

## **Appendix to Chapter 1 of *Emotion Talk Across Corpora* (Appendix 1)**

### Contents:

A 1.1	Design of BRC	2
A 1.1.1	Conversation	3
A 1.1.2	News reportage	5
A 1.1.3	Fiction	6
A 1.1.4	Academic discourse	8
A 1.1.5	Summary	9
A 1.1.6	Key words in the BRC sub-corpora	11
A 1.1.7	List of files included in BRC	13
A 1.2	Analyzing emotion terms	17
A 1.3	List of emotion terms	24
	<i>Notes</i>	35

## A 1.1 Design of BRC

The British Register Corpus (BRC) is a register-sensitive corpus with four distinct sub-corpora, which was compiled through a principled sampling from the BNC (British National Corpus). The following sections describe its design in detail in order to follow one of the most important principles in corpus linguistic research, namely to outline ‘[t]he design and composition of a corpus [...] with information about the contents and arguments in justification of the decisions taken’ (Sinclair 2004a: 8).

As noted, the compilation of the BRC was exclusively based on the BNC. This is a general corpus of British English, consisting of over 100 million words from different varieties of modern English (about ten million words of spoken and about 90 million words of written English). Not all of the various parts of the BNC were compiled to be representative of the register they represent, since the BNC was designed as a general corpus of English (representative of the British English language as a whole rather than representative of the different registers that make up the corpus). As pointed out by Aston & Burnard (1998), ‘[a]lthough the BNC distinguishes several different geographical, sociological, and generic varieties, it does not necessarily provide a reliable sample for any particular set of such criteria’ (Aston & Burnard 1998: 28). For more detailed information on the BNC see e.g. the BNC website at <http://www.natcorp.ox.ac.uk/what/index.html>, Burnard (1995), Crowdy (1995), Rayson et al (1997) or Leech et al (2001: 2-4). Since the BNC offers only a very broad categorization system for the texts included in the corpus (domain, context, socio-economic class), Lee’s (2001, 2002) much more detailed categorization system was used to select texts from the BNC to compile the BRC. For details of this system and some of the difficulties and complexities involved see Lee (2001, 2002). The design of the BRC thus resembles the design of the four-million word BNC baby (Burnard 2003, <http://www.natcorp.ox.ac.uk>) in that 1) both are four-part corpora, 2) both use the text classifications provided by David Lee. However, the BRC is about five times bigger than the BNC baby, and there is no complete overlap of files. Other differences include the sampling procedures adopted in selecting texts for each register. Readers particularly interested in the issue of corpus design may wish to compare the design of the BRC (as described below) with the design of the BNC baby (Burnard 2003) and the LSWE corpus used by Biber et al (1999) (the BRC registers are at times less broad and more homogenous than the LSWE registers and contain no American English).

### A 1.1.1 Conversation

#### *Design*

The BRC conversation sub-corpus is equivalent to the BNC spoken demographic section, which is made up of 153 files (4,206,058 words). This sub-corpus of the BNC comprises casual conversational British English, collected by 124 ‘respondents’ (aged 15 and older), who taped their conversations with other speakers during a certain time period. The respondents were selected with the help of demographic sampling, and a relatively high degree of representativeness was achieved by including ‘as far as possible, [...] equal numbers of men and women, equal numbers from each of the six age groups, and equal numbers from each of four social classes’ (Burnard 1995: 20).

To achieve a high degree of balance in terms of social group and gender is extremely important for the study of affect/emotion since it cannot be assumed *a priori* that both social group and gender do not have an effect on the usage of emotion vocabulary: ‘[i]n stratified societies, social groups are frequently perceived as having different affective styles, and class identification rests in part on the individual’s affective demeanor’ (Besnier 1990: 435).<sup>1</sup> Presumably, age difference is also potentially important in this respect (Schrauf & Sanchez 2004 note a difference in the emotion words that have psychological salience for young and old speakers), although it may be difficult to pinpoint exactly which social factor is responsible for a given variation, and such an analysis of social variables and affect/emotion was well beyond the scope of Bednarek (2008).

In total, the BRC conversation sub-corpus includes female and male respondents from four social groupings: AB (top or middle management, administrative or professional), C1 (junior management, supervisory or clerical), C2 (skilled manual), and DE (semi-skilled or unskilled). It also comprises respondents from six different age groups. Table A.1 below presents the breakdown of respondents by social class (according to Lee 2002), age (according to Burnard 1995) and gender (according to Burnard 1995), in number of words as well as in percentages (from Leech et al 2001: 3):

Table A.1 BRC conversation sub-corpus

<b>Respondent social class</b>	<b>Number of files</b>	<b>Number of words</b>	<b>Percentage</b>
AB	59	1,363,571	32.54
C1	36	1,097,023	26.02
C2	32	1,192,120	25.64
DE	19	515,981	14.88
Unknown	7	37,363	0.89
<b>Respondent Age</b>	<b>Number of files</b>	<b>Number of words (W-units)</b>	<b>Percentage</b>
0-14	26	265,716	6.3
15-24	36	668,947	15.88
25-34	29	847,236	20.11
35-44	22	839,026	19.92
45-49	20	956,474	22.71
60+	20	633,817	15.05
<b>Respondent Gender</b>	<b>Number of files</b>	<b>Number of words (W-units)</b>	
Male	73	1,732,731	41.14
Female	75	2,462,339	58.47
Unclassified	5	16,146	0.38
<b>Total</b>	<b>153</b>	<b>4,206,058</b>	<b>100</b>

### *Issues*

The BRC conversation sub-corpus is probably the most representative of all the BRC sub-corpora, since it was originally designed for maximum representativeness. It has been used in much linguistic research, and is argued to ‘[provide] an unparalleled resource for investigating, on a large scale, the conversational behaviour of the British population in the 1990s’ (Rayson et al 1997: 134). However, the sampling only relates to the respondents, rather than their interlocutors, and concerning cross-tabulation it is difficult to say whether there is indeed a balance for all combinations of factors (e.g. are there the same number of male speakers of age group 1 and social class 1 as there are female speakers of the same age and social class, and as there are male speakers of other age groups and social classes?). Furthermore, Table A.1 does not tell us much about the number of words that is actually produced by the different speakers. In fact, as pointed out by Rayson et al (1997) female speech is slightly overrepresented and male speech is underrepresented in the conversational sub-corpus of the BNC (compare Table A.2):

Table A.2 Female vs. male speakers in conversation sub-corpus

	Female speakers	Male speakers
Number of speakers	561	536
Number of turns	250,955	179,844
Number of words spoken	2,593,452	1,714,443
Number of turns per speaker	447.33	335.53
Number of words per turn	10.33	9.53

(from Rayson et al 1997: 136)

However, no emotion terms are among the most over or underrepresented lexical items listed by Rayson et al (1997: 136-139), so this appears relatively unproblematic at least for the study of affect/emotion.

### **A 1.1.2 News reportage**

#### *Design*

The news section of the BNC (that is, all files from newspapers) consists of 486 files (9,345,878 words). For the BNC news reportage sub-corpus, texts were sampled from across the various topic and section categories listed by Lee (2002). These are: arts/cultural material, commerce/finance, home/foreign news, science, lifestyle/leisure/belief & thought, and sports (coded by Lee as *arts*, *commerce*, *report*, *science*, *social*, *sports*). Texts were also sampled from both ‘quality’ broadsheet newspapers and ‘popular’ tabloid newspapers. The following British national daily newspapers are included in the corpus: *The Independent*, *The Guardian*, *The Daily Mirror*, and *The Daily Telegraph*, representing different readerships and political attitudes. Local and regional newspapers (as well as a miscellaneous (unclear) category and spoken news) were excluded, because researchers have found that ‘while the conditions for the formulation of media language are similar, since practitioners are bound by the strictures of their discourse community of media-makers, the results are realized differently in different local contexts’ (Cotter 2001: 429). Since the focus was on informative rather than persuasive discourse, newspaper editorials were also excluded, making up a distinct register with a different global purpose.

