

If Language Technology is the Solution, What, Then, is the Problem?*

Peter Bosch

IBM Deutschland Informationssysteme GmbH

Institute for Logic and Linguistics

PO Box 103068, D-69020 Heidelberg

bosch@vnet.ibm.com

By a technology I understand a set of technical solutions for a particular set or natural class of problems.

What makes me stop short agreeing that there is such a thing as "language technology", is not so much any disbelief or mistrust in the solutions that are offered under that name, but rather I have some difficulty identifying the class of real problems that these solutions are solutions for.

1 The Human Language Faculty as a Technology

For contrast with what people call "language technology" consider natural language itself. Natural language, i.e. the human language faculty, seems much closer to what I would be willing to call a technology. In this case a "wet" technology, if you wish. And it is a good technology that works reliably. Even though it shares with some man-made technology what we call the banana principle, i.e. it ripens with the customer, this is a process of ripening that reliably yields fair results. Natural language is one of those few technologies that adapt smoothly to whatever purpose they are put. It is user-friendly to the extreme: even the least intelligent specimen of the human race can use it for everyday purposes with no trouble at all.

The set of problems that this technology solves are, in the first instance, problems of *communication*. The transitory meaning-representation in the external physical medium of sound that is used in this technology is however itself capable of being represented persistently and thus naturally allows for other applications of Natural Language, in particular the *storage of information*.

Storage of information and communication are indeed real problems that Nature had to solve at a particular stage of evolution in order to ensure the survival of this race.

2 Problems for Language Technology

But what are the problems that man-made language technology is to solve ? I can see a choice of two basic options:

* This paper has originated from a statement for a panel discussion at the workshop on "Sprachtechnologie und Praxis der maschinellen Sprachverarbeitung" at the Werner-Reimers-Stiftung, Bad Homburg, 27th-29th May, 1991. I wish to express my gratefulness to the Werner-Reimers-Stiftung for their sponsorship and hospitality.

One is the problem of cloning, in soft- and hardware, a successful and established wetware product of our competitor Nature - up to now the unchallenged leader in this market.

The other problem language technology may be out to solve is to produce periphery, add-ons, and tools for Nature's successful product. And in some cases also fixes are required for Nature's errors or buggy implementations: aids for disabilities in reading, writing, hearing, and speaking - but this is only minor segment in the market.

In either case man-made language technology is dependent upon and secondary to Nature's language technology, and it is Nature who sets the standards. Our products must be Nature Standard or Nature compatible.

3 Nature's Adaptability

In a sense it is easy to make products that conform to Nature Standard, because Nature's technology is ever so adaptable. People who want to use speech recognition for practical purposes today are apparently quite happy to learn to pronounce words separately, with a short pause between them. And for many years anybody who worked with electronic computing gear and natural language didn't know any better than that these devices knew capital letters only and certainly wouldn't know any letters with funny dots or accent signs over them. Even nowadays there are millions of good people in the non-English speaking parts of the world who see their names misspelled every day for that very reason, and many of them have surrendered to misspelling their own name so as to adapt to a backward technology.

This may create the impression that whatever crap we produce, the customer will get used to it. And I believe this is indeed almost true.

There are only two limitations: One is that the customer will only adapt to a new tool when he has the feeling that it is a useful tool for his problems so that it is worth adapting to it.

The other is a more serious and very natural borderline to the language user's adaptability: he finds it very hard to grasp that language, for the machine, is a purely syntactic thing with no meaning or understanding whatsoever on the part of the machine.

This is not just due to lack of education, but to the very fact that even in science there is no conceptually clear separation between purely syntactic functions of human language and those that require the processing of non-linguistic or world knowledge. The fact of the matter is that Nature's language technology is rather, as hinted above, a language-and-information technology - with language and information inseparably intertwined, while current man-made language technology by and large ignores the information bit and treats language almost exclusively as a syntactic object.

For man-made language technology this means two things:

First, as long as we keep drawing the borderline between language and information in an ad-hoc manner (as we are forced to), practically all our products will yield the notorious 95-percent solutions. This is fine for many purposes, but unpredictable failure, even if only in 5 per cent of the cases, is unbearable in some applications. For simplicity's sake consider an example from an unrelated area of language technology: If a scanner recognizes only 95 per cent of the letters on a page correctly, this means, on an average typing page, 120 typing errors and is indeed quite scandalous.

