Observations on the phraseology of academic writing: local patterns – local meanings?[1]

Ute Römer

1. Introduction

The past few years have seen an increasing interest in studies based on new kinds of specialized corpora that capture an ever-growing range of text types, especially from academic, political, business and medical discourse. Now that more and larger collections of such specialized texts are becoming available, many corpus researchers seem to switch from describing the English language as a whole to the description of a number of different language varieties and community discourses (see, for example, Biber 2006; Biber, Connor and Upton 2007; Bowker and Pearson 2002; Gavioli 2005; Hyland 2004; and the contributions in Connor and Upton 2004; and in Römer and Schulze 2008).

This paper takes a neo-Firthian approach to academic writing and examines lexical-grammatical patterns in the discourse of linguistics. It is in many ways a tribute to John Sinclair and his groundbreaking ideas on language and corpus work. One of the things I learned from him is that, more often than not, it makes sense to "go back" and see how early ideas on language, its structure and use, relate to new developments in resources and methodologies. So, in this paper, I go back to some concepts introduced and/or used by John Sinclair and by John Rupert Firth, a core figure in early British contextualism, who greatly influenced Sinclair's work. Continuing Sinclair's (1996: 75) "search for units of meaning" and using new-generation corpus tools that enable us to explore corpora semi-automatically (*Collocate*, Barlow 2004; *ConcGram*, Greaves 2005; *kfNgram*, Fletcher 2002–2007), the aim of this paper is to uncover the phraseological profile of a particular sub-type of academic writing and to see how meanings are created in a 3.5-million word corpus of linguistic book reviews written in English, as compared to a larger corpus of a less specialized language.

After an explanation of the concept of "restricted language" and a discussion of ways in which meaningful units can be identified in corpora, the

paper will focus on a selection of common phraseological items in linguistic book review language, and investigate how specific (or how "local") these items are for the type of language under analysis and whether the identified local patterns are connected to local, text-type specific meanings. It will conclude with a few thoughts on "local grammars" and recommendations for future research in phraseology and academic discourse.

2. Taking a neo-Firthian approach to academic writing

The context of the analysis reported on in this paper is a large-scale corpus study of academic discourse. Central aims of the study are to investigate how meanings (in particular evaluative meanings) are created in academic writing in the discipline of linguistics, and to develop a local lexical grammar of book review language. The approach taken in the larger-scale study and described in the present paper is *neo-Firthian* in that it picks up some central notions developed and used by Firth (and his pupil Sinclair) and uses new software tools and techniques which lend themselves to investigating these notions but which Firth did not have at his disposal. The notions discussed here are "restricted language" (e.g. Firth [1956] 1968a), "collocation" (e.g. Firth [1957] 1968c; Sinclair 1991), "unit of meaning"/"meaning-shift unit" (Sinclair 1996, 2007 personal communication), "lexical grammar" (e.g. Sinclair 2004) and "local grammar" (e.g. Hunston and Sinclair 2000).

2.1. The discourse of linguistics as a "restricted language"

In the following, I will report on an analysis of a subset of the written English discourse among linguists regarded as a global community of practice. This type of discourse, the discourse of linguistics, is only one of the many types of specialized discourses that are analyzed by researchers in corpus linguistics and EAP (English for Academic Purposes). In Firthian terms, all these specialized discourses constitute "restricted languages".

As Léon (2007: 5) notes, "restricted languages ... became a touchstone for Firth's descriptive linguistics and raised crucial issues for early sociolinguistics and empiricist approaches in language sciences". Firth himself states that "descriptive linguistics is at its best when dealing with such [restricted] languages" (Firth 1968a: 105–106), mainly because the focus on limited systems makes the description of language more manageable. A

restricted language can be defined as the language of a particular domain (e.g. science, politics or meteorology) or genre that serves "a circumscribed field of experience or action and can be said to have its own grammar and dictionary" (Firth [1956] 1968b: 87). That means that we are dealing with a subset of the language, with "a well defined limited type or form of a major language, let us say English" (Firth 1968a: 98). A restricted language thus has a specialized grammar and vocabulary, "a *micro-grammar* and a *micro-glossary*" (Firth 1968a: 106, emphasis in original). An alternative concept to restricted language would be that of *sublanguage*. Sublanguage is a term used by Harris (1968) and Lehrberger (1982) to refer to "subsets of sentences of a language" (Harris 1968: 152) or languages that deal with "limited subject matter" and show a "high frequency of certain constructions" (Lehrberger 1982: 102). The concept of sublanguage also occurs in modern corpus-linguistic studies, for example in a study on the language of dictionary definitions by Barnbrook who considers the concept "an extremely powerful approach to the practical analysis of texts which show a restricted use of linguistic features or have special organisational properties" (Barnbrook 2002: 94). I will now turn to looking at the language of academic book reviews (a language of a particular domain with its own lexical microgrammar) and at some typical constructions in this sublanguage or restricted language.