On account of the complex process of news writing (for a summary see Bednarek 2006) multiple authors are involved, which means that the characteristics of the speaker/writer (age, sex, regional origin) that are important for conversation are somewhat irrelevant for news reportage. Table A.3 below shows the breakdown of news texts according to broadsheets/tabloids, topic/section (Lee 2002) and newspaper in terms of number of files and words:

Table A.3 BRC news reportage sub-corpus

Type	Topic/section	N° of files	Newspapers	N° of words
Broad-sheet	Arts/cultural material	51	<i>Independent, Guardian, Daily Telegraph</i>	351,811
	Commerce/finance	44	<i>Independent, Guardian, Daily Telegraph</i>	424,895
	Home/foreign news	49	<i>Independent, Guardian</i>	663,355
	Science	29	<i>Guardian, Daily Telegraph</i>	65,293
	Lifestyle/leisure/belief & thought	36	<i>Independent, Guardian, Daily Telegraph</i>	81,895
	Sports	24	<i>Independent, Guardian</i>	297,737
Tabloid	no further sub-categories	6	<i>Daily Mirror</i>	728,413
Total		239	<i>Daily Mirror, Daily Telegraph, Guardian, Independent</i>	2,613,399

### *Issues*

One of the disadvantages of the news reportage sub-corpus is that it includes only samples from four of the ten national British broadsheet newspapers: three broadsheets and just one tabloid newspaper (*Daily Mirror*). However, the newspapers at least represent different political stances, and the number of words sampled from the tabloid is relatively high. A more problematic issue relates to the fact that the tabloids are not sub-categorized in terms of different topics/sections/subject matter, which makes a more detailed analysis of intra-register variation difficult. In terms of tabloid newspaper language better large-scale corpora are clearly needed.

### **A 1.1.3 Fiction**

#### *Design*

In the BNC the genre *W\_fict\_prose* (novels and short stories) consists of 432 files (15,926,677 words). For the BRC fiction sub-corpus it was decided to focus on a less broadly defined and more homogenous register of relatively contemporary adult book fiction, comprising only fiction from 1985-1994 in book form. For this reason, files that indicated a collection or compilation of stories were excluded to filter out short stories. Texts were sampled with the aim of achieving balance of author sex (i.e. a similar number of male/female authors)

and author age (i.e. a similar number of younger/older authors). Audience sex was not taken into consideration, though the majority of files is addressed to a mixed audience. The breakdown for the included fiction texts by gender and age (according to Lee 2002) is given in Table A.4:

Table A.4 BRC fiction sub-corpus

<b>Gender</b>	<b>Number of files</b>	<b>Number of words</b>
male	86	3,326,231
female	84	3,297,300
Unknown	2	64,928
<b>Age</b>	<b>Number of files</b>	<b>Number of words</b>
25-34	14	532,585
35-44	28	1,091,130
45-59	26	1,011,365
60+	30	1,137,025
Unknown	74	2,916,354
<b>Total</b>	<b>172</b>	<b>6,688,459</b>

### *Issues*

Although the number of words for male/female and older/younger authors is relatively similar, there are still some imbalances when cross-tabulation is involved (the number of male/female speakers of a certain age). As Table A.5 shows, there are, for instance, more texts of male authors between 25-44 than female authors of that age, and more texts of female authors that are between 45-60+ than of male authors.

Table A.5 Cross-tabulation of age and sex

<b>Age</b>	<b>Number of files</b>		<b>Number of words</b>	
	<b>Female author</b>	<b>Male author</b>	<b>Female author</b>	<b>Male author</b>
25-34	5	9	209,311	323,274
35-44	7	21	251,777	839,353
45-59	16	10	635,887	375,478
60+	17	13	644,108	492,917
unknown	39	33	1,556,217	1,295,209

However, achieving complete balance would have meant a considerable reduction of the size of the corpus. Another disadvantage of the fiction sub-corpus is that no systematic sampling could be undertaken in terms of the different types or kinds of fiction (such as ‘mystery’, ‘romance’, ‘historical’, ‘adventure’, ‘science’, ‘general’) – which may well represent different registers – because this information is not explicitly given by Lee (2002) and would have had to be guessed from the contents or titles.

#### A 1.1.4 Academic discourse

##### *Design*

For the academic discourse sub-corpus of the BRC, samples of academic writing were taken from a wide range of (groups of) disciplines contained in the BNC (according to Lee 2002): the humanities, medicine, natural sciences, politics/law/education, social/behavioural sciences, and technology/computing/engineering (an approach similar to that adopted by Oakey 2002).<sup>2</sup> In order to achieve balance, about one million words from each of these ‘discipline’ groups were sampled, with the aim of including material that was as diverse as possible (taking into account the criteria of medium, keywords, author sex and type of authorship), e.g. research articles and book extracts (however, this was sometimes difficult since the material contained in the BNC is limited in variety). As a result, the academic discourse sub-corpus of the BRC represents a rather ‘broad’ register of academic writing, rather than a more specialized corpus of e.g. journal articles, book chapters or doctoral theses (for such studies see e.g. Hyland 1999, Thompson & Tribble 2001). The breakdown of texts according to subject area is listed in Table A.6.

Table A.6 BRC academic discourse sub-corpus

<b>Subject area</b>	<b>Number of files</b>	<b>Number of words</b>
Humanities	26	1,001,982
Medicine	14	1,001,196
Natural sciences	38	999,855
Politics, law, education	41	1,130,111
Social and behavioural sciences	38	1,141,785
Technology, computing, engineering	23	686,004
<b>Total written academic prose</b>	<b>180</b>	<b>5,960,933</b>

As becomes apparent, there is a relative balance in terms of the number of words sampled from different discipline groups, with the exception of technology/computing/engineering, which derives from the fact that the BNC contains only 686,004 words of this variety.

##### *Issues*

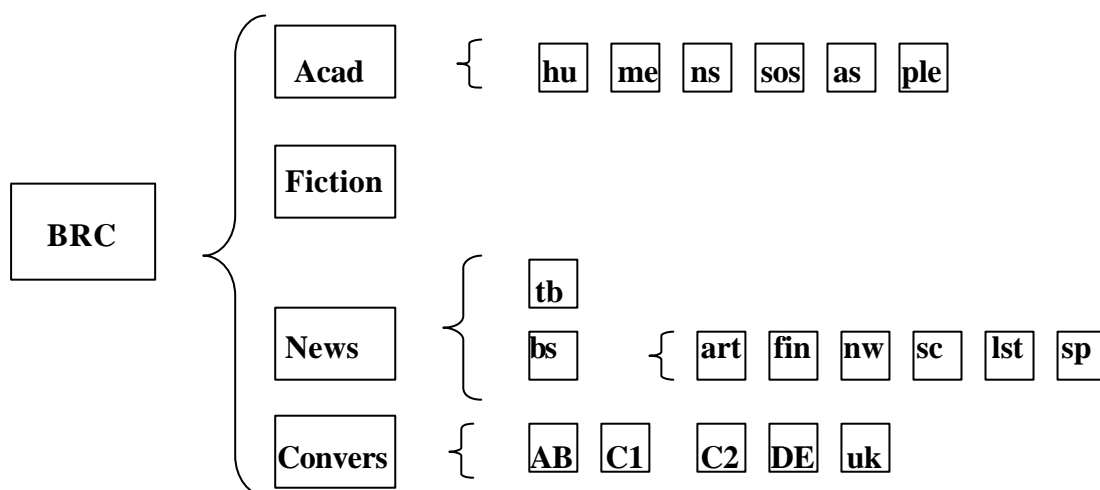
It is slightly problematic that different academic subjects such as politics, law and education or technology, computing and engineering are classified as one category rather than as several distinct categories. This means that there are fewer words concerning politics than e.g. medicine. In this, the academic discourse sub-corpus is not completely balanced. However, as with

fiction, the alternative (a more balanced corpus in these terms) would have resulted in a considerable reduction of corpus size (a sampling of far fewer than one million words from each academic subject).

