

John Benjamins Publishing Company



This is a contribution from *Functions of Language 15:1*
© 2008. John Benjamins Publishing Company

This electronic file may not be altered in any way.

The author(s) of this article is/are permitted to use this PDF file to generate printed copies to be used by way of offprints, for their personal use only.

Permission is granted by the publishers to post this file on a closed server which is accessible to members (students and staff) only of the author's/s' institute.

For any other use of this material prior written permission should be obtained from the publishers or through the Copyright Clearance Center (for USA: www.copyright.com).

Please contact rights@benjamins.nl or consult our website: www.benjamins.com

Tables of Contents, abstracts and guidelines are available at www.benjamins.com

Identification impossible?

A corpus approach to realisations of evaluative meaning in academic writing*

Ute Römer

University of Michigan

Evaluation is a pervasive element in spoken and written language but its identification poses serious problems to linguistic researchers, especially when they are dealing with larger amounts of text which require the application of computer-assisted analytic techniques. This article explores ways of identifying items of evaluative meaning in a three million word corpus of linguistic book reviews, a text type that is particularly rich in expressions of positive and negative evaluation. It discusses whether it is at all possible to capture evaluation in a corpus in a systematic way and which analytic strategies may be most promising in the search for a larger set of meaningful patterns. The paper ends with a discussion of some unresolved issues in the area of evaluation research and sketches tasks for future activities in the field.

1. Introduction

Of the functions that language can have, evaluation appears to be a particularly prominent one, maybe even the most prominent one. Evaluation, defined in a broad sense and largely in line with Thompson and Hunston (2000: 5) as a term for expressions of what stance we take towards an entity or a proposition, i.e. expressions of what we think of what we talk or write about, seems to be everywhere. We find expressions of evaluative meaning in speech and in writing, in different text types, and in the language of different groups of people with different social or cultural backgrounds. We hear and talk about “unfair decisions” or “absolutely wonderful news” or read and write about “corruption in politics” or “amazing discoveries”.

The centrality of evaluation in language has recently come to be acknowledged in a number of studies from different linguistic subfields, mainly genre studies, pragmatics, discourse analysis, and text linguistics (see, for example, Aijmer 2005, Hunston and Sinclair 2000, Hyland 2000, Martin 2000, Mauranen 2004, Stubbs

2001, Swales 2004). Most of these studies are based on corpora of varying sizes and text types (that is, on records of “used” language, in Brazil’s 1995 sense, stored on the computer), which means that multiple instances of particular evaluative items such as *corruption* or *absolutely wonderful* can be retrieved automatically by means of appropriate software packages, and that usage patterns of such items can easily be analysed in concordances that show the respective items in context.

Evaluative items are explorable in corpora without problems once you have determined them. One crucial problem that corpus researchers who wish to investigate evaluation encounter, however, is how such items can be identified in the first place. It is of course not a very difficult task to list, off the top of our heads, a handful of words and phrases (or maybe more) that express positive or negative evaluation, but “[i]dentifying evaluation in corpora is far from straightforward” (Mauranen 2004: 209; my emphasis). In addition, as rightly noted by Hunston (2004: 157), “the group of lexical items that indicate evaluative meaning is large and open”, and this is what makes a fully systematic and comprehensive account of evaluation extremely difficult, perhaps even impossible.

The aim of this article is to explore methods of tracing evaluative items in a corpus of academic writing, a particularly evaluative type of language (see the contributions in Del Lungo Camiciotti and Tognini Bonelli 2004, and Tognini Bonelli and Del Lungo Camiciotti 2005). The focus will be on evaluative expressions, rather than on acts of evaluation which do not always contain evaluative words/phrases but may still be understood as criticism or praise from the wider linguistic and/or situational context. I will assess the prospects for identifying evaluation in language by starting my investigations from the corpus itself, rather than from a single selected text or from a pre-determined list of expressions. However, before turning to strategies of tracking down items of evaluative meaning, I will first tackle the question of why evaluation is so prevalent in academic discourse and then briefly describe the composition of the corpus that has been compiled for the present analysis.

