

Karen Spärck Jones (1935–2007)

Yorick Wilks, *University of Sheffield*

Karen Spärck Jones contributed significantly to the information retrieval and natural language processing fields and, in her later years, was concerned with their relationship within general schemes of representation in

AI. Although she died on 4 April 2007, she worked up until a week before her death, and her major and most lasting contributions will almost certainly be her original PhD thesis and the inverse document frequency (idf) measure of the relevance of terms.¹ The latter is the notion that a document is relevant not only because key terms are frequent in it but because those terms are infrequent in other, nonrelevant, documents. This idea is now a basic part of information retrieval.

Born in 1935 in Huddersfield, Yorkshire, of English and Norwegian parents, Spärck Jones studied history at Cambridge but moved to philosophy (then called “moral sciences”) in her last year. Her first published conference paper was “The Analogy between Mechanical Translation and Library Retrieval,”² a title of great prescience in her career. At that time, it referred to using thesauri to resolve meaning problems in the two technologies, but the link preoccupied her all her life. She returned to this topic in “Information Retrieval and Artificial Intelligence,”³ arguing that AI in general, and natural language processing in particular, should make more use of information retrieval’s statistical methodology.

In 1962, after a brief spell of teaching, Spärck Jones accepted Margaret Masterman’s invitation to join the

Cambridge Language Research Unit (CLRU) and started working toward her doctorate. Under the supervision of Masterman’s husband, philosopher Richard Braithwaite, she wrote her thesis, “Synonymy and Semantic Classification.”⁴ It was the first application of statistical clustering methods to lexical data—in her case, the whole of Roget’s Thesaurus on punched cards—and was an ambitious attempt to create some notion of primitive concepts for machine translation on an empirical basis. Far ahead of its time,⁵ the work was not published until 20 years later in the Edinburgh University Information Technology series⁴—at which time Spärck Jones had to be persuaded it was still relevant.

This work is the ancestor of a range of empirical semantics research, from the semi-synonymous rows of terms (synsets) in WordNet to much later work on statistical clustering to determine semantic relationships. The historian in Spärck Jones added an extraordinary thesis appendix on artificial languages for coding meaning. She used the Theory of Clumps algorithms, which her husband Roger Needham developed and

used in his own thesis work on automatic classification. In 1968, the need for more serious computer facilities took Spärck Jones from the CLRU to the University Computer Laboratory. The director at the time would not allow work explicitly on AI or natural language processing, though he deemed information retrieval to be respectable and scientific. Spärck Jones, who had completed three years as a Research Fellow of Newnham College, then became a Royal Society Fellow. She again used the Theory of Clumps algorithms in her new career in information retrieval, a subject on which she became a world authority. Eventually, Need-



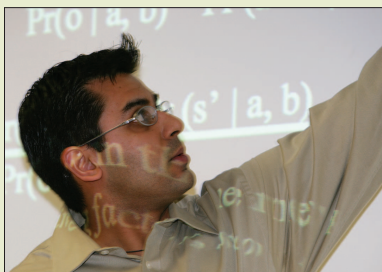
ham became director of the laboratory, and Spärck Jones was able to revisit her early interest in natural language processing. She took on students and produced major work in language front ends to databases, automatic summarization, content retrieval from video, evaluation methods, and belief revision.

Spärck Jones's academic promotion was slow in coming: most of her career was as an assistant director of research on grant money, and it was only in 1999 that she was awarded a personal professorship. Meanwhile, she had taken on a wider role. In 1985, she began managing the UK's Alvey Research Program. In 1992, she took leading roles in the US DARPA/National Institute of Standards and Technology evaluation projects and later was on the Advisory Committee for the DARPA TIDES program in language processing. In 1994, she was president of the Association for Computational Linguistics (ACL). Masterman remained an inspiration to her,⁶ and she thanked her, along with Needham, at the end of her acceptance speech for the ACL Lifetime Achievement Award. This speech⁷ remains the best overview of the many interleaved themes in Spärck Jones' work.

