



Jan Kochanowski University Press

This is a contribution from *Token: A Journal of English Linguistics*
Volume 2/2013.

Edited by John G. Newman and Sylwester Łodej.

© 2013 Jan Kochanowski University Press.

Patterns of lexical collocation in sermonic texts

Alexandra Uzoaku Esimaje

Benson Idahosa University

ABSTRACT

This study investigated lexical word associations in sermons to establish their types, patterns and behaviour. It adopted a corpus-computational technique in which 200 actual sermons were built into a corpus and compared to a reference corpus of contemporary English as a measure of normality. Data were analysed partly by computer techniques. The concord tool of Wordsmith 5 (Scott 1999) was used to specify the patterns of collocation of selected keywords while the statistical tools, log-likelihood test and mutual information score were applied to measure the reliability and validity of the results. The results showed significant differences in the collocation of words in the sermons and in other contexts: in associations, patterns and behaviour. In particular, different sets of collocates were retrieved for each of the selected items, new patterns of association were discovered and the words manifested a downward collocating style. The study recommends that new evidence revealed in lexical research such as this be incorporated into linguistic descriptions.

1. Introduction

The need to investigate the behaviour of English lexis in the specific context of religion is echoed by Crystal – Davy (1969), and Crystal (1995), who point out that the collocational idiosyncrasies which occur in religious English provide a discussion point in their own right. Claridge – Wilson (2002) also note that the sermon genre is interesting because it occupies the crossroads of orality and literacy. Yet sermons, as communicative events, have so far attracted very little linguistic attention. The paucity of research information on the linguistic significance of sermons is clearly evidenced by the minimal literature about it. So although mention is made of the distinctiveness of this genre, little effort has been made to investigate this claim by using a sermon

corpus against the background of a reference corpus, to reveal specific differences in the character of English lexis. This study is a step in filling this gap. This investigation- of the behaviour of English lexis in the context of Christian sermons- will serve the primary functions of illuminating this particular language event, enhancing its understanding, and contributing to our knowledge of language as a system.

The lexical investigation of meaning has led to the development of various theoretical positions on the subject. The three major ones, according to Cruse (2000), are: the componential approach, in which word meaning is viewed as a construction of smaller invariant units; the contextual approach, consisting of a) the holist view which holds that word meaning needs to be finitely specified independently of other words in the language. This idea is in direct contrast to b) the holistic view that the meaning of a word cannot be known without taking into account other words in its environment; and the conceptual approach, which defines the meaning of a word relative to the concepts it gives access to in the cognitive system.

This study is predicated on the holistic view, pioneered by Firth (1935-1951), and whose focus is the syntagmatic sense relations of words. This is understood to mean the collocational preferences or restrictions in language. It is on the basis of this behaviour that Firth argues that this phenomenon characterizes a level of language at which meaning is dispersed, the lexical level that is, and proposes it as an independent level of linguistic description alongside others. After Firth, this argument has garnered more strength through the works of: Sinclair in particular (1987, 1991, 1998, 2004); Halliday (1966, 1991); Stubbs (1993, 1996, 2002); Lewis (1993, 1997); and many others. The essential point is that language comprises lexical units which share horizontal or syntagmatic relations, such that the meaning of one is included in another or completed by another in its linguistic context, and this manifests as a major way through which meaning arises from text.

2. Conceptual background

2.1 Lexical patterns

A pattern is considered here as phraseology which frequently associates with a word or lexical item. This could be prepositions, groups or clauses that accompany a given word (Hunston – Francis 2000). Patterns and lexis are mutually dependent because each pattern is composed of a restricted set of lexical items, and each lexical item has a restricted set of patterns.

A pattern is closely associated with meaning in that different senses of words are distinguished by their patterns of occurrence and words which share a pattern tend also to share a meaning. Phraseology, then, can refer to the grammatical pattern which belongs to a word, and every word has its own pattern. A parallel idea to the one above is the phraseology of lexis, lexical patterns which account for the combinatorial tendency in language (Lyons 1981). This means that most language is a construct, not from *basic* structures and the lexicon, but from sequences of pre-arranged, pre-constructed or formulaic language (Sinclair 1998, 2004). When we speak of patterns in this sense, reference is made to patterns: of word co-occurrence, of collocation, and in text- textual patterns- and their implications for meaning. In this context then, a pattern describes the behaviour of a lexical item as evidenced in our data of language use. The axes of patterning are word, idiom (phrase) and collocation, which for the purposes of this study we shall call the *units of lexis*. Collocation is the focus of this study.