The restricted language I am dealing with here is captured in a 3.5-million word corpus of 1,500 academic book reviews published in *Linguist List* issues from 1993 to 2005: the *Book Reviews in Linguistics Corpus* (henceforth BRILC). The language covered in BRILC constitutes part of the discourse of linguistics (in an English-speaking world). BRILC mirrors how the global linguistic research community discusses and assesses publications in the field. For a corpus of its type, BRILC is comparatively large, at least by today's standards, and serves well to represent the currently common practice in linguistic review writing. However, the corpus can of course not claim to be representative of review writing in general, and certainly not of academic discourse in its entirety, but it helps to provide insights into the language of one particular discourse community: the community of a large group of linguists worldwide.

2.2. The identification of meaningful units in a corpus of linguistic book reviews

Continuing Sinclair's search for units of meaning, the question I would like to address here is: How can we find meaningful units in a corpus? Or, more specifically (given that BRILC contains a particularly evaluative type of texts), how can we find units of *evaluative* meaning in a corpus? Evaluation, seen as a central function of language and broadly defined (largely in line with Thompson and Hunston 2000) as a term for expressions of what stance we take towards a proposition, i.e. the expression of what a speaker or writer thinks of what s/he talks or writes about, comes in many different shapes, which implies that it is not easy to find it through the core means of corpus analysis (doing concordance searches or word lists and keyword lists). As Mauranen (2004: 209) notes, "[i]dentifying evaluation in corpora is far from straightforward. ... Corpus methods are best suited for searching items that are identifiable, therefore tracking down evaluative items poses a methodological problem". On a similar note, Hunston (2004: 157) states that "the group of lexical items that indicate evaluative meaning is large and open", which makes a fully systematic and comprehensive account of evaluation extremely difficult. In fact, the first analytical steps I carried out in my search for units of evaluative meaning in BRILC (i.e. the examination of frequency word lists and keyword lists, see Römer 2008) did not yield any interesting results which, at that point in the analysis, led me to conclude that words are not the most useful units in the search for meaning ("the word is not enough", Römer 2008: 121) and that we need to move from word to phrase level. So, instead of looking at single recurring words, we need to examine frequent word combinations, also referred to as collocations, chunks, formulaic expressions, n-grams, lexical bundles, phrase-frames, or multi-word units. In Römer (2008), I have argued that the extraction of such word combinations or phrasal units from corpora, combined with concordance analysis, can lead to very useful results and helps to highlight a large number of meaningful units in BRILC.

In the present paper, however, I go beyond the methodology described in the earlier study in which I only extracted contiguous word combinations from BRILC (n-grams with a span of n=2 to n=7), using the software *Collocate* (Barlow 2004). I use two additional tools that enable the identification of recurring contiguous and non-contiguous sequences of words in texts: *kfNgram* (Fletcher 2002–2007) and *ConcGram* (Greaves 2005). Like *Collocate*, *kfNgram* generates lists of n-grams of different lengths (i.e.

combinations of n words) from a corpus, e.g. 3-grams like *as well as* or *the book is*. In addition to that, the program creates lists of so-called "phrase-frames" (short "p-frames"). P-frames are sets of n-grams which are identical except for one word, e.g. *at the end of*, *at the beginning of* and *at the turn of* would all be part of the p-frame *at the * of*. P-frames hence provide insights into pattern variability and help us see to what extent Sinclair's Idiom Principle (Sinclair 1987, 1991, 1996) is at work, i.e. how fixed language units are or how much they allow for variation. Examples of p-frames in BRILC, based on 5-gram and 6-gram searches, are displayed in figure 1.

```
it would be * to                  101   10
it would be interesting to        44
it would be useful to             14
it would be nice to               11
it would be better to             9
it would be possible to           5
it would be helpful to            5
it would be fair to               4
it would be difficult to          3
it would be necessary to          3
it would be good to               3

it * be interesting to            58    3
it would be interesting to        44
it will be interesting to         8
it might be interesting to        6

it * be interesting to see        33    3
it would be interesting to see    23
it will be interesting to see     7
it might be interesting to see    3
```

Figure 1. Example p-frames in BRILC, together with numbers of tokens and numbers of variants (*kfNgram* output)

Together with the types and the token numbers of the p-frames, *kfNgram* also lists how many variants are found for each of the p-frames (e.g. 10 for *it would be * to*). The p-frames in figure 1 exhibit systematic and controlled variation. The first p-frame (*it would be * to*) shows that, of a large number of possible words that could theoretically fill the blank, only a small set of (mainly positively) evaluative adjectives actually occur. In p-frames two

and three, modal verbs are found in the variable slot; however not all modal verbs but only a subset of them (*would*, *will*, *might*).

