

Language personality of the month

Bob Clark
interviews
Eduard Hovy,
president
of the
Association
for Machine
Translation
in the
Americas

CLARK: Could you start right at the beginning? What led to your choosing a career in the field of computational linguistics?

HOVY: I was born in South Africa, in Johannesburg, of a Swiss mother and a South African father. So, at home, we spoke English to my mother and Afrikaans to my father. To one brother I speak Afrikaans, to the rest I speak English. At home, we were completely bilingual. We went to Afrikaans schools, some went to English universities, and some went to Afrikaans universities. We went to Boy Scouts in English, church in English. So the question of the language and the understanding of the play of words was always just part of our family culture. I don't think we even realised it. I think that's common. Many South Africans just grow up that way.

I was brought up in a household where there were various religious questions always going on. And eventually you start wondering, what is the real truth? And for me it was mathematics. So, I studied mathematics in Johannesburg at the Afrikaans University

you aren't allowed to sit down and you only drink tea for ten minutes and then you stand. I worked in a brewery; I made sandwiches in the Ministry of Fisheries and Agriculture in London, in Whitehall. All kinds of things, all over. But I did not have a work permit and eventually the police tracked me down and I had to leave.

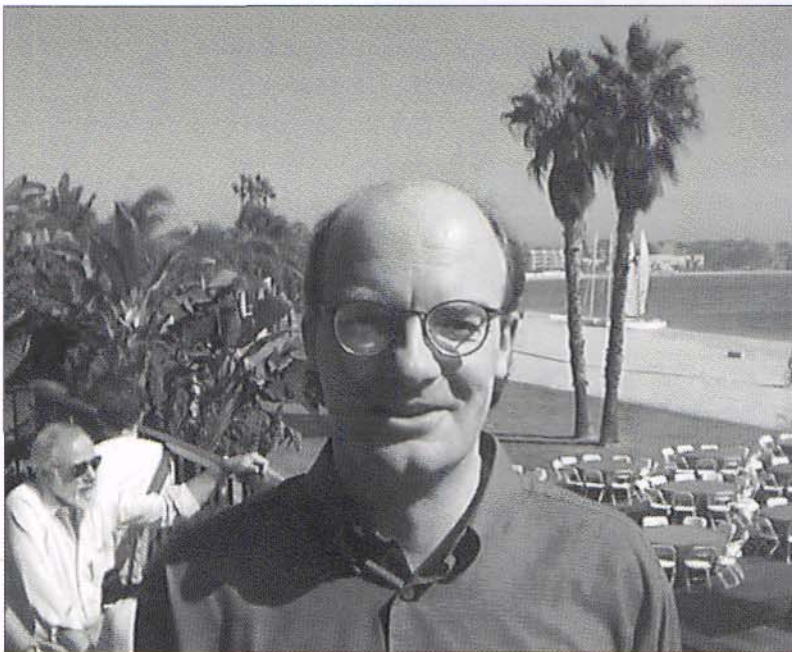
So, I went to Switzerland where I lived at my uncle's house and I did translation. I did German to English for a big company that was building a heavy water plant, a nuclear power plant in South America. So, our interlingua was English. We constructed two big dossier books with all the building plans. Down from the coat racks in the offices up to these big things like the cooling towers and stuff. So, I learned a lot of building terms, which I now have only in German. I can tell you what a gargoyle's little spout is in German but not in any other language. Certainly not in Afrikaans. And that gave me a sense of the excitement of translation, of really trying to get the meaning and, maybe, being allowed to change the form. Because, certainly the building office didn't care. They just wanted the meaning.

These documents were then sent to South America, Argentina, and translated there into Spanish. It was very interesting to think about, well, are they going to get the right thing through here? Because we were going via this interlingua, so to speak. I enjoyed that.

And then I started thinking more about studying again and I wanted to study computer science. Because it seemed to me that if you can build intelligence in a machine in a replical way then you can say this is what the truth is. This is what I am and this is where I'm going and so on. You know, what is the soul? And all those big questions you think about when you're 20.

So I went to the States for a month in 1980 and I travelled all around and visited various universities and applied. I went back to South Africa for some months, from October or so until September of the following year. I worked there in an oil company and had fun there. I did computing for them, I went on an oil rig and had various adventures and then I went to Yale.

I studied at Yale for five and a half years and I graduated with a PhD in Computer Science. My advisor was Roger Shank who is the originator of many



Eduard Hovy,
AMTA President

and I was a little disappointed. It seemed to me like games and so on. And then I looked at philosophy, I looked at psychology. I couldn't find anything that satisfied me. So I just gave up studies more or less in disgust after my honours degree and I went to Europe.

For 18 months, I worked in England. In a factory. I was a factory hand in Birmingham. You know, where

theories, conceptual dependency, semantic ideas, scripts and various things like this. It was an extremely interesting time, not always very happy. He is not a calm man, but he is a very interesting and exciting man to study with.

And then I got a job at the Information Science Institute in Los Angeles with the University of Southern California. So, there I joined a project that was focusing on natural language generation. My thesis was about producing variations of the same basic message in different ways, in different styles. So, once you speak to somebody that is antagonistic to you and who is angry and has little time, and another time you are writing a letter to your mother, and another time you are writing a telegram to somebody and you are neutral, you're a newspaperman or something. So, I had to define all kinds of parameters, social settings, and interpersonal levels. That was a lot of fun. Trying to do the same sort of thing at the Institute led me to working on discourse and questions of how do we construct more than single sentence communications.