2.2 Collocations

Collocation has been variously defined: as a relation of the probable co-occurrence of items (Malmkjaer 1991); “actual words in habitual company” (Firth 1957: 14); the co-occurrence of two or more words within a short space of each other (Sinclair 1991); the associations a word acquires on account of the meanings of words in its environment (1991); and the relationship a lexical item has with items that appear with greater than random probability (Hoey 1991). What resonates from all the definitions is the idea of association- that words prefer the company of some words rather than others.

It would seem then that collocation encapsulates the lexical item and its structures, that is, its patterns or phraseologies. It is in this extended sense that the term is used in this study- as a unit of language description parallel to the lexical or syntagmatic level of language. This is a substantive level of language which creates meaning, by the systematic association of words with certain other words.

3. Methodology: Model of extended lexical units and corpus method

The study utilised the model of lexical description proposed by Sinclair (2004: 29, 141) which presents five categories of description of any lexical item, two compulsory, three optional. The obligatory components are “the core, invariable evidence of the occurrence of the item ... and the semantic

prosody which is the determiner of the meaning of the whole". The optional categories serve to fine tune the meaning and cohesion of a whole text: collocation, colligation and semantic preference. In describing the lexis of sermons, this study utilized two of the categories: the *core* – the node word which is the invariable occurrence of a lexical item and *collocation* – the co-occurrence of words with no more than five intervening words.

As a further methodological basis, the study utilised a corpus-linguistic approach. Data for this study were sourced from two corpora. 200 sermons which were carefully selected for investigation were classified into twenty-five categories to form 25 sermon fields or data files. Each of the twenty-five files contains an equal number of sermons, i.e. eight sermons per text file. These constitute the main corpus which was named *Nigerian Sermons Corpus (NSC)*, and consists of 64,851 running words. All the sermons were written in contemporary English and published between 2004 and 2007.

Second, the *British National Corpus (BNC)* condensed sampler version was used as the reference corpus. This version contains two million words, a million words each of spoken and written English; this is labelled *BNC Sampler (BNCS)*. These corpora- *NSC* and *BNCS*- were then exploited and manipulated to derive the exact data for the analyses. First, the combined text was converted to plain text and, second, using the lexical software *Wordsmith*, both the frequency and keyword counts were obtained. *Wmatrix* software (Rayson 2008) was then applied to verify the results.

3.1 Population and sampling

The research population is made up of eight hundred published Pentecostal sermons. In the publications, there are twenty-five dominant topics. These became both the basis for the stratification of the sermons into text types and the guide for sampling. Because of the nature of sermon texts, it was impossible to take equal samples in terms of number of words since the sermons were of unequal lengths. So the criterion of sampling full sermon texts, which Halliday (1991), Sinclair (1991) and Biber – Conrad – Reppen (1998) recommend for the investigation of textual and association patterns, was adopted. Also, the selection of full sermons was necessary in order to widen the scope of search.

3.2 Significant collocations

Since the major objective of this study is the identification of the structural patterns of words- collocations in particular- a statistical measure was

required to identify which co-occurrences are significant. This was done using the formulae of mutual information score because it is compatible with the WordSmith 5 software tool chosen for the study. The mutual information score between any given pair of words compares the probability that the two occur together as a joint event (because they belong together) with the probability that they occur individually and so their co-occurrence is merely a factor of chance. The more strongly connected two items are, the higher their mutual information score, but if there is low-level co-occurrence, the mutual information score will be a negative number (McEney – Wilson 1996).

3.3 Choice of sermon keywords for analysis

A selection of keywords was made for analysis using the criterion suggested by Berber-Sadinha (1999): selecting either a simple majority (half + 1), or a significant subset (using a test of significance). The latter option is chosen in this instance. Some key words, out of the ones identified, were chosen based on their degree of statistical significance. Still, the number remained high, so manual grouping according to themes was done and then words drawn from them for further analysis. In Tabs. 1 and 2 below are shown the first and final selections made for analysis.