ConcGram allows an even more flexible approach to uncovering repeated word combinations in that it automatically identifies word association patterns (so-called "concgrams") in a text (see Cheng, Greaves and Warren 2006). Concgrams cover constituency variation (AB, ACB) and positional variation (AB, BA) and hence include phraseological items that would be missed by *Collocate* or *kfNgram* searches but that are potentially interesting in terms of constituting meaningful units. Figure 2 presents an example of a BRILC-based concgram extraction, showing constituency variation (e.g. *it would be very interesting*, *it should also be interesting*).

```
67  pon are backward anaphora, and it would be interesting to see how his theory can
68  spective of grammaticalisation it would be very interesting to have a survey of the
69   semantic transparency; again, it would be very interesting to see this pursued in
70  from a theoretical standpoint, it would be very interesting to expand this analysis
71  r future research, noting that it would be especially  interesting to follow the
72  ift from OV to VO in English.  It would be particularly interesting to see if this
73  ook as exciting as I had hoped it might be, although Part 4 was quite interesting,
74  d very elegantly in the paper, it would be interesting to discuss the
75  s in semantics. In my opinion, it would be interesting to see how this ontological
76  oun derivatives are discussed, it would be interesting at least to mention verbal
77  felt most positively" (p. 22). It should be noted that some interesting results
78  rs also prove a pumping lemma. It woul!  d be interesting to see further
79  pear in Linguist List reviews: it wouldn't be very interesting, I didn't make a
80   is given on this work, though it seems to be very interesting for the linguist's
81   and confined to the endnotes. It would also be interesting to set Hornstein's view
82  n of a book title was omitted. It would also be interesting to see if some of the
83  erative work on corpora. Maybe it would also be interesting to test the analyses in
84   second definition. Of course, it would also be interesting to find out that
85  ub-entries, for instance). So, it should also be interesting to find, among the
86  olved in dictionary-making and it should also be interesting to all dictionary
87  n a constituent and its copy.  It might however be interesting to seek a connection
88  ity and their self-perception. It might prove to be interesting to compare the
89  iteria seem fairly reasonable. It would, however, be interesting to study the
90  is, rhetoric, semantics, etc.  It would certainly be very interesting to see what
91  nages to carry out the action. It would most certainly be interesting to look at
92  CTIC THEORY" by Alison Henry). It seems to me that it would be  interesting to
```

Figure 2. Word association pattern (concgram) of the items *it* + *be* + *interesting* in BRILC (*ConcGram* output; sample)

All three tools (*Collocate*, *kfNgram* and *ConcGram*) can be referred to as "phraseological search engines" as they facilitate the exploration of the phraseological profile of texts or text collections.

The extraction of n-grams (of different spans), p-frames and concgrams was complemented by manual filtering of the output lists and extensive concordancing of candidate phraseological items. These semi-automatic BRILC explorations resulted in a database of currently a little over 800 items (i.e. types) of evaluative meaning. Part of these items are inherently evaluative (e.g. *it is not clear*, *wonderful*, or *a lack of*), while others appear "neutral" in isolation but introduce or frame evaluation (e.g. *at the same*

time or *on the one hand*). This type of implicit or "hidden" evaluation is much more pervasive than we would expect and will be focused on in the remainder of the paper. In the next section, we will look at items that prepare the ground for evaluation to take place and examine their use in linguistic book reviews. The items that will be discussed are all frequent in BRILC and appeared at the top of the n-gram and p-frame lists.

3. Uncovering the phraseological profile of linguistic book reviews

3.1. Central patterns and their meanings

Before I turn to some of the high-frequency n-grams from my lists and their use in BRILC, I would like to look at an item that came up in a discussion I had about evaluation with John Sinclair (and that is also quite common in BRILC, however not as common as the other items that will be described here). In an email to me, he wrote: "Re evaluation, I keep finding evaluations in what look like "ordinary" sentences these days. ... I came across the frame "the – – lies in – –"" (Sinclair 2006, personal communication). I think *lies in* is a fascinating item and I am very grateful to John Sinclair for bringing it up. I examined *lies in* in my BRILC data and found that gap 1 in the frame is filled by a noun or noun group with evaluative potential, e.g. *the main strength of the book* in example (1). Gap 2 takes a proposition about action, usually in the form of a deverbal noun (such as *coverage*), which is pre-evaluated by the item from the first gap.

