

What do you think of that: A pilot study of the phraseology of the core words of English

Antoinette Renouf

1. Background

There is a continuing need in the field of applied linguistics, and particularly in language teaching and the hybrid area of text processing software development, for reliable criteria for assigning priorities in the selection of lexis. Corpus-based research points to frequency of word occurrence as an important factor, since a frequency profile of the lexicon of a text of any kind indicates that a small minority of words, unchanging at the very core, make the major contribution in text creation.

In 1984, as one of the advocates of a lexically-based language syllabus that took frequency into account (Sinclair- Renouf 1988: 140-160), I worked with a team within the Co build Project to identify the central features of the top 650 word forms in the Birmingham Corpus, for inclusion in such a syllabus (Renouf 1987: 167 -178).

The lexical environments of those words were observed to include low frequency items; high-frequency words do not, of course, combine exclusively with each other. Linguists such as Lyons (1969) have pointed out that they are low in information content; and so they alternate with less common and semantically richer items, forming "collocational frameworks" (Renouf - Sinclair 1991: 128 -143) to support them.

Since the 1984 study, the phenomenon of alternation between common and less common words, which I term "oscillation", has remained a matter of personal curiosity. In particular, I have wondered about the effect of reducing the degree of oscillation; or, to put it another way, what the nature is of the exclusively high-frequency language environment.

Draft

302 Antoinette Renouf

2. Aim

In this paper, we are therefore seeking to report on a study of what native speakers do with the language in the higher range; what the recycling of highly-frequent words produces; and, from a pedagogic point of view, what the learner of English can do with just the core words at his or her disposal.

We shall look briefly at the nature of the longest strings that are produced, then move on to study in some detail the semantics and grammar of shorter, more recurrent strings.

3. Data

3.1. The corpora

The data for the study consisted of an 18 million word written corpus and a 1.3 million word spoken corpus, the two components of the Birmingham Corpus. In this report, we shall refer to them as "corpora" rather than "sub-corpora". From these two corpora were drawn lists of the top 150 word forms, subdivided into the top 50, 100 and 150 respectively.

3.2. The word lists

In creating the word lists, a number of measures were taken in the interests of simplicity. Orthographic representations of nonverbal elements, such as *er* and *urn* were removed, because they generally function as interpolations within otherwise unbroken lexical strings. A number of single characters, representing a range of functions, were also removed. They chiefly represent markers of participant change, but include the word *p* for the monetary denomination. On the other hand, contracted forms such as *don't* and *there's* were allowed, since these are single graphic forms. No attempt was made at homographic separation.

Table J. Top 150 word forms in the Birmingham Corpus

Written Component (18 m)			Spoken Component (1.3 m)		
I-50	51 - 100	101-150	1 -50	51-100	101-150
the	up	made	the	don't	than
of	out	too	and	an	course
and	more	before	of	got	good
to	who	such	a	just	into
a	me	many	to	by	there's
in	them	us	I	them	should
that	about	through	that	when	only
was	has	little	in	get	did
it	will	also	you	that's	could
I	into	don't	it	then	fact
is	can	go	is	see	where
he	some	still	have	has	want
for	could	should	this	oh	its
with	like	own	yes	say	they're
as	do	years	we	going	over
on	then	good	they	were	work
had	only	must	but	some	here
be	time	never	was	more	make
you	now	those	for	out	you're
his	than	being	on	who	lot
at	other	make	think	been	first
but	its	come	be	up	actually
not	your	long	are	had	come
they	very	off	as	mean	down
by	over	work	what	other	us
have	two	day	well	go	back
this	even	think	it's	really	take
from	people	old	so	me	rather
are	any	again	one	much	those
or	our	world	at	will	does
she	these	between	do	their	years
her	down	it's	with	your	being
were	first	here	very	time	you've
all	most	right	not	I'm	most
which	just	take	or	my	her
we	back	life	there	how	put
one	after	say	if	she	many
an	way	might	about	thing	look
there	because	another	which	these	I've
their	did	come	all	any	question

Table 1. continued

Written Component (18 m)			Spoken Component (1.3 m)		
1-50	51-100	101-150	1-50	51-100	101-150
would	much	thought	know	sort	same
so	how	going	he	things	find
if	know	away	no	way	year
said	may	great	can	right	also
been	new	got	people	quite	kind
my	man	went	because	said	we're
him	see	men	now	something	point
when	well	same	would	two	must
what	where	children	like	our	can't
no	too	while	from	his	may

The definition of high-frequency item includes, as can be seen in Table 1, not just items traditionally thought of as grammatical but also lexical words, notably core verbs.