Table 1. The typology of sermon keywords

N	Historical words	Qualities and activities words	Commonly used specifically religious words	Technical words	Words used in religion and in other registers
1	2	3	4	5	6
1	God	Love	Heaven	Watchman	Spirit
2	Jesus	Power	Anointing	Sacrifice	Wisdom
3	Bible	Prayer	Jehovah	Flesh	World
4	Abraham	Healing	Devil	Blood	Man
5	Christians	Praise	Miracle	Body	Knowledge
6	Christ	Truth	Brethren	Word	People
7	Manna	Confession	Covenant	Ministry	Heart
8	Israel	Mercy	Sin	Tongues	Speak
9	Church	Godliness	Believer	Sow	Life

1	2	3	4	5	6
10	Ark	Glory	Hell	Fruit	Promotion
11	Satan	Faith	Seedtime	Sword	Name
12	Gospel	Patience	Testimony	Ministry	Walk
13	Crucified	Forgiveness	Demon	Message	You
14	Salvation	Worship	Satan	Seed	Victory
15	Pharisees	Counsel	Grace	Wilderness	Works
16	Pharaoh	Obedience	Hallelujah	Harvest	Virtue
17	Nazareth	Holy	Saviour	Gospel	Prosperity
18	Disciple	Great	Revelation	Scripture	Poverty
19	Resurrection	Righteousness	Lord	Saved	Father
20	Genesis	Boldness	Judgement	Challenges	Wealth

From this table, a total of twenty lexical items were randomly selected for analysis as shown in Tab. 2 below.

Table 2. Lexical items selected for analysis

N	Historical words	Qualities and activities words	Commonly used specifically religious words	Technical words	Words used in religion and in other registers
1	God	Mercy	Miracle	Word	Life
2	Salvation	Faith	Anointing	Fruit	Walk
3	Christians	Love	Sin	Saved	Name
4	Crucified	Praise	Covenant	Challenges	People

4. Analysis

The purpose here is to identify those words which habitually occur with the sermon keywords, and to describe their patterns of combination to underline any peculiarities in their behaviour in this context. In order to reduce the robustness of this paper, only an example analysis of the collocational behaviour of a category of sermon lexis is presented here, and in comparison to their behaviour outside the sermon context.

4.1 Sample analysis of the personal qualities and activities keywords

The four keywords in this group- *love, praise, faith and mercy*- are shown in the table below to confirm their keyness status, and to indicate the extent of their spread in the texts.

Table 3. Keyness and dispersion

KEYWORD	FREQ. 1	C %	FREQ. 2	RC %	KEYNESS	P VALUES	DISPERSION
Love	105	0.16	486	0.02	203.890	0	0.311
Praise	19	0.03	16	0	84.06	0	0.719
Faith	186	0.29	88	0	947.41	0	0.598
Mercy	14	0.02	38	0	38.6	0	0.64

All four items represented in this group demonstrate higher frequencies in the sermons than in general English. This is proved by their relative frequencies in both corpora. For example, *love*, which has a frequency of 105 in *NSC* and 486 in *BNCS*, accounts for 0.16% of the sermons, while it accounts for only 0.02% of general English, proving it is more important to sermons than to general English. The other words do not have a sufficient number of occurrences and so lack any proportional percentage rating in *BNCS* to validate their keyness status in *NSC*.

The dispersion ratings for each of these words show that only *love* is unevenly spread in the sermon texts. So although *love* has the highest keyness value, it manifests the lowest dispersion to show that dispersion is only based on textual spread. In the following table we shall see other keywords which co-occur with these key keywords in a number of texts, called associates.

Shown above are those words that associate with the main keywords, because they are also keywords in the same texts. *Faith* demonstrates more productivity in this regard. Each of the sets of words forms a clump which gives information about the keyword. Take *faith* as an example; its associates seem to suggest that it is religion based. The words differ appreciably in the nature of associates they have to support the individuality status of each keyword.

