KAREN SPÄRCK JONES

Karen Ida Boalth Spärck Jones
1935–2007

KAREN SPÄRCK JONES was born on the 26 August 1935 in Huddersfield, Yorkshire, the only child of A. Owen Jones, and Ida Spärck. Owen Jones was a chemistry lecturer at a local technical college: Ida Spärck was a Norwegian who later went on to work for the Norwegian government in exile during the Second World War. During this period, with her mother in London and since her father often had to teach in the evenings, for two or three years Karen was looked after by another family, but other than this she spent her entire childhood in Huddersfield. Karen was educated at the local girls' grammar school and had characteristically determined at the age of 12 that she wanted to go to Cambridge. She indeed went on (1953) to win a place at Girton College, Cambridge, to read history. Girton was one of two Cambridge colleges to admit (only) women at that time (New Hall, a third, was founded in 1954) and in later years Karen maintained that 'there was something to be said for being in an all-girls' school and being at an all-girls' college. You were very serious about your education. It was a *privilege* to come to Cambridge; you didn't mess around.'[1] After graduating in History in 1956, she stayed on to read the third year of the Moral Sciences tripos, the term then used in Cambridge for philosophy.

After this enjoyable year, Karen spent a short period teaching, in a position found for her by her Girton Director of Studies via the old girls' network. She did not enjoy the experience and later also came to realise that there was something a little scandalous in obtaining a position, for

[1] Unpublished interview with Janet Abbate, dating from 2001: a corrected version dated 2006 formerly appeared on Karen's web site. The quotation is on page 4.

Proceedings of the British Academy, **166**, 273–298.

which she had no training, simply by coming from the right college! A lucky suggestion by a fellow undergraduate in Maths and Philosophy, Roger Needham, led to her joining the Cambridge Language Research Unit (CLRU) in 1958. CLRU was an independent research unit with no official status within the university, existing for about twenty years on 'soft' money. CLRU was devoted to the 'analytic investigation of language, and in particular with a correlative study of the descriptive-linguistic, logical, algebraic and other notational characteristics of natural languages and of translation between natural languages'.[2] It was housed in an outbuilding in the garden of a private house in Millington Road, and had grown out of a discussion group led by Margaret Masterman, a charismatic former student of Wittgenstein. CLRU employed some researchers on grants, and also had associate members from within the university: these included Michael Halliday, subsequently to become a very influential linguistic theorist, and Roger himself, who went on to become an eminent computer scientist, head of Cambridge University's Computer Laboratory for many years, and later the first director of Microsoft's Cambridge research centre. (For a description of the unit at this time see Margaret Boden's 'Mind as machine'.[3] Margaret Boden, a former Vice-President of the British Academy, was also associated with CLRU).

Roger and Karen were married in 1958 and continued to collaborate in research until Roger's death in 2003. Other members of CLRU included Yorick Wilks and Martin Kay, both of whom, like Karen, would later receive the Association for Computational Linguistics Lifetime Achievement Award: an astonishing tribute to the influence that this small unit proved to have on the discipline.

Mechanical translation and meaning

Karen undertook a Ph.D. thesis while working at CLRU: since this was not an academic department she was technically supervised by the philosopher Richard Braithwaite (Masterman's husband), although the topic itself was heavily influenced by Masterman's own interests in dictionaries and thesauri.[4] Karen later recalled her infrequent meetings with Braithwaite

[2] *Mechanical Translation*, 3(2) (Nov. 1956), 36–7: <http://www.mt-archive.info/MT-1956-CLRU.pdf>.
[3] Margaret Boden, *Mind as Machine: a History of Cognitive Science* (Oxford, 2007).
[4] M. Masterman, 'The thesaurus in syntax and semantics', *Mechanical Translation* 4(1/2) (1957), 35–43.

as pleasant but contributing little to her work. Her main influence—indeed the dominating influence in CLRU—was Masterman. The thesis was completed in 1964, but not published until 1986.[5]

One of the main lines of research at CLRU was mechanical translation, which was one of the most obvious applications of computers to non-numerical tasks. Although two small-scale prototype systems for machine translation had been developed in 1954, one a collaboration between Georgetown University and IBM (Russian to English) and one at Birkbeck College (English to French), the idea of using computers to process language was still more science fiction than science fact. Masterman had emphasised the role of semantics in translation (in opposition to the current Chomsky-inspired emphasis on syntax) and was developing various ideas about this which Karen took over and developed in her thesis.

The thesis begins with a short discussion of the requirements of precision in getting computers to deal with language. It is worth remembering that, at this time, almost all discussion of language took place within the discourse of literary criticism, linguistics, or philosophy of language. Literary criticism is of course not noted for mathematical precision. Within linguistics, Chomsky's *Syntactic Structures* had just been published in 1957 and had shown for the first time that it was possible to describe the grammatical structure of natural languages in a formal and rigorous way, even to the extent of being able to prove theorems about the properties of the descriptive devices needed. Philosophers of language aimed for precision of argument, to be sure, but were not yet 'formal' in the way they described the structures of language. Thus the notion of a logical and algorithmic treatment of any aspect of language was still relatively novel.

Even the idiom and mindset of computer programming was only just emerging. The first reprogrammable computer had only been built nine years before (by Maurice Wilkes and his team in the Cambridge University Mathematical Laboratory, subsequently the Computer Laboratory) and the science of computing was in its infancy. Few people were aware of the skills needed to programme a computer, even where the subject matter (for example, mathematics or engineering) was well understood: where the task involved replicating some aspect of human ability, such as translation, abstracting or summarising, the level of precision and detail needed would have seemed daunting (as indeed it is). Computers are completely unforgiving and utterly literal-minded: every little thing has to be spelled out for them explicitly, and even one small error may mean that no aspect

[5] As Karen Spärck Jones, *Synonymy and Semantic Classification* (Edinburgh, 1986).

of a programme works. Hence the emphasis, which must have been quite unusual at the time in studies of natural language semantics, on procedures that could be reproduced algorithmically.

Karen's thesis introduces several ideas that have found their way into the mainstream of work in computational linguistics. One of the main themes of the thesis is the role of a thesaurus in a possible machine translation system, influenced by Margaret Masterman's 1957 paper on the topic. One of the problems a machine must solve in translating from a source language to a target language is deciding which sense an ambiguous word has in the current context. For example, the English word 'bank' will translate into French as 'bord' or 'banque', according to whether the meaning is water or money related. While people seldom even notice ambiguities like this, the task of getting a machine to choose the correct sense is very difficult, and still unsolved in the general case.