The commoner a word is in these lists, the more likely it is to dominate the word strings that it and the other words enter into. For this reason, an anticipatory look was taken at the top 50 in each case, to see what they suggested in relation to language use.

The top 50 written words contain the pronouns *that, it, I, he, you, his, they, this, she, her, which, we, one, their, my, him* and *what*, that play a range of grammatical roles. There is a very sparse selection of verbs: simply *was, is, had, be, have, are, were, would, said* and *been*. Of these, *said* is the only lexical verb. This would suggest that the writers will be concerned with the fact or possibility of someone being or doing something, and on reporting what others said.

The top 50 spoken words include fewer pronouns, only *I, that, you, it (it's), this, we, they*, and *he*, in that order. Together with *people*, and in most cases the multi-class *all, what* and *which*, these can all occupy a grammatical subject position. There are few verbs available: *is, have, was, think, be, are, do, know, can, would* and *like*. The indications are that the 50 words will combine to facilitate interaction between speakers. Expression of opinion, and requesting of opinion and information are suggested. The words *yes* and *well* will allow agreement or qualification and, together with *but*, hedging and polite disagreement.

4. Methodology

Strings (string tokens) made up of the two word lists were extracted from the different corpus components. They were sorted into string types in order of length, beginning with the longest. The longest strings inevitably subsumed shorter ones, but the frequency count given to shorter string types was not modified to take that into account, and so understates somewhat their significance. This was not felt important for a preliminary study.

The strings were studied in isolation and in context, and the following findings emerged.

5. Findings

5.1. The longest word strings

We were interested to see just how long language users can go on stringing high-frequency words together, and what these strings consist of. Altenberg (1991: 1 - 26) observes that "the longer, less common [combinations] have a higher-incidence of open-class items", but this will not be the case where open-class words are largely prohibited. The results are shown in Tables 2 and 3.

From Table 2, we can see that the writer of English occasionally keeps going surprisingly long on a small vocabulary. We also see that this propensity increases with a greater lexical range. Strings are longer still, however, in the spoken data, as can be seen in Table 3.

The strings in Tables 2 and 3 are statistically interesting, but they are not of overwhelming interest to the linguist. One can observe that they are nonce formations. Whilst incomplete, they are largely regular in grammatical terms, made up of a series of linked or overlapping units. Their length is partly attributable to lexical reiteration. This is expected in speech (10, 15), but is perhaps surprisingly evident in the written examples, (1-6). The strings are largely lacking in idioms, but (4) is a

(mis)quotation of the Lincoln aphorism beginning: "You can fool all the people some of the time and some of the people all the time..."

Table 2. Longest word strings in the written corpus

No.	Word Rank	No. of Words	
(1)	1-50	12	he was and what he was and he was not to be
(2)		12	it not to be there but it was there and when the
(3)		12	it was for you it was all for you you said I
(4)	51-100	19	all of the people some of the time and some of the people all the time but could not
(5)		19	if we can have two one for me and one for you like I said and we can have
(6)		18	that you know it and when you do not to know that you do not know that is
(7)	101 -150	21	for a way to get through it I would think only about where I was I thought not what I was
(8)		21	it was for the little man that he should do so he had to get him out of the way at
(9)		21	that it was not for me even if it was or might have been for her even if it was her

Table 3. Longest word strings in the spoken corpus

No.	Word Rank	No. of Words	
(10)	1-50	18	and then I think she said that she will do it but I mean I don't know how
(11)		18	you know you know you know if there was some way if there was some way that we
(12)	51-100	22	in there and say you know right this is what you have to say and and this is what you have to
(13)		19	I know but I mean some people are I think I got my I think I got quite a
(14)	101-150	24	really want to take them I said to her well I don't know what to do and she said well if I were you
(15)		23	I mean I think you can there are I mean I think it's good to go into that sort of thing with the

Predictions as to the effect of the top 50 items in each corpus are not borne out by these data. We had anticipated that the presence of forms of the lexeme *be* in the written corpus would lead to existential statements.

as in (1, 2), but in these strings we see that *be* is also used in phrasal combinations such as (3), (8) and (9). We thought that *said* would characterise the written strings, in introducing quotation; in fact, it is

more influential in the spoken extracts. However, these discrepancies may be explicable in terms of the singular nature of very long strings.