### A 1.1.5 Summary

To sum up, the British Register Corpus (BRC) is made up of four sub-corpora: conversation, news reportage, fiction, and academic discourse with different sub-categories for academic discourse, conversation and news reportage according to topic, discipline and social class.<sup>3</sup>

#### BRC (British Register Corpus):



The BRC contains samples rather than whole texts, which is not completely in line with a functional approach (Butler 2004: 152). However, many of the samples are rather substantial (full chapters, paragraphs, whole articles, conversations); thus, the standard size in the BNC of a sample from a book is 40,000 words (Kilgarriff 1997a: 138). It is also worth pointing out that the corpus only includes English up to the early 1990s, and perhaps no longer represents contemporary English as spoken at the beginning of the 21<sup>st</sup> century. Finally, it is important to emphasize that the BNC was not originally designed for studies of individual registers or register variation (see above, Aston 2001, and Oakey 2002: 115). Designing the BRC on the basis of the BNC hence violates one of the principles established by Sinclair: ‘Only those components of corpora which have been designed to be independently contrastive should be contrasted’ (Sinclair 2004a: 3). Sinclair continues,

the existence of components differentiated according to the criteria discussed below, or identified by archival information, does not confer representative status on them, and so it is unsafe to use them in contrast with other components. Now that with many corpus management systems it is possible to ‘dial-a-corpus’ to your own requirements, it is important to note that the burden of demonstrating representativeness lies with the user of such selections and not with the original corpus builder. It is perfectly possible, and indeed very likely, that a corpus component can be adequate for representing its variety within a large normative corpus, but inadequate to represent its variety when freestanding. This point cannot be overstated; a lot of research claims authenticity by using selections from corpora of recognised standing (Sinclair 2004a: 3).

However, I tried to be as careful and balanced in the design of the BRC as possible, and to provide maximum transparency regarding its design, so that the degree of representativeness that the BRC achieves can be judged by fellow researchers.

Furthermore, the only alternative is to design one’s own corpus, which results in a number of problems because of time and other (e.g. financial) constraints where large-scale corpus research is concerned. And what the design of the BRC loses in representativeness, it gains in replicability. In any case, a corpus that fulfils all the design criteria mentioned by Sinclair remains yet to be compiled, and ‘[w]hile these [representativeness and balance] are not precisely definable and attainable goals, they must be used to guide the design of a corpus and the selection of its components’ (Sinclair 2004a: 9). This was clearly the case in the design of the BRC: Each of its four sub-corpora was designed to be as representative as possible of the register it represents, and as balanced as possible, while at the same time being as large as possible. The important phrase here is *as possible*, which means that some compromise had to be made, as has become apparent in the discussion above. Much more could be said on representativeness and corpus design, which is a thorny issue in corpus linguistics (for discussions compare e.g. Hunston 2002, Mahlberg 2004, Wynne 2005), but this is not the place to do so. Instead, it must be noted that the results from the BRC are to be considered as indicative, but not definitive (compare Sinclair 2004b: 81), and that no claims of guaranteed representativeness are made. Clearly, the corpus is not ‘representative’ of the chosen registers as such, rather it is *as representative as possible* given the material contained in the BNC. As Sinclair concludes:

It is important to avoid perfectionism in corpus building. It is an inexact science, and no-one knows what an ideal corpus would be like. [...] [C]ompilers make the best corpus they can in the circumstances, and their proper stance is to be detailed and honest about the contents. From their description of the corpus, the research community can judge how far to trust their results, and future users of the same corpus can estimate its reliability for their purposes. We should avoid claims of scientific

coverage of a population, of arithmetically reliable sampling, of methods that guarantee a representative corpus. The art or science of corpus building is just not at that stage yet (Sinclair 2004b: 81).

## 1.6 Key words in the BRC sub-corpora

100 *Wordsmith* key words in the BRC sub-corpora (reference corpus BRC, log-likelihood test, max p value 0.1, minimum frequency 3):

	<b>Academic</b>	<b>News</b>	<b>Fiction</b>	<b>Conversation</b>
1	#	#	EQUO	I
2	OF	POUND	BQUO	YOU
3	THE	MR	HER	YEAH
4	IS	YESTERDAY	SHE	T
5	IN	YEAR	HE	S
6	WHICH	HAS	HIS	IT
7	BY	BY	HAD	OH
8	ARE	MDASH	WAS	WELL
9	MAY	LAST	HIM	GOT
10	PATIENTS	WILL	HELLIP	ER
11	OR	BRITISH	LOOKED	MM
12	BE	NEW	SAID	KNOW
13	THESE	GOVERNMENT	EYES	WHAT
14	SUCH	CENT	MY	DON
15	FORMULA	MILLION	ME	NO
16	BETWEEN	BRITAIN	FACE	ERM
17	NDASH	WORLD	BACK	VE
18	ALSO	PER	KNEW	THAT
19	HOWEVER	FOR	MAN	DO
20	EXAMPLE	MARKET	HEAD	THEY
21	DATA	SHARES	ROOM	YES
22	SOCIAL	THE	HERSELF	RE
23	THIS	COMPANY	INTO	GET
24	AS	LONDON	DOOR	LL
25	STUDY	ENGLAND	TURNED	COS
26	LSQB	EUROPEAN	COULD	WE
27	RSQB	SEASON	FELT	GO
28	SYSTEM	AFTER	VOICE	JUST
29	THUS	CITY	SEEMED	THINK
30	HAS	CHAIRMAN	ASKED	THERE
31	THEREFORE	TEAM	WOULD	GONNA
32	FORM	FORMER	MDASH	RIGHT
33	AN	UNION	FATHER	LIKE
34	FIG	DAVID	HAND	MEAN
35	ANALYSIS	MINISTER	STOOD	SO
36	LANGUAGE	PRESIDENT	AT	CAN
37	NON	LEAGUE	AWAY	M
38	CELLS	WHO	SMILED	ALRIGHT
39	DISEASE	DOLLAR	HIMSELF	YOUR
40	SHOWN	ITS	HANDS	REALLY
41	WITHIN	PLAYERS	CAME	ONE