(1) *The main strength of the book lies in its wide coverage of psycholinguistic data and models ...*

This is a neat pattern, but what type of evaluation does it mainly express? An analysis of all instances of *lies in* in context shows that 16 out of 135 concordance lines (12 %) express negative evaluation; see examples in (2) and (3). We find a number (27.8 %) of unclear cases with "neutral" nouns like *distinction* or *difference* in gap 1 (see examples [4] and [5]), but most of the instances of *lies in* (80, i.e. 60.2 %) exhibit positive evaluation, as exemplified in (1) and (6). The BRILC concordance sample in figure 3 (with selected nouns/noun groups in gap 1 highlighted in bold) and the two *ConcGram* displays of word association patterns in figure 4 serve to illustrate the dominance of positively evaluative contexts around *lies in*. This means that a certain type of meaning (positive evaluation) is linked to the

lies in pattern. In section 3.2 we will see if this is a generally valid pattern-meaning combination or whether this combination is specific to the restricted language under analysis.

(2) *The obvious defect of such an approach lies in the nature of polysemy in natural language.*

(3) *Probably, the only tangible limitation of the volume lies in some typographical errors ...*

(4) *The main difference lies in first person authority ...*

(5) *This distinction lies in the foregrounded nature of literary themes.*

(6) *The value of this account lies in the detail of its treatment of the varying degrees and types of givenness and newness relevant to these constructions.*

```
89        outstanding contribution made by Saussure lies in his theory of general linguisti
90         tions. Kennedy concludes that a solution lies in maintaining a purely syntactic
91         e (to y)' (J.K's ex. (8a)). The solution lies in the exploitation of Generalized
92         mplex). Evidence for the above statement lies in the following linguistic facts:
93         s of word geography is a task that still lies in future'' (p. 405). In ''Diachro
94         the scope of the term. Hinkel's strength lies in the fact that she led her resea
95         entation is convincing, and its strength lies in that it concentrates on one lan
96         n book. As a textbook, its main strength lies in the presentation of the details
97          curious aspect of this agreement system lies in the fourth available agreement
98          biological systems and semiotic systems lies in the key concepts of instantiati
99         opsis, a major strength of this textbook lies in the integration of essential se
100        ish linguistic history through its texts lies in part with the wealth of textual
101        selective loss. One explanation for that lies in the hypothesis that identificat
102        of the strengths of "Language in Theory" lies in offering an opportunity for dis
103        facing the author of a work such as this lies in where to set the limits of scho
104        s covert. Suranyi's explanation for this lies in the nature of the features at t
105        he importance of providing this training lies in the fact that simultaneous inte
106        eculiar trait of the hymn as a text type lies in "the degree of 'openness' of te
107        iscernible stress. Aguaruna's uniqueness lies in the following two properties ha
108        s true for the present volume. Its value lies in the fact that we can select fro
109         related fields of study. Its true value lies in its compact though penetrating
```

Figure 3. BRILC concordance sample of *lies in*, displaying predominantly positive evaluation

```
1  The strength, then, of The Korean Language, lies in its encyclopedic breadth of cover
2      The main strength of this book probably lies in the fact that it incorporates int
3      A particular strength of Jackson's book lies in its relevant biographical informa
4  synopsis, a major strength of this textbook lies in the integration of essential sema
5 ransformations. The strength of this chapter lies in the discussion where the authors
6    ASSESSMENT The main strength of this book lies in the personal testimonies and stor
7        SUMMARY The main strength of the book lies in its wide coverage of psycholingui
8 rgumentation is convincing, and its strength lies in that it concentrates on one langu
9       book. As a textbook, its main strength lies in the presentation of the details.
10 s the scope of the term.  Hinkel's strength lies in the fact that she led her researc
```

```
1 the preface that the    value of the reader lies in bringing together work from VARIOU
2  and in my view the main value of the paper lies in the mono- and multi-factorial anal
3that is terribly new in this book; its value lies rather in how it selects, organizes a
4  there is an intellectual value in exposing lies and deceptions, and here I think even
5 addressed. The value of his    contribution lies in the realization of the power imbed
6  startling claim. The value of this account lies in the detail of its treatment of the
7 enge, a further added value of this chapter lies in the close link with Newerkla's ch
8 ins strong"(471).  The value of this volume lies a) in its bringing together in one pl
9 syntacticians.  The real value of this book lies in its treatment of the larger issues
10 al related fields of study. Its true value lies in its compact though penetrating dis
11 not an easy read.  Despite this, its value lies in how it still manages to demonstrat
12 lds true for the present volume. Its value lies in the fact that we can select from t
13 the compound prosodic word), and its value lies mainly in demonstrating how some rece
```

