can optimise the link between the two technologies. It was fascinating to see that we could use the same technology of automatic translation for controlled English. So we developed a checker for controlled English which is called the LANTMASTER system which is based on the technology of automatic translation. In reality, we are translating from English into controlled English. We are the only company in the world able to come up with a proposal for controlled English. Most of the systems can give a diagnosis of what's going wrong. They can even localise the error. We can localise it and we can correct it automatically. This is thanks to the fact that we have this rich technology of automatic translation. You have to make a full analysis of your sentence to control language. It takes years to develop full analysis. We use this analysis from our automatic translation system. So we go a little bit further than what is usually called controlled language. Normally, what they have in mind with controlled English is readability, understandability of the text. It comes from the aeronautics industry where they use English all over the world. More than 80 percent of the technicians that work on the aeroplanes at the airports don't speak English as a native speaker. Therefore, the people in the field of aeronautics came up with the idea of keeping it as simple as possible. They haven't chosen the multilingual solution. Instead they use English because they have all that documentation, more than 500,000 pages for an aeroplane. Imagine, for one aeroplane, the technical documentation to be translated into all these

languages. So they keep it in English but then they have to write it in a very simple way that everybody in the world can understand. They came up with rules, the AECMA rules. The basis of controlled language. And now, more and more what I see is that the other domains are taking over the idea of controlled language, also for the idea of readability and understandability. But what we see, and I think that we are the first, is integrating a set of rules into the controlled language for translatability, which have nothing to do with readability. To give you an example, for an automatic translation system it's not easy to disambiguate the homographs. Is this a noun or a verb? It's not that difficult for a human being, but for a machine it's hard to disambiguate, especially in English. So, in order to facilitate a machine to disambiguate it's better to put an article in front of the noun, then it's clear that it's a noun. If you put this as a rule, when it's a noun put an article in front of it. It has to do with translatability and this is new. What we are doing with General Motors is working very well. We are working together in using language technology in streamlining the whole process of the documentation.

CLARK:

How many people does it take to make all this happen? What size team do you have here at LANT?

CAEYERS:

It's growing each day. So yesterday evening someone asked how many employees there were. I didn't know, I had to count them. We have over 30 now. It's growing that fast now. On one side we have the further development of the basic modules and then we have our projects to customise them for different solutions. The more projects we have the more people we need. The problem is that it's hard to find good computational linguists. The universities in Europe are not following the future of what language technology can be at the moment. So we have to have internal training. So people coming from university, here in Belgium, are not well prepared for performing computational linguistic work. It's not easy at the moment. Here in Belgium you study engineering or mathematics or linguistics. And, when you have to combine them there is no opportunity to do so as an undergraduate. Maybe as post graduate students. I think it's time to let people start taking computational linguistics when they are 18 years old. ■