Two observations of particular interest emerged from the data in Tables 2 and 3. The first is that the longest of the restricted word strings in the written corpus consist of written representations of speech from novels, such as (3); while, as intimated in the last paragraph, the longest strings of spoken words consist of quotation, of self and others, as in (10), (12) and (14).

The second observation is that these long and grammatically regular strings do not inspire one to internal investigation, but rather to speculation about their lexical completion. Two examples, (1) and (12) will be shown in their original contexts by way of illustration:

for (1) "He liked *where he was and what he was and he was not to be* lured into any associations that might compromise him,"

for (12) "No I mean, really, er I'm going to prime people before I go in there and say, you know, "Right, this is what you have to say and this is what you have to ask.""

The second of these contexts is notable in that, with the exception of the rarer word *prime*, a slight relaxation in lexical restriction would have allowed it to qualify as a single string.

5.2. The most recurrent word strings

The longer a word string, the more likely it is to be a single occurrence. The real interest comes when the phrases become recurrent. *Bona fide* dual instances of strings begin to occur, though sparsely, at around 7 words in both the written and the spoken corpora. They become established, along with multiple occurrences, at six words.

In Tables 4 and 5, the 30 most frequent six-word strings are provided, again made up of the top 150 words of each of the two corpora. The frequency count for each string in the particular corpus is provided (as it will be in all subsequent tables).

There are fewer instances of each phrase type among the spoken strings in Table 5 than among the written in Table 4. This is because the spoken corpus is approximately fourteen times smaller. The difference should be borne in mind when comparing the two batches; rank is probably a more reliable indicator of significance.

Table 4. Top 30 six-word strings in the written corpus

48	in such a way as to	10	and at the same time the to
25	at the same time as the what	9	be one of the most
24	you do and what you all you	9	it was going to be a
19	have to do is	9	for the first time in their at
18	in the same way as the had it	9	the same time he was the
17	not been for the	8	first of these is the
16	on the first day of the	8	it was one of the most
15	in such a way that the what	8	it is one of the most
13	are you going to do this is	8	in the same way that the I
12	not to say that	8	don't know what to do for
12	in much the same way as for	8	the first time on the but it
11	the first time in years and at	8	was not to be when we got
11	the same time to of the way	7	back to the when I got
10	in which the for the first	7	back to the what do you
10	time in the	7	think of the

Table 5. Top 30 six-word strings in the spoken corpus

6	what do you think of the	3	in a way but you see
6	can we have our first question	3	I think it would be an
5	what do you think about the may	3	I think it would be a
5	we have our first question to me	3	I do not think that the
4	to be a very	3	and I think that's a very and I
4	if you're going to be a	3	think one of the
4	if you see what I mean	3	and I think it is very
4	at the same time as the	3	and all that sort of thing yes I
4	and this is one of the	2	think the question is yes I
4	and one of the things that and all	2	think it's a very
4	this sort of thing	2	yes I think it was a
4	and all that kind of thing	2	with the question of how the
3	to me that one of the	2	which I think does to some
3	this is the first of a	2	what do you want to say what
3	there are quite a lot of	2	do you think of that

The strings shown are interesting in ways that the long strings are not. They are fairly heavily recurrent, which makes them more representative of the language as a whole. They are also less grammatically predictable from their parts, on the whole, so are worthy of internal investigation: we shall examine the idiomatic nature of these strings in more detail in table 11.

They also reveal recurrent semantic features. A tentative semantic regrouping of the two batches might be as shown in Tables 6 (written) and Table 7 (spoken).

Table 6. Semantic groupings in top 30 six-word strings, written corpus

Placing Things in Time/Order

25 at the same time as the 16
on the first day of the

11 for the first time in years 11

and at the same time to

10 for the first time in the 10 and

at the same time the

9 for the first time in their 9 at

the same time he was 8 the first

of these is the 8 for the first

time on the 7 when we got back

to the 7 when I got back to the

Complete/Interactive Phrases

19 all you have to do is

13 what are you going to do

8 I don't know what to do

7 what do you think of the

Indicating Manner

48 in such a way as to

IS in such a way that the 10 of

the way in which the

Making Comparisons

18 in the same way as the 12 in

much the same way as

8 in the same way that the

Placing Things in Importance/Specifying

9 to be one of the most 8 it

was one of the most 8 it is

one of the most

Making Contrasts

17 had it not been for the 12

this is not to say that

Prefacing/Concluding

9 it was going to be a 8

but it was not to be

Fragments

24 what you do and what you

, Some instances of the string-types containing *at the same time* may not express notions of time, but belong in the "contrasts" group.