2. Academic discourse and evaluation

It is probably not a coincidence that a number of recent studies on evaluation focus on text types that capture the spoken and written discourse of academics (cf. e.g. Bamford 2005, Conrad and Biber 2000, Diani 2004, Fortanet Gómez 2004, Freddi 2005, Hunston 1993 and 2005, Hyland 1999, Shaw 2004, Swales 2004). Academic genres such as the research article, the university lecture, and the book review have proven to be particularly rich in expressions of speaker/writer attitudes. In their interactions with other scholars and novices, be they spoken or written, researchers make prolific use of a range of words and phrases that denote their approval

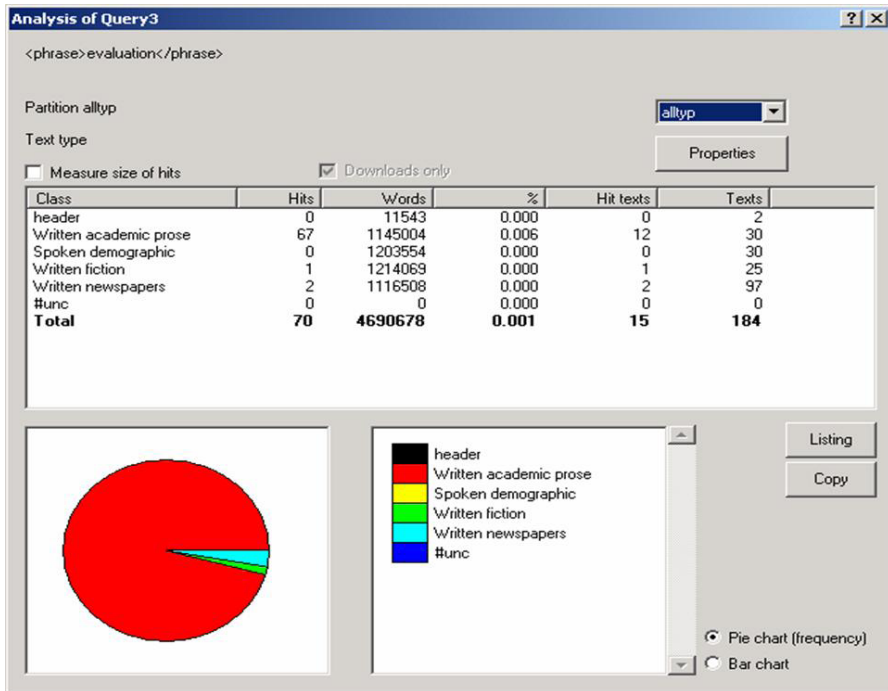


Figure 1. The distribution of the item *evaluation* across registers in BNC-baby

or disapproval of earlier work in their field of expertise, thus paving the way for advances to take place. This constant assessment of other people's contributions to the field is perhaps the main difference between academic and non-academic discourse. This was noted by Wolfgang Teubert during a round table discussion at the first "Evaluation in Academic Discourse" conference in Pontignano, Italy, in June 2003. An explanation Teubert gave for the distinct characteristics of academic communication and its evaluative nature was that in academic discourse we negotiate our views of something with other members of the community, whereas in 'normal' non-academic discourse we talk about reality.

Interesting in the context of discussions about the highly evaluative nature of academic discourse is the observed prevalence of the lexical item *evaluation* and related forms (e.g. *evaluate*, *evaluations*) in written academic English (see also Oakey 2005). Figure 1 displays the distribution of the search word *evaluation* across registers in BNC-baby, a corpus of four million words extracted from the British National Corpus and split up into four sub-sections of equal size: "written academic prose", "spoken demographic", "written fiction", and "written newspapers" (cf. Burnard 2003). As the numbers in the "hits" column in Figure 1 show, the clear majority of all instances of *evaluation* in the corpus (67 out of 70) comes from the "written academic prose" section of BNC-baby. This indicates that

evaluation is a topic of central importance in academic writing and that the members of the academic discourse community are certainly aware of this fact. For my corpus-driven explorations on identifying evaluation, I have hence selected a text type from the area of academic discourse: the academic book review.

3. The empirical basis: BRILC — a corpus of linguistic book reviews

One problem that I see with research on evaluation in language is that the majority of studies are based on comparatively small collections of data. By “small” I mean corpora of sizes between several thousand and several hundred thousand words that only consist of a limited number of texts. Such studies can no doubt lead to important results and provide valuable insights into evaluation, but they do not facilitate quantifications of more complex lexical-grammatical patterns and thus cannot enable really systematic accounts of the topic.

Since the central aim of the project reported on in this article is to work towards a maximally comprehensive lexicon of evaluative expressions which are quantified and systematically described, I decided to compile a larger corpus that is to a high degree representative of the text type it contains. This corpus, the Book Reviews In Linguistics Corpus (BRILC), contains, as its name implies, reviews of publications from the discipline of linguistics. BRILC is a collection of reviews that were all published online in issues of the Linguist List discussion forum between 1993 and 2005.¹ It is designed as a monitor corpus, which means that the corpus has no pre-specified upper limit; new texts are added successively as they become available. The findings discussed in the following sections go back to the July 2005 version of BRILC, which consisted of 1,330 book reviews that made up slightly over three million words. In the meantime, 170 reviews have been added, and the corpus now has a size of 3,502,674 words (October 2005). 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. This 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.