She gained many other honors, some of which she did not live to receive (although she recorded acceptance speeches): Fellowships of the American and European AI societies, the Fellowship of the British Academy, the Lovelace Medal of the British Computer Society, the SIGIR Salton Award, the American Society for Information Science and Technology's Award of Merit, and the ACM-AAAI Allen Newell Award.

Spärck Jones was as active as ever in retirement, returning to issues of representation and to her early interest in semantic primitives.⁸ However, she was always mindful of her powerful slogan "Words stand only for themselves." She remained finely balanced on the issue of whether natural language processing can help information retrieval, conscious that most claimed nonstatistical advantage can be reproduced statistically. And yet, she wanted NLP to matter: Although she had attributed statistical influence on AI to information retrieval, she knew well that it was above all Jelinek's machine translation research at IBM that had driven natural language processing to take up statistical methods. But she remained skeptical that the tasks of machine understanding could all been seen as

Jay Modi (1975–2007)



Intelligent Systems Advisory Board and recently received the prestigious National Science Foundation's Career award //Correct?//. He was a valuable member of the Computer Science department, a great friend, and a wonderful human being. He will be dearly missed.

It is with profound sadness that I inform you that Jay Modi passed away on 9 April 2007. Modi was a professor in the Department of Computer Science at Drexel University. He received his PhD from the University of Southern California and held a postdoctoral research position at Carnegie Mellon University. He was an outstanding teacher and a rising star in the AI community. Last year, he was selected as one of the top 10 young AI researchers by the IEEE

— Jeremy Johnson, Drexel University

"recovery processes," in the way the answer recovers the question, the document recovers the original query, and the transcription recovers the speech signal. In 2005, she asked, can we *really* see machine translation of Shakespeare into Spanish as recovering his hidden Spanish within the English? Furthermore, she produced a stimulating late paper on how the Semantic Web movement addresses these questions. She never forgot that Masterman had been a student of Ludwig Wittgenstein, so she was therefore only one step away from him—evidenced by how close her own slogan was to his demand to look not for the meaning but the use.

Spärck Jones also campaigned hard for more women to enter computing and was conscious that she, like Masterman before her, had a husband with a more powerful formal role. We can now examine, in both, on which side the more creative achievements lay. She was, with Needham, an accomplished sailor, and they built their house themselves; she made wonderful things from objets trouvés. ■

References

1. K. Spärck Jones, "A Statistical Interpretation of Term Specificity and its Application in

Retrieval," *J. Documentation*, vol. 28, 1972, pp. 11–21.

2. M. Masterman, R.M. Needham, and K. Spärck Jones, "The Analogy Between Mechanical Translation and Library Retrieval," *Proc. Int'l Conf. Scientific Information*, Nat'l Academy of Sciences—Nat'l Research Council, 1959, pp. 121–140.
3. K. Spärck Jones, "Information Retrieval and Artificial Intelligence," *Artificial Intelligence*, vol. 114, 1999, pp. 257–281.
4. K. Spärck Jones, *Synonymy and Semantic Classification*, doctoral dissertation, Univ. of Cambridge, 1968; reprinted, Edinburgh Univ. Press, 1984.
5. Y. Wilks and J. Tait, "A Retrospective View of Synonymy and Semantic Classification," *Charting a New Course: Natural Language Processing and Information Retrieval, Essays in Honour of Karen Spärck Jones*, Springer, 2005, pp. 1–13.
6. M. Masterman, *Language, Cohesion and Form*, Y. Wilks, ed., Cambridge University Press, 2005.
7. K. Spärck Jones, "Some Points in a Time," *Computational Linguistics*, vol. 31, no. 1, 2005, pp. 1–14; www.cl.cam.ac.uk/~ksj21/ksjdigipapers/compling05.pdf.
8. K. Spärck Jones, "Semantic Primitives: The Tip of the Iceberg," to be published in *Words and Intelligence: Part II: Essays in Honour of Yorick Wilks*, K. Ahmad, C. Brewster, and M. Stevenson, eds., Springer, 2007; www.cl.cam.ac.uk/~ksj21/ksjdigipapers/wilksbook07.pdf.