A Retrospective view of Synonymy and Semantic Classification
Yorick Wilks and John Tait

1.   Introduction

Karen Sparck Jones' Cambridge PhD thesis of 1964 has had an interesting and unusual history. Entitled Synonymy and Semantic Classification (henceforth SSC) it was reproduced only in the simple mimeo book form then used by the Cambridge Language Research Unit where she worked. It was finally published in 1986, in an Edinburgh University Press series. Even that late publication managed to be ahead of a great deal of later work that recapitulates aspects of it, usually from ignorance of its existence. There is no doubt that SSC was developing statistical and symbolic techniques for the use of what we now call language resources so far ahead of other work that it was almost impossible for contemporary researchers to understand the book or to relate it to their own activity. At the time SSC was being written, Olney and Revard (1968) were exploring the content of Webster's Third Dictionary quantitatively on punched cards at Systems Development Corporation (where Sparck Jones also was by chance at the time and joined in their work with her 1967) and their work met a similar lack of reception, it, too, being twenty to thirty years ahead of its time.

2.   A brief overview of Synonymy and Semantic Classification

SSC begins with a review of the implications of the use of the computer as a tool to study natural language text. It discusses the need for precision of representation (in dictionaries, grammars and thesauri) for automatic processing, but rapidly moves to a deeper discussion of meaning, focussing in particular on the claim that in the context of a coherent text (fragment) different words will be used in senses with related meanings. There is a developed example in which canal and road are cited as means of communication. Roget's thesaurus is then put forward as a means of operationalising this intuition. The approach adopted finds strong echoes in much later work by Morris and Hirst (1991) and Ellman and Tait (2000), which do not really share the same intellectual heritage.

Sparck Jones then moves on to consider the notion of semantic relations between words: her focus is on synonymy and on Lyons (1961), although a range of other relations (including antonymy, hyponymy, logical implication, Lyon's incompatibility and so on) and a range of other authors are mentioned. Chapter two ends with a proposal to test a notion of synonymy based on substitution, reducing a subtle and complex notion to an empirically testable notion, without losing sight of the limitations of the test.

Chapter Three and Four develop this notion of synonymy based on substitution by using the notion of a row or set of close synonyms (cf. Wordnet synsets). The discussion is sophisticated in many ways, but suffers from the use of an obscure notion of a ploy (a kind of semantic interpretation), from considering the context of the use of a word only in terms of the sentence (and not more broadly), and from the attempt to move between a specific word-use and a word-sign (string of characters) without any intermediate notion of morphology (strictly graphology) or intermediate word senses. This is not to say that taking on these notions would necessarily simplify the discussion: but they sometimes make the discussion hard to follow to at least one of the current authors' eyes. The chapter moves on to a fascinating discussion of a

notion of semantic distance (likeness) between words (and then between phrases) based on similarity of their occurrence patterns in thesaurus rows.

Chapter Six describes a series of Practical Experiments, using an analysis of Richards' book "English through Pictures" which reports some success in building a simple prototype system of the kind described in the previous chapters.

The thesis concludes with some manual experiments concerning the feasibility of discovering the semantic relationships between words in coherent text and then argues that, taken together, these experiments support the notion that there is conceptual repetition in discourse (p 200).

A brief summary like this is inevitably unfair to the original. Some passages, even now, reveal a deep understanding of aspects of language which we have yet to fully face up to in Computational Linguistics. For example, some of the discussion of metaphor in Chapter Two and Three, and the kind of conflict between specific use and overtones of a word derived from its whole range of uses (Sparck Jones, 1986, p86).

We now return to the strengths and weaknesses of SSC which we outlined at the beginning of this section. We will then pass on to highlight some aspects of this work which resonate with more recent developments in Computation Linguistics and Information Retrieval, despite the fact that it is now over forty years old.