42	C	NATIONAL	SAT	WANT
43	FIGURE	WEST	D	AH
44	INFORMATION	WIN	DOWN	GOING
45	B	JOHN	MOTHER	THEN
46	STUDIES	SPOKESMAN	DARK	PUT
47	P	PROFITS	OUT	BIT
48	ET	INTERNATIONAL	THOUGHT	ME
49	RESULTS	PARTY	AGAIN	SEE
50	SECTION	COMPANIES	TOOK	DIDN
51	G	HONG	NODDED	OOH
52	EFFECT	MONTH	QUOT	HAVEN
53	DIFFERENT	KONG	SAW	COME
54	USING	LABOUR	MOMENT	NICE
55	NUMBER	DIRECTOR	NEVER	THEM
56	IMPORTANT	SECRETARY	NOTHING	ISN
57	TREATMENT	EAST	WOMAN	MUM
58	DEVELOPMENT	CLUB	SUDDENLY	LOOK
59	PARTICULAR	GAME	SMILE	SAY
60	ACT	AGAINST	SHOOK	HAVE
61	CASES	MATCH	TOO	DID
62	AREAS	EUROPE	HAIR	OKAY
63	LEVEL	TAX	BEGAN	GOTTA
64	STATE	BID	BEHIND	UP
65	GROUPS	UNITED	ACROSS	DOING
66	AL	STAR	STARED	SORT
67	FUNCTION	SPORT	UP	THING
68	BOTH	TV	WALKED	BLOODY
69	EVIDENCE	BANK	HEARD	HERE
70	EACH	YEARS	TOLD	GOOD
71	PROCESS	CONFERENCE	FEET	TWENTY
72	SPECIFIC	PAGE	LOOKING	GOES
73	SYSTEMS	WEEK	SIR	D
74	EDUCATION	PRICE	MOUTH	HA
75	DNA	GERMANY	LAUGHED	IF
76	STRUCTURE	MANAGER	BED	FIVE
77	ORDER	CORRESPONDENT	AM	ALL
78	CONTROL	EDITION	CORBETT	LOT
79	USE	SALES	HOUSE	HUNDRED
80	ACTIVITY	CUP	AROUND	ANYWAY
81	CASE	THEIR	BESIDE	WHY
82	POSSIBLE	CORPUS	LIPS	DOESN
83	ACID	CASH	ARMS	NOW
84	VALUES	LEADER	SOMEONE	AYE
85	DESCRIBED	INDUSTRY	REPLIED	ACTUALLY
86	ASSOCIATED	UNIVERSITY	ROSE	WON
87	E	GERMAN	WINDOW	WANNA
88	SPECIES	SMITH	TELL	INNIT
89	SIGNIFICANT	FIRST	OPENED	FUCKING
90	H	POLICE	PULLED	DONE
91	GASTRIC	SHARE	SILENCE	MY
92	INDIVIDUAL	AMERICAN	TOWARDS	HE
93	POPULATION	SOVIET	WONDERED	SAYS
94	WOMEN	INVESTORS	FLOOR	SAID
95	SIMILAR	GROUP	WANTED	DAD
96	CHAPTER	FINAL	GIRL	MHM
97	PER	BILLION	GLANCED	SOMETHING

98	GENERAL	TORY	HADN	AIN
99	X	CHAMPION	MRS	WOULDN
100	REQUIRED	FINANCIAL	ATHELSTAN	EH

## 1.7 List of files included in BRC

### *Conversation sub-corpus*

Spoken_AB	Spoken_C1	Spoken_C2	Spoken_DE	Spoken_Unclass
KB0	KB9	KB5	KB1	KNS
KB3	KBG	KB7	KB2	KNT
KB8	KBH	KBA	KB4	KNU
KBC	KBJ	KBB	KB6	KNV
KBD	KBL	KBF	KBE	KNW
KBK	KBV	KBN	KBP	KNY
KBM	KBY	KBT	KC2	KPO
KBR	KC6	KBX	KC5	
KBS	KCC	KC1	KC7	
KBU	KCK	KCE	KCA	
KBW	KCN	KCF	KCJ	
KC0	KCP	KCG	KCU	
KC3	KCS	KCL	KD6	
KC4	KD0	KCM	KDF	
KC8	KD2	KCT	KDN	
KC9	KD5	KCX	KDS	
KCB	KD7	KCY	KE5	
KCD	KDB	KD1	KPS	
KCH	KDG	KD3	KSS	
KCR	KDJ	KD8		
KCV	KDK	KD9		
KCW	KDM	KDA		
KD4	KDP	KDC		
KDL	KDV	KDD		
KDR	KDW	KDE		
KDU	KDY	KDH		
KE1	KE0	KDT		
KNR	KE3	KDX		
KP2	KE4	KE2		
KP3	KP1	KE6		
KP4	KP5	KP7		
KP6	KPJ	KPD		
KP8	KPM			
KP9	KPR			
KPA	KPU			
KPB	KR0			
KPC				
KPE				
KPF				
KPG				
KPH				
KPK				
KPL				
KPN				
KPP				
KPT				

KPV				
KPW				
KPX				
KPY				
KR1				
KR2				
KSN				
KSP				
KSR				
KST				
KSU				
KSV				
KSW				

*News reportage sub-corpus:*

<b>News_ Arts</b>	<b>News_ Commerce</b>	<b>News_ Report</b>	<b>News_ Science</b>	<b>News_ Social</b>	<b>News_ Sports</b>	<b>News_ Tabloids</b>
A1D	A1E	A1G	A1M	A1L	A1N	CH1
A1K	A1S	A1J	A3Y	A1X	A22	CH2
A1R	A21	A1V	A82	A2C	A2E	CH3
A20	A26	A1Y	A8A	A3M	A2S	CH5
A24	A2H	A28	A8E	A3X	A33	CH6
A25	A2V	A2A	A8R	A48	A3L	CH7
A2B	A37	A2M	A92	A4M	A40	
A2D	A3J	A2P	A9B	A5V	A4B	
A2G	A3K	A2X	AAG	A7Y	A4P	
A2R	A3S	A30	AH9	A8B	A52	
A2U	A43	A3D	AHD	A8M	A5C	
A32	A4F	A3G	AHL	A8Y	A5U	
A35	A55	A3U	AHP	A98	A80	
A36	A5G	A3W	AHV	A9Y	A8C	
A3H	A5S	A46	AHY	AHE	A8N	
A3R	A5T	A49	AJ4	AHH	A90	
A42	A7T	A4H	AJ7	AHM	A99	
A4A	A7U	A4K	AJB	AHS	A9H	
A4E	A85	A4N	AJE	AHW	A9R	
A4L	A86	A4X	AJK	AJ1	AA0	
A4S	A8G	A50	AJS	AJ5	AA7	
A51	A8H	A57	AK0	AJC	AAE	
A54	A8U	A59	AK3	AJG	AAN	
A5B	A8V	A5M	AK7	AJL	AAW	
A5E	A94	A5R	AKA	AJT		
A5F	A9D	A7V	AKF	AJW		
A7S	A9L	A7W	AKN	AK1		
A83	A9U	A87	AKW	AK8		
A8F	AA3	A88	AL4	AKC		
A8S	AAA	A8J		AKG		
A93	AAJ	A8K		AKK		
A9C	AAS	A8W		AKP		
A9K	AHB	A8X		AKT		
A9T	AHJ	A95		AKX		
AA2	AHT	A96		AL1		
AA9	AJ2	A9E		AL5		
AAH	AJ9	A9F				

AAR	AJH	A9M				
AHA	AJP	A9N				
AHG	AJX	A9V				
AHR	AKD	A9W				
AJ0	AKL	AA4				
AJ8	AKU	AA5				
AJF	AL2	AAB				
AJN		AAC				
AJV		AAK				
AK4		AAL				
AKB		AAT				
AKJ		AAU				
AKS						
AL0						