Second, all applications are excluded that regularly require knowledge processing, or if you please, AI. And this is an area where I believe that we do have a set of genuine problems that a language-and-information technology still to come could solve: the control and penetration of

complex systems by their user, merely by means of the user's mother tongue. - But here language is used as a language in the full natural sense: not just a syntactic medium, but a semantically interpreted medium for representing and processing information.

4 Natural Language Interfaces and AI

The area where I see the most promising applications for computational linguistics coming up thus is not an area where we use language now. Rather it is in interaction with machines, particularly complicated ones, and more particularly, devices of information technology, where ergonomic (as well as economic) reasons plead for NL Interfaces, for the processing of spoken as well as written language, for text as well as dialogue.

But there are two problems here:

The first is that although it would be nice to do more with NL interfaces, the market has not yet understood that there is a need for them. In real terms: there is no recognized problem (even though we might have a solution)

The second is that we do not really have a full solution, but only parts of a solution. Computational Linguists cannot build such interfaces. NL interfaces are not just a matter of linguistics, but also of AI: you don't want to give your machine the parse tree or DRS, but you want the poor thing to understand and communicate with the user. But an explanation component of an expert system that was designed with no NL in mind will never talk comfortably to the planner of your NL generation component and parse trees or DRSs are pretty indigestible for current AI systems.

I suppose it is only fair to admit that there is no technology available in this area. Isolated solutions and many brilliant ideas in research: yes. But this is not what I mean by a "technology". Even conceptually we are far from a good understanding of the relation between language and knowledge and even further from an implementation of such understanding in terms of linguistically sensible Knowledge Representation Formalisms.

This is however the direction which I believe we ought to go. Not only because it is an exciting area to work in, but also because here we are working towards the solution of real problems for which there is a market.

5 Technology Creates Practice Creates Technology

What I am saying, I suppose, is rather trivial: new technology is created in tandem with a new practice. As long as there is no such practice and there are no products that this practice requires, all our linguistic engineering and computational linguistics is a bunch of nice and fancy ideas, but has little to do with technology.

One could go one step further and not only dispute the existence of language technology but also dispute that it has any right of existence. For the simple reason that there is neither a natural class nor even a significant set of practical problems that it could solve.

The class of problems I mentioned, one might argue, is catered for by what we call information technology and it is in this context where linguistic solutions are required, but as an integrated part of information management, as part of Knowledge Based Systems, and not as free floating parsers.

6 Market-Driven Research

Sure enough, we will never develop a technology in this area, worthy of the name, unless we make the market believe that we already have a technology. In a market-driven economy (with market-driven research) we are practically forced to claim that we can turn, if you will pardon me, shit into gold in order to be given a fair chance to prove that we can turn zinc and copper into brass.

In this sense then, I am, after all, glad to pronounce that I believe in the existence of language technology. I only ask your permission to modify this credo slightly: I believe in language-and-information technology. Here we do have a genuine set of problems to which this technology will - in due course - provide the required solutions.

7 But This is all AI and not Really Linguistics

True. And I have also ignored, except for a very short mention, what I called above the area of periphery, add-ons, tools, and fixes to Nature's language technology. But this is exactly the area of which I am claiming that there is neither a natural class of problems nor do the solutions that are available to the various odd and isolated problems in this area have very much in common that would justify calling this motly bunch of tricks a "technology".

This would all be different if we had at our command something like a generic natural language engine that incorporates the best insights of linguistic theory and can provide solutions for linguistic problems on this basis. The fact that no such natural language engine is available is due, next to the enormous difficulty of the subject matter, to the problem of separating, within the functioning of human language, the genuinely linguistic contributions from the contributions of the conceptual machinery and world knowledge, i.e. the problem of separating language and information. I am not sure that this problem can be solved. But what we can do is try to come to a better understanding of the problem, i.e. a better understanding of the interactions of language and information in the ordinary functioning of human language. – I don't care much whether this study is conceived of as part of Cognitive Psychology, Linguistics, or AI. But it is the problem we must focus on if we ever want to get anywhere near something worth being called Language Technology. It is only on the basis of this understanding that we can develop the concepts that can eventually provide generic solutions - and hence a technology in the proper sense - rather than a motly bunch of tricks. The step that is required is comparable to the step from Alchemy to Chemistry.