Figure 4. Word association patterns (concgrams) of the items *lies* + *in* + *strength* and *lies* + *in* + *value* in BRILC (*ConcGram* output; sample)

Let us now take a closer look at three items from the frequency-sorted n-gram and p-frame lists*: at the same time, it seems to me* (*it seems to* *) *and on the other hand*. In linguistic book review language as covered in BRILC, *at the same time* mainly (in 56 % of the cases) triggers positive evaluation, as exemplified in (7) and in the concordance sample in figure 5. With only 5 % of all occurrences (e.g. number [8]), negative evaluation is very rare. In the remaining 39 % of the concordance lines *at the same time* is used in its temporal sense, meaning "simultaneously" (not "also"); see example (9).

(7) *Dan clearly highlights where they can be found and at the same time provides a good literature support.*

(8) *At the same time, K's monograph suffers from various inadequacies ...*

(9) *At the same time, some new words have entered the field ...*

```
142  e animal world. At the same time, it includes a careful and honest discussion of wh
143  n October 1997. At the same time, it is a state-of-the-art panorama of the (sub-)fi
144  e at times, but at the same time it is an almost encyclopedic source of information
145  s corpus data). At the same time, it is clear that not every author has been using
146  ghout the book. At the same time, it is flexible enough in organisation to allow th
147  ian and Hebrew; at the same time it is never the case that, say, accomplishments sh
148  ard Macedonian. At the same time, it is notable that Mushin's results are consisten
149    by _his_, but at the same time it is the subject of the Japanese predicate phrase
150  anguage change. At the same time, it may equally be used by college teachers who wi
151     of the base. At the same time, it must be no larger than one syllable (as discus
152      taste, and, at the same time, it provides the research with steady foundations
153  osition itself. At the same time, it was cliticised to an immediately following ver
154  ertainly rigid. At the same time, King claims, we can easily account for such utter
155  a events, while at the same time leading to interesting questions about the often
156  ge history, but at the same time maintains an engaging and entertaining style throu
```

Figure 5. BRILC concordance sample of *at the same time*, displaying predominantly positive evaluation

The next selected item, *it seems to me*, prepares the ground for predominantly negative evaluation (281 of 398 instances, i.e. 70.5 %), as exemplified in (10) and the concordance sample in figure 6. Positive evaluation, as

shown in (11), is rare and accounts for only 4.9 % of all cases. About 24.6 % of the BRILC sentences with *it seems to me* constitute neutral observations, see e.g. (12).

(10) *Finally, it seems to me that the discussion of information structure was sometimes quite insensitive to the differences between spoken and written data.*

(11) *In general, it seems to me this book is a nice conclusion to the process started in the Balancing Act ...*

(12) *It seems to me that it is a commonplace that truth outstrips epistemic notions ...*

```
11  new. It seems to me, nevertheless, that there are some difficulties related to this
12  lem; it seems to me, rather, that precedence is always transitive; it is the particu
13  ail, it seems to me that a more explicit definition of word would be needed to handl
14  68). It seems to me that a high price has been paid in terms of numbers of categorie
15  ude, it seems to me that as for theoretical results, much more [...] should be said
16  ies. It seems to me that both fields would benefit from acting a little more like th
17  ath; it seems to me that Copper Island Aleut is not a good example of such process
18  per. It seems to me that in some cases this could lead M to certain misinterpretatio
19  ry). It seems to me that it would be interesting to examine such problems in a more
20  3c). It seems to me that M-S conjures up notions of abstract constructs that are not
21   Yet it seems to me that one can likewise make a strong case for claiming that espec
22  VIEW It seems to me that one of the central questions being analyzed in this book is
23  ary. It seems to me that some additional topics could have been incorporated into th
24  ded. It seems to me that such a term is used in more than one sense, having to do bo
25  lly, it seems to me to be a weakness of this approach that it will not easily handle
```

Figure 6. BRILC concordance sample of *it seems to me*, displaying predominantly negative evaluation

Finally, if we look at *on the other hand*, positive evaluation follows the 4-gram in only 8 % of the 567 BRILC examples, as in (13). Negative evaluations (54 %) and neutral observations (38 %) are considerably more frequent. This is illustrated in figure 7 and in examples (14) and (15) below.