*Fiction sub-corpus:*

A0L	BPA	F9C	GVP
A6J	C86	F9X	GW2
A6N	C8D	FAB	GWF
A74	C8E	FAJ	GWG
AB9	C8S	FB0	H7F
ABW	C8T	FB9	H7H
AC2	C98	FNT	H7P
AC3	CA0	FNU	H7W
ACE	CAM	FP0	H84
ACK	CB5	FP3	H85
AD9	CCD	FP6	H8A
ADS	CCM	FP7	H8B
ADY	CD2	FPF	H8X
AE0	CDB	FPK	H90
AEA	CDE	FPM	H98
ALJ	CEB	FPX	H9C
AN7	CEC	FR3	H9G
ANL	CEH	FRF	H9N
ANY	CFY	FRS	HA2
APM	CHG	FS8	HGL
APR	CJA	FSC	HGN
APU	CJF	FSE	HGU
AR2	CK9	FSF	HH0
AR3	CKB	FYV	HH5
ARK	CKC	FYY	HHC
AS7	CKD	G04	HJH
ASD	CKE	G0A	HNJ
ASE	CKF	G0N	HR8
ASN	CLD	G0S	HR9
ASS	CMJ	G0X	HTH
AT7	CN3	G10	HTM
ATE	CR6	G15	HTR
B1X	CRE	G16	HTT
B20	EA5	G17	HTU
B3J	ECK	G1L	HTW
BMN	EDN	G1S	HU0
BMR	EEW	GUE	HW8

BMW	EF1	GUG	HWE
BNC	EV1	GUK	HWL
BP0	EVC	GV2	HWN
BP1	EVG	GV6	HWP
BP7	EWH	GV8	JY3
BP9		GVL	K8R
			K95

*Academic discourse sub-corpus:*

<b>Academic_ humanities</b>	<b>Academic_ medicine</b>	<b>Academic_ natural science</b>	<b>Academic_ Politicslaw education</b>	<b>Academic_ socialscience</b>	<b>Academic_ techengin</b>
A6G	B0X	EV6	B12	AN3	B2M
A6U	B30	ARY	CJG	B16	BP2
ARD	B33	FTB	EF3	CGF	CA4
F98	EWX	FTC	GW1	CGT	CG7
GWM	CAN	FTD	HPX	CMN	CG8
J0V	HU2	FTE	HXT	CMR	CG9
J1A	HWU	K5N	HXW	CRF	CGA
A04	EA0	K5P	HYB	ECB	CHF
A05	HU3	K5R	J6R	ECE	EES
A07	HWV	K5S	J76	FRL	EUS
CG0	EA1	K5T	FBS	J14	EWV
EA7	HU4	K5U	FBT	J7H	FE6
EE2	HWW	K5V	FBU	J7L	FNR
EEE	HWS	K5W	FBV	AS6	FPG
G1N		K5X	FBW	EB1	G3N
G1R		B2J	FBX	EWA	H0U
GUW		B2K	FBY	HP2	H7R
HY6		CMA	FC0	J7F	HGR
A6B		CMH	FC1	J7G	HGX
B2C		E9X	FC2	J7J	HR3
CFK		EV9	FC3	J7K	HRK
CK1		EVW	FC4	J7P	HX9
CM2		FEF	FC5	K93	K90
CM8		AML	FC6	ALM	
H7Y		AMM	FC7	ALN	
H8V		FU0	A5Y	ALP	
		FU9	A62	B2X	
		G1E	A64	FBH	
		GU5	A6F	GWJ	
		GU8	A6M	B1G	
		GV0	ABP	CJ1	
		GW6	ACJ	CLH	
		H79	ADD	FST	
		H8K	AM6	G08	
		H9R	AN5	G1J	
		H9S	APE	GVA	
		HRG	APN	H0J	
		J12	ASB	H8D	
			HTF		
			HXD		
			HXE		

Further information on these files is provided in David Lee's excel spread sheet, available to download at <http://clix.to/davidlee00>.

## A 1.2 Analyzing emotion terms

One of the central problems in the analysis of affect/emotion is how to establish a list of emotion terms. *Affect* is used in Bednarek (2008) as a cover term for feelings, emotions, moods, affect, etc and thus includes more than just emotion, but how much more? Where is the line between an emotional state of mind and a state of mind that does not involve emotion? Like other semantic-conceptual categories, affect is a fuzzy category, with no clear boundaries, presumably organized with the help of prototypes as cognitive reference points and *family resemblances* (e.g. Shaver et al 1987, Russell 1991, Oatley et al 2006: 184). For example, does *confusion* denote an emotion or not? How can we delineate the semantic domain of affect?<sup>4</sup> What Ortony et al (1987) call the *affective lexicon* comprises words that implicate emotions in various ways, as has also been noted by other emotion researchers. Words can denote

expressions of emotions (e.g. laughter, smiling, crying, tears, frown), bodily states associated with emotions (e.g. strong, tiredness), properties of emotion (e.g. deep, positive, negative, expressive, mixed, disturbed, uncontrollable, turbulent), characteristics of behaviour motivated by emotion (e.g. sincerity, giving, helping, sharing, violence), personality traits related to emotion (e.g. outgoingness, gentleness, sensitive, stubbornness, hardness, vulnerability, hyperactive), states of mind associated with emotions (e.g. confusion, uncertainty, arousal, control, conflict, thinking, meditating, alert), and cognates and superordinates of emotion (e.g. reactions, responsive, state, communication, expression). (Johnson-Laird & Oatley 1989: 87-88)

As Fiehler notes: 'Es gibt sicherlich einen Kernbereich prototypischer Emotionen, bei denen auch intersubjektiv große Übereinstimmung herrschen wird, daß es sich um Emotionen handelt [...] . Es gibt aber einen breiten Übergangsbereich, in dem die Urteile differieren.' (Fiehler 1990: 56).<sup>5</sup> For instance, there is no clearly defined border between affective and other mental processes (Fiehler 1990: 56), and many existing taxonomies involve differences of degree (see, for example, Ortony et al 1987: 353). Consequently, while the list of 1060 emotion terms used in Bednarek (2008) contains many lexical items that would unanimously be judged as prototypical emotion terms by many researchers (e.g. *hate*, *love*), it also includes terms that may be excluded by some researchers, and excludes terms that may be included by

other researchers. As noted in Chapter 1 (Note 6) different lists of emotion vocabulary have very different lengths.

