REPRESENTING WORD SEMANTICS FOR IR BY CONTINUOUS FUNCTIONS

Péter Wittek
School of Computing, National University of Singapore
wittek@comp.nus.edu.sg

Sándor Darányi
Swedish School of Library and Information Science, Göteborg University
Sandor.Daranyi@hb.se

Keywords: Information retrieval, language representation, word semantics, signal analysis

Abstract: Information representation is an important but neglected aspect of building text information retrieval models. In order to be efficient, the mathematical objects of a formal model, like vectors, have to reasonably reproduce language-related phenomena such as word meaning inherent in index terms. On the other hand, the classical vector space model, when it comes to the representation of word meaning, is approximative only, whereas it exactly localizes term, query and document content. It can be shown that by replacing vectors by continuous functions, information retrieval in Hilbert space yields comparable or better results. This is because according to the non-classical or continuous vector space model, content cannot be exactly localized. At the same time, the model relies on a richer representation of word meaning than the VSM can offer.

1 Introduction[1]

This is a first report on interdisciplinary work in progress. Since language as the carrier of meaning (a.k.a. information) is crucial to IR, we are interested in its modelling. With the respective linguistic field of study being called semantics, meaning in language comes in two major kinds: word meaning or word semantics, and sentence meaning or sentence semantics. This paper focuses on the mathematical representation of the former variant only. Further, here we do not move beyond applying Lyons' word semantics to index terms in the vector space model (VSM) and its derivate using continuous functions.

A word of warning is in place here. Since the study of meaning has been the preoccupation of scholars for millennia now, we cannot and will not attempt to offer new answers to the nature of this phenomenon. Instead, we will concentrate on showing that as much as vocalized and written character strings, represented by vectors, continuous functions can be used as carriers of word semantics as well. In this respect, we build on the seminal work of Hoenkamp, but also go beyond his predictions, at least in terms of theoretical implications. There will be two of these. First, what he has shown was that square integrable functions from signal processing such as the Fourier transform, JPEG or wavelets, are unitary operators based on the $\mathcal{L}^2$ norm in Hilbert space, and thereby suitable for latent semantic indexing (LSI) (Hoenkamp, 2001). However, the reason why these can be applied to IR at all is that there exists a fitting model of word semantics, mapping the meaning of index terms onto continuous functions. Secondly, we conjecture that by departing from a special representational blend of word semantics, the mathematical equivalent of the Heisenberg uncertainty principle in quantum mechanics (QM) applies to the results.

This proof-of-concept paper is structured as follows. We start with the information representation potential of vectors (Section 2), explaining types of word meaning important for our experiment. We depart from the working assumption that a document or query vector in the traditional VSM is a discrete sample of a continuous "content signal", represented either by periodical (Section 3) or by non-periodical functions (Section 4) and resulting in the two model

[1] In: Dominich, S. - Kiss, F. (Eds.). Studies in Theory of Information Retrieval. Proceedings of the ICTIR07 Conference, Budapest, 18-20 October 2007 (Foundation for Information Society) Budapest. pp. 149-155.

variants named Continuous Model A (CMA) and Continuous Model B (CMB). Section 5 introduces a similarity function which enables both the CMA and CMB to perform information retrieval. First results in an information retrieval scenario indicate that both CMA and CMB tend to outperform the traditional (”discrete”) vector space model but part of the improvements goes back to the exploitation of quasi-referential meaning (Section 6). Finally, we discuss two theoretical implications of the non-classical approach to the VSM: IR by continuous functions can be interpreted in terms of semantic fields in general linguistics, and the mathematical uncertainty principle applies to such models. (Section 7).

2 Word Semantics in the VSM and in Function Space

In this paper, we will distinguish between semantic content as contrasted with e.g. image content in content-based image retrieval (CBIR), a misnomer from the point of linguistics.

In language, words with a meaning do not exist alone but in relation to one another. Such a complex structure is divided into regions according to similarities in meaning, and is called a semantic field (Trier, 1934; Lehrer, 1975) or a lexical field (Löbner, 2002). In a semantic or lexical field, there are gaps, as not every location in the field carries semantic content (Lyons, 1968).