The key observation (originally from Masterman) that begins the thesis is that ambiguous words like 'road' or 'canal' each have among their senses something like 'means of communication', along with others sometimes apparently unrelated. If we find these words together in the same text this repetition of one shared sense encourages us to interpret them in this way, because texts are usually coherent, not random sequences of sentences on different topics. The words may not mean exactly the same thing but they express the same general concept. We disambiguate by assigning the words to this concept: this is what Karen calls 'semantic classification'. A model of such a system of semantic classification is given by a thesaurus, such as Roget's, in which words are listed under the concepts or 'heads' they express. An ambiguous word will be listed under each of the heads which captures part of its meaning, along with other words that are synonymous with that sense of the word.

This suggests a simple algorithm for word sense disambiguation, using the observation that coherent texts will repeat ideas: simply replace each ambiguous word by all the thesaurus heads in which it is listed, and then look for heads that recur. Karen illustrates this with the sentence 'Let's play a game.' 'Play' occurs in eight Roget heads, and 'game' occurs in thirteen. The only head that recurs for both words is 'amusement' which is indeed the appropriate sense of the words in this context.

The thesaurus headings, she argues, could also be used to disambiguate words for automatic translation purposes. The assumption is that thesaurus heads, while stated in a particular language, are not language dependent—of course, there may be some concepts for which a particular language does not have a word, but for the most part this will be the case.

Now we have the basis for an algorithm to choose the appropriate target language translation for an ambiguous source language word. We look at the context in which the word occurs in the source language and apply the procedure described above to identify which thesaurus heading classifies this use of the word. Then we choose as our target language translation a word which appears under the same heading in the target language thesaurus.

Karen tested this procedure by hand using Roget's thesaurus and by simulating a target language thesaurus. While promising, there were many cases where the procedure failed, either because the word in question did not appear at all, or did not appear under an appropriate heading, or where the intersection of the heads containing the word and its context was empty. Karen's conclusion is that what is needed is a better thesaurus (and perhaps a more sophisticated intersection procedure).

The thesis now turns to the question of how to construct a better thesaurus. To build a new thesaurus, preferably automatically, we need to find a way of deriving the appropriate heads, and a way of assigning words to those heads. Some notion of similarity of meaning of words is required, but this has to be something that is empirically testable rather than being simply a question of intuition. And of course we have already seen that many words are ambiguous, so we really need a notion of similarity of senses of words. Karen takes the notion of two *sentences* having the same meaning or use (she calls it a 'ploy', a mode of emPLOYment) as a primitive notion and one not reducible to word synonymy. Rather, synonymy between words is defined operationally: two words are synonymous if they could be exchanged in a sentence without changing the 'ploy' of that sentence. While aware of the multitude of philosophical and linguistic questions such a definition raises—many of which she discusses—her conclusion is that, with suitable qualifications, this conceptual apparatus can be taken as a building block for the construction of a thesaurus.

A 'row'[6] of words which are mutually replaceable in a sentence without changing its ploy can be regarded as synonyms. We could choose one member of the row as a representative label, or we could simply let the row itself serve as a thesaurus heading. A row might be something like 'job, career, vocation', or 'career, swerve, skid'. Of course, the criterion for membership implies that all members of a row will be of the same syntactic category.

[6] 'Row' was Margaret Masterman's word and referred to the smallest unit in Roget, namely a group of words separated by semicolons.

This is a robustly practical way of approaching the difficult issue of word sense synonymy, and exactly the same approach was taken some twenty-five or more years later by the creators of WordNet,[7] the public domain English thesaurus. Their system is based on the notion of 'synsets' (synonym set) which is exactly the same idea as Karen's rows.

The notion of rows also makes possible a procedure for automatically deriving wider thesaurus style groupings of words that are semantically related, but not in such a direct way as synonymy. Rows which contain some of the same words may be regarded as semantically similar in some way, provided some kind of (empirically determined) threshold is applied to the density of recurring row elements. The thesis describes a number of experiments trying out different measures of similarity between rows in order to find naturally emerging groups of them. Karen was inspired to try these grouping methods—we would call them now clustering techniques—because they were the subject of Roger's own thesis work, developed with A. F. Parker Rhodes, another CLRU member, though Roger applied them to other types of information such as records of archaeological artefacts. To the extent that they could be tested on the very limited computational resources available at the time, the techniques were not wholly successful, almost certainly because of what would now be called 'data sparseness', that is very low frequencies of particular combinations of words. However, in subsequent work in the last ten or fifteen years on very much larger data sets it has been shown that what are essentially variants of Karen's methods do indeed lead to the kinds of results she was hoping for.[8] In fact, these distributional methods for discovering word senses were not the only ideas in the thesis which were only able to be properly tested much later on. Interestingly, in some of Karen's discussions in the later chapters of the thesis one can discern the seeds of ideas which were to become much more powerful in a different setting. For example (page 175) a measure akin to what was later called 'inverse document frequency' (see below) is introduced to capture the fact that rows sharing a relatively rare word are more likely to form a valid group than rows which share frequently occurring words.

[7] <http://wordnet.princeton.edu/>; 'WordNet: An Electronic Lexical Database', ed. Christiane Fellbaum (MIT Press, 1998).

[8] e.g. Hinrich Schütze, 'Automatic word sense discrimination', *Computational Linguistics*, 24 (1998), 97–123; James R. Curran and Marc Moens, 'Improvements in automatic thesaurus extraction', in *Proceedings of the Workshop on Unsupervised Lexical Acquisition* (SIGLEX), 2002, pp. 59–66; James R. Curran, 'From Distributional to Semantic Similarity', Ph.D. thesis, University of Edinburgh (2004).

In the final chapter Karen returns to the problem of word sense disambiguation and compares, informally, two strategies for achieving this: what she calls the 'route finding' or 'semantic distance' metric and a version of the original intersection of heads method. 'Semantic distance' is defined in terms of shared row components. If words A and B co-occur in a row, their distance is zero. If A co-occurs with C in one row and B co-occurs with C in another row the distance between A and B is one, etc. Both of these strategies in later work with bigger resources have led to successful word sense disambiguation algorithms. Karen leans in favour of the intersection route as being simpler and at least as accurate, although it is probably true to say that with larger datasets the semantic distance measure proves more robust.

