

Obituary

Karen Spärck Jones

John I. Tait*

University of Sunderland

Karen Spärck Jones died peacefully on 4 April 2007 after a number of months battling cancer. Karen was President of the Association for Computational Linguistics in 1994 during a difficult period, and saw through a transitional phase in which a solid basis for the long-term health of the association was established.

Karen was an exceptional individual who made very substantial contributions in two separate fields, computational linguistics (CL) and information retrieval (IR), as well as in computing and artificial intelligence more broadly. In IR, her greatest contribution was probably her invention (1972) of the concept of inverse document frequency (IDF), in which the importance of terms is weighted according to the proportion of documents in the corpus in which they occur; the intuition being that terms which occur in many documents are poor index terms. This is the partial basis of all weighting schemes adopted by widely used Internet search engines, and will undoubtedly be a lasting contribution to the field.

Her primary inspiration and interest was always language, especially language in practical use. Her long involvement with IR arose (as someone who subsisted for an inordinately long time on soft money) by the need to find a new line of research in the aftermath of the ALPAC Report and the subsequent difficulties in getting machine translation work funded. However, she was always well qualified to work in IR, a topic addressed in her very early publications (see, for example, Masterman, Needham, and Spärck Jones 1958), although in a rather different context to her later work.

Given this focus on language it is therefore perhaps odd that her main contributions to CL or natural language processing are harder to pin down. They include contributions to summarization, natural language interfaces to structured databases, dialogue, semantics, and more. In the long term perhaps her greatest contribution will come to be seen to be her Ph.D. thesis (1964), which was far ahead of its time. It brought together statistical or machine learning approaches with the use of an existing resource (*Roget's Thesaurus*—punched onto cards!) and still can be read with profit today (see Wilks and Tait 2005). It can be viewed as the forerunner of a whole range of more recent attempts to derive semantic information on an empirical basis.

It is perhaps as an empiricist that Karen will be best remembered. Her long associations with the US TREC initiatives and the more recent DUC summary evaluations are well known, as is her earlier association with DARPA TIDES, but her desire to work on well-founded experimental bases runs through from her earliest days, initially through involvement with Cyril Cleverdon and the Cranfield project, and then later in the work reported, for example, in Spärck Jones and van Rijsbergen (1976) and Spärck Jones and

* University of Sunderland, School of Computing Engineering and Technology, Tom Cowie Campus at St. Peter's, Sunderland, SR6 0DD, UK.

Galliers (1996). Perhaps this derived from her being a long associate of Wittgenstein's student Margaret Masterman (a connection Karen acknowledged in her acceptance speech for the ACL Lifetime Achievement Award in 2005) and consequently from the focus of Wittgenstein's philosophy on language use.

Karen was born in Huddersfield, Yorkshire, England in 1935 of English and Norwegian parents and was very much a child of Second World War Britain. She initially read History at Cambridge, but graduated in Philosophy (Moral Sciences). Following a brief spell as a schoolteacher she joined the Cambridge Language Research Unit in 1957 under Masterman and so began a very lengthy period on soft money. The University of Cambridge finally recognized her standing by awarding her a Readership in 1994 and a personal chair in 1999.

Karen received very many awards during her career. These include Fellowships of the American and European Artificial Intelligence societies, Fellowship of the British Academy, the ACL Lifetime Achievement Award, the Lovelace Medal of the British Computer Society, the ACM SIGIR Salton Award, the American Society for Information Science and Technology's Award of Merit, and the ACM-AAAI Allen Newell Award. Finally she was awarded the ACM Women's Group Athena Award, which sadly she did not live to receive formally, although she bravely recorded an acceptance lecture a few weeks prior to her death.

Although Karen felt she was never overtly discriminated against as a woman, it must have taken considerable courage and tenacity to develop a leading academic career at a time when society's expectations of women went little beyond their roles as wives and mothers. She acknowledged the support of her parents and then her late husband (and long-time academic collaborator) Roger Needham in her obtaining a good education and developing a career. She noted that in professional circles she was very frequently the only woman, and made very active efforts, especially in recent years, to interest young women and girls in careers in computing, believing that "computing was too important to be left to men." Karen remained as active as ever after her formal and undesired retirement (see *The Guardian's* Letters Page, 5 October 2002¹).

These academic achievements should not be allowed to eclipse Karen's well-roundedness as a person. She and her late husband Roger were keen sailors, owning, and maintaining themselves for many years, an 1872 Itchen Ferry Cutter. They also built their first house with their own hands. Karen could sketch well, made works of art from everyday objects, and had an enormous knowledge of natural history and church architecture. She was candid and passionate and could be enormously generous in her support. She will be greatly missed.

References

- Masterman, Margaret, Roger M. Needham, and Karen Spärck Jones. 1958. The analogy between mechanical translation and library retrieval. In *Proceedings of the International Conference on Scientific Information (1958)*, volume 2, pages 917–935, Washington, D.C.
- Spärck Jones, Karen. 1964. *Synonymy and Semantic Classification*. Ph.D. thesis, University of Cambridge.
- Spärck Jones, Karen. 1972. A statistical interpretation of term specificity and its application in retrieval. *Journal of Documentation*, 28:11–21. Reprinted in *Journal of Documentation*, 60:493–502, 2004.
- Spärck Jones, Karen. 2005. "Some points in a time." Acceptance speech for ACL Lifetime Achievement Award. *Computational Linguistics*, 31(1):1–14.

1 Available at http://www.guardian.co.uk/letters/story/0,,805186,00.html#article_continue

- Spärck Jones, Karen and Julia R. Galliers.
1996. *Evaluating Natural Language Processing Systems*. Number 1083 in *Lecture Notes in Artificial Intelligence*. Springer-Verlag, Berlin.
- Spärck Jones, Karen and Keith van Rijsbergen.
1976. Information retrieval test collections. *Journal of Documentation*, 32:59–75.
- Wilks, Yorick and John Tait. 2005. A retrospective view of synonymy and semantic classification. In John Tait, editor, *Charting a New Course: Natural Language Processing and Information Retrieval, Essays in Honour of Karen Spärck Jones*, pages 1–11. Springer, Berlin.