IR has been trying to model semantic fields as distributions of index terms by different variants of the VSM (Salton et al., 1975; Wong et al., 1985; Deerwester et al., 1990). The underlying idea was to use locations in a geometry as, metaphorically speaking, vehicles of some charge, so that charged locations should represent index terms in vector space, whereas empty locations should stand for lexical gaps. The pursuit of this goal has led from utilizing term frequencies from unweighted to weighted variants for creating term vectors to including sources of word semantics other than term occurrence, such as their sense structure (Banerjee and Pedersen, 2002; Budanitsky and Hirst, 2006) or their definitions (Lesk, 1986). As a result of such experiments, vector space can be shown to comply with different major theories of word meaning, briefly discussed below.

There exist quite a few theories of word meaning (Nöth, 1990), but in IR, the typical assumption, going back to language technology , is that the distributional hypothesis (Harris, 1970), also called the contextual hypothesis (Miller and Charles, 1991), is the sole source of word meaning available for automated extraction from document texts. Recently, Sahlgren has examined this model and discussed how its application, by means of statistics and geometry, can lead to the creation of word-spaces (Sahlgren, 2006). The central idea underlying this hypothesis is that if we consider words or morphemes A and B to be more different in meaning than A and C, then we will often find that the distributions of A and B are more different than the distributions of A and C. In other words, difference of meaning correlates with difference of distribution (Harris, 1970). There exists also a philosophical argument for a contextual theory of meaning by Wittgenstein, stating that “Meaning is use” (Wittgenstein, 1967), also advocated by (Blair, 1990; Blair, 2006). This argument can be interpreted to hold in a vector space environment either as: (1) meaning is contextual, where contextuality stands for co-occurrence of observation values on some variables, discarding false assumptions about senses of an expression and leaving the true assumption only; and (2) meaning is term occurrence rate based, i.e. the consolidation of contextuality (the solidity of meaning) is a function of the co-occurrence of words related to other ones, the latter interpretation underlying the former. According to Aitchison, such a disambiguating role for context is a commonplace in structural linguistics: “Language can (...) be regarded as an intricate network of interlinked elements in which every item is held in its place and given its identity by all the other items” (Aitchison, 1999).

The VSM, being term occurrence-based, contains and can utilize contextual meaning only. Let there be M index terms, one vector of the canonical basis of $\mathbb{R}^M$ is assigned to each index term. The order of assignment is *arbitrary*. Let a_{ij} be the weight of term i in document j. Thus a document vector $\mathbf{a}_j$ is a linear combination of the basis vectors:$\mathbf{a}_j = \sum_{i=1}^{M} a_{ij}\mathbf{e}_i$.

The representation of documents ignores any semantic relation between the index terms, as orthogonal vectors are assigned to the index terms. In fact index terms are not independent, a discrepancy much investigated, and an essential point in this work.

In our current attempt we relied on a combination of sources of word semantics richer than the one based on tfidf, using WordNet (Fellbaum, 1998). This blend is computed from the hyponymic hierarchy of word senses and their distances, as if we dealt with *quasi-referential* word semantics. This terminus technicus covers the following:

1. Apart from contextual word meaning, Lyons – among others– distinguishes between referential meaning and sense (Lyons, 1968). The former equals the relation between a word in language and an element of physical reality outside lan-

guage. The latter connects two words in language, sense typifying the kinds of relations possible between any word pairs such as incompatibility, synonymy, homonymy, antonymy, hypo- and hypernymy, mero- and holonymy, etc.;

2. Because our respective algorithm, based on the Lesk similarity (Banerjee and Pedersen, 2002), computes the distance of two index terms as WordNet glosses from their hyponymic hierarchy, both the glosses and their sense hierarchies being intra-linguistic, this type of meaning cannot be regarded referential because such hierarchies are not elements of physical reality. On the other hand, they are definitely not contextual either, as the sense structure of a vocabulary is external to index term distributions.