4. Attempts to identify evaluation in BRILC

Let us now take up the challenge formulated in the introduction, namely to identify a large number of items which express evaluative meaning in a corpus, in our

case BRILC. I will hence attempt to do what, according to Hunston (2004: 158), “many writers on the topic avoid [...] altogether.”

Since what is aimed for is a systematic rather than an eclectic account of evaluation, the preferred evaluation detection strategy cannot just be to simply read a few of the 1,330 BRILC reviews, compile a list of evaluative language items, and analyse the distribution of a small set of, let’s say, negative evaluative adjectives (as done in Römer 2005). Instead, methods have to be employed which provide insights on a higher level and help to uncover a larger number of linguistic means, the function of which is to evaluate. What I intend to carry out is an automatic instead of a manual analysis of the selected corpus.² The following sections will discuss the analytic steps that have been carried out in the exploration of evaluative items in BRILC.

4.1 Frequency wordlists and keyword analysis

Since the book review constitutes a particularly evaluative genre, I assumed that my book review corpus should contain a particularly evaluative lexis, and that it might be fruitful to examine which words are particularly or unexpectedly frequent in BRILC. I hence started the process of tracing evaluative items with the compilation of a frequency wordlist and a keywordlist based on BRILC.

The top 60 items of the BRILC frequency wordlist are displayed in Table 1. Perhaps contrary to expectations, none of the 60 most frequent words in the corpus is in itself evaluative (e.g. *good* or *unclear*) or can be said to provide an obvious link to an expression of evaluation (like *however* or *somewhat* would).³ The items that we find among the most common content words in BRILC, such as *language* (rank 15), *English* (26), *languages* (34), and *linguistic* (45), clearly point to the concerns of the books under review rather than to the reviewers’ praise or criticism.

To find out which words in the review corpus are particularly “key”, i.e. more frequent than we would expect them to be on the basis of general language use, the BRILC frequency wordlist was compared to a frequency wordlist from a much larger and less specialised corpus of English writing, the 90 million word written part of the British National Corpus (BNC_written). The software used to retrieve these lists was Mike Scott’s *WordSmith Tools* (version 3.0; Scott 1999). Table 2 shows the 60 highest-ranking keywords in BRILC, in order of their keyness values. The only words in this list that may be said to be evaluative or point to expressions of evaluative meaning are *cannot* (rank 40) and *argues* (rank 50). The remaining 58 items look rather ‘innocent’ evaluation-wise and mainly refer to the topics and issues dealt with in the books under review (“language”, “linguistics”, “discourse”, “syntactic”, “grammar”, etc.).

Table 1. The 60 most frequent words in BRILC

N	Word	Frequency	N	Word	Frequency
1	THE	195,945	31	I	8,053
2	OF	132,433	32	P	8,034
3	AND	91,087	33	MORE	7,709
4	IN	79,105	34	LANGUAGES	7,675
5	TO	66,060	35	CAN	7,500
6	A	61,640	36	THEIR	7,131
7	IS	49,951	37	THEY	7,119
8	THAT	37,817	38	BETWEEN	7,006
9	AS	28,945	39	ALSO	6,919
10	FOR	24,407	40	SOME	6,766
11	ARE	23,052	41	OTHER	6,728
12	THIS	21,369	42	SUCH	6,669
13	ON	21,161	43	THESE	6,657
14	WITH	20,766	44	TWO	6,545
15	LANGUAGE	20,554	45	LINGUISTIC	6,216
16	BY	17,550	46	HAS	6,068
17	BE	17,265	47	HE	6,014
18	IT	16,505	48	THERE	5,878
19	NOT	14,451	49	E	5,734
20	WHICH	14,183	50	ANALYSIS	5,724
21	FROM	13,826	51	DIFFERENT	5,414
22	AN	13,534	52	ITS	4,942
23	CHAPTER	12,973	53	ALL	4,894
24	OR	11,320	54	HOW	4,835
25	BOOK	8,959	55	WHAT	4,768
26	ENGLISH	8,230	56	ONLY	4,698
27	ONE	8,139	57	FIRST	4,693
28	AT	195,945	58	USE	4,568
29	HAVE	132,433	59	THEORY	4,528
30	BUT	91,087	60	THAN	4,489

Unfortunately, with respect to identifying items of evaluative meaning in the corpus, the first two, essentially word-based, analytic steps did not lead to the results I was hoping for. What I inferred at this point was that, after all, the word may not be the ideal unit of study when our main interest is in meaning.