Information retrieval

In the early 1960s Karen began work in the field of 'Information Retrieval', the process of retrieving (electronic) documents in response (usually) to a few key words typed in by the user. It is tempting to think of this as an early form of search engine like Google, and in a sense it is, although at the time this research was quite closely related to one of the traditional concerns of librarianship, namely classification of documents under subject headings and their subsequent retrieval, which can be thought of as an analogous process. While Karen's association with the CLRU continued until 1968, in 1965 she took up a three-year research fellowship at Newnham College, Cambridge, but became increasingly based at the Computer Laboratory, which was her home for the rest of her career.

The move from machine translation to information retrieval must have seemed like a very good decision when in 1966 the Automatic Language Processing Advisory Committee report (the 'ALPAC Report') was published in the USA.[9] This report had been commissioned in order to assess the state of the art in automated translation; it was thoroughly scathing about the results achieved by machine translation systems to date and, worse, pessimistic about the likelihood of improvement in the near term. There is, of course, considerable debate about whether these conclusions were merited, but whether merited or not the effect of this report was to

[9] John R. Pierce, John B. Carroll *et al.*, *Language and Machines: Computers in Translation and Linguistics*. ALPAC Report, National Academy of Sciences, National Research Council (Washington, DC, 1966).

cut off most US funding for research in machine translation for the next decade or so, and the effects of this extended winter were also felt in the UK.

Independently of funding considerations, there were in fact rather natural extensions of Karen's work in thesaurus construction to information retrieval, which CLRU had explored in an earlier paper.[10] If you are searching for a document using a particular set of keywords or index terms, you would want that search to locate also a document which contained a synonym of one of those keywords, particularly if the original keyword search produced few or no matches. In a search for documents giving advice about careers you would also want those that happened not to contain that precise word but a variant like 'job' or 'profession'. It might also be useful to extend the search to documents containing words or phrases with other semantic relations like hyponymy and hypernymy ('inclusion of meaning': 'dog', 'cat' etc. are hyponyms of 'animal', which is one of their hypernyms). Karen also points out that non-semantic relations may also be useful: for example, words with a high co-occurrence probability with search terms are likely to signal a relevant document.

All of these possibilities can be explored given an automatic method of obtaining what Karen called 'classifications', thesaurus-like groups of keywords having these specifiable loose semantic relations. Such explorations were the topic of the research she carried out in the latter half of the 1960s, summarised in her book *Automatic Keyword Classification for Information Retrieval*.[11] Roger Needham's 'theory of clumps' was the starting point for ways to derive automatically networks of words displaying different kinds of semantic relations to each other. Given such networks, the utility of different types of retrieval thesaurus could be measured precisely. Information retrieval in the 1960s was considerably more sophisticated in its evaluation methodology than machine translation was, although the latter is perhaps a more challenging task: certainly the question of how to evaluate the quality of machine translation systems has not yet been settled to everyone's satisfaction.

Performance of an information retrieval system, given a fixed test set of documents, can be measured in terms of 'recall' and 'precision'. Given

[10] Presented at a conference in 1958, but not published until the following year: Margaret Masterman, Roger M. Needham and Karen Spärck Jones, 'The analogy between mechanical translation and library retrieval', in *Proceedings of the International Conference on Scientific Information, Washington, D.C., November 16–21, 1958* (Washington, DC, 1959), pp. 917–35.

[11] Karen Spärck Jones, *Automatic Keyword Classification for Information Retrieval* (London, 1971).

a search request from a user, and assuming that we know exactly which of the documents in the set are relevant to it, that is, the documents that should be retrieved, then 'recall' is defined as the number of relevant documents retrieved, divided by the actual number of relevant documents; 'precision' is defined as the number of relevant documents retrieved, divided by the total number of documents retrieved. Colloquially, recall measures how many documents you got of those that were there to be got, and precision measures how many of the documents that you got were correct. The system will have high recall if it gets most of the truly relevant documents, but of course it would be easy to achieve this by just retrieving *all* the documents, relevant or not. The system will have a high precision if it retrieves *only* relevant documents (even if it only retrieved a few of these). Ideally, of course, we would retrieve *all* and *only* the relevant documents, but in all practical systems the best combination of precision and recall is usually dictated by the use to which the system is put.

Karen was able to carry out precise experiments measuring the effect on precision and recall of different classification schemes using a collection of documents assembled by Cyril Cleverdon at (what was then) Cranfield College of Aeronautics (now Cranfield University). Cleverdon, a librarian, and his team had begun work on this collection with US funding in 1957 and, in effect, helped to lay the foundation for all later work in information retrieval, in particular by providing these benchmark data that enabled competing ideas to be rigorously tested. The pattern of Karen's experiments was as follows. Starting with a list of 712 manually specified index or search terms provided by Cleverdon, and a machine readable version of Roget's thesaurus—originally input via punched cards—various clustering algorithms were used to derive automatic classifications of related words along various dimensions of similarity, characterised informally in terms of the geometric shape yielded by the similarity measure (for example, a 'string' consisted of a set of terms each connected to at most one other, a 'star' consisted of one term connected to all the others, but where none of the others had further connections, etc.). Now it was possible to compare the results of searches made with just the original index search terms against those made with the classifications associated with these extra terms, along with various adjustments for frequency of occurrence of terms in the document set. Another variable to be tested was the distribution of the original search terms: Karen experimented with a restricted subset of terms characterised by the fact that they were relatively uncommon in the document collection. By restricting the vocabulary to those terms that occur relatively infrequently in the document collection, intuitively you are homing in on the most important and informative of the index terms. A term that occurs in almost all documents will have little discriminatory power.

The findings from these experiments were that performance was indeed better using automatically derived classification networks instead of just terms, depending on the precise values chosen for the relevant parameters, but with the important proviso that only classifications derived from the restricted vocabulary of terms rather than the full set were used. This finding, already prefigured in Karen's thesis work, was later to form the basis of one of the most widely used principles in information retrieval, the notion of 'inverse document frequency'. This principle is still widely used in almost all search applications, and was described (although not by that name) in her 1972 paper 'A statistical interpretation of term specificity and its application in retrieval',[12] still one of the most frequently cited papers in the field. The inverse document frequency measure is usually used in combination with a measure of term frequency (how many times the term appears in a particular document) in the celebrated *tf*idf* formula. Put simply, if you are searching for a document using keywords $W_1 \ldots W_n$ then a document will count as relevant in proportion to, for each W_i, the number of times W_i occurs in the document (usually normalised by the length of the document to avoid long documents being unfairly preferred), multiplied by the informativeness of W_i, its inverse document frequency (the reciprocal of the number of documents the word appears in). Almost all information retrieval systems and search engines today use some version of this formula, usually along with other factors. A related idea, based on weightings derived from relevance feedback (judgements from a user about how relevant were the results returned from a preliminary search), was further developed in collaboration with Stephen Robertson,[13] and this technique, refined after many experiments demonstrating its usefulness, was also an important contribution to the field.