As we conjecture that semantic fields are by nature continuous, albeit discontinuities in the form of lexical gaps exist in them, continuous functions are more suitable to model this nature than vectors are. This is the reason why we conceived the continuous VSM.

3 A Continuous Model for Information Retrieval

The continuous model of document representation and information retrieval is based on the classical vector space model. It is non-classical inasmuch as the coordinates of real-valued document vectors are interpreted as a subset of the range of continuous functions. The functions are constructed by inverse Fourier transformation. The continuity of semantic content is a prerequisite for this IR model, provided by term clustering.

In order to reproduce the continuity of semantic content, one has to deal with the following problem. A word is a hypernym if its meaning encompasses the meaning of another word of which it is a hypernym; it is more generic or broader than another given word. Sometimes hypernymy is referred to as the “is-a” relation . Its opposite relation is hyponymy. A word is a hyponym (Lyons, 1977) if its semantic range is included within that of another word; it is also referred to as the “instance-of” relation. For example, *ship* is an instance of *vessel*, so *ship* is a hyponym thereof, whereas *icebreaker* is a *ship*, *ship* here being a hypernym to *icebreaker*. For the most comprehensive general domain ontology with a hypernym hierarchy for English, see the WordNet database (Fellbaum, 1998).

The assignment of canonical basis vectors to index terms is arbitrary in case of the classical VSM. As the continuous model heavily utilizes term interdependence, such arbitrary assignment should be avoided. Instead, the clustering of index terms arranges them in a semantically continuous fashion, with occasional gaps between ”islands of similar meaning”. Related words are assigned to subsequent vectors of the canonical base. Secondly, consider the hypernymic terms *vessel*, *ship*, and *icebreaker*. As these vectors need to reproduce, apart from their related meaning, a respective hierarchy as well, first they should be assigned to $\mathbf{e}_{i-1}$, $\mathbf{e}_i$, and $\mathbf{e}_{i+1}$ for some i, before we transform them to the $\mathcal{L}^2$ base of function space. We will return to this point below, after briefly discussing k-means clustering as applied for the arrangement of our index terms.

For the semantic arrangement of index terms, two k-means clustering methods based on cosine dissimilarity and Lesk similarity were tested, the former on tfidf, the latter on the above specified quasi-referential sources of word meaning. Lesk similarity has proved to be superior to other similar distances for the task of word sense disambiguation (Lesk, 1986; Pedersen et al., 2005). The so-called adapted Lesk algorithm counts overlaps between the definitions of terms: the more words they share the more similar the terms are (Lesk, 1986). It also compares the hypernymic, hyponymic, holonymic, meronymic and troponymic attributes of each word in a term pair for which similarity is computed (Pedersen et al., 2005). Compound words and their definitions are included in the comparison (Pedersen et al., 2005). A term usually has more than one sense, but it is computationally not feasible to identify (disambiguate) the senses in each processed document for indexing by senses instead of terms. As a simplification, only the first, most frequent senses and their definitions were considered in this experiment. Since, however, the canonical basis cannot reproduce sense hierarchy even after semantic rearrangement, one has to opt for the $\mathcal{L}^2$ basis instead.

Lately, Hoenkamp has pointed out that the $\mathcal{L}^2$ space can be used for IR when he introduced a Haar basis for the document space and identified three properties for desirable transformation operators. Such operators (1) should be unitary (to preserve cohesion of the documents), (2) they should lead to dimension reduction (to preempt the lexicon problem), and (3) they should be computationally inexpensive (Hoenkamp, 2001).

