

MY COMPUTER UNDERSTANDS ME

Piek Vossen is a bit like a supercomputer himself. He is a linguist and a computer programmer, a disciplined researcher and an efficient project manager. He works hard and successfully on computers that have a thorough understanding of language.



By Tom Arends and Rianne Lindhout

It all started with Deep Blue, the chess computer that beat world champion Gary Kasparov in 1997. Then the computer Watson won the prestigious American quiz show Jeopardy against strong competition from two high-ranking human contestants in January 2011. The IBM supercomputer understands spoken language and answers complex questions quickly ... much better than Anna, the online computer-assistant that answers customers' questions on the IKEA website, for example. What's next?

Maybe a computerized agony aunt (or uncle). This intelligent avatar (you can choose the gender yourself) not only sympathetically listens to your problems but can also interpret your facial expressions, the tone and volume of your voice and really put you on track to find a solution. Such supercomputers master all aspects of human communication, and give you all the attention you need at a reasonable price. That's a very interesting option, in view of the growing pressure all healthcare workers undergo nowadays. Piek Vossen (born 1960), Professor of Computational Lexicology in the Faculty of Arts at VU University Amsterdam, studies how computers can understand language. He believes it will take quite a long time before Windows offers a high-quality psychotherapeutic avatar as a standard app, but adds: "Researchers are making great advances in this field, investigating how computers interpret all kinds of human input from facial expressions and tone of voice to word choice and context. At a given moment it becomes possible to put all the pieces together. The visible results so far are mainly fringe applications with largely an entertainment value; the IKEA avatar is fairly primitive, for example. But if you've got the IT muscle of IBM and you understand the social aspects of language, you can get computers to communicate in an increasingly human way."

"Chinese people talk very differently the environment, from Dutch people"

SUPERMARKET LOYALTY CARD

Piek Vossen's activities can be compared with those of a supermarket loyalty card. Consumers think that those loyalty cards are just there to offer them a friendly little bonus, but what the

supermarket is really doing is recording all our purchases, analysing them and charting our consumer behaviour. This process is known as data mining. What Prof. Vossen is called text mining: the discovery of structures and meanings in the texts we write. "Language is very ambiguous," he explains. "Understanding a passage of written text involves more than just knowing the individual words. For example, Chinese people talk very differently the environment, from Dutch people. The environment is quite a touchy-feely topic in the Netherlands, but not in China. If you translate an environmental report the way Google Translate does – and that's not too bad at all nowadays – this will not be enough to bridge the gap between the Chinese and Dutch way of thinking completely." Language consists of words plus a context – which may be the message the speaker wants to get across, for example, or his political, cultural or personal background. Prof. Vossen's work is about recognizing contexts and taking them into account during the interpretation process.

Take the global climate problem, for example. It will be easier to solve if scientists from around the world can understand one another. Prof. Vossen is the coordinator of the EU Kyoto project, which aims to set up an international Wiki portal on the environment and ecology. The technology used here is a system that allows a certain community – in this case the community of climate scientists – to define the terminology and concepts it uses in such a way that a computer can also use them. This enables the computer to extract the relevant facts from the texts fed into it. You can then see at a glance when the glaciers in the Himalayas are expected to melt as a result of global warming according to the experts, who made these predictions, and what predictions have been made about other glaciers around the world. The system already masters seven languages, and can easily be expanded to handle others. If this works, it will also be possible to develop a similar system for other topics.

HISTORY RECORDER

Another one of Piek Vossen's projects is the development of a quadrilingual history recorder for economic news, for which he received a grant of nearly three million euros from the EU in May 2012. He is working with Italian and Spanish colleagues and the information broker LexisNexis, among others, to write computer software for a system that will 'read' and index the news daily,

keeping an accurate record of what happened when, and who was involved. It will then generate a single coherent description of the past from these individual news items. The underlying idea here is that humans are no longer able to consider all the information from (news) sources, because there are too many: who ever looks beyond the first ten hits in a Google search? We need more advanced assistance to extract the relevant information. Such a program must be able not only to select the right facts, but also to interpret them. If it cannot do that, the results it generates are not only unreliable but also useless as a basis for decision-making. Developing computer software that will interpret information is the key objective of Prof. Vossen's work.

AN EXAMPLE OF A HISTORY RECORDER: VOLKSWAGEN & TAKE-OVER

If you use Volkswagen take-over as a search term in Google, you will get thousands of hits relating to reports of take-overs over the years in which Volkswagen was involved. An intelligent survey of these reports will reveal the existence of a take-over war in the automobile industry in recent years. You will find many reports dating from the period round 2008 and 2009 indicating that Porsche planned to take over Volkswagen. They were buying Volkswagen shares, which led to speculation that a take-over was imminent. And then the situation was completely reversed in the course of a year, partly due to the banking crisis of 2008. Volkswagen started buying Porsche shares, Porsche's shareholder sued the company's management and Volkswagen emerged from this whole affair as the big winner. However, this story (and many like it) cannot be derived directly from the Google search results. It only emerges after the reports selected have been interpreted by someone with expert knowledge of the situation, or by a whole series of sophisticated computer programs. Each term has to be analysed to determine which event it refers to, who was involved (Porsche, Volkswagen or the management) and when it took place – or whether it really did take place. And then all these events have to be linked up to form a history. The whole set of computer programs involved is what we call a history recorder.