Table 2. Keywords in BRILC (reference corpus: BNC_written)

N	Word	Keyness	N	Word	Keyness
1	LANGUAGE	90,361.50	31	COGNITIVE	8,695.70
2	CHAPTER	53,066.60	32	PHONOLOGY	8,611.20
3	LANGUAGES	36,811.80	33	ACQUISITION	8,588.10
4	LINGUISTIC	30,603.20	34	PHONOLOGICAL	8,390.90
5	BOOK	25,792.40	35	STRUCTURE	8,376.90
6	ENGLISH	22,782.80	36	SPEECH	8,359.20
7	LINGUISTICS	19,812.20	37	DISCUSSION	8,019.80
8	AUTHOR	15,605.00	38	PP	7,961.60
9	P	14,964.50	39	MORPHOLOGY	7,887.20
10	OF	14,348.80	40	CANNOT	7,595.00
11	DISCOURSE	14,348.10	41	VOLUME	7,582.20
12	LEXICAL	14,248.90	42	SPEAKERS	7,482.30
13	ANALYSIS	13,803.60	43	WH	7,115.80
14	SYNTACTIC	13,547.00	44	THEORETICAL	6,932.80
15	CHAPTERS	13,165.40	45	ISSUES	6,889.30
16	SEMANTIC	13,056.20	46	TEXT	6,789.70
17	GRAMMAR	12,727.80	47	DATA	6,638.80
18	IS	12,636.70	48	HTTP	6,511.60
19	THEORY	11,885.10	49	MEANING	6,421.40
20	VERB	11,871.30	50	ARGUES	6,360.20
21	AUTHORS	11,778.70	51	WORD	6,355.90
22	LINGUIST	11,627.80	52	HTML	6,282.00
23	VERBS	10,738.10	53	BILINGUAL	6,140.20
24	SYNTAX	10,584.90	54	LINGUISTLIST	6,115.00
25	DISCUSSES	9,928.90	55	IN	6,111.20
26	CORPUS	9,589.70	56	TRANSLATION	6,095.40
27	SEMANTICS	9,206.80	57	G	5,867.40
28	E	8,966.80	58	FEATURES	5,833.20
29	CONSTRUCTIONS	8,884.40	59	SPANISH	5,832.90
30	GRAMMATICAL	8,727.10	60	PAPERS	5,757.40

4.2 ‘The word is not enough’ — Looking at larger units

The rather disappointing results of the wordlisting steps indicate that the word is perhaps not the most useful unit of analysis in the search for evaluative meanings. I therefore decided to expand the scope of analysis and look at word combinations instead of single words, hoping that the extraction of larger recurring units from the corpus would highlight at least a few evaluative items.

Table 3. Most frequent 5-grams in BRILC

N	5-gram	Type	N	5-gram	Type
1	at the end of the	S	21	it should be noted that	E
2	the book is divided into	S	22	the chapter concludes with a	S
3	the second part of the	S	23	the relationship between language T and	T
4	on the other hand the	E	24	second part of the book	S
5	at the end of each	S	25	on the part of the	E
6	the end of the book	S	26	the title of the book	S
7	on the basis of the	E	27	the author points out that	E
8	the point of view of	E	28	it seems to me that	E
9	from the point of view	E	29	contribution to the study of	E
10	on the one hand and	E	30	and at the same time	E
11	the rest of the book	S	31	at the same time the	E
12	the first part of the	S	32	it would be interesting to	E
13	at the beginning of the	S	33	the book under review is	E
14	the end of each chapter	S	34	the papers in this volume	S
15	with a discussion of the	E	35	the book as a whole	S
16	due to the fact that	E	36	an important contribution to the	E
17	the end of the chapter	S	37	in the first part of	S
18	in the second part of	S	38	the rest of the chapter	S
19	English as a second language	T	39	in the light of the	E
20	book is a collection of	S	40	does not seem to be	E

To retrieve such recurring units from BRILC, I used *Collocate* (Barlow 2004), a program for finding collocations in a corpus. The *Collocate* command that I selected from a range of functions, the “full extract — n-gram” command, works along the lines of a simple word frequency list, but instead of listing the most common words in a corpus, it finds the most common bigrams, trigrams, 4-grams, 5-grams, etc., depending on how “n” is specified in the search options window. In the analysis of BRILC, I defined spans ranging from $n=2$ to $n=7$, i.e. the output were lists of bigrams (such as “of the” and “the book”) to 7-grams (e.g. “in the second part of the chapter”), and found the displays of 4-grams and 5-grams to be the most revealing in my search for meaningful units. The following combinations, with frequencies of occurrence in BRILC between 473 and 118, were found among the top items in the 4-gram list: *on the other hand*, *as well as the*, *at the same time*, *the fact that the*, *on the one hand*, *that there is no*, *it would have been*, *that there is a*, and *it is not clear*. Unlike the top-ranking items in the (key)wordlists, these combinations have a strong evaluative potential, which may not always be readily visible (as is the case with *as well as the* or *on the other hand*) but which becomes apparent when we look at the wider lexical context that these 4-grams occur in.