The inverse document frequency principle plays a part in millions of people's lives these days, via its incorporation in most, if not all, internet search engines. This originally happened almost by accident: in a later 1994 technical report Stephen Robertson and Karen summarised all of their work on these and other measures in a nicely self-contained and

[12] Karen Spärck Jones, 'A statistical interpretation of term specificity and its application in retrieval', *Journal of Documentation*, 28 (1972), 11–21.
[13] Stephen E. Robertson and Karen Spärck Jones, 'Relevance weighting of search terms', *Journal of the American Society for Information Science*, 27 (1976), 129–46.

accessible way.[14] A former Computer Laboratory Ph.D. student, Michael Burrows—co-inventor of the widely used Burrows–Wheeler text compression algorithm—then working at the Digital Equipment Corporation (DEC) research laboratory in Palo Alto, California, was given the paper by Roger when he was describing his need for searching large amounts of text. The techniques described in the paper were subsequently incorporated by Burrows and his co-designer in the first general purpose search engine, Alta Vista, which DEC launched in 1995.

During the 1970s, Karen became increasingly interested in the methodology of information retrieval experimentation, and frustrated at the relatively small amount of benchmark material available for precise experiments. She had always aimed in her own approach to information retrieval to treat this kind of experimental work as a serious scientific enterprise, with all the variables and parameters being carefully controlled. In collaboration with Keith van Rijsbergen and others she developed a design for a test collection of documents that would be capable of supporting a variety of information retrieval experiments in a more robust and scalable way than anything currently available.[15] Unfortunately, it proved impossible to secure funding in the UK for such an ambitious project and, as with so many areas of computer science, it was left to the USA to mount an equivalent program, albeit fifteen years later, with the 'Tipster' and subsequently the 'Text Retrieval Conference' (TREC, beginning in 1991) series of meetings, which developed and used the necessary infrastructure, and which has led to enormous progress in many areas of text retrieval and processing. Karen was one of the few UK advisers and contributors to this activity during the 1990s, and there must have been a considerable mixture of 'if only . . .' and 'I told you so . . .' feelings. Who knows whether with more generous funding at an earlier time we might not now have a British Google? The results of Karen's interest in the details and principles of information retrieval experimentation—along with most other major researchers in the field—were recorded in her 1981 edited book *Information Retrieval Experiment* [*sic*],[16] the only book devoted to this topic until the results of the TREC projects were published in

[14] Stephen E. Robertson and Karen Spärck Jones, *Simple, Proven Approaches to Text Retrieval*, University of Cambridge Computer Laboratory Technical Report no. 356 (Cambridge, 1994).
[15] Karen Spärck Jones and Keith van Rijsbergen, 'Information retrieval test collections', *Journal of Documentation*, 32 (1976), 59–75.
[16] Karen Spärck Jones, *Information Retrieval Experiment* (London, 1981).

2005.[17] Karen fittingly contributed an epilogue 'Metareflections on TREC' to this volume.

Back to computational linguistics

Since the end of the Newnham research fellowship in 1968 Karen had been supported at the Computer Laboratory by a series of temporary fellowships funded by, among others, the British Library, the Royal Society and the General Electric Company. This gave her considerable freedom to choose the directions for her research and by the late 1970s and early 1980s, perhaps in reaction to the disappointing state of information retrieval funding at the time, her interests began to turn back to computational linguistics or, as it had by now also come to be called, natural language processing. (It is possible to associate an ideological distinction with the different labels: natural language processing tends to have less linguistics in it, and more engineering.) Work in the area had continued, despite the ALPAC report and, in the UK, the 'Lighthill Report' of 1973.[18] This latter, written by Sir James Lighthill for the Science Research Council (ancestor of today's Engineering and Physical Sciences Research Council), evaluated academic research in the emerging field of artificial intelligence and gave a very pessimistic forecast for its likely success, being particularly negative about the field's ability to scale up from small, slightly artificial, domains to larger real-world problems (as if this was not a difficulty that had occurred to those in the field).

Although indeed limited to narrow domains, natural language processing systems like Terry Winograd's 1971 'SHRDLU'[19] and William Woods's 1973 'Lunar'[20] represented huge breakthroughs in the field. SHRDLU allowed the user to interact with a (simulated) robot in a landscape occupied by blocks and pyramids of varying sizes and colours, allowing for conversations such as the following.

[17] Ellen M. Voorhees and Donna K. Harman (eds.), *TREC: Experiment and Evaluation in Information Retrieval* (Boston, MA, 2005).
[18] James Lighthill, 'Artificial Intelligence: a general survey', in *Artificial Intelligence: a Paper Symposium*, Science Research Council. Reprinted in *Artificial Intelligence: Critical Concepts, Volume 3*, ed. Ronald Chrisley (London, 2000), pp. 497–518.
[19] *Procedures as a Representation for Data in a Computer Program for Understanding Natural Language*, MIT AI Technical Report 235 (Cambridge, MA, Feb. 1971); T. Winograd, *Understanding Natural Language* (New York, 1972).
[20] William Woods, 'Progress in natural language understanding—an application to lunar geology', in *AFIPS Conference Proceedings*, 42 (1973), 441–50.

> Person: Pick up a big red block.
> Computer: Okay (does it).
> Person: Grasp the pyramid.
> Computer: I don't understand which pyramid you mean.
> Person: Find a block which is taller than the one you are holding and put it into the box.
> Computer: By 'it', I assume you mean the block which is taller than the one I am holding.
> Computer: Okay (does it).
> Person: What does the box contain?
> Computer: The blue pyramid and the blue block.
> etc.

Woods's Lunar system used a general-purpose syntactic and semantic analyser to answer questions, sometimes involving some further reasoning and computation, about a NASA database containing the results of chemical analysis of moon rock samples brought back by the recent Apollo missions:

> Give me all lunar samples with magnetite.
> What is the average potassium/rubidium ratio in basalts?
> In which breccias is the average concentration of titanium greater than 6%?

While the range of things these systems could talk about was limited (we are a long way from passing any Turing tests here) the syntactic and semantic constructions implemented were truly impressive (even if the actual performance of the systems was somewhat fragile). Communicating with a computer in fairly natural English about the composition of rocks recently brought from the moon must have been incredibly exciting as a demonstration of what contemporary science and technology was capable of. (The 1968 film *2001—A Space Odyssey* had featured HAL, an artifically intelligent speaking computer.)