3.   Strengths and Weaknesses of SSC

SSC has three great strengths. First, SSC brought together IR methods with linguistic semantics and CL for the first time, a link that is now accepted and productive (as well as the subject of her 1999 AIJ article (Sparck Jones 1999), and thus an interest spanning her career). In saying that, we do not imply SSC is about IR, but that the underlying clustering algorithm she applied in it to thesaurus rows was the so called Theory of Clumps (1962) of Needham and Parker-Rhodes, a development in automated classification of Tanimoto's (1958) original idea for derived clusters as a basis of IR. The principal originality of SSC was to take an IR clustering algorithm and apply it to features that described not physical objects or documents but other words or features at the same level as the classifiers themselves, and to which they were bound by a defined relationship of semi-synonymy. The kinds of associative nets/clumps she derived have been rediscovered many times since by others, probably in part because her thesis was not published e.g. Schvaneveldt's Pathfinder networks (1990) which were patented for IR.

Secondly, SSC's use of Roget's Thesaurus is possibly the first use of an established machine-readable linguistic/lexical resource in CL, apart perhaps from the roughly contemporary quantitative computations with Websters' Third Dictionary by Olney and Revard at SDC mentioned above. The widespread use of linguistic resources, such as machine-readable dictionaries, as a basis for NLP did not become commonplace until the late Eighties, when among the earliest contributors were members of her own laboratory, such as Bran Boguraev (Boguraev and Briscoe, 1989).

Thirdly, SSC shows an appreciation of the need to evaluate ideas about language processing by experiments on realistic samples of language using well-defined tasks, a matter we now take for granted but, when SSC was written, AI was still in the heyday of its toy systems with tiny sets of examples.

However, the principles underlying SSC as well as its implementation and evaluation, unfair as it perhaps is to raise these modern notions for work done 40 years ago, still

do give rise to real problems and we set out some problems with SSC that were always evident and have not changed with time (as its virtues have in the list above).

There are serious short comings in the discussion of the experiments which are very hard to interpret: there is a lack of detail (for example algorithmic descriptions) allowing judgements to be made about the scalability of the algorithms. There is also a lack of clarity about the experimental set ups: some are clearly manual, some apparently automatic, one probably semi-automatic. The general notion underlying the experiments is very clear: namely, applying the theory of clumps to features, where in SSC the features were words whose features were being-corow-members, which should have resulted in clumps of words associated by the clump algorithm. But the matrix inversions required for that computation were very large and almost certainly not tractable over a data base the size of Roget. The whole of Roget's thesaurus was put onto punch cards by Betty May, but only a sample can have been used in the experiments described in SSC. Chapter 6, note 14 clearly implies the adaption of the ideas to the practicalities of computing with then available machines.

One of the problems in interpreting SSC today is confusion between what was achieved with the then available computing engines and knowledge of software engineering, and what could have been achieved if Sparck Jones had had today's computers and software engineering and was doing this work with the benefit of the insights on language we have gained over the intervening 40+ years.

There is also a failure to grasp the problems posed by basing the synonymy analysis on the use of words in context, presumably in the case of the meaning of this word in this sentence (for simplicity let us confine ourselves to writing) in this text at this time to this reader, as opposed to a word sense in a dictionary. Indeed, there is a slide between the two, with the definition of word-use on page 79 being concerned with "ployed" sentences versus page 122 where word-use is defined by existence in a row. The discussion in Chapter 4 seems to show Sparck Jones is aware of the problem, but she shies away from the introduction of a intermediate layer between word-use and word-sign, one that corresponds to what we all now refer to, without much sign of scepticism, as a word-sense. One might (then or even now) put forward the objection that this is introducing an artificial abstract notion into a system which is otherwise entirely dependent on directly observable phenomena in language. But might not the avoidance of this (conventional) abstraction be the reason the system has the problems it undoubtedly has, dealing with the more complex relations like hyponymy or antonymy?

These last arise from a definition of likeness (SSC page 102) seems to gloss over the previous distinction between word-use and signs. Sparck Jones is clearly aware that there are complex relations between notions of substitution, hyponymy, homonymy, and synonymy, but only the first is given an operational definition with any plausibility.