Table 4. Associates

Faith		Love	Mercy	Praise
God	Healed	Supernatural	Sin	Spiritual
God's	Blood	Thee	Sins	Spirit
Your	Corinthians	World	Righteousness	You
Unto	Christ	You	Spiritual	Word
Holy	Condemn	Your	Of	Your
Life	Accuse		Says	
Word	Scripture		Yourself	
Romans			Romans	
Jesus			Spirit	
Spirit			Unto	
Says			Word	
Hath			Your	

We shall now examine these words to identify words which collocate with them, within our specified span of five words on either side. The table below shows the top 20 collocates found for each of the node words.

Table 5. Top 20 collocates

Mercy	Love	Praise	Faith
1	2	3	4
Abundant	Manifest	Naked	Produced
Abounding	Receiving	Celebration	Application
Mercy	Hearkening	Praise	Versus
Cries	Love	Tower	Faith
Merit	Unchangeable	Rejoiced	Eleven
Sprinkle	Conquers	Sweating	Concept
Obtain	Sincerity	Profusely	Impart
Cover	Reaffirmed	Rising	Conquering
Toward	Wound	Continual	Subdues
Merciful	Testify	Sacrifice	Starting
Seat	Translation's	Continually	Adding
Deeds	Emulate	Worship	Outlandish
Nevertheless	Universal	Thank	Lastly
Forgiveness	Reciprocate	Offer	Bile
Animals	Expresses	Glorious	Lick
Behalf	Fickle	Glory	Hopelessness

1	2	3	4
Song	Ultimate	Strong	Erring
Grace	Responses	Refuse	Emulate
Devil's	Surpasses	Flesh	Proving
David	Compared	Greater	Respect

Once again, it is observed that each word has its own class of collocates and there do not seem to be overlaps in these classes. From these lists therefore we can derive such collocations as *abundant mercy, abounding mercy, mercy cries; manifest love, unchangeable love, candid love; strong praise, glorious praise, praise sacrifice; and conquering faith, outlandish faith and produces faith*. In the following section, the strength and directionality of these collocations will be assessed.

Table 6. Strength and directionality of collocation in NSC

Word	with	Mi	Word	with	Mi
Hearkening	Love	10.27	Application	Faith	9.446
Love	Love	9.284	Versus	Faith	9.446
Surpasses	Love	9.271	Produced	Faith	9.446
Tempered	Love	9.271	Faith	Faith	8.536
Purge	Love	9.271	Eleven	Faith	8.446
Responses	Love	9.271	Engaged	Faith	8.446
Compared	Love	9.271	Erring	Faith	8.446
Compassion	Love	9.271	Grant	Faith	8.446
Works	Love	9.271	Finisher	Faith	8.446
Emulate	Love	9.271	Exact	Faith	8.446
Word	with	Mi	Word	with	Mi
Naked	Praise	11.74	Cries	Mercy	12.18
Celebration	Praise	11.74	Abundant	Mercy	12.18
Praise	Praise	11.74	Mercy	Mercy	12.18
Tower	Praise	11.74	Merit	Mercy	12.18
Rejoiced	Praise	11.74	Sprinkle	Mercy	12.18
Sweating	Praise	11.74	Abounding	Mercy	12.18
Profusely	Praise	11.74	Obtain	Mercy	11.18
Rising	Praise	11.74	Merciful	Mercy	11.18
Continual	Praise	10.74	Cover	Mercy	11.18
Sacrifice	Praise	9	Toward	Mercy	11.18

As can be seen above, *hearkening*, *naked*, *application*, and *cries* are among the first 10 words which strongly associate with each of our search words. This is authenticated by the mutual information scores of: 10.27 for *hearkening love*; 11.74 for *naked praise*; 9.446 for *application faith*; and 12.18 for *mercy cries*. To assess the strength of this relationship requires us to determine the reciprocity of this association by examining how the search words themselves relate to these other words. The results show that in all cases there is no reciprocity in the strength of association because, although the search words also collocate with these words, they collocate with very many other words. So the strength is unequal. The search words are therefore weaker since they are freer and less predictable in their behaviour and in each case the direction of collocation is shown to be downward. The results are summarized in the following table.