More importantly, rather than being just an individual tour de force, these early systems provided well-defined formalisms like Augmented Transition Networks (a computationally oriented grammar formalism), which could be reimplemented and used by other researchers. Aided by a talented group of Ph.D. students (in particular Branimir Boguraev, Hiyan Alshawi and John Tait) who were able to supply the software skills necessary to such implementations, Karen was able to embark on a series of projects exploring natural language interfaces to databases as well as combinations of natural language processing and information retrieval.

In a series of papers Karen explored the relation between linguistic processing and non-linguistic knowledge, both in the context of building practical systems that would enable you to interrogate a database by typing in English sentences (instead of having to learn a specialised database query programming language like SQL), and also with attention to the problem of customising or transporting such systems to new databases and types of information.[21] The kinds of problem to be faced can be illustrated with a simple example. Assume we have a relational database with, among others, a <person, location> relation, stating which office an employee is located in, and various other relations like <location, car-park?>, <location, near-hotel?>, detailing properties of the various office locations like whether they have a car park or are near a hotel. There are many problems to be solved allowing natural language access to even such a simple database, but the two most prominent are (a) the variety of ways in which a request can be phrased, and (b) the amount of reasoning that may be needed to go from the literal content of the natural language request to something that will be accepted as a valid database query. Consider, for example, the number of ways in which one might ask where an employee is based:

Where is Smith?
Where is Smith based?
Which city is Smith in?
Give me Smith's location.
What is Smith's location?
Where's Smith's office?
Which office is Smith based in?
etc.

Common practice at the time was to build the natural language analyser in such a way that essentially the same meaning representation would be assigned to all of these variants. While a workable and practical solution, this approach nevertheless requires you to redesign the natural language analyser when adapting the system to a new database. This is rather inefficient, and completely out of line with intuition: the syntax and semantics of English surely stay the same whatever factual domain we are talking about. However, the alternative of trying to use a general-purpose English analyser causes other problems to occur. For example, on purely syntactic grounds a request like 'Show me employees in city offices with car parks'

[21] Branimir Boguraev and Karen Spärck Jones, A 'General Semantic Analyser for Data Base Access' *Proceedings of the International Joint Conference on Artificial Intelligence (IJCAI)* 1981: 443–5; Branimir Boguraev and Karen Spärck Jones: 'How to drive a database front end using general semantic information', in *Proceedings of the 1st Applied Natural Language Processing Conference (ANLP),* Association for Computational Linguistics, 1983: 81–8; Karen Spärck Jones, 'Shifting meaning pepresentations', *IJCAI 1983*: 621–3.

can be interpreted as a request for city-based employees who own car parks, indicated informally by bracketing constituents thus:

> 'Show me [[employees in city offices] with car parks]'
> instead of the more natural one:
> 'Show me [employees in [city offices with car parks]]'.

It is not a sensible strategy to just rule out some syntactic possibilities by fiat, fixing the analyser so that only one syntactic analysis is delivered, because an apparently similar request 'Show me employees in city offices with computer skills' is in fact interpreted in exactly this 'wrong' way. It is not asking for offices with computer skills.

The second type of problem arises when there is a mismatch between the literal content of the sentence and the structure that is needed for a valid database query. So for example if we had a question like 'Show me all employees with car parks' which has no relevant alternative syntactic analysis to the one which is talking about 'employees with car-parks', we will be faced with an apparently impossible database query, since there is no <employee, car-park> relation in our database. In fact, of course, the query is perfectly interpretable as a kind of coercion or metonymy with the meaning 'Find me all employees in locations with car parks.' But in order to arrive at this well-formed database query we must make the inference that since employees are in locations and it is locations rather than employees (in this database of least) that have car parks, then we must ask a query that finds the set of employees that are in locations that have car-parks.

Work continued throughout the 1980s and early 1990s on these issues in several further projects, culminating in one with Ann Copestake (who was later to become a lecturer and colleague of Karen's in the Computer Laboratory). In a series of papers,[22] Karen described several experimental implementations which attempted to address these issues. Although the systems gave reasonable performance, in that some methods were shown to be possible in principle—in so far as this could be assessed in the absence of a thorough evaluation regime—the work did not have a particularly big impact on the field. Although the problems Karen was addressing have not gone away, and in fact recur in many other types of envisageable natural language processing application, this particular

[22] e.g. Bran Boguraev, Ann Copestake and Karen Spärck Jones, 'Inference in natural language front ends', in R. A. Meersman and A. C. Sernadas (eds.), *Data and Knowledge* (Amsterdam, 1988); Ann Copestake and Karen Spärck Jones, 'Natural language interfaces to databases', *Knowledge Engineering Review*, 5:4 (1990), 225–49.

application has become rather unfashionable. The use of natural language as a means of accessing information held within databases has led to some commercial systems, but none of them has been very successful. Rather, much of the kind of information which at one time was only available in such databases is now often freely available on the world wide web for anyone with a browser, and for professional users of commercial relational databases progress in the development of graphical interfaces has made access to such databases much easier for anyone prepared to spend a little time getting familiar with these tools.

'Heavy duty public service'

However, in other respects Karen's activities in the early 1980s did have a very considerable effect on the field. This was a good time in the UK in terms of funding for natural language processing, and artificial intelligence in general, although for political reasons to do with the legacy of the Lighthill report 'artificial intelligence' was renamed as 'intelligent knowledge-based systems' in government funding programmes. The Japanese had announced in 1982 their 'Fifth generation computer' programme for the development of powerful new computing hardware, and intelligent software to run on it. This was intended to provide a stimulus to the already powerful Japanese electronics industry and its announcement rang alarm bells in Europe and, to a lesser extent, the US. Japan was already feared as an industrial competitor. The UK government's response was to set up the 'Alvey' program in 1983, which was similarly aimed at encouraging research progress and collaboration between academic departments and industry via jointly funded projects. Karen was an adviser to the Alvey programme and it is almost certainly largely due to this period of what she referred to as 'heavy-duty public service' that research in natural language processing was able to flourish in the UK in centres like Cambridge and Edinburgh during the remainder of the 1980s and into the 1990s. Two particular factors enabled Karen to build up Cambridge as a centre for this type of work. In 1983, in collaboration with Frank Fallside, Professor of Information Engineering in the Department of Engineering at Cambridge University, an expert in the burgeoning field of automatic speech recognition, Karen secured funding for a new interdisciplinary one-year M.Phil. course in Computer Speech and Language Processing, taught jointly between the Computer Laboratory and the Engineering Department, and a 'new blood' lectureship in each department to support the teaching. (Steve Young was hired in the Engineering Department and the present author in the Computer Laboratory.) This course, which took around twenty students a year from 1985 until its last entry in 2009, produced many students who went on to distinguished academic careers (including one Cambridge and two MIT professors) as well as being represented in almost every industrial research laboratory with interests in speech or language. Karen taught a component of this course almost every year until her retirement, usually presenting an overview of implementations of historically important natural language processing systems. It should be said that Karen was not particularly comfortable as a teacher, but carried on gamely despite occasionally unfavourable student feedback. One factor in this may have been Karen's habit of continuing to use handwritten transparencies, long after everyone else had moved to computer-generated presentations. Her handwriting had the curious property of appearing from some distance to be very neat, while being almost illegible once close enough to read. Karen's notoriously high rate of words per minute when speaking also made no concessions to the non-native speakers in the audience.

