

**Karen Sparck Jones**

**References for Grace Hopper Lecture, University of Pennsylvania, April 2002**

(work specifically mentioned, and some generally-pertinent items)

M. Banko, V. Mittal, and M. Witbrock, 'Headline generation based on statistical translation', *ACL 2000: Proceedings of the 38th Annual Meeting of the Association for Computational Linguistics*, 2000, 318-325.

A. Berger and J. Lafferty, 'Information retrieval as statistical translation', SIGIR 99, Proceedings of the 23rd International ACM-SIGIR Conference on Research and Development in Information Retrieval, 1999, 222-229.

B. Boguraev and C. Kennedy, 'Salience-based content characterisation of text documents', in Mani and Maybury, 1999.

H. Borko, 'Research on computer-based classification systems' in *Classification research*, (Ed. P. Atherton), Copenhagen: Munksgaard, 1965.

P.F. Brown et al. 'Class-based n-gram models of natural language', *Computational Linguistics*, 18(4), 1992, 467-680.

K.W. Church and P. Hanks, 'Word association norms, mutual information, and lexicography', *Computational Linguistics*, 16(1), 1990, 22-30.

W.B. Croft (ed.), *Advances in information retrieval*, Dordrecht: Kluwer.

S.F. Dennis, 'The design and testing of a fully-automatic indexing- searching system for documents consisting of expository text', in *Information retrieval - a critical view*, (Ed. G. Schechter), Washington, DC: Schechter, 1967.

L.B. Doyle, 'Semantic road maps for literature searchers', *Journal of the ACM*, 8(3), 1961, 404-417.

S.T. Dumais, 'Latent semantic indexing (LSI): TREC-3 report', TREC-3, Proceedings of the Third Text REtrieval Conference, (Ed. D.K. Harman), Special Publication 500-225, Gaithersburg, MD: National Institute of Standards and Technology, 1995, 219-230.

U. Germann et al. 'Fast decoding and optimal decoding for machine translation", *ACL 2001, Proceedings of the 39th Annual Conference of the Association for Computational Linguistics*, 2001, 228-235.

K.E. Harper, 'Measurement of similarity between nouns', RM-4532-PR, Rand Corporation, Santa Monica, 1965.

K. Knight and D. Marcu, 'Statistics-based summarisation - Step 1: sentence compression', *AAAI-2000, Proceedings of the 17th National Conference on Artificial Intelligence*, 2000, 703-710.

Knight - see Germann et al, 2001.

Lee, L. and F. Pereira, 'Distributional similarity models: clustering versus nearest neighbour', *ACL 99, Proceedings of the 37th Annual Conference of the Association for Computational Linguistics*, 1999, 33-40.

H.P. Luhn, 1957 quotation, p 96 in Schultz, 1968; papers, see Schultz.

I. Mani and M.T. Maybury (eds), *Advances in automatic text summarisation*, Cambridge, MA: MIT Press, 1999.

Marcu - see Knight and Marcu, 2000.

M. Masterman, R.M. Needham and K. Sparck Jones, 'The analogy between mechanical translation and information retrieval', *Proceedings of the International Conference on Scientific information*, (1958), Washington, DC: National Academy of Sciences - National Research Council, 1959, 917-935.

R. Mihalcea and D. Moldovan, 'A method of word sense disambiguation of unrestricted text', *ACL 99, Proceedings of the 37th Annual Conference of the Association for Computational Linguistics*, 1999, 152-158.

Pereira - see Lee and Pereira, 1999.

D. Prescher, S. Riezler and M. Rooth, 'Using a probabilistic class-based lexicon for lexical ambiguity resolution', *COLING 2000, Proceedings of the 18th International Conference on Computational Linguistics*, 2000, 649-655.

Robertson - see Sparck Jones et al, 2000.

- Rooth - see Prescher et al, 2000.
- N. Sager, 'Natural language information formatting: the automatic conversion of texts to a structured database', *Advances in Computers* 17, (Ed. M. Yovits), New York: Academic Press, 1978, 89-162.
- G. Salton (ed), *The SMART retrieval system*, Englewood Cliffs, NJ: Prentice-Hall.
- H. Schuetze, 'Automatic word sense discrimination', *Computational Linguistics*, 24(1), 1998, 97-123.
- C.K. Schultz (ed), *H.P. Luhn : Pioneer of information science*, New York: Spartan, 1968.
- K. Sparck Jones, *Synonymy and semantic classification*, (1964), Edinburgh: Edinburgh University Press, 1986.
- K. Sparck Jones, 'Notes and references on early classification work', *ACM SIGIR Forum*, 25(1), 1991, 10-17.
- K. Sparck Jones and P. Willett (eds), *Readings in information retrieval*, San Francisco: Morgan Kaufmann, 1997.
- K. Sparck Jones, 'Automatic summarising: factors and directions', in Mani and Maybury, 1999.
- K. Sparck Jones, G. Gazdar and R. Needham (eds), 'Computers, language and speech: formal theories and statistical data', *Philosophical Transactions of the Royal Society of London, Series, A*, Vol. 358, Number 1769, 2000, 1225-1431.
- K. Sparck Jones, S. Walker and S.E. Robertson, 'A probabilistic model of information retrieval: development and comparative experiments. Parts 1 and 2', *Information Processing and Management* 36 (6), 2000, 779-840.
- M.E. Stevens, *Automatic indexing: a state-of-the-art report*, Monograph 91, Washington, DC: National Bureau of Standards, 1965.
- H.E. Stiles, 'The association factor in information retrieval', *Journal of the ACM*, 8, 1961, 271-279.
- Witbrock - see Banko et al, 2000.