Table 7. Collocation strength

NODE	FREQ.	NO. OF COLL.	NODE	FREQ.	NO. OF COLL.
Hearkening	1	9	Love	105	269
Naked	1	10	Praise	19	74
Application	1	10	Faith	186	252
Cries	3	19	Mercy	14	72

4.2 Phraseology

Each of the four sermon words selected in this group demonstrated the ability to form multi-word items of various lengths. Among these, the dominant patterns identified seem to be N + V, N + N, V + N, and ADJ + N. For example, *faith filled*, *faith inspired*, *weak faith*, *love conquers*, *love based*, *love God*, *praise God*, *praise sacrifice*, *praise worship*, *mercy cries*, *abundant mercy*, and *mercy deeds*. From the three- and four-word structures, the dominant patterns seen are in prepositional phrases or noun phrases; for example, *love of God*, *faith in God*, *praise to God*, and *mercy on us* and *praise be to God*.

5. Discussion of results: The patterns of collocations in the sermons

For each of our search words, collocates were found in large numbers. However, because we are constrained by both space and time it became

necessary to select a sample for closer examination. For each keyword therefore, the first 50 collocates, sorted on their mutual information (MI) scores, were retained; these are those words that are most strongly associated with the keywords such that they are highly primed to co-occur. Our discussion here will centre on the behaviour of the keywords selected for in-depth study, in terms of the collocations they form- their strength, directionality and phraseology- to underline peculiarities in comparison to general English.

Each of the twenty keywords demonstrated a very high tendency to co-occur with certain words, each manifested a high association strength (an MI score of 7 or above), proving that they are very much attracted to each other and, as such, are confirmed collocates of the nodes. However, the results show that when placed on the collocation strength clines of strong, medium and weak, all but one of them illustrated weak collocating strength, showing that there is lack of reciprocity in the degrees of association since they are unpredictable in their partners and so are unreliable as indicants of fixed meaning. It was also seen that each establishes its own individuality in that each has more or less its own class of collocates, as was the case with their associates. At the same time, there was more or less uniformity in the collocating style of the nodes: a downward collocating pattern in which the node chooses collocates less frequent than itself.

The implication is that each word derives meaning from its immediate environment, defined by the words constantly in proximity to it. Therefore, the contextual meaning of a word is a function of its co-text, its syntagmatic relations rather than a function of choice alone, by paradigmatic means as standard grammars hold. This confirms Firth's assertion that meaning inheres the horizontal plane of language, and bears out previous research found in the literature of lexical studies, for example, Halliday (1991), Sinclair (2004), Stubbs (2002), Hunston (2002), and Biber (1999). Some examples of the collocations found in the sermons are *God partner, Christians testify, Lord crucified, salvation tragedy, praise God, salvation helmet, love responses, rising praise, faith heroes, mercy deeds, forbidden fruit, sin saved, joyful challenges, soil word, prophet anointing, root sin, miracle seed, and victorious walk.*

In terms of phraseology or patterns, the study shows that the keywords exhibit both grammatical and lexical patterns. However, our interest is in the lexical patterns, to investigate constancy and peculiarity. Ten (10) major patterns of collocation were found as shown in the table that follows.

Lexical collocations are formed from combinations of lexical word classes: nouns, verbs, adverbs and adjectives; and Benson – Benson – Ilson

(1997) identified 7 types as: V+N/P/PP; V+N; ADJ+N; N+V; N1 of N2; ADV+ADJ; and V+ADV. Of these, the first five were found in the sermons while the last two were not; rather, some other kinds of collocation structures occurred, for example, N+N, ART+N, NP, V+V, and V+NP. The results call for extension of the index of lexical collocation types to accommodate the ones revealed in the study.

Table 8. Structural types of collocations found in the sermons

S/n	Collocation Structure	Examples of Collocations
1	ADV + N	Forbidden Fruit, Heavy Anointing, Abundant Mercy, Continual Anointing
2	N1 of N2	Kingdom of God, Agent of Salvation, Word of God, Blood of Jesus
3	V + V	Confesses Saved, Baptized Saved, Recorded Crucified, Redeem Crucified
4	V + N	Seek Salvation, Praise God, Conquering Faith, Committing Sin
5	V + NP	Take the Helmet of Salvation, Receive your Salvation, Praise the Lord
6	N + V	Faith Subdues, Prophet Anointing, Grace Saved, Praise Offer
7	N + N	Mercy Deeds, Faith Heroes, Trials Challenges, Substance Fruit
8	ADJ + N	Strong Praise, Universal Love, Rich Mercy, Backslidden Christians
9	ART + N	the Anointing, a Believer, a Watchman, the Faith, the Word
10	NP	the Body of Christ, the Kingdom of God, the Name of Jesus