The second result of the Alvey programme was the establishment of SRI International's Cambridge Computer Science Research Centre. SRI International (formerly known as Stanford Research Institute, based in Menlo Park, California), is a distinguished not-for-profit research organisation founded by Stanford University to carry out research and development of a type usually a little closer to practical applications than that typically undertaken within universities. Its research and development work is mainly funded by government, health or defence agencies. Karen had had many contacts with SRI over the years—as she had with many other US research labs—and had a high opinion of their natural language processing and artificial intelligence research groups. At the time, she and Roger had facilitated an arrangement whereby SRI would second a senior researcher to Churchill College, Cambridge for a year collaborating with Karen and her group in the Computer Laboratory. When SRI considered opening up research labs outside the US—it already had consultancy offices in London and Frankfurt—Cambridge was the natural place to go, and the Alvey programme, with its aim of encouraging collaborations between industry and academe, was the natural place to look for some initial investment. Karen was working closely with David Thomas, director of the Alvey intelligent knowledge-based systems programme, who was very enthusiastic about setting up an SRI Cambridge laboratory with real industrial involvement in natural language processing research (up to that time, all the Alvey natural language processing projects were purely academic). Bob Moore, the SRI Churchill Fellow at that time, and the first director of SRI Cambridge, recalls that Karen's help was invaluable in persuading the industrial members of the consortium (British Aerospace, British Telecom, Hewlett-Packard, ICL, Shell, Phillips, and Olivetti) to sign up to a three-year project to build a 'Core Language Engine', a general-purpose syntactic and semantic analyser for English. SRI Cambridge, after a year of so of planning and raising research funding, got started in 1986 and continued for the next fifteen years or so to collaborate with the Computer Laboratory on a variety of projects in natural language processing, as well as other areas of computer science, providing a source of employment for several of Karen's Ph.D. students.

Although Karen continued to support SRI, the direction that natural language research took there became less congruent with her interests, which had by the early 1990s now turned more towards tasks like automatic summarisation and less well-researched areas of language understanding like discourse and dialogue—the study of the rhetorical structure of text and conversation. However she continued to be actively involved in the Core Language Engine project (1986–9), and its successor project from 1989 to 1992, and as her own contribution worked on a study of evaluation methods for natural language processing systems, eventually published some years later.[23] This was the first book devoted to this topic and it is still highly relevant today.

Unlike the information retrieval community, natural language processing had been slow to develop standard ways of testing its systems. However, in the world of automatic speech recognition, by the late 1980s statistical methods were sweeping all before them, consistently outperforming the more traditional handcrafted systems, and providing new paradigms of rigorous methods of performance assessment. Such systems train a probabilistic model on a large corpus of data, and given such a corpus it is easy to measure performance by training a model on one part of the corpus and testing it on another unseen portion. This allows for perfectly precise and objective measures of performance. Of course, in the case of speech recognition the criteria for success are quite clear—how many of the words that were actually spoken did the system correctly recognise? In the case of natural language processing these criteria are less clear—in her book Karen advocated the need for well-constructed test materials analo-

[23] Karen Spärck Jones and Julia Rose Galliers, *Evaluating Natural Language Processing Systems, an Analysis and Review* (New York, 1996).

gous to those being used by the speech community (as she had previously urged for information retrieval) but pointed out that it was also important to evaluate natural language processing systems in the context of the tasks they were intended to carry out. For example, a database query system of the type we sketched earlier might well score very highly on a purely syntactic measure, getting perfectly correct analyses for input sentences, but that would be completely useless if not one of these led to queries that produced the correct answer. Likewise it would be possible in principle for a system to score highly on a metric counting the proportion of correctly answered queries, while performing relatively poorly in terms of linguistic accuracy, since it might be that just the presence of a few keywords was sufficient to deduce the relevant query expression.

The book contains a comprehensive survey of previous attempts at evaluation, and a discussion of the emerging 'bake-off' evaluation methodology then being developed by the US Defense Advanced Research Projects Agency (DARPA) and the US National Institute of Standards and Technology. The term 'bake-off' (not official terminology!) refers to a small-town bakery competition popular in the United States, and used in this context because the typical DARPA evaluation regime (for example in the Tipster, TREC, and Message Understanding Conference (MUC) programmes) was as follows: the test is defined with a clear measure of accuracy, and a set of example data is released, which will have the correct answer associated with each example (often produced by hand). For example, in the case of the MUC task of 'information extraction' the task is defined to be the recognition of some basic 'who, what, when, where, why …' chunks of information from a newspaper text. A set of example texts with the correct values of these chunks would be released so that the competing research groups could customise and refine their analysis systems. After a period some similar previously unseen test data are released and the performance of the different systems is measured by a neutral observer, usually a piece of software written for the purpose. Finally, the results of the bake-off are announced, with possibly adverse reputational and even funding consequences for those groups at the bottom of the list.