In order to introduce a wavelet-based continuous model of IR in $\mathcal{L}^2$ space, we approach the problem from a different angle than Hoenkamp did, observing semantics as a constraint on IR. Therefore apart from tfidf as the raw material for testing, we also utilize the Lesk similarity and thereby quasi-referential sources of word meaning, in addition to his above observations. The Whittaker-Shannon formula (Equation 1) describes a way to reconstruct the function from discrete values (Weaver, 1988). It uses the unnormalized sinc function (sinus cardinalis), defined by $\mathrm{sinc(x)} = \frac{\sin(x)}{x}$. The function has a removable singularity at zero, its value at this point is specified explicitly as 1. The sinc function is analytic on $\mathbb{R}$. The Whittaker-Shannon formula is defined by

$$f(t) = \sum_{k=0}^{n-1} f(k)\mathrm{sinc}(2\pi\omega_c(t-k)). \tag{1}$$

The sinc function is not periodic and is of finite power and the formula does not assume periodicity of the original signal.

$$f_j(t) = \sum_{k=0}^{M-1} a_{kj}\mathrm{sinc}(2\pi\omega_c(t-n)). \tag{2}$$

The bandwidth ω_c is unknown; in fact, it is more like a parameter than a derived constant, as long as the conditions of the sampling theorem are satisfied. By changing ω_c, the width of one sinc impulse is controlled. The width of a sinc function determines how much of the semantic relatedness between terms (Section 7) is considered.

The Whittaker-Shannon formula is not the only way of restoring a signal of discrete samples. The next section introduces another formulation, hence the above model will be referred to as Continuous Model A (CMA).

4 A Variant of the Continuous model

Since it is possible to use both wavelets as non-periodical continuous functions (Model CMA), and continuous periodical functions to construct and reconstruct document and query vectors, we designed a respective model called Continuous Model B (CMB) as well.

Following clustering, we perform Fourier transform on the document vectors of the matrix A. Let d_{nj} denote a Fourier coefficient. For the sake of simplicity, the elements in the matrix are indexed starting from 0.

$$d_{nj} = \frac{1}{M}\sum_{k=0}^{M-1} a_{kj}\exp(-\imath n\frac{2\pi}{M}k)),$$
$$0 \le n \le M-1, 0 \le j \le N-1,$$

where a_{kj} is an entry in the vector space model ($\mathbf{a_j}$ is the jth document vector), M is the number of features (i.e. the number of keywords used to index the documents), N is the total number of documents, and d_{nj} is a Fourier coefficient. The inverse transform which reconstructs the original document vector is consecutively defined by

$$a_{kj} = \sum_{n=0}^{M-1} \mathrm{Re}(d_{nj})\cos(n\frac{2\pi}{M}k) - \mathrm{Im}(d_{nj})\sin(n\frac{2\pi}{M}k),$$
$$0 \le k \le M-1, 0 \le j \le N-1.$$

Assuming that the initial vector space had only real entries, the number of elements to be summed is reduced.

$$a_{kj} = \mathrm{Re}(d_{0j}) + 2\sum_{n=1}^{M/2} \mathrm{Re}(d_{nj})\cos(n\frac{2\pi}{M}k) - \tag{3}$$
$$\mathrm{Im}(d_{nj})\sin(n\frac{2\pi}{M}k),$$
$$0 \le k \le M-1, 0 \le j \le N-1.$$

If Formula (3) is extended from the discrete set of $k = 1, 2, \ldots, M$ to the set of real numbers, periodic functions with period M are created:

$$f_i(x) = \mathrm{Re}(d_{0i}) + 2\sum_{n=1}^{M/2} \mathrm{Re}(d_{ni})\cos(n\frac{2\pi}{M}x) - \tag{4}$$
$$\mathrm{Im}(d_{ni})\sin(n\frac{2\pi}{M}x), \quad 0 \le j \le N-1.$$

The continuous document functions are the above functions restricted to the $[0, M-1]$ interval. If a document vector is regarded as a sample of a hypothetical continuous signal, Equation (4) is the reconstruction of that signal.

5 Information Retrieval

In the VSM, the document retrieval process is based on the inner product of the Euclidean space as the similarity measure, where document and query vectors represent a formal description of a users information need and a set of documents satisfying that need (van Rijsbergen, 1979; Dominich and Kiezer, 2007). The continuous model follows the same principle: it maps natural language documents to a feature space (that is $\mathcal{L}^2([0, M-1])$), and uses the usual inner product of this Hilbert space to express similarity between documents, or a document and a query for retrieval.