In modern terms, there may be a parameter of the machine learning algorithm in which every occurrence of every sign (to use SSC's terminology) has its own row (cf. SSC p90) and a much smaller collection of rows emerge, but there is a danger of hyponyms and synonyms occurring in the same row. Antonymy, too, must be part of word meaning but the structure seems unable to take account of this.

The output from unsupervised methods is notoriously hard to interpret: given clumps of row-associated words from the program, why would they be better clumps than those provided by the Thesaurus heads themselves? No answer to this could have been

expected at the time, and is barely available now: there is awareness in Chapter 7 of the need to measure this output against some operational task, such as machine translation, though that was of course beyond the scope of SSC.

The basis of the property of co-row-ness (for words) is that of substitution-preserving-some-property: Sparck Jones discusses this notion and its evident circularity yet she goes on to adopt it and then identify that with Thesaurus rows. She refers to, and is clearly aware of, Quine's critique of any such notion as circular (1953). There is a double sleight of hand here: even if substitution does provide a test of rowness, why should we accept Roget's rows as passing it, as she clearly does in order to get a data set? One could say that SSC's rows are ambiguous between an emergent property of language use (corresponding to unsupervised learning in more modern usage, and Parker-Rhodes & Needham's clumping) and artefacts extracted from a human constructed resource: such as Roget's rows, and, later, Miller's (2000) synsets (corresponding to supervised learning perhaps).This is perhaps best illustrated by considering the question: if the practical experiments of Chapter Six had produced row systems quite unlike Roget, what would this have meant for the (implicit) hypothesis of the whole thesis ?

SSC is presented explicitly as a search for emergent semantic primitives; but how do (or could) these emerge from these computations? Yet, by using Roget she already assumes such a set (the 1000 heads of Roget): so why is that set worse (or better) than any she derives, or might derive with further computation? Perhaps what is lacking, in modern terms, is an understanding of the need for an objective function, allowing us to distinguish more and less optimal solutions the need for which is now so well understood in unsupervised machine learning.

This might seem a long list of problems. However in view of the groundbreaking nature of the work, the intellectual tradition from which it sprang, the extraordinarily limited computational environment in which it was undertaken, they are comparatively minor, and in no way detract from the major strengths.

4.   A further Twenty Years Later: Sparck Jones's view of SSC after 20 years.

Sparck Jones wrote a new introduction to SSC when it was finally published, 20 years late, in the Edinburgh University IT series, run by one of the present authors. Perhaps the most striking feature of her retrospective, as compared to the original SSC, is the emphasis on semantic primitives and the explicit opening claim that "The thesis proposes a characterisation of, and a basis for deriving, semantic primitives, i.e. the general concepts under which natural language words and messages are categorized (p.1)"

This view of SSC is not one that a reader of the original thesis would necessarily come to from its text, although it makes perfect sense if we take semantic primitives to mean the topic markers that are the 1000 or so Roget heads, such as 324 SOFTNESS. However, and as noted in the previous section there are some problems with reconciling this notion of predefined primitives and truly emergent ones. In her retrospective discussion Sparck Jones widens comparisons at this point, describing such primitives as domain dependent (e.g. SHIP-as-a-type) by contrast with more general notions of semantic primitive in the work of Katz (1972), Wilks (1975) and Schank (1975), and which was criticised by Lewis (1970), Pulman (1983) and others.

These latter primitives (usually equivalent to notions such as such as human, physical object, movement etc.) she takes as being general rather than domain dependent,

which suggests the two types could all be fitted together in some very semantic hierarchy with physical object near the very top and types of ship at the bottom; and this is something like what one gets in Wordnet and indeed in the hierarchy Roget himself offers at the start of his Thesaurus. That Sparck Jones sees these two types of primitive as closely related, as is shown by the original appendix to SSC on Thesauri and Synonym dictionaries, a historical excursus that covers both types of primitive and remains for some the best thing in the book.