Also, as cross-checks with the written part of the BNC show, these items are typical of book review language. The highly frequent 4-gram *it is not clear*, for example, occurs 36.1 times per one million words in BRILC but only 4.5 times per one million words in BNC_written.

If we turn to the most common 5-grams, we get a similar picture. Table 3 shows the 40 most common 5-word combinations in the entire corpus. Apart from two topic-related 5-grams in the top-40 list ("English as a second language" and "the relationship between language and"; both labelled "T"), all items can be classified as either "structure-related", meaning that they refer to structural aspects of the work under review (see label "S" in Table 3) or "evaluative", which means that they either express evaluation themselves or appear in highly evaluative contexts (labelled "E"). When we review a book, it is obviously very common to provide detailed information about its composition and internal structure, and we might ask how much evaluative meaning hides behind a reviewer's comments on structural aspects of a book. This will be an interesting topic for future explorations of BRILC. For the time being, however, I will concentrate only on a few of the abovementioned 4-grams and on some of the 5-grams in Table 3 that have been labelled "evaluative".

4.3 Isolating meaningful patterns

In order to determine patterns of evaluative meaning in BRILC and to extract instances of evaluation from the corpus, concordances of selected high-frequency 4-grams and 5-grams were compiled and carefully examined.

user of this book it would have been extremely helpful to be able to look up all
 ing the help files. It would have been extremely useful if the book had provided a
 r and section, but it would have been far more useful in the chapters to point di
 ce" (p.134). Here, it would have been good if Harris had taken a more critical po
 ves are very brief. It would have been good if the authors and editors had follow
 of linguists. Hence it would have been good to include work by linguists, not onl
 ys. First, I think it would have been good to include some discussion of the "phoi
 ithin the region. It would have been good to have noted its value in this earlie
 familiar with LFG it would have been helpful if the book contained a brief intro
 ts are progressing. It would have been helpful to have an additional chapter or st
 sion. For example, it would have been helpful if, in the discussion of appositiv
 1 chapter 4; yet, it would have been helpful in chapter 3 for a better comprehen
 ade by bilinguals. It would have been helpful to see more in-depth discussion on
 , for some readers, it would have been helpful to explain the formalisms used for
 ry attractive, but it would have been helpful (at least to this reviewer) if he
 ough the later has. It would have been helpful for the readers if the former term
 racter of the book, it would have been helpful to include connecting passages that
 itively means, but it would have been helpful to provide a fuller explanation bec
 s part of the case. It would have been helpful if the author had been more explic
 3 years of contact. It would have been interesting if Hammond had made some refer
 titude and memory. It would have been interesting to have included one or two stu
 took place in 1996, it would have been interesting that the author made some expl
 y very interesting, it would have been interesting to also compare with results in
 imited to English. It would have been interesting to include transcription exerc
 ne points, however, it would have been interesting to see how she would have incor
 sions" [p. 185]. - It would have been interesting to learn about the reception of
 iew. In particular, it would have been interesting to examine the combinability of
 ular structure. It would have been interesting to look at the individual data
 istic of this area, it would have been interesting to have examples from other co
 kornai (chapter 10) it would have been interesting to provide more details about 1

Figure 2. Extract from BRILC concordance of *it would have been*

ly by force. It should be noted that not all data are as clear as Postal makes the
 nevertheless, it should be noted that not every Nivkh clause is marked for such a h
 her things, it should be noted that not all minimalist work assumes chains as ing
 these units. It should be noted that one of the most disputable problems in gramma
 (p. 251-275). It should be noted that the name of certain chapters does not always
 s about SLA. It should be noted that the individual contributions can be read inde
 nown abroad. It should be noted that the book presupposes some elementary knowledg
 and grammars. It should be noted that the only corpora that have been directly used
 counted for. It should be noted that the description of semiotics is in no sense s
 ic) factors. It should be noted that the author talks only about the interpretatio
 research. It should be noted that the book, on the whole, does not attempt to p
 nevertheless, it should be noted that the two patterns are not alike. The causat
 eclination). It should be noted that the authors discuss negative accusative const
 rs. Finally, it should be noted that the majority of papers concentrate on the des
 ers'. Still, it should be noted that the standard assumption according to which 'I
 importantly, it should be noted that the alternative without interpretive parsing
 (p.709). It should be noted that the approach demonstrated in this chapter seem
 still, it should be noted that the author neglects a number of important sour
 conclusion, it should be noted that the book "Thinking Italian translation" is of
 e described. It should be noted that these ideas are very much similar with Y.D. A
 ealisations. It should be noted that this scheme clearly distinguishes between fun
 e study. It should be noted that this chapter contains some doubtful, vague and
 e moving on, it should be noted that this introductory chapter makes a more than a
 is omission. It should be noted that this omission is less visible with discussion
 sion, though it should be noted that what counts as an antonym of a given word may