As for this inner product, let λ be the Lebesgue measure on $\mathbb{R}$, then

$$(f_i, f_j) = \int_{[0,M-1]} f_i(x) f_j(x) d\lambda(x), \tag{5}$$
$$f_i, f_j \in \mathcal{L}^2([0, M-1]).$$

Thus, information retrieval can be performed by Equation (5) in a similar fashion as in the VSM.

Cluster Size	Bandwidth (ω_c)				
	0.001	0.01	0.1	1	10
12.5	0.110	0.130	0.124	0.105	0.105
25	0.102	0.112	0.107	0.114	0.134
50	0.137	0.112	0.109	0.126	0.098
100	0.110	0.138	0.111	0.132	0.097

Table 1: Mean Average Precision Values of Continuous Model A, Cosine Clustering

Cluster Size	Bandwidth (ω_c)				
	0.001	0.01	0.1	1	10
12.5	0.102	0.112	0.119	0.116	0.145
25	0.129	0.125	0.122	0.106	0.088
50	0.116	0.118	0.119	0.144	0.111
100	0.123	0.112	0.120	0.139	0.131

Table 2: Mean Average Precision Values of Continuous Model A, Lesk Clustering

6 Results

Experimental results for both the CMA and CMB models were computed on the ADI test collection[2]. The subject of the collection is Information Science and it consists of 82 documents and a relatively larger number of queries, 35. A range of parameters were tested for both models.

The standard 11-pt recall-precision values were calculated to estimate performance. The eleven values were averaged to get a single-number description (mean average precision). The mean average precision of the VSM was 0.131.

6.1 Results of the Continuous Model A

The results for CMA are consistently better than for CMB. Non-periodical functions outperformed the ”discrete” VSM in 4 out of 20 measurements based on contextual word meaning, i.e. cosine clustering (Table 1), and in 3 out of 20 measurements based on quasi-referential meaning, i.e. Lesk clustering (Table 2). The best mean average precision, 0.145, was reached by a combination of bandwidth 10 and cluster size 12.5, in other words where bandwidth best approximated cluster size.

The detailed results of the best three parameter settings with Lesk clustering are plotted on a standard 11-point recall versus precision diagram (Figure 1).

[2] http://ir.dcs.gla.ac.uk/resources/test_collections/

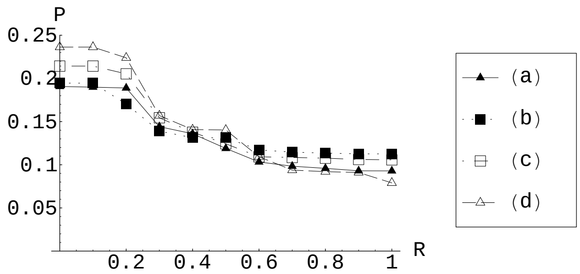

Figure 1: 11-point Recall versus Precision Diagram, Continuous Model A; (a) Vector Space Model; (b) Clustering: Lesk Similarity, Average Cluster Size: 100, Bandwidth: 1; (c) Clustering: Lesk Similarity, Average Cluster Size: 50, Bandwidth: 1; (d) Clustering: Lesk Similarity, Average Cluster Size: 12.5, Bandwidth: 10

Cluster Size	Cut-off Frequency				
	M/4	5M/16	3M/8	7M/16	M/2
12.5	0.131	0.129	0.124	0.118	0.131
25	0.108	0.104	0.116	0.117	0.127
50	0.140	0.126	0.131	0.131	0.130
100	0.116	0.119	0.116	0.129	0.130

Table 3: Mean Average Precision Values of Continuous Model B, Cosine Clustering

6.2 Results of the Continuous Model B

The efficiency of CMB, measured by the mean average precision, is close to but worse than that of the ”discrete” vector space model based on contextual word meaning, and is worse based on quasi-referential word meaning. Table 3 and 4 present the results for various cut frequencies and average cluster sizes.

