The AI vision then and now:
what does natural language processing tell us about it?
(Summary)

Karen Sparck Jones

Computer Laboratory, University of Cambridge

Where's the work on natural language gone?

I am asking this question as a lead into what was exciting about AI fifty years ago, and what's exciting now.

There were three events that throw light on how things were then: the International Conference on Scientific Information in 1958, the Conference on the Mechanisation of Thought Processes in 1958, and the Conference on Machine Translation of Languages and Applied Language Analysis of 1961. The first was about the growth of scientific information and what you could do about it computationally. Could you manipulate linguistically expressed information automatically, develop general theories for doing this? The second meeting ranged widely from core AI concerns to automating government administration, and again language processing figured. The third focused on language processing. All three shared the growing excitement about computing. But it wasn't about computing as, in the 1955 Dartmouth proposal's word, *simulation*: it was about computing as *emulation*. Automatic word classification wasn't called learning; the grit grinding in parsing for translation wasn't cognitive modelling. But doing things with language was important because language matters for humans.

So language processing was central in AI through the 60s and 70s; and maybe you could do it better if you did approach it with more attention to humans and modelling. Trying to do this led to a rich stream of research, and argument. But there were problems: effective natural language processing systems need a lot of hard graft on specific language detail which hasn't anything manifestly to do with knowledge about things and events in the real world, or about human attitudes to these. When people began to be able to build modest inquiry systems, verb tenses mattered and beliefs, desires, intentions and plans didn't much.

Language processing changed in the 80s: research in computational linguistics was delivering better tools, notably types of grammar, and ideas about what the dictionaries, on which all language processing rests, should be like. During the 80s, AI and language processing moved apart: language processing seemed to have less need of AI, and AI increasing interest in non-language areas.

But in the 90s, something else happened, that changed the relation between AI and language processing again. Knowledge of language (grammar, dictionary, etc) is essential for this processing, but facts about actual language usage also matter. The language apparatus may eliminate many interpretations, but not all. How then to select? Take the more common. Frequencies in usage data are needed to supplement, or complement, symbolic processes. More grandly, statistical language processing has a part to play. Fortunately in the 90s machine-readable text appeared in ever-growing quantities, far greater than anyone earlier had imagined. Natural language *information* processing systems now combine symbolic and

statistical processes, in varying ways according to the task: more statistics for retrieval, more symbolics for question answering or summarising, all tasks the 50s wanted to automate and all tasks then believed to require intelligent systems. But now, as then, these systems are envisaged as emulating, not simulating, humans.

Yet these systems also emphasise something new. In exploiting the text data, they are exploiting the outcomes of many particular language users, and while that is something that any individual human also does, the system exploitation is on a scale beyond the individual's power. Moreover, since language use is intrinsically social, is there something about the collective behaviour, as expressed in the evolving data, that is itself intelligent? Can one talk about intelligent *collective* behaviour, where this collective behaviour is a response to the collective data? AI has been traditionally concerned with modelling (or replicating) individual intelligent behaviour, even if that model itself is a community of agents. The question that language information processing, in the face of the vast quantities of data that computing is making available, now raises is whether there is such a thing as *social intelligence*, and whether the processes it applies are like or unlike those of individuals.

In the 50s, in language processing, we saw something like translation as an individual process, though it invoked dictionaries and grammars that represented collective knowledge. We did not think about the implications of this relationship, even if we assumed, for example, that the process applied to construct a semantic classification was the same for individual and social data. Now, with so much vaster quantities of data (and not only language data), we have the challenge and the opportunity of thinking about intelligent behaviour that is necesarily beyond the individual.