5.1 Result of collocation analysis in BNCS

The examination revealed the dissimilarities between collocates in these groups and those in NSC and immediately pointed to different usages. It became obvious that we could not hypothesize the same attitudinal meanings for these words in this context. This suggests that these words differ significantly from general English words and may characterize only sermon texts. We find that the asymmetry between collocates of each word in both contexts is overwhelming. For example, each word manifested a difference in meaning and, as a consequence, in the words that co-occur with it. For instance, *love* was predominantly used in a personal, natural sense; *faith* was used to mean general belief in anything including the public confession of

faith; *praise* was used to refer mainly to personal praise or approval; and *mercy* was used to indicate human and political mercy (pardon). As expected therefore, their collocates also differed in both contexts to confirm our earlier conclusion that definite behaviours characterize lexical items in different contexts of use.

6. Findings

The analyses revealed differences in the usage of words in the sermons and in general English: in associations, patterns and behaviour. In the first place, the words examined portrayed noteworthy differences in the kinds of words they co-occur with: different sets of collocates were retrieved for each of the selected items in sermons and outside of sermons. Therefore, different usages were posited for them.

In terms of patterns, the study showed 10 main kinds of patterns of word combination: 5 out of the 7 lexical collocation types known to occur in general English were found, and another 5 types, uncommon in ordinary usage, were also found. This suggests that, beyond meaning relations, patterns of collocation may distinguish texts, in addition to leading to some generalisations in language.

We find then that the previous knowledge of patterns of lexical collocation is extended by our results, and this constitutes essential information for the learner and teacher of English as a second language.

In terms of behaviour too, notable differences were recorded. All but one of the examined words showed weakness in collocational strength and all manifested downward collocation to illustrate the non-fixedness of word meaning and the importance of the immediate environment (co-text) in meaning making. So the principle of co-selection of lexis, which avoids lexical dissonance or semantic clash (Cruse 2003), was supported by the results.

7. Implications

As the findings show, the sermon words collocated with sets of words different from their usual collocates and formed new structures of combination. The implication of the unknown collocations found is that new shades of meaning will emerge which did not previously exist and as such the collocations need to be taken into account in linguistic description when their meanings will

enhance understanding of sermon communications both for its immediate and extended hearers. Also, findings on the behaviour and patterns of English words constitute vital information for non-native speakers of English (Scott – Tribble 2006). This is so because it enables them to know not only what is grammatically possible in the language, but also what is appropriate and what happens in specific contexts. Therefore, knowledge of lexical patterns, like those provided in this study, has great import for the achievement of communicative competence.

8. Conclusion

The study has revealed the different character of English in sermons. This would have been impossible without the evidence provided by the sermon corpus. In view of the findings, the study recommends that future descriptions of English take cognisance of the new evidence provided by corpus research, especially in the area of lexical patterning. In order to enhance the pedagogic value of lexical research, further studies should build on the present state of lexical research, by going beyond the presentation of knowledge of lexical structures or patterns to establish general conclusions concerning the uses of the various forms.

REFERENCES

Sources

British National Corpus
<http://www.natcorp.ox.ac.uk/>
Nigerian Sermons Corpus

Special studies

- Benson, Morton – Evelyn Benson – Robert Ilson
 1997 *The BBI Dictionary of English Word Combinations*. Amsterdam/
 Philadelphia: John Benjamins.
- Biber, Douglas – Susan Conrad – Randi Reppen
 1998 *Corpus Linguistics: Investigating Language Structure and Use*.
 Cambridge: Cambridge University Press.