7 Discussion

The proposed models are theoretically interesting because of their linguistic relevance and alternative

Cluster Size	Cut-off Frequency				
	M/4	5M/16	3M/8	7M/16	M/2
12.5	0.115	0.126	0.125	0.126	0.128
25	0.123	0.118	0.120	0.118	0.125
50	0.115	0.117	0.121	0.128	0.131
100	0.120	0.114	0.124	0.125	0.127

Table 4: Mean Average Precision Values of Continuous Model B, Lesk Clustering

approach to modelling semantic content, i.e. using continuous functions instead of vectors. Although (Hoenkamp, 2001) proposed to use such continuous functions for LSI, and we are experimenting with the coupling of theories of word meaning and continuous LSI as well, in this paper we have limited ourselves to the more general question of language representation by them.

There are two implications of CMA and CMB. The first regards the question, why is it possible to use mathematical objects in Hilbert space for IR at all? If we consider the fact that an infinite number of continuous functions may exist with only the potential, but not the chance, to signify, i.e. mean anything, it becomes evident that what matters is the *coupling* between language and Hilbert space. The elements thereof can mean anything based on human convention only. Once this convention is called a language, the next step is to find a theory of word semantics which happens to map reasonably well onto the geometry of Hilbert space. As we have indicated above, such theories exist, and can be traced back to the inherent notion of the continuity of semantic content.

Since the 18th century, this continuity gave rise to two related concepts, that of a semantic field and of a lexical gap (Lyons, 1977). The two together essentially describe something evolving in a well-documented fashion, following socio-historic changes in language use, and can be conceived as an n-dimensional texture of semantic content, i.e. a geometry with locations used vs. not used for the representation of semantic content. Whereas we note in passing that lexical semantics is working on the above concepts and their status is far from being finalized, their existence is not debated. This offers a solid opportunity for IR to cooperate with linguistics on elaborating a joint synoptic understanding.

Secondly, as also discussed by (Park, 2003), there is a mathematical phenomenon that closely resembles the Heisenberg uncertainty principle in quantum mechanics, and is also referred to as the mathematical uncertainty principle (Papoulis, 1963).. Roughly speaking, the more tightly localized an f signal (function) is, the less localized is its Fourier transform $\hat{f}$.

The duration of f is defined as follows:

$$D_t^2 = \int_{\mathbb{R}} t^2 |f(t)|^2 d\lambda(t)$$

The bandwidth of f is defined as follows:

$$D_\omega^2 = \int_{\mathbb{R}} \omega^2 |\hat{f}(\omega)|^2 d\lambda(t)$$

Time-Frequency Uncertainty Principle: If $\lim_{t\to\pm\infty} \sqrt{t} f(t) = 0$, then $D_t D_\omega \geq \sqrt{\frac{\pi}{2}}$.

Applying a low-pass filter or defining a low cut-off frequency, as in our experiment, therefore means that in the time domain (which is now the axis of terms), a document function is no longer localized precisely. Hence as a representation it no longer resembles the ”discrete” VSM with its exact localization, instead document content is ”smeared out” reflecting the fuzziness of information. A more “smeared out” representation is due to, and contains more quasi-referential meaning. However, as a trade-off, a document is no longer localized precisely, that is, its original vector cannot be fully reconstructed by the continuous function. This trade-off is captured in the above uncertainty principle.

8 Conclusions

We departed from the working hypothesis that the classical VSM, when it comes to information representation - with special regard to the representation of word meaning -, is approximative and this has a bearing on its efficiency.

We have found an indication that the mathematical uncertainty principle, a property of the continuous VSM model, may have significance to IR utilizing more advanced combinations of semantic content. As is known, this principle is a general form of the Heisenberg uncertainty principle, and it states that content cannot be exactly localized.

In our experiment, the same VSM IR model, run on the ADI test database, performed in some cases better once the same information was represented by appropriate continuous functions instead of vectors in it. We regard these first results inconclusive because of the small size of the test database, and keep working on collecting more evidence to understand whether there is a fundamental trade-off between IR effectiveness –the success rate of an algorithm to retrieve relevant documents– and the localization of content in an IR model.

9 Acknowledgement

We wish to thank for an unknown referee for his constructive criticism; and to Pirkko Suihkonen (Department of General Lingustics, University of Helsinki) and Barnabás Garay (Institute of Mathematics, Budapest University of Technology and Economics) for discussions on semantic fields and the Heisenberg uncertainty principle, respectively.