Figure 3. Extract from BRILC concordance of *it should be noted that*

- (7) However, it may have been more helpful for readers unfamiliar with Bakhtin or dramaturgy, if these had been arranged closer together in the text.

The revised pattern, which accounts for the observed variation in the position of *would*, can hence be given as follows:

it would have been ADJ_{pos} to INF
 could if X (had) V-ed
 might
 may

A further evaluative pattern that is apparently typical of book review language is illustrated in the concordance extract in Figure 3. What we find exemplified here is an expansion of the frequent 5-gram *it should be noted that*. A search for *it * be noted that* shows that *should* can be replaced in this pattern by *must*, but these instances are rather rare. On the right hand of the string we most commonly find *the* (or sometimes *this*, *some*, or *not all*) followed by a noun (e.g. *book*, *author*) or noun group (e.g. *standard assumption* or *majority of papers*). Hence, the following pattern can be formulated:

it should be noted that the N
 must this
 some
 not all

While the combination *it would have been* is always part of a critical reviewer's comment in our data, the picture is less clear concerning the semantics of utterances introduced by *it should be noted that*. What we find in the concordance of *it * be noted that* are both instances of positive evaluation, as in (8), and of negative

evaluation, as in (9). A closer analysis of the concordance revealed, however, that in the large majority of instances, *it should/must be noted that* introduces negative evaluation (of over 80 examples only a handful were of the positive type). This pattern is apparently used as a means to emphasise criticism and can perhaps be said to function as an eyecatcher for the reader.

- (8) **It should be noted that** some interesting results based on the Corpus-driven framework are already available [...]
- (9) **It should be noted that** the author talks only about the interpretation of 'pronominal' anaphors, [...]

Examples of other typical patterns of book review language that either express or prepare the ground for evaluative meanings and that could be identified by means of the n-gram analysis described above, are given in (10) to (15). Concordance examinations show that *on the other hand* and *it seems to me (that)* are usually part of negative evaluative comments, whereas with *at the same time* we are a lot more likely to find praise than criticism. One of the rare counter examples in which *at the same time* introduces negative evaluation is given in (15).

- (10) **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.
- (11) **On the other hand**, the discussion says relatively little about the actual behavior of lexical items and about cases in which they are really used.
- (12) **It seems to me** that some additional topics could have been incorporated into the book without making it unmanageable, since it is quite short.
- (13) [...] **it seems to me** a bit unbalanced to devote an entire chapter, out of nine, to verbal number.
- (14) **At the same time**, I feel K does a good job of presenting the material in a form that should be accessible to readers who do not have a strong background in statistics.
- (15) **At the same time** this work can be criticized for attempting too much too soon.

In my examinations of concordances I moved on from an automatic computer-based to a mainly manual but in part computer-assisted (I still made use of some sorting and collocation-highlighting functions of the software) type of analysis. This way of approaching the data helped me to find additional items of evaluation in the left- and right-hand context of the initial search items. It can be expected

that the concordancing of these items is likely to highlight further evaluative patterns so that the overall set of patterns may grow steadily.

5. Conclusion and outlook

In the introduction to this article, the systematic identification of evaluation in a corpus has been described as an “extremely difficult, perhaps even impossible” task. I have tried to demonstrate that, even though some difficulties may have to be overcome, it is in fact possible to access a corpus in such a way that a significant number of evaluative expressions can be retrieved semi-automatically, and that these items can in turn be used as starting points for the identification of further expressions of evaluative meaning in the corpus. Of course, manual searches may highlight additional items of evaluation, but the suggested automatic extraction procedure (combined with extensive concordancing) leads to those expressions that are most frequent and most typical of the (specialised) text type under analysis.

It became clear in the process of corpus exploration that, although I was dealing with a text type that can be assumed to contain a particularly evaluative lexis, an analysis of frequent words and keywords in BRILC did not bring the expected (or hoped for) results. Very revealing, however, was the subsequent, essentially phraseology-driven, approach that centred on recurring n-grams in the corpus. This approach facilitated the isolation of patterns which express or introduce evaluative meanings and which are obviously typical of book review language. The results thus neatly confirm what Stubbs (2001: 215) states about evaluative meanings, namely that they “are conveyed not only by individual words, but also by longer phrases and syntactic structures”. Perhaps we need to go one step further and say that they are conveyed more commonly by longer phrases than by individual words.