In the structures associated with the LDOCE dictionary (Procter, 1978) and both types are given as quite separate hierarchies (of semantic and domain terms) and dictionary entries are decorated with both as features independently. Again, in much recent work on word-sense disambiguation (e.g. Yarowsky (2000), Wilks and Stevenson (1997)) both types of hierarchy have been used as separate information sources, combined ultimately by the algorithm, but where it can be seen that one type tends to disambiguate verbs and the other nouns. None of these considerations are definitive as to whether there are two levels or types of primitives or not, or whether the difference is merely one of degree and domain. Sparck Jones certainly distinguishes two roles for primitives, as do many authors, namely being definitional of sense (as in a dictionary) and being selective for particular senses (as in a disambiguation program) but that distinction has no implications for the one- or two levels of primitive issue.

5.   The view of semantics embodied in SSC

As noted above, SSC was perhaps the first attempt to capture computationally the elusive notion of linguistic relations or fields, one well established in the descriptive literature (Lyons, 1961) but with no formal or computational basis up to that point. It is notion close to some Continental notions of text structure and meaning, ones that have received wide popular discussion, and in which the meaning of any symbol depends, by a relation of contrast, on its relation to other symbols, rather than to objects in the world, as in the basic, rather simple minded, version of Anglo-Saxon philosophy.

Of course linguistic or semantic fields are a subtle and complex subject. A later review (Lyons, 1977, Chapter 8) points out some commonalities, but also contradictions and contrasts between different field theories. In the most accessible form of the theory, there is postulated some sort of meaning surface lying between the lexemes of a language and the world of language use. Particular lexemes are then related to areas of this meaning surface. Most field theorists are concerned with changes in the meaning of language over time, and this creates an odd contrast with SSC, which, like almost all computational work which followed it, takes a rather static or at least snap shot, view of language. As we have noted, all field theorists share a focus on lexical semantics, in terms of the relations between words and other words or the whole vocabulary, which is presumably what made the approach attractive to Sparck Jones, but they also share a difficulty in formalising the notion of field in a consistent and useful way.

Much of the discussion of SSC shies away from putting forward anything which cannot be directly observed in text. In the end SSC resorts to concepts as additional, artificial, constructs lying outside observable language. One might say the work is caught between Skinner and Saussure, having on the one side the poverty of sticking to the merely observable and on the other problem of subjecting the abstract to empirical verification. In a later overview of her work in IR (Sparck Jones, 2003) she refers to a

simple principle underlying everything she does as ´ « taking words as they stand» a position already present in SSC, before Sparck Jones began her distinguished career in IR, namely a reluctance to decorate words with logical, primitive semantic, or other linguistic codings (as opposed to relations). This was something shared, in an interdisciplinary way, with linguistic field theorists and their Continental counterparts. Against this, it could be argued that, by accepting, as she did, the overarching a priori architecture of Roget, all derived from a single mind by intuition, Sparck Jones was accepting a great deal of decoration beyond the words themselves. Conversely, it can be argued, changing sides as it were, that nothing violates that principle in using a thesaurus or a dictionary because the decorations are only more words, as are the Thesaurus heads of Roget themselves.

6.   Other resonances between SSC and more modern work in MT, NLP and IR.

The discussion of the likeness between words and phrases in Chapter Three of SSC, referred to in section 2 above as a form of semantic distance, finds many echoes in later query expansion techniques, like pseudo-relevance feedback or local feedback (Xu and Croft, 1996). These techniques presume that terms which co-occur in documents with query terms are semantically related to query term uses. They rely on the implicit existence of an empirically derived thesaurus, or clump dictionary, on which similarity calculations of the sort described in SSC can be computed.

The introductory material to Chapter Five contains a couple of oddities which hide really quite deep insights.

First, there is discussion of the very large number of rows in which a word might be placed reflecting the very fined grained distinctions of sense which might be required for high-quality machine translation. However, oddly, there is no discussion of how one might link these to another natural language. Was Sparck Jones perhaps thinking that some form of parallel corpora would solve this problem given the automatic procedure? Or was the problems posed by the need to link the source and target languages simply missed?