This evaluation regime has the effect of forcing evolution, as clearly successful techniques will tend to be quickly adopted by all the teams. The criticism is that techniques which at a particular point do not seem competitive but which might be winners later on after some more research might get squeezed out in favour of a quick win, especially if the funding of your group is at risk. Karen takes no stand on this, but points out that evaluation methods like these, while undoubtedly valuable and which have led to important advances in the field, nevertheless ignore most aspects of what she calls the 'setting' of the task, by which she means the intentions and interests of the end-user of a system. To stick with the information extraction example, the MUC measures are in terms of recall and precision (or a weighted mean of the two), analogous to the information retrieval task. But in particular settings recall and precision might not be equally valued. For example, if I am an intelligence analyst scanning through the output of such a system that has been processing material on fundamentalist Islamic websites, looking for hints of future terrorist activity, I am likely to value recall over precision. I may not mind numerous false positives, but missing even one true positive could be a disaster. On the other hand a similar system in a different application, such as the detection of descriptions of protein interactions in a large molecular biology document collection like Medline (the US National Library of Medicine's online database of medicine-related scientific articles) might well place a premium on precision. There is not time to follow up every possibility, so an investigator will want to be sure that every hypothesised interaction has a high probability of being a genuine one. Thus if the setting of information extraction as a task was properly taken into consideration, then the ability to vary precision and recall ratios with ease might be a significant factor in evaluating one system as more effective than another.

In the remainder of the 1990s Karen continued to work on these topics, along with various aspects of 'user modelling', that is, the incorporation of knowledge about various aspects of a system user—their interests, level of expertise, etc.—in order to produce more natural or helpful output. Of necessity, since this work presupposed a level of maturity in natural language processing systems that had not then, and has still not, been achieved, much of this work was carried out in a 'gedanken experiment' fashion, in which she had to imagine not just the system, but also the entire environment in which it might be used.

She also continued her pioneering work on automatic summarisation, again largely a pencil and paper exercise, in which the output of imagined future natural language processing systems was simulated as a way of trying to develop at least an abstract picture of the types of operation that would be needed to carry out summarisation. While this work did not lead to implementations, several aspects of it have become very influential. Her identification of factors over and above the purely linguistic measures influencing evaluation of summary quality, such as the interests of the intended recipient, the purposes for which the summary is intended, and others, were a major influence on the definition of the various tasks and their evaluation which the US Document Understanding Conference (DUC), a forum for summarisation and other types of text processing in the style of TREC and MUC, was to develop. Karen became an outspoken member of the advisory board of DUC and had a large influence on the direction this research and evaluation methodology developed. A second major contribution was her work on the use of rhetorical structure theory to identify the main points of an article or other text theory.[24] This was also innovatory and influential, and even though this work too did not lead to implementations, the questions that Karen was addressing in her user modelling and summarising work remain valid ones, and her careful conceptual ground-clearing in all these areas will, when the technology has caught up, be seen to be a valuable and visionary contribution to the field. Her 2000 paper remains a key paper for anyone working on automatic summarising.[25]

Karen also had a series of fruitful collaborations in more implementation-oriented projects with the speech recognition group in the Department of Engineering, taking part in several large-scale collaborative efforts on topics like retrieval from news audio or multi-modal documents. Recent progress in automatic speech recognition meant that a transcript of spoken news reports could be obtained which, while by no means perfect, would be sufficiently accurate that document retrieval techniques could be applied, and the original recording returned as the result of the search: 'find me some news report describing Princess Diana's wedding'. Karen much enjoyed these collaborations, and particularly the excitement of being involved in building relatively large-scale systems that worked, and 'did something new'.

Becoming an honest woman

By the 1980s Karen was already an internationally respected figure, but had no permanent position in Cambridge University. In some ways this was no disadvantage to her research since she had few teaching and no administrative distractions. She also felt it would be slightly awkward to

[24] W. C. Mann and S. A. Thompson, 'Rhetorical Structure Theory: toward a functional theory of text organization', *Text*, 8:3 (1988), 243–81.

[25] Karen Spärck Jones, 'Automatic summarising: factors and directions', in I. Mani and M. T. Maybury (eds.), *Advances in Automatic Text Summarisation* (Cambridge, MA, 1999), pp. 1–12.

have a position in a department which was led by her husband. Nevertheless, her colleagues felt increasingly that her research reputation merited some kind of official recognition and so after existing for over twenty years—until the age of about 50—on a series of research grants, Karen was finally appointed to a permanent post in 1988 as Assistant Director of Research (a kind of junior lectureship) in the Computer Laboratory. Roger referred to this appointment as 'making an honest woman of her'.

In 1994 she was promoted to Reader in Computers and Information, and in the same year became President of the Association for Computational Linguistics (ACL). This is of course a great honour and is normally a position with more prestige than hard work associated with it. However, Karen took over at a time of great change for the Association, which had for many years in effect been run by its Secretary–Treasurer, Don Walker. Sadly, Don had died in November 1993 after a long battle with cancer, and the process of finding a successor took an unexpectedly long time, during which it transpired that much of the information needed to operate the ACL effectively had been carried only in Don's head. Karen characteristically took on a vigorous reorganisation of the Association, insisting that the finances were put on a proper footing and that well-defined procedures for things like conference organisation were established and followed.

In 1995 she was elected as a Fellow of the British Academy, and went on to serve as Vice-President from 2000 to 2002. She is pictured in the large portrait of nine female Vice-Presidents in the Mall Room at 10 Carlton House Terrace. In 1999 she was appointed to a personal chair (Professor of Computers and Information), and she officially retired in 2002, although this made no noticeable difference to her work.

In her long research career Karen produced over 200 publications, including nine books. She received many awards and honours: the Association for Computing Machinery (Special Interest Group in Information Retrieval) Salton Award (named after Gerard Salton, another early Information Retrieval pioneer, whom Karen had known well) in 1988; the American Society for Information Science and Technology Award of Merit in 2002; the Association for Computational Linguistics Lifetime Achievement Award in 2004; the joint Association for Computing Machinery and Association for the Advancement of Artificial Intelligence Allen Newell Award in 2007; the British Computer Society Lovelace Medal; the Athena Lecturer Award for Women in Computing (she received the news of both awards on the same day in February 2007, but did not live to collect either of them); and she was a Fellow of the Association for the Advancement of Artificial Intelligence (previously known as the American Association for Artificial Intelligence).

In 2005, her friends and colleagues organised a Festschrift volume and a day of lectures in Karen's honour which took place in Downing College, Cambridge:[26] many attendees, some themselves by now rather frail, travelled from the USA for the occasion.