Despite the fact that n-grams have proven to be extremely useful in identifying evaluation, it may not be wise to rely solely on them. A problem I see is that n-grams might not be flexible enough since they cannot account for pattern-internal variation. To work around this problem, I often used wildcard characters in the process of concordancing and, for example, searched for *it * be noted that* instead of *it should be noted that* which was the initially retrieved 5-gram. Another useful option is to compile concordances of shorter strings than are given in the examined n-gram lists (for instance to search for *it seems* and not for the longer *it seems to me* or *it seems to me that*) and then apply different sorting commands, in this case preferably right-sorting. This method certainly leads to the identification of interesting patterns, but it also involves considerably more manual analytic work.

No matter how promising our corpus-driven exploration of evaluative expressions in a particular type of academic discourse may appear, there is undoubtedly still a lot of work waiting to be done in this area. To answer the question posed in the title of this paper, I would say that the identification of evaluative language items is not impossible, but it is not without problems either (as also convincingly discussed by Hunston 2004). Future research on the topic will have to involve the examination of a much larger set of evaluative items and patterns than has been described here. As Stubbs (2001: 216) notes, “it is not yet clear how many lexical items and syntactic structures express evaluative meanings.” My explorations of BRILC indicate that the number of such items and structures is very large, at least in the language of linguistic book reviews, but we need more corpus research on evaluative meaning in different types of language to gain more certainty about the inventory and the distribution of items that speakers and writers use to express evaluation.

Apart from the work that still needs to be done on **identifying** evaluation, a considerable amount of work lies ahead on **classifying** the identified expressions (cf. also Mauranen 2004: 214). It will be an important task for the future to organise evaluation in some way and, ideally, to work towards a system or framework that can account for all, or at least a large number of evaluative items that we may find in a particular text type. In this context I would argue that, until we have collected much more evidence and analysed larger amounts of data, it may not be entirely safe to build such a framework. My suggestion is thus to postpone model-building for the time being, and to attack it as soon as we have made more progress on the identification side and gained further insights into the ways in which evaluative meanings are created in the language.

Notes

* I would like to thank my audiences at the “Evaluation and Text Types” conference in Augsburg, Germany, 22–23 July 2005, and at “Phraseology 2005 — The Many Faces of Phraseology” in Louvain-la-Neuve, Belgium, 13–15 October 2005, for stimulating questions and helpful comments after my presentations. I am also grateful to Nicholas Groom for insightful comments and helpful suggestions on an earlier draft of this article.

1. The reviews included in BRILC can be accessed via the Linguist List website at <http://www.linguistlist.org> (consulted: 01.11.05).
2. An example of a “manual” rather than an automatic analysis of a small corpus of book reviews is Suárez Tejerina’s (2005) study. In this study, which focuses on the macro-structure of the book review, the individual texts in the corpus were examined for instances of evaluation without the help of computer-based corpus-analytic techniques.

3. The item *but* may be considered a possible exception, since one of its uses is to signal a switch of evaluative direction, as illustrated by the following example (taken from BRILC): “This may indeed contribute to ease of reading, but it can also create confusion”.
4. For an account of comparable evaluative patterns and a draft of a local grammar of evaluation, see also Hunston and Sinclair (2000).