REFERENCES

Aitchison, J. (1999). *Linguistics*. Hodder & Stoughton, London.

Banerjee, S. and Pedersen, T. (2002). An adapted Lesk algorithm for word sense disambiguation using WordNet. *Proceedings of the Third International Conference on Intelligent Text Processing and Computational Linguistics*, pages 136–145.

Blair, D. (1990). *Language and representation in information retrieval*. Elsevier North-Holland, Inc. New York, NY, USA.

Blair, D. (2006). *Wittgenstein, language and information:" Back to the rough ground!"*. Springer, Dordrecht.

Budanitsky, A. and Hirst, G. (2006). Evaluating WordNet-based Measures of Lexical Semantic Relatedness. *Computational Linguistics*, 32(1):13–47.

Deerwester, S., Dumais, S., Furnas, G., Landauer, T., and Harshman, R. (1990). Indexing by latent semantic analysis. *Journal of the American Society for Information Science*, 41(6):391–407.

Dominich, S. and Kiezer, T. (2007). A measure theoretic approach to information retrieval. *Journal of the American Society for Information Science and Technology*, 58(8):1108–1122.

Fellbaum, C. (1998). *WordNet: An Electronic Lexical Database*. The MIT Press, Cambridge, Mass.

Harris, Z. (1970). *Distributional structure*. Humanities Press, New York.

Hoenkamp, E. (2001). Unitary operators for fast latent semantic indexing (FLSI). *Proceedings of the 24th annual international ACM SIGIR conference on Research and development in information retrieval*, pages 400–401.

Lehrer, A. (1975). *Semantic fields and lexical structure*. North-Holland. New York: American Elsevier.

Lesk, M. (1986). Automatic sense disambiguation using machine readable dictionaries: how to tell a pine cone from an ice cream cone. *Proceedings of the 5th annual international conference on Systems documentation*, pages 24–26.

Löbner, S. (2002). *Understanding Semantics*. Arnold, London.

Lyons, J. (1968). *Introduction to theoretical linguistics*. Cambridge University Press, Cambridge.

Lyons, J. (1977). *Semantics*. Cambridge University Press, Cambridge.

Miller, G. and Charles, W. (1991). Contextual correlates of semantic similarity. *Language and Cognitive Processes*, 6(1):1–28.

Nöth, W. (1990). *Handbook of Semiotics*. Indiana University Press, Bloomington.

Papoulis, A. (1963). *The Fourier Integral and its Applications*. McGraw-Hill, New York.

Park, L. (2003). *Spectral Based Information Retrieval*. PhD thesis, University of Melbourne.

Pedersen, T., Banerjee, S., and Patwardhan, S. (2005). Maximizing Semantic Relatedness to Perform Word Sense Disambiguation. *Supercomputing institute research report umsi*, 25.

Sahlgren, M. (2006). *The Word-Space Model: using distributional analysis to represent syntagmatic and paradigmatic relations between words in high-dimensional vector spaces*. Institutionen för lingvistik. Department of Linguistics, Stockholm University.

Salton, G., Wong, A., and Yang, C. (1975). A vector space model for information retrieval. *Journal of the American Society for Information Science*, 18(11):613–620.

Trier, J. (1934). Das sprachliche Feld. Eine Auseinandersetzung. *Neue Jahrbücher für Wissenschaft und Jugendbildung*, 10:428–449.

van Rijsbergen, C. J. (1979). *Information Retrieval*. Butterworths, London.

Weaver, H. (1988). *Theory of Discrete and Continuous Fourier Analysis*. John Wiley & Sons, New York.

Wittgenstein, L. (1967). *Philosophical investigations*. Blackwell Oxford.

Wong, S., Ziarko, W., and Wong, P. (1985). Generalized vector spaces model in information retrieval. In *Proceedings of the 8th annual international ACM SIGIR conference on Research and development in information retrieval*, pages 18–25. ACM Press New York, NY, USA.