Table 7. Semantic groupings in top 30 six-word strings, spoken corpus

Expressing Opinion

3 I think it would be an
 3 I think it would be a
 3 I do not think that the
 3 and I think that's a very 3 and I
 think one of the
 3 and I think it is very
 2 yes I think the question is 2 yes I
 think it's a very
 2 yes I think it was a
 2 which I think does to some

Complete/Interactive phrases

6 can we have our first question 5 may
 we have our first question 2 what do
 you want to say

Requesting Opinion;

6 what do you think of the
 5 what do you think about the 2 what
 do you think of that

Generalising/Being Vague

4 and all this sort of thing 4 and
 all that kind of thing 3 there are
 quite a lot of 3 and all that sort of
 thing

Specifying

4 and this is one of the
 4 and one of the things that

Placing Things in Time/Order

4 at the same time as the 3 this
 is the first of a

Requesting Agreement

iii. I agree with what I mean

ii.

4 to me to be a very
 4 if you're going to be a
 3 to me that one of the
 3 in a way hut you see
 2 with the question of how the

i The third type in the "requesting opinion" group is a complete phrase, and so belongs also in the "complete/interactive" group

ii These fragments are suggestive of the semantics of other groups, but not interpretable without further context.

The semantic groupings in Tables 6 and 7 are of necessity provisional, but they serve to indicate that certain functions of language are characteristic among the high-frequency words, at least in the Birmingham Corpus. The most recurrent function within the written strings is the placing of a thing or event in time or order. Within the spoken strings, it is the expression of opinion; and four of the fragmentary spoken string types are also redolent of this pragmatic area.

There is some correspondence between the two corpora, particularly in the groups of "complete and/or interactive phrases". This may be attributable to the written representations of speech in the written corpus. But the similarity is only in the lexical content; in discourse terms, three of the written phrase types are "problem-solution" signals (Hoey, 1983 [1991: passim]), whilst this function is absent in the spoken types. There is one identically-worded string in the two groups - *what do you think of the* - which might or might not be significant. Svartvik (1991: passim) says, of multi-word items, that "most types do not appear to be characteristic of either the spoken or written varieties"; this is not borne out by the recurrent string types that we have looked at, but they are, after all, a particular subset of the totality of language.

They are nevertheless important for what they do show, which is that strings of common words bring meaning to the text. In speech, they are not just a product of the interaction; they create it.

5.3. Phrasal contexts

It will be apparent even from the limited data in Tables 4 and 5 that longer strings are in turn made up of shorter, more recurrent strings. If we extract these, within their longer contexts, we shall gain an insight into the context of phrases. Tables 8 and 9 show some patterns that emerge. Table 8 looks within six-word strings; Table 9 looks outside them.

As we see in Table 9, the lexical restriction imposed in our study allows us access only to recurrent contexts that are short and largely abbreviated. Nevertheless, these short contexts are worthy of consideration. What we see within each string type in Table 8 are the beginnings of collocational sets that represent the potential for internal variation in those strings. But it may be necessary to turn to unrestricted context outside these string types, if we are to reveal more about the relationship between the lexical variation and the semantics of these strings; this might, for ex

ample, show when *and all that sort of thing* is synonymous with *and all that kind of thing*, and when not.

Surrounding each phrase in Table 9, rather than within it, are the rudiments of grammatical units or, lexically speaking, phrases; in other words, we see colligational information as to the use of each phrase. In some cases, the context is complete, as in *in years* or in the incipient collocational set, *in his/her life*.

These preliminary findings point a way forward in relation to how we may look at phraseology. If progressively greater oscillation, or lexical range, is allowed, we shall be able to monitor accurately the manner in which collocational sets, of both single words and phrases, accumulate at each stage, or in each frequency band, within and around the phrase.