Karen's last years were clouded by illness. She recovered from cancer in 2002, but ironically Roger, while nursing Karen through this episode, was himself diagnosed with cancer, from which he died in 2003. In her acceptance speech for the ACL Lifetime Achievement Award in 2004 she paid tribute to Roger 'not only because we worked and published together at particular times, but because I could always talk to him about my research and he always encouraged me'.[27] Karen's cancer returned in 2006 and she died at her home on 4 April 2007. However, she continued working right up to the end, producing a video version of her Allen Newell Award acceptance speech, and video versions of the lectures she was to have given as a Lovelace Medallist and Athena Lecturer. Her final computational linguistics paper appeared in the September 2007 issue of the journal *Computational Linguistics*:[28] the same issue carried her obituary by her former student and long term collaborator, John Tait. Her final information retrieval paper appeared a few months later:[29] in her introduction to the volume, the then chair of the ACM SIGIR, Liz Liddy, referred to Karen's ACM Athena Award video that had been shown in a special session at the most recent meeting in Amsterdam: 'As we viewed the video, we all were again reminded of both the creative insight and practicality that characterized Karen's outstanding contributions to our field' (p. 2). There are few people who have made contributions of this magnitude to two different (if related) fields.

A striking figure in every sense, Karen was very tall, thin, and upright, with a shock of grey (originally blonde) hair, and in hot weather usually sporting a dashing hat of some kind. Famously devoid of small talk, her opening conversational line was usually to do with one's work, or a book

[26] J. Tait (ed.), *Charting a New Course: Natural Language Processing and Information Retrieval: Essays in Honour of Karen Spärck Jones* (New York, 2005).

[27] Karen Spärck Jones, 'Some points in a time', *Computational Linguistics*, 31:1 (2004), 13.

[28] Karen Spärck Jones, 'Computational linguistics: what about the linguistics?', *Computational Linguistics*, 33:3 (2007), 437–41.

[29] Karen Spärck-Jones, Stephen E. Robertson and Mark Sanderson, 'Ambiguous requests: implications for retrieval tests, systems and theories', *ACM SIGIR Forum December 2007*, Volume 41, Issue 2 (2007), pp. 8–17 (ISSN: 0163-5840).

she had recently read. Initially, this could be off-putting, but her enthusiasm and complete absence of malice or 'side' would always draw her conversationalist in. Although speaking with crisp received pronunciation, Karen's conversation was sprinkled with Americanisms like 'buddy' and 'boondoggle': after a six-month stay in California in 1966, she and Roger had made many American friends, and maintained and grew their contacts and collaborations there.

Karen and Roger never owned a television (Karen also refused to have a computer at home until Roger's Microsoft appointment made this unavoidable: Karen kept quite a strict work–life separation), and during the time they lived in Brook Lane, Coton, just to the west of Cambridge (in a Scandinavian-style cedar shingle bungalow they had built together as Ph.D. students), her extensive library occupied an entire terraced house just across the road (originally bought for Karen's elderly mother until she became too infirm to live alone: Roger built an extension to their house for her). In 1997 they moved to Willingham: the increasing noise of the nearby M11, exacerbated by the death of what had been a sound-insulating, and much lamented, group of elm trees some years earlier, prompted this. Roger and Karen were keen sailors and had bought their first boat in 1961. They later bought and restored an 1872 vintage Itchen Ferry Cutter named 'Fanny of Cowes', which they kept on the Essex coast. They had no children.

There are relatively few women in computer science. Karen was used to being the only woman in meetings, particularly in Cambridge. There was a time when Roger Needham was able to say truthfully, if with heavy irony, that every woman who had ever applied for a job in the Computer Laboratory had got one. Karen worked hard to try to improve the position of women in computing and to attract more women to the discipline. She was a founder member of the 'women@cl' network based at the Computer Laboratory and was always unstinting with her time when women students and researchers asked her advice, as indeed Margaret Masterman had been, and she was a powerful role model for women on both sides of the Atlantic. Professor Kathleen McKeown, of Columbia University, an eminent researcher in computational linguistics, and another former President of the ACL, says 'I would count her as the most important role model I had in the field. … I was so impressed with her style of talking and her focus on work. … she became a great source of inspiration and a friend [who had] an impact on younger women world wide.'

Karen was invited to speak at the first Grace Hopper Conference in Washington in 1994—a celebration of women in computing, named after

Rear Admiral Grace Hopper, an early computer pioneer—and gave the Grace Hopper Lecture at the University of Pennsylvania in 2002. Karen accurately diagnosed many of the factors that make computing seem unattractive to girls—the 'geek' label easily acquired, and the impression given by much 'information and communication technology' teaching in schools that computing consists of little more than the ability to operate spreadsheets. Karen thought that this was a calamitous state of affairs: apart from the fact that such a bias means that many women forgo what could be a satisfying and profitable career in computing, she thought that computing offered a way of thinking about many scientific or societal problems, as well of course as human abilities like language and reasoning, that was valuable in itself. She also wanted to communicate the excitement of being able to create something entirely new, which was an opportunity that few disciplines offered—and anyway, as she put it, 'computing is too important to be left to men'.[30]

Karen's interests were wide. She was a knowledgeable amateur naturalist, would always read about the history and architecture of places she visited, was interested in textiles—she made many expertly constructed wall-hangings, she expertly reupholstered chairs—and she took great pleasure in sketching. Martin Kay recalls that 'She was always "sensibly" attired, but occupied much of her time in meetings very ably doodling sketches of women in very elegant gowns. My wife Iris was looking for a scarf of a very particular silk on one occasion. Karen was able to direct her to a basement shop in Soho where she found exactly what she was looking for. Clothing was one of many things for which she had a quite unexpected appreciation.'

Deeply interested in and appreciative of the art and artefacts produced by 'primitive' cultures, Karen was frequently to be seen in the Archaeology and Anthropology Museum in Cambridge (until the Computer Laboratory moved from its Corn Exchange Street site to the new Gates Building in West Cambridge this was of course very near to her office) or the Pitt Rivers in Oxford. For many years there hung in Roger's Computer Laboratory office a haunting face fashioned from wire that looked for all the world like an ethnic art work: Karen had made it from a coat-hanger. She wore Touareg jewellery, had a Berber rug in her office, and had accumulated a large collection of baskets from all over the world (now donated to the Archaeology and Anthropology museum): her criterion for

[30] Interview with Brian Runciman, British Computer Society, available at <http://www.bcs.org/server.php?show=ConWebDoc.10791>.

collection was simple—'they didn't have to do that', meaning that some piece of work had gone beyond the merely functional to something that had an aesthetic value.

S. G. PULMAN
Fellow of the Academy

Note. I am grateful to Ann Copestake, Martin and Iris Kay, Kathy McKeown, Bob Moore, Derek McAuley, Stephen Robertson, John Tait, and Yorick Wilks for their help in writing this memoir.