(13) *Other chapters, on the other hand, provide impressively comprehensive coverage of the topics ...*

(14) *but on the other hand, it is obvious that the book under review fails in various regards to take into account major developments in research into Indian English over the last 25 years.*

(15) *Prepositional clauses, on the other hand, do not allow stranding.*

```
85  M., on the other hand, comes to the opposite conclusion on the same point. It would
86  on, on the other hand, concerns the marking of event sequences through lexical and s
87  ts, on the other hand, consider the legitimate explanations to be those that do not
88  on. On the other hand, context in CA is not a priori but something that emerges from
89  nd. On the other hand, corpus linguists who want to develop their own tailor-made so
90  t). On the other hand, C denies the existence of the notion of "subject" as a univer
91  un. On the other hand, C importantly neglects other hypotheses on the origin of pers
92  re, on the other hand, denote type shifted, generalized quantifier-like or ,>-type e
93  an, on the other hand, despite its importance in the United States, left very little
94  acy on the other hand develops more slowly, influenced by production ease, salience,
95  ns, on the other hand, do not contribute to the truth conditions of the utterance bu
96  R , on the other hand, do provide support for D, and O and L (2002), while finding f
97  en, on the other hand do seem to have such restrictions, lengthening such words and
98   B, on the other hand, does not view optimization of language very seriously. Instea
99      On the other hand, [...], articles are missing on subjects that FG did attend to
```

Figure 7. BRILC concordance sample of *on the other hand*, displaying examples of negative evaluation and neutral observation

3.2. Corpus comparison: How "local" are these patterns and meanings?

The items we have just analyzed clearly show interesting patterns and pattern-meaning relations. Their existence in BRILC alone, however, does not say much about their status as "local" patterns, i.e. patterns that are characteristic of linguistic book review language as a restricted language (in Firth's sense). In order to find out how restricted-language-specific the above-discussed phraseological items (*lies in*, *at the same time*, *it seems to me*, *on the other hand*) are, I examined the same items and their patterns and meanings in a larger reference corpus of written English, the 90-million word written component of the British National Corpus (BNC_written).

In a first step, I compared the frequencies of occurrence (normalized per million words, pmw) of the four items in BRILC with those in BNC_written. As we can see in table 1, all units of evaluative meaning are more frequent in BRILC than in BNC_written, which may not be all that surprising if we consider the highly evaluative type of texts included in BRILC. Moving on from frequencies to functions, the next step then involved an analysis of the meanings expressed by each of the phraseological items in BNC_written. For *lies in* I did not find a clear preference for one type of evaluation (as in BRILC). Instead, there was a roughly equal distribution of examples across the three categories "positive evaluation" (34.5 %), "negative evaluation" (32.5 %) and "neutral/unclear" (33 %). While negative evaluation was rather rare in the context of *lies in* in the book review corpus, the item forms a pattern with nouns like *problem* and *difficulty* in BNC_written, as the concordance samples in figure 8 show.

Table 1. Frequencies of phraseological items in BRILC and BNC_written

	BRILC	BNC_written
lies in	38 pmw	19 pmw
on the other hand	162 pmw	57 pmw
at the same time	100 pmw	73 pmw
it seems to me	19 pmw	5 pmw

```
360  s without saying;  my difficulty lies in knowing how defensible they are in the fo
361  s. The chief cause of difficulty lies in the fact that confessions are typically o
362  olved.  The practical difficulty lies in deciding how to value the external effects
363   on CD-ROM.  The real difficulty lies in the fact that CD-ROM can only process one
364  n Fig. 8.5.  A second difficulty lies in the uncertainty in our knowledge of the to
365  , on the contrary the difficulty lies in obtaining sufficient evidence to identify
366  ists the cause of the difficulty lies in an institution,  central planning, which c
367  opinion.  Part of the difficulty lies in the developments which have taken place in
368  d spellcheckers.  The difficulty lies in building real quality into the products. D
369  um, is apparent.  The difficulty lies in providing an adequate theoretical framewor
370   experimentally.  The difficulty lies in heating the fuel to temperatures of about
371  y extraordinary.  The difficulty lies in finding an acceptable implied limitation.
372  ilm. At present,  the difficulty lies in understanding how this relates &mdash if a
373.. mess of things.  The difficulty lies in convincing yourself of that!  If all is we

1007 e root of the innovation problem lies in a dilemma : &bquo Curriculum innovation re
1008 rations.  &equo The main problem lies in the amount of translation the software wil
1009 p to my Martin D-16.  My problem lies in the fact that I can &bquo t get the same o
1010 way from the house.  One problem lies in the fact that the space is considerably wi
1011 asonably good;  the only problem lies in reaching the RAM upgrade s lots, which are
1012 rly where the particular problem lies in the case of this ruler. Books about Mary t
1013 ould argue that the real problem lies in the fact that shares had been overvalued f
1014 rpose,  whereas the real problem lies in the adjustment of the model 's control lin
1015 equo. Perhaps the Met 's problem lies in the present state of museum affairs, where
1016 If Radiohead 's singular problem lies in the sheer obviousness of their line of att
1017 taining prose.  Here the problem lies in the generality of the terms &quot descript
1018 believes the root of the problem lies in a fault with the child 's immune cells in
1019 FAO) 1985).  Part of the problem lies in the fact that much of this produce is expo
1020 umour? &equo Part of the problem lies in his opening statement : &bquo Eighty seven
```