The focus on financial and economic news in this project was not chosen at random. It is a field in which a great deal of information is available – which makes it a good test case – and one in which a wide variety of professionals are under considerable pressure to take key decisions involving huge sums of money. They would welcome such a system with open arms.

"I was a real science geek when I went to secondary school"

AN OUTLOOK THAT COMBINES SCIENCE AND THE HUMANITIES

How did Piek Vossen achieve his prominent position in this field? It is because he has an outlook that combines science and the humanities – very rare in a researcher. "I was a real science geek when I went to secondary school. But then I got interested in music and literature. Not just reading and listening, but writing and composing too. And I became fascinated by the creative process: what is actually going on when you write lyrics or compose a tune? How do your ideas get converted into words, and your feelings into melodies? I became socially involved too, and wanted to become a writer. I decided to study Dutch at university. And then I discovered mathematical generative

grammar: that was the moment when science and humanities sides suddenly merged – it all came together, you could say." Piek Vossen wanted to understand how cognitive processes of communication and thinking take place in our brains, and how they can be recorded in a logical form. That is the core of his work. "Linguistics has only recently become an empirical science in which we analyse data and model this through software." He will have achieved his ultimate goal when he has written a computer program that takes over the task of interpreting language from us humans; then we will be one step further along the road towards a psychotherapeutic avatar as a standard app in Windows.

SPIDER'S WEB OF WORDS

When you know what you want to know, how do you turn this aspiration into successful research? One of Piek Vossen's success factors is that he managed to develop a model that was warmly welcomed by the research world. He was funded by the EU to build WordNets in eight languages. A wordnet is a linguistic database, a kind of spider's web containing all the words of a given language with their meanings, grouped into synonym classes and linked by semantic relationships. Wordnets can be used in artificial intelligence systems like the one Prof. Vossen is working on. Up to 1996, the only wordnet in existence was for English. Piek Vossen took the initiative to expand the database to include eight European languages. This was not just a question of cutting and pasting, since each language has a structure of its own and each one of us is more or less 'caught' in his or her own language. "The project confronted me with some fundamental questions," said Prof. Vossen: "For example, what is a word and what is a concept? Are all imaginable compositions of meaningful elements words and concepts? Can concepts exist without words to describe them? Which words are universal? There is no single word for 'pet' in Spanish or Italian, for example. The availability of such a word in English tells us something about the English way of talking about the world. Does this also imply a different way of thinging? Not necessarily."

"There is no single word for 'pet' in Spanish or Italian, for example"

When the project finished, the framework could not be continued through funding. "That would mean the end of the network of people and methods we had spent so much time and effort building up," Piek Vossen said. "And I didn't want that to happen. This whole subject is too important for that." If that organizational network was lost, it would no longer be possible to construct wordnets for different languages along the same principles and to link them or compare them with one another. That is why he founded the Global WordNet Association in 2000, together with some like-minded researchers. He is currently co-chair of this association. The main aim of this non-profit organization is to develop wordnets in a wide variety of languages and link them together. This project is also sometimes called the Global Wordnet Grid. It allows you to jump from any given wordnet to any other via the meaning of the words. How many world languages have a separate word for the concept of a pet, for example? Italian is one that doesn't – though it does have quite a nice expression for this concept: animale da compagnia, an animal for company.

>>



RISKS

A feeling for both exact sciences and humanities, a model that has been warmly welcomed by the academic community and a worthwhile initiative are not the only secrets of Piek Vossen's success. His way of looking for the essence of things is also a key ingredient. "I like to look at things in a different way, not step by step in a predetermined sequence from beginning to end. I often skip steps that others might regard as essential. I compare my approach to the way my children look at things. When my children were little, we often played the computer game Myst together. You had to solve puzzles in a mysterious, disorganized virtual world, and to my surprise my children were much better at it than I was. They took an unconventional approach to the game and dared to make moves that were not obvious. I try to do the same thing in my linguistic research. You run some risks if you work like that, but you can achieve a lot if you have a sound grasp of the underlying methodology."

PROUD OF VU UNIVERSITY AMSTERDAM

Piek Vossen is glad that he ended up at VU University Amsterdam. "There is an excellent climate for cooperation here, even across faculty boundaries. I have been in other places where most researchers prefer to be left in peace. The organizational set-up here is very supportive: people are not blinkered, and the research atmosphere is human and enjoyable."

Even so, this linguist with a strong feeling for science does sometimes feel a bit of a square peg in a round hole in the Faculty of Arts. "If I need to analyse a large amount of text, I can't do that here – though I do go around a bit like a prophet trying to persuade my Faculty colleagues that we need high-grade IT tools for linguistic analysis, for the students as well as the researchers." Fortunately, the Faculty of Sciences is not far away, just across the campus. That is where the supercomputer power Prof. Vossen needs for his studies is located. He welcomes the favourable climate for interdisciplinary research that prevails at VU University Amsterdam. Interdisciplinary research institutes have a very important role to play here. For example, Prof. Vossen works closely together with researchers from the Faculties of Sciences and Social Sciences of VU University Amsterdam from his position at the Network Institute.