In Bednarek (2008), the *Encarta Thesaurus (ET)* was used as the basis for establishing a list of emotion terms, but it was not adopted on a one-to-one basis. While some words had to be excluded for purely methodological reasons (Bednarek 2008: Section 1.6.3), other decisions related to the *ET* itself. For instance, its categorization is sometimes not consistent (e.g. *laugh* and *weep* are not included but *close to tears*, *crying*, *sigh* and *crack up* are), and it is not clear on which methodological principles its categorization is based. In some cases it looks as if one word if included as adjective was automatically included as noun, verb etc even if this made no sense. Other words that are problematic include *canonise*, *oracle* or *martyrdom* where the connection to affect seems controversial. A possible solution to this problem would be to use the *ET* list as basis for a prototype rating of emotion terms, working with a representative sample of native speakers. However, to establish the status of almost 2000 words was not feasible (and existing lists of prototype ratings for emotion terms were considered too small). Instead, it was decided to adopt a relatively broad approach to affect, in effect following the compilers of the *ET*, but excluding terms that seemed to refer to states of mind or character rather than emotions, and focusing on emotion talk rather than emotional talk (Bednarek 2008: Section 1.3). For instance, the following categories of the *ET* were excluded, as *expressing* rather than *denoting* emotions/states of mind, as referring to states of mind rather than to emotions, or as referring to psychological and related phenomena:<sup>6</sup>

- (1) Expressions of surprise (e.g. *gee*, *oh my* etc), regret (e.g. *alas*), uncertainty (e.g. *apparently*, *seemingly*)
- (2) Calmness, confidence, and composure (e.g. *accustomed*, *calm*), Pensiveness and interest (*attentive*, *pensive*), Prejudice (e.g. *bias*, *chauvinism*), Neutrality and indifference (e.g. *absently*, *apathy*), Bore and fail to interest (e.g. *bore*, *weary*), Soothe and calm (e.g. *appease*, *sweeten*), Appeal to and arouse interest (e.g. *captivate*, *inspire*), Encourage (e.g. *fortify*, *hearten*), Change of mood and composure (e.g. *adapt*, *take heart*), Experience and encounter (e.g. *suffer*, *undergo*), Ignorance (e.g. *untrained*, *ignorant*), Knowledge and wisdom (e.g. *academic*, *bookish*), Uncertainty (e.g. *agnostic*, *doubt*), Certainty (e.g. *certainly*, *belief*)
- (3) Eccentricity and irrationality (e.g. *barmy*, *raving*), Fears and phobias (e.g. *acrophobia*), Fads, fetishes, and idolatry (e.g. *craze*, *cult*), Devotees and addicted people (e.g. *addict*, *patriot*)

- (4) The will and willingness (e.g. *cooperative, submissive*), Unwillingness and stubbornness (e.g. *averse, difficult*), Rebelliousness and disobedience (e.g. *radical, turbulent*), Uncooperative or rebellious person (e.g. *activist, warmonger*)

In Bednarek (2008), establishing a list of emotion terms was a methodological issue, rather than an issue that I was interested in from a more theoretical view point. The final list of more than 1000 emotion terms certainly covers a large variety of different emotion terms in British English.

After having established a list of emotion terms to investigate in the corpus, decisions had to be made as to which of their usages to include as affect. As mentioned, the frequency analyses reported in Chapter 2 are the result of a semi-automated count of the meaning of emotion terms, involving a manual deletion of some concordance lines. Only the ‘affective’ meanings and usages of emotion terms were counted, with ‘non-affective’ meanings excluded. This means that the analysis is much more ‘subjective’ than a purely automatic frequency analysis that is not sense-sensitive (à la Leech et al 2001). Ideally, such a meaning-sensitive analysis would be as detailed as the analyses undertaken by West (1953), based on a corpus of 2½-5 million words. Compare Table A.7 on page 20:

Table A.7: Analysis of meanings of emotion terms in West (1953)

<p><b>Nouns</b></p>	<p><b>love (2059 n, v):</b>  <u>affection:</u>  <i>Love of/for a friend (19%)</i>  <i>The Goddess of Love (25%)</i>  <u>Phrases:</u>  <i>in love with, fall in love, my love (4%)</i>  <i>Love-affair, love-letter, etc (2%)</i>  <b>hope (1283 n, v):</b>  <i>Have some/no/a hope of success, lose hope; raise the hopes of, build/fix one's hopes on (48%)</i>  <i>The hope of his party, our last hope (3%)</i>  <b>fear (846 n, v):</b>  <u>emotion:</u>  <i>A feeling of fear; the fear of death (27%)</i>  <i>You need have no such fear (13%)</i>  <u>anxiety:</u>  <i>Did not go for fear that he might be hurt; was in fear of his life (8%)</i>  <i>(fear of god 2%; no fear!, no fear of that 0%)</i>  <b>feeling (feel, v 1615)</b>  <u>sense of touch:</u>  <i>I've no feeling in my fingers (1%)</i>  <u>pleasure, pain and emotion:</u>  <i>feeling of pleasure, pain, hope</i>  <i>feeling of ease, equality etc</i>  <i>showed such nice feeling in sending these flowers</i>  <i>good/ill feeling (14%)</i>  <u>sensibilities:</u>  <i>You've hurt my feelings (6%)</i>  <u>attitude, conviction:</u>  <i>I cannot go against the feeling of the nation</i>  <i>No particular feeling in the matter</i>  <i>A feeling that it's going to be a success (11%)</i>  <b>shock (174 n, v)</b>  <u>The shock of battle, a blow, or explosion:</u>  <i>earthquake shock</i>  <i>electric shock (20%)</i>  <u>Figurative:</u>  <i>a shock to my feelings</i>  <i>her death was a great shock to me (23%)</i>  <u>Special medical use:</u>  <i>cases of shock after air-raids (8%)</i>  <i>shock-absorber etc (1%)</i>  <b>surprise (470 v, a, n)</b>  <i>I felt some surprise at seeing him there (24%)</i>  <u>Phrases:</u>  <i>What a surprise (3%)</i>  <i>Take a person by surprise (3%)</i></p>
<p><b>Adjectives</b></p>	<p><b>happy (721)</b>  <u>contented, joyful:</u>  <i>be happy at school, feel happy (82%)</i>  <u>fortunate:</u>  <i>Happy New Year, many happy returns of the day</i>  <u>apt</u>  <i>a happy answer (9%)</i>  <b>sad (226)</b>  <i>feel sad, a sad look, a sad event (65%)</i>  <i>It is sad for/that (26%)</i>  <b>anxious (203)</b>  <u>uneasy:</u>  <i>feeling very anxious about the future; an anxious moment (48%)</i></p>

	<u>eager:</u> <i>I am most anxious to please you (52%)</i> <b>angry (155)</b> <u>irate:</u> <i>I'm very angry; angry at/with (67%)</i> <u>of mood, easily angered:</u> <i>He's in an angry mood (8%)</i> <u>figurative:</u> <i>angry sky, sea, wound (25%)</i>
<b>Adverbs</b>	<b><i>happily</i></b> <u>in happiness:</u> <i>lived happily every after (35%)</i> <u>luckily (50%)</u> <u>aptly (15%)</u>
<b>Verbs</b>	<b><i>love (2059, n, v)</i></b> <u>have friendship (16%)</u> <u>between the sexes:</u> <i>he loves her; 'Tis better to have loved and lost (13%)</i> <u>appreciate:</u> <i>love reading, good wine etc (8%)</i> <u>with a verb:</u> <i>I love to hear you sing, you singing (2%)</i> <b><i>worry (94 v, a, n)</i></b> <u>cause minor irritation to:</u> <i>Does the noise of my typewriter worry you? (19%)</i> <u>cause anxiety:</u> <i>I'm very worried about my son's health (28%)</i> <u>give way to anxiety:</u> <i>Don't worry (23%)</i> <b><i>enjoy (434)</i></b> <i>enjoy a book/reading/oneself etc</i> <b><i>care (1134 n, v)</i></b> <u>feel anxious for:</u> <i>He cares only for his own interests; I don't care (11%)</i> <u>idea of responsibility:</u> <i>The child has been well cared for (8%)</i> <u>wish:</u> <i>Would you care to read this?</i> <i>Do you care to come out for a walk?</i> <i>I don't much care for dancing (14%)</i> <b><i>hate (242 n, v)</i></b> <i>I hate him</i> <i>I hate to trouble you</i> <i>I hate to hear good music badly played (78%)</i> <b><i>admire (140)</i></b> <u>observe with pleasure and surprise (96%)</u>

However, because over 1000 emotion terms were analyzed, such a level of detail was not possible. Instead, decisions were made with respect to meanings/usages of each emotion term, simply as to whether to include or exclude them from the frequency count (see Appendix A 2.3 online for examples). Such decisions necessarily entail some element of subjectivity, especially where the difficult nature of evaluative/emotion meaning is concerned (compare also Wallace & Carson 1973: 6, Moore et al 1999: 541, Whitelaw et al 2005).