Figure 8. BNC_written concordance samples of *lies in*, displaying patterns of negative evaluation

For *at the same time* we also find a lower share of positive contexts in the BNC_written than in the BRILC data. While authors of linguistic book reviews use the item predominantly to introduce positive evaluation, this meaning is (with 9 %) very rare in "general" written English (i.e. in a collection of texts from a range of different text types). An opposite trend can be observed with respect to *it seems to me*. Here, positive contexts are much more frequent in BNC_written than in BRILC, where negative evaluation dominates (with 70.5 %; only 30 % of the BNC_written examples express negative evaluation). Finally, with *on the other hand* positive evaluation or a positive semantic prosody is (with 33 %) also much more common in BNC_written than in BRILC (see [16] and [17] for BNC_written examples). For book reviews, I found that *on the other hand* mostly introduces negative evaluation and that only 8 % of the BRILC

concordance lines express positive evaluation. These findings indicate that the examined patterns and their meanings are indeed quite "local", i.e. specific of the language of linguistic book reviews. Not only do we find certain phraseological items or patterns to occur with diverging frequencies across text types and to be typical of a particular kind of restricted language, we also observe that the same items express different meanings in different types of language.

(16) *Jennie on the other hand was thrilled when the girls announced wedding plans* <BNC_written: B34 914>

(17) *On the other hand, he at last gains well-deserved riches and a life of comfort.* <BNC_written: ADM 2192>

4. Concluding thoughts

Referring back to the groundbreaking work of John Firth and John Sinclair, this paper has stressed the importance of studying units of meaning in restricted languages. It has tried to demonstrate how a return to Firthian and Sinclairian concepts may enable us to better deal with the complex issue of meaning creation in (academic) discourse and how corpus tools and methods can help identify meaningful units in academic writing or, more precisely, in the language of linguistic book reviews. We saw that the identification of units of (evaluative) meaning in corpora is challenging but not a hopeless case and that phraseological search-engines like *Collocate*, *kfNgram* and *ConcGram* can be used to automatically retrieve lists of meaningful unit candidates for further manual analysis. It was found to be important to complement concordance analyses by n-gram, p-frame and concgram searches and to go back and forth between the different analytic procedures, combining corpus guidance and researcher intuition in a maximally productive way. In the analysis of high-frequency items from the meaningful unit candidate lists, it then became clear that a number of "innocent" n-grams and p-frames have a clear evaluative potential and that apparently "neutral" items have clear preferences for either positive or negative evaluation.

The paper has also provided some valuable insights into the special nature of book review language and highlighted a few patterns that are particularly common in this type of written discourse. One result of the study was that it probably makes sense to "think local" more often because the isolated patterns were shown to be actually very restricted-language-

specific. In a comparison of BRILC data with data retrieved from a reference corpus of written English (the written component of the British National Corpus), we found that not only the patterns but also the identified meanings for each of the patterns (and their distributions) are local. I would suggest that these local patterns be captured in a "local lexical grammar" which "is simply a logical extension of the concept of pattern grammar" (Hunston 1999) in that it, being text-type specific, covers the patterns that are most typical of the text type (or restricted language) under analysis and links these patterns with the most central meanings expressed in the specialized discourse. I think that a considerable amount of research on disciplinary phraseology still needs to be done, and see the development of local lexical grammars based on restricted languages as an important future task for the corpus linguist. These text-type specific grammars will help us get a better understanding of how meanings are created in particular discourses and come closer to capturing the full coverage of Sinclair's (1987) idiom principle.

Notes

1 I would like to thank the participants at the symposium on "Chunks in Corpus Linguistics and Cognitive Linguistics: In Honour of John Sinclair", 25–27 October 2007, at the University of Erlangen-Nuremberg for stimulating questions and suggestions after my presentation.

References

Barlow, Michael
2004 *Collocate 1.0: Locating Collocations and Terminology*. Houston, TX: Athelstan.

Barnbrook, Geoff
2002 *Defining Language: A Local Grammar of Definition Sentences*. Amsterdam: John Benjamins.

Biber, Douglas
2006 *University Language: A Corpus-based Study of Spoken and Written Registers*. Amsterdam: John Benjamins.

Biber, Douglas, Ulla Connor and Thomas A. Upton
2007 *Discourse on the Move: Using Corpus Analysis to Describe Discourse Structure*. Amsterdam: John Benjamins.