Secondly, initially she proposes to distinguish every sentence position of every use of a word, but this abandoned on grounds of efficiency. However retaining this position would imply the learning not only of a synonym dictionary but also of a corresponding grammar in some sense. Further it might imply a finite model of language (in the absence of a generative component). It is hard to believe these restrictions were an oversight in view of the sophistication of the discussion elsewhere. Sparck Jones and her collaborators clearly understood such a process might imply learning or deriving a grammar stored in the thesaurus (Masterman, Needham and Sparck Jones 1958), but perhaps not its implications for the underlying model of language.

7.   What was the SSC computation/algorithm?

It is clear that Sparck Jones in SSC made use of the Theory of Clumps, an unsupervised classification algorithm, deriving ultimately from Tanimoto and refined by Roger Needham (her husband) and Frederick Parker-Rhodes at CLRU. The Theory of Clumps (from now on TC) which she found ultimately unsatisfactory for her purposes (see the quotation above) was an algorithm that took a set of objects x classified by a set of features y and produced clumps or sets drawn from x which expressed natural subsets of x in terms of the assigned features. An aspect of TC

which Sparck Jones liked and drew attention to (as did Roger Needham) was that it had a feature close of Wittgenstein's notion of family resemblances namely that subsets, so found, did not need to share any common feature at all and hence this notion was not at all part of the old Necessary and Sufficient Conditions tradition for being a thing of a certain sort.

Roger Needham's thesis was classic application of TC, outside IR that is, and he took a set of Greek pots classified by a range of features (colour, handles, decorative figures etc.) and produced plausible sets of pots based on the core notion of TC that things should be seen as alike if they tended to have the same features, or had separately the features that other things had as common features etc. It was thus an associative rather than definitional model of similarity and would have fallen under Firth's phrase about words and "knowing them by the company they keep" Things in the same clump would tend to keep the same company in terms of features.

Sparck Jones's application of TC was thus more original than taking objects and features as quite different sorts of thing: she realised that both could be words and that words as features could be used to classify words as objects Thus her classification relationship was that of appearing in the same row in Roget's thesaurus. Elsewhere in this paper we discuss the implications of that assumption of classification as a form of synonymy but here we simply note that, in TC terms, co-row words were features of any given member word, where the co-row members were derived from the OED by seeking in entries for semi-synonyms and testing their substitutability (intuitively) within the example sentences given in the dictionary.

Given this assumption, TC could proceed, which meant first a matrix of features against objects was constructed, notionally at least, and here, since the matrix is symmetrical (both sides being in principle the whole vocabulary of Roget words) we can imagine a matrix with something of the order of 50K rows and columns. At this point forms of the TC algorithms come into play, of which the most basic is a measure of how close any two rows (derived as above) are. Sparck Jones adopts a rough and ready measure of the number of common words in two rows divided by the total number of distinct words in both i.e. their intersection divided by their union.

The main TC algorithm then runs and produces tentative clumps of objects based on the object-feature associations established in this way. These clumps should, being empirically based on associations in a corpus (the OED) yield better groups than Roget heads (considered as groupings of semi-synonyms). On p.183 she writes of assuming that we now have a better thesaurus than Roget's, but one of the same kind, and one that might be tested against Roget in simple MT experiments.

This shows clearly that the clump output from SSC was of the same type as Roget heads themselves and, at one point, she discusses a possible recursive procedure for organizing the clumps produced by the program into a flat hierarchy more like Roget itself.

The account above must be treated with caution because of the different way experiments were handled and described then and now, and in large measure because, as we noted earlier, the 50K square matrix could not be constructed with the computers then available, nor were there as many techniques then for representing large very sparse matrices in alternative, more compressed, forms. Hence whatever experiments she did were necessarily on very small samples and she shows a sample of 500 rows and describes an experiment based on 180 rows (p.170), the maximum number her program could handle. An output clump is given as a set of rows deemed

sufficiently close (p.172) and there is an extensive discussion (pp.176-181) of how this should be evaluated (by comparing it with Roget or by doing say machine translation and getting a better result than with Roget) which shows the attention to detail in and general importance of evaluation procedures which is most striking for its time.