Altenberg (1987: 133 -143) created a diagrammatic representation of high-frequency combinations in the London-Lund Corpus (LLC) to indicate their functional distribution in text. The presentation was similar to that of Tables 8 and 9, although we have not attempted to apply the

Table 8. Patterns of internal variation in six-word strings

(prep) the same (X)					
25	at	the	same	time	as the
11	and	at	the	same	time to
10	and	at	the	same	time the
9		at	the	same	time he was
18		in	the	same	way as the
8		in	the	same	way tha the
					t
in such a way (X)					
48	in	such	a	way	as to
15	in	such	a	way	that the
and all (det) (X) of thing					
4	and	all	this	sort	of thing
3	and	all	that	sort	of thing
4	and	all	that	kind	of thing
I think (pron) (be) (det)					
3		I	think	it	would be an
3		I	think	it	would be a
3	and	I	think	that	's a very
3	and	I	think	it	is a very
2	yes	I	think	it	's a very
2	yes	I	think	it	was a

Table 9. Patterns of variation in the context of strings

for the first time in					
~11		for	the	first	time in
11		for	the	first	time in years
10		for	the	first	time in the
9		for	the	first	time in their
6		for	the	first	time in a
6		for	the	first	time in his life
4		for	the	first	time in my life
4		for	the	first	time in its
7	and	for	the	first	time I
4	and	for	the	first	time he
3		to me	for	the	fi rst time
what do you think					
7	what	do	6	you	think
	what	do	5	what	think of the
	do	3	what	do	2
	what	do	2	what	you think you
	do		you	you	think about that
			you	think	of
			you	think	of
could do with					
6		I	could	do	with a
4		you	could	do	with a
4			could	do	with a
1	and	he	could	do	with more of it but this

functional type of analysis to our data in any formal way at this stage. The LLC Corpus is relatively small, so yielded few instances of each phrase, on average. The Birmingham Corpus is larger and inevitably generates more instances of each phrase type, as shown in Tables 4 and 5, and would be an appropriate test-bed for Altenberg's (inspired by Halliday 1985: Chapter 3) analysis of functional position.

5.4. Lexicalising tendency of high-frequency word forms

Pawley and Syder (1983: 191-225) use the term "lexicalised" of units "of clause length or longer whose grammatical form and lexical content is wholly or largely fixed". We apply it somewhat differently; to the phenomenon of high-frequency words taking on a lower-frequency function of one kind or another.

When the degree of oscillation within word strings is limited, we note that the lexicalising phenomenon becomes more pronounced. It manifests itself in different ways, which we shall identify and illustrate with reference to our written corpus data.

5.4.1. Lexicalisation of grammatical words

Some of the grammatical items among the top 150 words take on a rarer lexical role, as shown in Table 10.

Table 10. Lexicalised grammatical words

he (pronoun to noun)

1 that this he was not only a he but a he in the

in (preposition to adjective)

3 with the in (breath)

5.4.2. Lexicalisation of high-frequency verbs

On a previous occasion (Sinclair - Renouf 1988: 153), we have talked of high-frequency verbs as being "delexical", of having the tendency to relinquish their independent meaning and enter into shared meaning with lower-frequency words. An example would be *make* in the lexical unit *make an arrangement*, where it cooperates with *arrangement* to mean *arrange*; and *arrangement* is the main contributor of meaning.

In our realm of restricted lexis, we find evidence of an opposing tendency. The very common verb is seen to shift role on occasion, and become a less frequent noun. This act of grammatical transformation is accompanied by a move into an idiomatic unit. Given the lexical restriction, we find the transformed item itself being the main contributor of meaning to the idiom. Some examples, taken from the written corpus, are presented in Table 11.

The important point demonstrated by Table 11 is that we have uncovered a grammatical shift that is a feature of everyday, educated English. It is one of the means by which speakers ring the stylistic changes whilst not shifting out of a high-frequency register; it allows variety.

A rarer lexicalised alternative is also an insurance against potential ambiguity. The unmarked *I know* is pressed into service very regularly and to many different semantic and pragmatic ends. To convey that one is a member of a small elite who knows something that others don't, there is always *I know* as a potential realisation, but it only conveys this meaning when pronounced in a particular way. The marked *I am ill the know* is unambiguous.

Table 11. Grammatical transformation from high-frequency verb to low-frequency noun in the written corpus

in the know - (extract)

4 in the know

2 in the know and

I are in the know

I in the know that

I is in the know

I those in the know

I all in the know of

I by those in the know

I not in the know the

I of those in the know

know how

2 the know how of

I have the know how to

I on the know how of

I but the know how of the

I for the know how they will

I through the know how of old

I to see but the know how had also been

(be) a must

4 is a must

I a must if your

I are a must for

I as a must before

I is a must a

I is a must for

I which is a must

I is a must in our

I to be is a must

I it is a must for any

on the make

3 on the make

I men on the make

I he was on the make had long

I people on the make but I just then was

(while/if/when) the going (be) good

2 while the going was good

while the going is good a

if the going is good he will

him when the going was good and he had

The further significance of this finding is that it indicates a move towards nominalised structures. This is a trend that is conventionally associated with writing in technical domains, and formal and academic registers (although we prefer to see it in relation to discourse purpose). Here we have found nominalisation in everyday speech and in nontechnical writing.