Bowker, Lynne and Jennifer Pearson
2002 *Working with Specialized Language: A Practical Guide to Using Corpora*. New York/London: Routledge.
Cheng, Winnie, Chris Greaves and Martin Warren
2006 From N-gram to Skipgram to Concgram. *IJCL* 11 (4): 411–433.
Connor, Ulla and Thomas A. Upton (eds.)
2004 *Discourse in the Professions: Perspectives from Corpus Linguistics*. Amsterdam: John Benjamins.
Firth, John R.
1968a Descriptive linguistics and the study of English. In: *Selected Papers of J. R. Firth 1952–59*, Frank Robert Palmer (ed.), 96–113. Bloomington: Indiana University Press. First published in 1956.
Firth, John R.
1968b Linguistics and translation. In: *Selected Papers of J. R. Firth 1952–59*, Frank Robert Palmer (ed.), 84–95. Bloomington: Indiana University Press. First published in 1956.
Firth, John R.
1968c A synopsis of linguistic theory. In: *Selected Papers of J. R. Firth 1952–59*, Frank Robert Palmer (ed.), 168–205. Bloomington: Indiana University Press. First published in 1957.
Fletcher, William H.
2002–07 *KfNgram*. Annapolis, MD: United States Naval Academy.
Gavioli, Laura
2005 *Exploring Corpora for ESP Learning*. Amsterdam: John Benjamins.
Greaves, Chris
2005 *ConcGram Concordancer with ConcGram Analysis*. HongKong: Hongkong University of Science and Technology.
Harris, Zellig S.
1968 *Mathematical Structures of Language*. New York: Interscience Publishers.
Hunston, Susan
1999 *Local Grammars: The Future of Corpus-driven Grammar?* Paper presented at the 32nd BAAL Annual Meeting, September 1999, University of Edinburgh.
Hunston, Susan
2004 Counting the uncountable: Problems of identifying evaluation in a text and in a corpus. In: *Corpora and Discourse*, Alan Partington, John Morley and Louann Haarman (eds.), 157–188. Bern: Peter Lang.

Hunston, Susan and John McH. Sinclair
2000 A local grammar of evaluation. In: *Evaluation in Text: Authorial Stance and the Construction of Discourse*, Susan Hunston and Geoff Thompson (eds.), 74–101. Oxford: Oxford University Press.

Hyland, Ken
2004 *Disciplinary Discourses: Social Interactions in Academic Writing*. Ann Arbor, MI: University of Michigan Press.

Lehrberger, John
1982 Automatic translation and the concept of sublanguage. In: *Sublanguage: Studies of Language in Restricted Semantic Domains*, Richard Kittredge and John Lehrberger (eds.), 81–106. Berlin: Walter de Gruyter.

Léon, Jaqueline
2007 From linguistic events and restricted languages to registers. Firthian legacy and Corpus Linguistics. *Henry Sweet Society Bulletin* 49: 5–25.

Mauranen, Anna
2004 Where next? A summary of the round table discussion. In: *Academic Discourse: New Insights into Evaluation*, Gabriella Del Lungo Camiciotti and Elena Tognini Bonelli (eds.), 203–215. Bern: Peter Lang.

Römer, Ute
2008 Identification impossible? A corpus approach to realisations of evaluative meaning in academic writing. *Functions of Language* 15 (1): 115–130.

Römer, Ute and Rainer Schulze (eds.)
2008 *Patterns, Meaningful Units and Specialized Discourses* (special issue of *International Journal of Corpus Linguistics*). Amsterdam: John Benjamins.

Sinclair, John McH.
1987 The Nature of the evidence. In: *Looking Up: An Account of the COBUILD Project in Lexical Computing*, John McH. Sinclair (ed.), 150–159. London: HarperCollins.

Sinclair, John McH.
1991 *Corpus, Concordance, Collocation*. Oxford: Oxford University Press.

Sinclair, John McH.
1996 The search for units of meaning. *Textus* IX (1): 75–106.

Sinclair, John McH.
2004 *Trust the Text: Language, Corpus and Discourse*. London: Routledge.

Thompson, Geoff and Susan Hunston
2000 Evaluation: An introduction. In: *Evaluation in Text: Authorial Stance and the Construction of Discourse*, Susan Hunston and Geoff Thompson (eds.), 1–27. Oxford: Oxford University Press.

Corpora

BNC *The British National Corpus*. Distributed by Oxford University Computing Services on behalf of the BNC Consortium. URL: http://www.natcorp.ox.ac.uk/.

BRILC *Book Reviews in Linguistics Corpus*. Compiled by the author of this paper.