8. The Practical Experiments and their relation with modern experimental work in IR and MT/NLP

We now take for granted the need to verify hypotheses in language processing by conducting large scale experiments. TREC and DUC (elsewhere in this volume), SENSEVAL (Edmonds and Kilgarriff, 2002), and DARPA's MT competitions, are all modern examples of this approach. They rely on enormous volumes of data, and careful standardisation of tasks to produce their results. They provide an opportunity to compare systems, theories and approaches which was denied earlier language workers. They also could not be carried out without the computer's ability to process large volumes of language data in a verifiable and repeatable manner.

Despite the shortcomings in the experimental work and its description noted above, there is no doubt that SSC represents an early land mark on the journey that led to these modern destinations. The scale of the experiments is puny by today's standards. 533 rows were extracted from English through Pictures, and experiments were conducted on 500 rows extracted from the Oxford English Dictionary (compared to 28000 bigram collocations in the 1.2 Million documents processed for Sunderland's TREC 2002 experiments (Stokoe, Oakes and Tait, 2003)). However, given the puny computational resources available at the time, these experiments must have seemed daunting in the extreme. Particularly creditworthy is the clear understanding of the limitations of the experiments shown by the small sample size (p178) although oddly it now appears that skewedness of sense distributions in even very large samples might be a feature of real language (Krovetz and Croft, 1992; Stokoe and Tait 2003).

It was of course impossible for Sparck Jones to know this in 1964.

9. Concluding Remarks

We can say that while the theory of clumps was not wholly satisfactory in itself, it has been of importance for other reasons. It was intended to be a theory of classification that explicated our intuitive idea of a set of things that are somewhat loosely related by family resemblances, which was the basis of the notion of conceptual classes of the kind that seemed appropriate to retrieval (Sparck Jones, 1971).

We have highlighted three great strengths of this work: the bringing together of automatic classification with linguistic semantics and computational linguistics; the use of a pre-existing machine readable resource (Roget); and the appreciation of the need for experimental work on the largest scale feasible. These three aspects of the work make it far ahead of its time. In rereading the book we have found many insights which would have seemed profound and far sighted in the 1980's or even the 1990'.

Synonymy and Semantic Classification is not a widely read or referenced work: this is perhaps a product of it only having been properly published over twenty years after its original acceptance as a thesis. Despite the inevitable shortcomings of such early work. It can still be read with profit by any student of the relevant fields, and the material covered, and issues raised, are as central to the study of these fields as it was in 1964. The Appendix on the history of artificial languages remains of strong independent

interest. The thesis was to serve as worthy foundation of a long and successful career, and it provided themes followed by Karen throughout that period. We hope this review will stimulate wider reading of the book so it can finally achieve the recognition it deserves.

Acknowledgement

References

Boguraev, B. and T. Briscoe (eds.) Computational Lexicography for Natural Language Processing, Longman, London, 1989.

Edmonds, P. and A. Kilgarriff Introduction to the Special Issue on evaluating word sense disambiguation systems Natural Language Engineering 8(4) 2002.

Ellman, J. and Tait, J.I. "On the Generality of Thesaurally derived Lexical Links" Proceedings of JADT 2000, the 5th International Conference on the Statistical Analysis of Textual Data eds. M. Rajman & J.-C. Chappelier. March 2000, Swiss Federal Institute of Technology, Lausanne, Switzerland. Ecole Polytechnique de Federale Lausanne, Switzerland. Pp 147-154.

Katz, J.J. Semantic Theory, Harper and Row, New York, 1972.