References

- Aijmer, K. (2005) Evaluation and pragmatic markers. In E. Tognini Bonelli and G. Del Lungo Camiciotti (eds.) *Strategies in academic discourse*. Amsterdam: Benjamins. 83–96.
- Bamford, J. (2005) Subjective or objective evaluation? Prediction in academic lectures. In E. Tognini Bonelli and G. Del Lungo Camiciotti (eds.) *Strategies in academic discourse*. Amsterdam: Benjamins. 17–29.
- Barlow, M. (2004) *Collocate 1.0: Locating collocations and terminology*. Houston, TX: Athelstan.
- Brazil, D. (1995) *A grammar of speech*. Oxford: Oxford University Press.
- Burnard, L. (2003) *Reference guide for BNC-baby*. Oxford: Oxford University Computing Services.
- Conrad, S. and D. Biber (2000) Adverbial marking of stance in speech and writing. In S. Hunston and G. Thompson (eds.) *Evaluation in text: Authorial stance and the construction of discourse*. Oxford: Oxford University Press. 56–73.
- Del Lungo Camiciotti, G. and E. Tognini Bonelli (eds.) (2004) *Academic discourse: New insights into evaluation*. Bern: Peter Lang.
- Diani, G. (2004) Evaluation in academic review articles. In A. Partington, J. Morley, and L. Haarman (eds.) *Corpora and discourse*. Bern: Peter Lang. 189–203.
- Fortanet Gómez, I. (2004) Verbal stance in spoken academic discourse. In G. Del Lungo Camiciotti and E. Tognini Bonelli (eds.) *Academic discourse: New insights into evaluation*. Bern: Peter Lang. 99–119.
- Freddi, M. (2005) From corpus to register: The construction of evaluation and argumentation in linguistics textbooks. In E. Tognini Bonelli and G. Del Lungo Camiciotti (eds.) *Strategies in academic discourse*. Amsterdam: Benjamins. 133–151.
- Hunston, S. (1993) Evaluation and ideology in scientific writing. In M. Ghadessy (ed.) *Register analysis: Theory and practice*. London: Pinter. 57–73
- Hunston, S. (2004) Counting the uncountable: Problems of identifying evaluation in a text and in a corpus. In A. Partington, J. Morley, and L. Haarman (eds.) *Corpora and discourse*. Bern: Peter Lang. 157–188.
- Hunston, S. (2005) Conflict and consensus: Construing opposition in applied linguistics. In E. Tognini Bonelli and G. Del Lungo Camiciotti (eds.) *Strategies in academic discourse*. Amsterdam: Benjamins. 1–15.
- Hunston, S. and J. McH. Sinclair (2000) A local grammar of evaluation. In S. Hunston and G. Thompson (eds.) *Evaluation in text: Authorial stance and the construction of discourse*. Oxford: Oxford University Press. 74–101.
- Hyland, K. (1999) Disciplinary discourses: Writer stance in research articles. In C. Candlin and K. Hyland (eds.) *Writing: Texts, processes and practices*. London: Longman. 99–121.

- Hyland, K. (2000) *Disciplinary discourses: Social interactions in academic writing*. Harlow: Longman.
- Martin, J. R. (2000) Beyond exchange: APPRAISAL systems in English. In S. Hunston and G. Thompson (eds.) *Evaluation in text: Authorial stance and the construction of discourse*. Oxford: Oxford University Press. 142–175.
- Mauranen, A. (2004) Where next? A summary of the round table discussion. In G. Del Lungo Camiciotti and E. Tognini Bonelli (eds.) *Academic discourse: New insights into evaluation*. Bern: Peter Lang. 203–215.
- Oakey, D. (2005) Academic vocabulary in academic discourse: The phraseological behaviour of EVALUATION in economics research articles. In E. Tognini Bonelli and G. Del Lungo Camiciotti (eds.) *Strategies in academic discourse*. Amsterdam: Benjamins. 169–183.
- Römer, U. (2005) “This seems somewhat counterintuitive, though...”: Negative evaluation in linguistic book reviews by male and female authors. In E. Tognini Bonelli and G. Del Lungo Camiciotti (eds.) *Strategies in academic discourse*. Amsterdam: Benjamins. 97–115.
- Scott, M. (1999) *WordSmith tools (Version 3.0)*. Oxford: Oxford University Press.
- Shaw, P. (2004) How do we recognise implicit evaluation in academic book reviews? In G. Del Lungo Camiciotti and E. Tognini Bonelli (eds.) *Academic discourse: New insights into evaluation*. Bern: Peter Lang. 121–140.
- Stubbs, M. (2001) *Words and phrases: Corpus studies of lexical semantics*. Oxford: Blackwell.
- Suárez Tejerina, L. (2005) Is evaluation structure-bound? An English-Spanish contrastive study of book reviews. In E. Tognini Bonelli and G. Del Lungo Camiciotti (eds.) *Strategies in academic discourse*. Amsterdam: Benjamins. 117–132.
- Swales, J. (2004) Evaluation in academic speech: First forays. In G. Del Lungo Camiciotti and E. Tognini Bonelli (eds.) *Academic discourse: New insights into evaluation*. Bern: Peter Lang. 31–53.
- Thompson, G. and S. Hunston (2000) Evaluation: An introduction. In S. Hunston and G. Thompson (eds.) *Evaluation in text: Authorial stance and the construction of discourse*. Oxford: Oxford University Press. 1–27.
- Tognini Bonelli, E. and G. Del Lungo Camiciotti (eds.) (2005) *Strategies in academic discourse*. Amsterdam: Benjamins.

Author's address

Ute Römer
 University of Michigan
 English Language Institute
 500 E. Washington Street
 Ann Arbor, MI 48104
 USA

uroemer@umich.edu