The software tool for analyzing emotion terms was the Zurich BNCweb interface (<http://escorp.unizh.ch/>), which allows different types of searches of the BNC and the BRC.

The type of search that was most frequently used was the ‘lemma query’, permitting a part-of-speech sensitive search for emotion terms (e.g. *love* as verb or noun). For the analysis of lexical distribution the correctness of the lemmatization was not checked. The non POS-sensitive ‘standard query’ was also used in some cases. For instance, in order to consider spelling variations/mistakes, the search was for the forms *bad-tempered* (lemma query) und *bad tempered* (standard query); for the lemma *overwrought* search terms were *overwrought*, *overwrought* (both lemma query) and *over wrought* (standard query) and so on. With standard queries only the correct POS occurrences were included in the frequency count. Standard queries were also always necessary whenever more than one orthographic word was involved or a specific form of a word was looked for (e.g. *champing at the bit*, *guilty conscience*, *second thoughts*, *willies*).

As said, occurrences that did not refer to emotion were excluded. For example, *depress* has four different meanings (from OALD – *Oxford Advanced Learner’s Dictionary*):

- 1 ‘to make sb sad and without enthusiasm or hope’: *Wet weather always **depresses** me.*
- 2 ‘to make trade, business, etc. less active’: *The recession has **depressed** the housing market.*
- 3 ‘to make the value of prices or wages lower’: *to **depress** wages / prices*
- 4 ‘to press or push sth down, especially part of a machine’: *to **depress** the clutch pedal*

Needless to say, only occurrences of meaning 1 were included in the frequency count. In cases where it was easy to differentiate between meanings of affect, judgement and appreciation (Bednarek 2008: Section 1.4), appreciation and judgement were excluded (e.g. *it’s a pity*, *what a pity* vs. *out of pity*); in other cases, this was not easily possible, and the figures may include appreciation/judgement meanings of emotion terms.

Further, on account of the large number of words investigated, such decisions were mostly based only on the immediate context – usually the sentence in which the potential emotion term occurred, rather than the source text itself. If in doubt, the instance was included, rather than excluded. In some cases, differences between meanings were very small, with meanings shading into each other (often metaphorical and literal meanings). The attempt was made to be as consistent as possible in the analysis, but with so many occurrences, there is a margin of error because of human fallibility and an incapacity for complete consistency across the analysis of 1000 words: even if one emotion term only occurred ten times on average in the entire corpus this would mean looking at 10,600 occurrences. The OALD (which gives definitions for emotion term meanings and usages) was used as a help in the

analysis but the final decision was mine. A problem was also that run-on entries,<sup>7</sup> in particular emotion adverbs (compare Appendix A 2.3 online), are not described very thoroughly in the OALD.

While it is not possible to detail decisions made with respect to every single emotion term, some decisions are more general and will briefly be listed:

1. Identical or very similar occurrences were counted only once, though not all repeated instances may have been detected.
2. Emotion terms in titles and names were excluded (e.g. *Lady Happiness, Felicity, Turkish/Angel Delight* etc), but emotion terms in quotes (e.g. *He said 'I'm delighted'*) were included.
3. Different spellings were included (e.g. *self-pitying/self pitying; ise/ize*).
4. When there were more than 500 occurrences of an emotion term in one of the sub-corpora, the figures for this sub-corpus were extrapolated from a random subset of 10-20%. 20% (every 5<sup>th</sup> concordance line examined) were used as a subset if 20% of the total number of occurrences was itself not more than 500 occurrences; otherwise 10% were used (every 10<sup>th</sup> concordance line). If even 10% made up more than 500 occurrences, no manual analysis was undertaken across all sub-corpora, and the resulting figures are not sense frequencies but word frequencies. This happened only with three emotion terms: *like* (V),<sup>8</sup> *want* (V),<sup>9</sup> and *sorry* (A),<sup>10</sup> which were in fact excluded from the analysis of lexical distribution (compare Bednarek 2008: Section 2.2.1).
5. Dispersion (Leech et al 2001: 18) was not analyzed systematically. This is slightly problematic because '[p]articularly for the lower frequency words, a large proportion of the occurrences might take place in one conversation, and this would skew the results.' (Rayson et al 1997: 149). This was partly levelled out by the large amount of words looked at.
6. Occurrences for each emotion term also include negated occurrences. For example, figures for *angry* include *angry* and *not angry*.

Summing up, readers should be aware that the reported figures are sense frequencies, not word frequencies, and that there are methodological implications of this fact. All figures should be interpreted as tendencies rather than precise and stable figures: 'Preliminary investigations into the stability of outcomes in lexical semantics suggests that it is severely lacking.' (Kilgarriff 1997b: 98). On the other hand, the analysis does have some advantage over a

pure automatic calculation: ‘A human parser, of course, does not attain 100 per cent, because some instances are inherently indeterminate and humans also make mistakes; but there is still a critical gap between what the machine can achieve and what the human can achieve.’ (Halliday 2005: 67).

### A 1.3 List of emotion terms (alphabetic)

The listed words only have an ‘affective’ meaning in some of their usages, and only these were counted. Spelling variations were also considered, i.e. for the lemma *bad-tempered* the search was for the forms *bad-tempered* and *bad tempered*; for the lemma *overwrought* search terms were *overwrought*, *over-wrought* and *over wrought*. For more details see A 1.2 above.

abash	Verb	agonised	Adjective
abashed	Adjective	agony	Noun
abhor	Verb	alarm	Noun
abhorrence	Noun	alarm	Verb
abjection	Noun	alarmed	Adjective
abjectly	adverb	amaze	Verb
abominate	Verb	amazed	Adjective
ache to	Verb	amazement	Noun
aching (for)	Noun	ambition	Noun
acrimony	Noun	amorous	Adjective
addled	Adjective	amuse	Verb
addlepated	Adjective	amused	Adjective
admiration	Noun	anger	Noun
admire	Verb	anger	Verb
admiring	Adjective	angry	Adjective
adoration	Noun	angst	Noun
adore	Verb	angst-ridden	Adjective
adoring	Adjective	anguish	Noun
adulation	Noun	anguished	Adjective
affection	Noun	animosity	Noun
affinity	Noun	animus	Noun
affronted	Adjective	annoy	Verb
aflutter	adverb	annoyance	Noun
afraid	Adjective	annoyed	Adjective
afterglow	Noun	antagonise	Verb
aggravate	Verb	antagonism	Noun
aggravated	Adjective	antagonistic	Adjective
aggravation	Noun	anticipation	Noun
aggrieve	Verb	antipathetic	Adjective
aggrieved	Adjective	antipathy	Noun
aghost	Adjective	antsy	Adjective
agitate	Verb	anxiety	Noun
agitated	Adjective	anxious	Adjective
agitation	Noun	anxiousness	Noun
agog	Adjective	apologetic	Adjective