Krovetz, R and W.B. Croft Lexical Ambiguity and Information Retrieval ACM Transactions on Information Systems 10(1). 1992

Lewis, David, `General Semantics, Synthese, 22, 18-67, 1970

Lyons, J. A structural Theory of Semantics and its Application to some Lexical Sub-Systems in the Vocabulary of Plato, Ph.D. Thesis, University of Cambridge, 1961. Published as Structural Semantics, No. 20 of the the Publications of Philological Society, Oxford, 1963.

Lyons, J, Semantics, Cambridge University Press, Cambridge, England. 1977.

Masterman, M., R.M Needham and K. Sparck Jones The Analogy between Mechanical Translation and Library Retrieval Proceedings of the International Conference on Scientific Information. Vol. 2 917-935. Washington DC. 1958

Masterman, M., Needham, R.,Sparck Jones, K., Mayoh, B. Agricola Terram Dimovit Aratro, ML 92, Cambridge Language Unit, Cambridge, England,1958.

Morris, J. and G. Hirst, Lexical Cohesion Computed by Thesaural Relations as an Indicator of the Structure of Text Computational Linguistics 17(1). 1991.

Miller, G.A., Chodorow, M., Fellbaum, C., Johnson-Laird, P., Tengi, R., Wakefield, P. and Ziskind, L. WordNet - a Lexical Database for English, Cognitive Science Laboratory, Princeton University., 2000

Needham, R. and Parker-Rhodes, F. A method for using computers in information classification, in Proc. IFIP62, The Hague, 1962.

Olney, J., Revard, C., and Ziff, P. Some monsters in Noah's Ark. Research memorandum, Systems Development Corp., Santa Monica, CA 1968.

Procter P.et al. The Longman Dictionary of Contemporary English. Longman, Burnt Mill, Herts.1978

Pulman, S. Word meaning and belief, Croom Helm, Beckenham, Kent, 1983.

Quine, W.V.O. From a logical point of view, Cambridge, MA, 1953.

Schank, R. Conceptual Information Processing, North Holland, Amsterdam, 1975.

Schvaneveldt, R. Pathfinder Networks, Ablex, Norwood, NJ, 1990.

Sparck Jones, K, Dictionary Circles, SP-3304, System Development Corporation, Santa Monica, CA 1967.

Sparck Jones, K. Theory of Clumps in Encyclopaedia of Library and Information Sciences, (eds.) A. Kent and M. Lacour. Vol.5. New York: Marcel Dekker, 1971.

Sparck Jones, K. Information retrieval and artificial intelligence, Artificial Intelligence 114 (1999) 257-281.

Sparck Jones, K. Document Retrieval: Shallow Data, Deep Theories, Historical Reflections, Potential Directions, in (Advances in Information Retrieval ECIR 2003), (ed.) F. Sebastini, Springer, Berlin, 2003, 1-11.

Stokoe, C., M. Oakes and J. Tait Word Sense Evaluation in Information Retrieval Revisited Proceedings of the 26th ACM SIGIR Conference on Research and Development in Information Retrieval (SIGIR 2001), Toronto, July 2003. 159-166

Tanimoto, T. An elementary theory of classification and prediction, IBM Research, Yorktown Heights, 1958.

Wilks, Y. Primitives and Words, in (eds.) R. Schank and B. Nash-Webber, Theoretical Issues in Natural Language Processing, BBN Inc., cambridge, MA, 1975.

Wilks, Y. and Stevenson, M. Sense tagging: Semantic tagging with a lexicon. In Proceedings of the SIGLEX Workshop Tagging Text with Lexical Semantics: What, why and how? Pp.47-51, Washington, D.C., 1997

Yarowsky, D. Hierarchical Decision Lists for Word Sense Disambiguation. Computers and the Humanities, 34(2):179-186, 2000.

Xu, J. and W.B. Croft Query Expansion using Local and Global Document Analysis Proceedings of the 19th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval (SIGIR 96), H.P. Frei, D. Harman, P. Schauble and R. Wilkinson (eds.), Zurich, Switzerland. 1996.