- Biber, Douglas
 1999 "Lexical bundles". In: Biber et al. (eds.) *Longman Grammar of Spoken and Written English*. London: Longman.
- Claridge, Claudia – Andrew Wilson
 2002 "Style evolution in the English sermon". In: T. Fanego and E. Seoane (eds.) *Sounds, Words, Texts, and Change*. (Current Issues in Linguistic Theory 224.). Amsterdam: John Benjamins, 25-44.
- Cruse, D. Alan
 2000 *Meaning in Language*. Oxford: Oxford University Press.
- Crystal, David
 1995 *The Cambridge Encyclopaedia of the English Language*. Cambridge: Cambridge University press.
- Crystal, David – Derek Davy
 1969 *Investigating English Style*. London: Longman
- Firth, John Rupert
 1935-1951 *Papers in linguistics 1934-1951*. London: Oxford University Press.
 1957 "A Synopsis of linguistic theory, 1930-55". In: F. R. Palmer (ed.) *Selected Papers of J.R. Firth 1952-59*. [Repr. 1968] London and Harlow: Longman.
- Halliday, Michael Alexander Kirkwood
 1966 "Lexis as a linguistic level". In: C. E. Bazell, J. C. Catford, M. A. K. Halliday, and R. H. Robins (eds.) *In Memory of J.R. Firth*. London: Longman, 148-161.
 1991 "Corpus studies and probabilistic grammar". In: K. Aijmer and B. Altenberg (eds.) *English Corpus Linguistics: Studies in Honour of Jan Svartvik*. London: Longman, 30-41.
- Hoey, Michael
 1991 *Patterns of Lexis in Texts*. Oxford: Oxford University Press.
- Hunston, Susan – Gill Francis
 2000 *Pattern Grammar: A Corpus Driven Approach to the Lexical Grammar of English*. Amsterdam: John Benjamins.
- Leech, Geoffrey
 1991 "The state-of-the-art in corpus linguistics." In: K. Aijmer and B. Altenberg (eds.) *English Corpus Linguistics: Studies in Honour of Jan Svartvik*. London: Longman, 8-27.
- Lewis, Michael
 1993 *The Lexical Approach*. Hove: Language Teaching Publications.
 1997 *Implementing the Lexical Approach*. Hove: Language Teaching Publications.
- Lyons, John
 1981 *Language, Meaning and Context*. Suffolk: The Chaucer press.
- Malkmjaer, K. (ed.)
 1991 *The Linguistic Encyclopaedia*. London: Routledge.

McEnery, Tony

- 2006 *Swearing in English: Bad language, Purity and Power from 1586 to the present*. London: Routledge.

McEnery, Tony – Andrew Wilson

- 1996, 2001 *Corpus Linguistics*. Edinburgh: Edinburgh University press.

Rayson, Paul

- 2008 *Wmatrix: A Web-based Corpus Processing Environment*. Lancaster: Computing Department, Lancaster University.
<http://ucrel.lancs.ac.uk/wmatrix/>.

Sadinha, Berber Tony

- 1999 *Using Keywords in Text Analysis: Practical Aspects*. (Direct Working Papers 42.) Sao Paulo and Liverpool.
<http://www2.lael.pucsp.br/direct/DirectPapers42.pdf>.

Scott, Mike

- 1999 *WordSmith Helpline*. Liverpool: Lexical Analysis Software.
2006 "Concordances". In: M. Scott and C. Tribble *Textual Patterns: Key Words and Corpus Analysis in Language Education*. Amsterdam: John Benjamins, 33-57.

Scott, Mike – Christopher Tribble

- 2006 *Textual Patterns: Key Words and Corpus Analysis in Language Education*. Amsterdam: John Benjamins.

Sinclair, John

- 1987 "Collocation: A progress report". In: R. Steele and T. Threadgold (eds.) *Language Topics: Essays in Honour of Michael Halliday*. Vol. 2. Amsterdam: John Benjamins, 319-332.
1991 *Corpus, Concordance, Collocation*. Oxford: Oxford University Press.
1998 "The lexical item". In: E. Weigand (ed.) *Contrastive Lexical Semantics*. Amsterdam: John Benjamins, 1-24.
2004 *Trust the Text: Language, Corpus and Discourse*. London: Routledge.

Stubbs, Mike

- 1993 "British traditions in text analysis: From Firth to Sinclair". In: M. Baker, G. Francis, and E. Tognini-Bonelli (eds.) *Text and Technology: In Honour of John Sinclair*. Amsterdam: John Benjamins, 1-36.
1996 *Text and Corpus Analysis*. Oxford: Blackwell.
2002 *Words and Phrases: Corpus Studies of Lexical Semantics*. Oxford: Blackwell.