6. Conclusion

This was a preliminary study, intended to reveal the phraseological relationships entered into by just the commonest 150 words. We have discovered that at this level of restriction, there is relatively little which can be expressed fully. To this extent, we have not yet hit the most fruitful vein, which would support a categorical statement about the language, nor are we yet in a position to offer any comprehensive data or advice to language learners.

Nevertheless, we have been able to observe what can be achieved at this level. The recurrent strings that occur are seen to be anticipating and setting up the conditions for different sorts of discourse activity, such as the requesting and giving of opinions, or the narration of events. They also manage to create some complete utterances, albeit short ones.

Looked at from a lexico-grammatical point of view, such strings reveal the beginnings of high-frequency collocational sets as part of their internal variation. The external context of recurrent strings has also been examined, and found to contain rudimentary colligational information about the use of phrases, rather than just words.

Finally, we have been able to uncover a tendency among the high-frequency words, both grammatical and lexical, to take on new roles as lexical items in low-frequency idioms.

The natural next step will be to increase the lexical range of our study. We have a strong suspicion, based on inspection of the data during this study, that with even a minor addition to the word lists, say of another 50 or 100 words, the strings will become significantly longer before hitting the oscillation barrier, and more often phraseologically complete.

The less restricted lexicon will also bring us a clearer view of phrasal contexts; fuller collocational sets will emerge within and around phrases.

The sets will still not be complete, at this high-level range, but they will begin to reveal more about the interchangeability of high-frequency items, such as the modals and other verbs.

References

- Aarts Jan - Willem Meijs (eds.)
1990 Theory and practice in corpus linguistics. Amsterdam: Rodopi.
- Karin Aijmer - Bengt Altenberg (eds.)
1991 English corpus linguistics: In honour of Jan Svartvik. London: Longman.
- Altenberg, Bengt
 1987 "Speech as linear composition", in: Graham Caie *et al.* (eds.), 133-143.
- Altenberg, Bengt - Mats Eeg-Olofsson
 1991 "Phraseology in spoken English: Presentation of a project", in: Jan Aarts - Willem Meijs (eds.), 1-26.
- Caie, Graham - Kirsten Haarstrup - Arnt Lykke Jakobsen - Jorgen Erik Nielsen - Jorgen Sevaldsen - Henrik Specht - Arne Zettersten (eds.)
1990 Proceedings from the Fourth Nordic Conference for English Studies (Elsinore 11th-14th May 1989). Copenhagen: University of Copenhagen, Department of English.
- Carter, Ron - Michael McCarthy
 1988 *Vocabulary and language teaching*. London: Longman.
- Halliday, Michael A. K.
1985 An introduction to functional grammar. London: Edward Arnold.
- Höey, Michael
1983 On the surface of discourse. (Reprinted 1991, University of Nottingham.)
 London: George Allen and Unwin.
- Lyons, John
1969 Introduction to theoretical linguistics. Cambridge: Cambridge University Press.
- Pawley, Andrew - Frances Hodges Syder
 1983 "Two puzzles for linguistic theory: Native-like selection and native-like fluency", in: Jack C. Richards - Richard W. Schmidt (eds.), 191-225.
- Renouf, Antoinette
 1987 "Moving on", in: John Sinclair (ed.), 167-178.
- Renouf, Antoinette - John Sinclair
 1991 "Collocational frameworks", in: Karin Aijmer - Bengt Altenberg (eds.), 128-143.
- Richards, Jack C. - Richard W. Schmidt (eds.)
1983 Language and communication. London: Longman.
- Sinclair, John (ed.)
1987 Looking up: An account of the COBUILD project in lexical computing and the development of the Co/ins COBUILD English Language Dictionary. London: Collins.
- Sinclair, John - Antoinette Renouf
 1988 "A lexical syllabus for language learning", in: Ron Carter - Michael McCarthy, 140-160.
- Svartvik, Jan
 1991 "What can real spoken data teach teachers of English?". [Paper read at the Georgetown University Round Table on Languages and Linguistics, April 1-4, 1991.]