CLARK: How did you eventually become head of the Information Science Institute?

HOVY: After some ups and downs with funding, my boss retired and various people left and we had a very small group in 1990. Just two people. The funding had dropped from a million dollars a year to 150 thousand dollars a year in a matter of two or three months. So, it was a bit of a challenge to learn what the funding game is all about. The fact that research involves as much the funding issues and structuring the future as it does doing the actual work. That was an early lesson to learn. I felt I would have liked to do more research before being thrown into it at that stage of one's life. But it was lucky too that, at that point, the Defence Advanced Research Agency, DARPA, was interested in Machine Translation. They talked to us and we were able to form a consortium with Carnegie Mellon University and New Mexico State University. Very famous people like Yorick Wilks, Sergei Nirenburg and Jaime Carbonell. So I was like the junior little tot, trotting on behind and learning everything I could. It was a wonderful experience. I was very lucky to be with the giants of the field.

And so, through the years, we were in the 'Statistics Wars' time when, on the one hand, DARPA funded the pure statistics system at IBM and, on the other hand, they funded the more traditional linguistics Machine Translation system of our group. And in the middle they funded something called LINGSTAT from Dragon Systems, the speech people, to mix linguistics and statistics, which turns out to be, today, the best answer. And they put us into competitive evaluations

every year and they had a lot of companies taking part. It was a very, very interesting time to see this aggressive, cut-throat kind of approach of the IBM crowd and the argumentative and very clever style of Carbonell and Wilks and people like this. It was a very good thing, throwing into the maelstrom of research, really bringing into computational linguistics this new paradigm of the statistical processing. It was wonderful to see that. It gave me a sense of awareness of the history, of how the ebb and flow of thought in history how it shapes the field, even a relatively small one like computational linguistics or Machine Translation.

So, from that I became interested in the idea of hybridising techniques from all the different fields to get the best result. Under evaluation, so that you are not just doing blue-sky research with interesting ideas that don't actually pan out necessarily. I suppose I became more of the engineer type scientist rather than the cognitive type scientist that looks at brain research for their validation or the intuition type scientist that looks at internal intuitions for their validation. And I find more and more that I'm comfortable in this position. I like that.

CLARK: And where do you think things are headed?

HOVY: Recently, what's happened that's exciting is that, in addition to machine Translation, more and more work is being done in automated summarisation. Where you put the text in the machine or on a scanner or you get it from the Web and you say, 'I want two paragraphs, or I want just the keywords, or I want a summary of a hundred documents'. You can imagine how wonderfully useful that would be, especially coupled with multilingual and information retrieval, Netscape and all this stuff together. Multimedia too, if you could put pictures in there and summarise the picture. There's some interesting thoughts in that too. So, stepping a little bit wider and looking at this whole field, this whole pot pourri of ideas on one table, through one lens, you can imagine. Just like technology. Suddenly within ten years, there's this burst of things and you can't live without it, as you see with computers or with fax machines.

I imagine that we are going to see very soon, within the next ten to twelve years, a proliferation of language-based tools in the average person's home. You are going to have your emailer, and more and more people have this. You are going to have your Netscape access, or whatever, to the Web. And in there and in your letter writer, your WordPerfect engine or whatever, you are going to find multilingual capabilities, information retrieval and summarisation. And it's going to be crude, it won't be wonderful but it

will do. It will give the ordinary person access. When I think of this, it is so exciting. I feel so privileged to be in a field that is doing something to really change society. Really there, on the information wave, now, today. Rather than, say, physics, which had its heyday in the early part of the century with the Bomb and all the physics. Or biochemistry, which had its plastics and things and now the medicines and things that are being developed. Really, Information Science and Language Information Science is exciting. I really love it. So, I feel sometimes that I don't know that one should be having so much enjoyment.

Yes, there are downsides too, right? I have to worry about funding. A lot. I have to worry about keeping the research groups alive. I have to ensure that our group does well-placed research that really speaks to the future and doesn't dream and get lost in various directions. I don't think, with the current situation and the shortage of funds and the competition, increasingly from Asia, especially from China, that we can afford many blind alleys. So, one has to combine your best technical judgement with some intuition, with some luck, with, maybe, some historical sense. It's nice to think again, in the sense of the history. What happens to fields and how can you try something but still hedge your bets? With a mixture of conservative and risky approaches,

so that always there is something that succeeds.

To conclude, coming to Machine Translation as an anchor point, just as summarisation is an anchor point, information retrieval is an anchor point, speech recognition is an anchor point. Of all those things Machine Translation is the oldest one. It has its 50th anniversary this year. Listening to the earliest pioneers and looking at the demonstrations at the MT Summit VI conference and seeing how their early intuitions and wishes and dreams have been realised and how excited some of them are, I was privileged to have dinner with Victor Yngve, one of the very early pioneers, and listening to what he said and how he was surprised at how some of these ideas have crystallised.

He can recognise in what we think of as the newest invention the seeds back to his days in 1950. It's really nice to see that and to know that a hundred years from now, if I was still alive, I would be able to go in there and recognise the seeds of the thoughts that we are now trying in the lab there, worked out and changed, wonderful and beautiful, but still the seeds are here. That gives you an excitement that you cannot believe. Whether or not you know the Truth in the religious sense, of who am I and all this, you've been part of the fabric of the intellectual growth of humanity, so to speak. I like that. That really gives me a lot of satisfaction. ■