apoplectic	Adjective	blackness	Noun
appal	Verb	bliss	Noun
appalled	Adjective	blissfully	adverb
appetite	Noun	blow away	Verb
appreciate	Verb	blue	Adjective
appreciation	Noun	blues	Noun
appreciative	Adjective	bother	Verb
apprehension	Noun	bothered	Adjective
apprehensive	Adjective	bowl over	Verb
apprehensiveness	Noun	bowled over	Adjective
approval	Noun	brightness	Noun
approving	Adjective	broken-hearted	Adjective
ardour	Noun	broodily	adverb
ashamed	Adjective	broodiness	Noun
aspiration	Noun	broody	Adjective
aspire	Verb	browned-off	Adjective
astonish	Verb	bruised	Adjective
astonished	Adjective	brutalise	Verb
astonishment	Noun	bug	Verb
astound	Verb	bullishness	Noun
astounded	Adjective	buoyancy	Noun
attachment	Noun	buoyant	Adjective
aversion	Noun	buoyantly	adverb
avid	Adjective	burdened	Adjective
avidity	Noun	bursting (to do)	Adjective
awe	Noun	butterflies	Noun
awe-stricken	Adjective	caginess	Noun
awe-struck	Adjective	care	Verb
awkwardness	Noun	carried away (get/be)	Adjective
bad	Adjective	cast down	Verb
bad blood	Noun	chafe	Verb
bad mood	Noun	chagrin	Noun
bad temper	Noun	chagrined	Adjective
bad-tempered	Adjective	champing at the bit	Adjective
baffle	Verb	chariness	Noun
baffled	Adjective	charmed	Adjective
bafflement	Noun	chasten	Verb
bask	Verb	cheer	Noun
be burning to	Verb	cheer up	Verb
be spoiling for	Verb	cheerful	Adjective
bedazzle	Verb	cheerfully	adverb
bedazzled	Adjective	cheerfulness	Noun
befuddlement	Noun	cheeriness	Noun
begrudge	Verb	cheery	Adjective
bemusement	Noun	cheese off	Verb
bent on	Adjective	cheesed off	Adjective
besotted	Adjective	cherish	Verb
bewilder	Verb	chill	Noun
bewildered	Adjective	chill	Verb
bewilderment	Noun	chipper	Adjective
bewitched	Adjective	chirpy	Adjective
bile	Noun	cholera	Noun
bitter	Adjective	choleric	Adjective
bitterness	Noun	chuffed	Adjective

circumspection	Noun	demotivate	Verb
cock-a-hoop	Adjective	deprecatory	Adjective
compassion	Noun	depress	Verb
compulsion	Noun	depressed	Adjective
compunction	Noun	depression	Noun
concern	Noun	desire	Noun
concerned	Adjective	desire	Verb
confound	Verb	desirous	Adjective
confusion	Noun	desirously	adverb
conscience-stricken	Adjective	desirousness	Noun
consternation	Noun	desolate	Adjective
contempt	Noun	desolately	adverb
content	Adjective	desolation	Noun
content	Verb	despair	Noun
contented	Adjective	despairing	Adjective
contentment	Noun	desperate	Adjective
contrite	Adjective	desperately	adverb
contriteness	Noun	desperation	Noun
contrition	Noun	despise	Verb
covet	Verb	despondency	Noun
covetous	Adjective	despondent	Adjective
cow	Verb	detest	Verb
cowed	Adjective	detestation	Noun
crabbed	Adjective	devastate	Verb
crabbiness	Noun	devastated	Adjective
crabby	Adjective	devotedly	adverb
crankiness	Noun	devotion	Noun
cranky	Adjective	disaffect	Verb
crave	Verb	disappoint	Verb
craving	Noun	disappointed	Adjective
crazed	Adjective	disappointment	Noun
crazy	Adjective	disapprobation	Noun
crestfallen	Adjective	disapproval	Noun
cross	Adjective	disapprove	Verb
crotchety	Adjective	disarm	Verb
crush	Noun	discombobulate	Verb
cut up	Adjective	discomfit	Verb
daunt	Verb	discomfited	Adjective
daunted	Adjective	discomfiture	Noun
daze	Noun	discomfort	Noun
dazzle	Verb	discomposure	Noun
defeatism	Noun	disconcert	Verb
deflated	Adjective	disconcerted	Adjective
deify	Verb	disconsolate	Adjective
dejected	Adjective	discontent	Noun
dejection	Noun	discontented	Adjective
delectation	Noun	discontentment	Noun
delight	Noun	discourage	Verb
delight	Verb	discouraged	Adjective
delighted	Adjective	discouragement	Noun
delirious	Adjective	disdain	Noun
demoralisation	Noun	disdain	Verb
demoralise	Verb	disenchant	Verb
demoralized	Adjective	disenchanted	Adjective

disenchantment	Noun	edginess	Noun
disfavour	Noun	edgy	Adjective
disgruntle	Verb	elate	Verb
disgruntled	Adjective	elated	Adjective
disgust	Noun	elation	Noun
disgust	Verb	electrified	Adjective
disgusted	Adjective	embarrass	Verb
dishearten	Verb	embarrassed	Adjective
disheartened	Adjective	embarrassment	Noun
disillusioned	Adjective	embitter	Verb
disillusionment	Noun	embittered	Adjective
dislike	Noun	enamoured	Adjective
dislike	Verb	enchanted	Adjective
dismay	Noun	enchantment	Noun
dismayed	Adjective	enjoy	Verb
dispirited	Adjective	enjoyment	Noun
displease	Verb	enmity	Noun
displeased	Adjective	enrage	Verb
displeasure	Noun	enraged	Adjective
disquiet	Noun	enrapture	Verb
disquieted	Adjective	enraptured	Adjective
disrespect	Noun	enthral	Verb
dissatisfaction	Noun	enthused	Adjective
dissatisfied	Adjective	enthusiasm	Noun
dissatisfy	Verb	enthusiastic	Adjective
distaste	Noun	entrance	Verb
distraught	Adjective	envious	Adjective
distress	Noun	enviousness	Noun
distress	Verb	envy	Noun
distressed	Adjective	envy	Verb
disturb	Verb	esteem	Noun
disturbed	Adjective	esteem	Verb
doldrums	Noun	euphoria	Noun
doleful	Adjective	euphoric	Adjective
dolefulness	Noun	exaltation	Noun
doting	Adjective	exasperate	Verb
dotty (about)	Adjective	exasperated	Adjective
down	Adjective	exasperation	Noun
down in the mouth	Adjective	excited	Adjective
downcast	Adjective	excitedly	adverb
downhearted	Adjective	excitement	Noun
dread	Noun	exhilarate	Verb
dread	Verb	exhilarated	Adjective
drive insane	Verb	exhilaration	Noun
drive round the bend	Verb	expectancy	Noun
drive up the wall	Verb	expectant	Adjective
druthers	Noun	expectantly	adverb
dying for	Adjective	expectation	Noun
dying to	Adjective	exultant	Adjective
eager	Adjective	exultation	Noun
eagerness	Noun	fall for	Verb
ecstasy	Noun	fanatical	Adjective
ecstatic	Adjective	fancy	Verb
edgily	adverb	fascination	Noun

faze	Verb	glad	Adjective
fear	Noun	gladden	Verb
fearful	Adjective	gladdened	Adjective
fearfully	adverb	gladness	Noun
fearfulness	Noun	glee	Noun
fed up	Adjective	gleeful	Adjective
feel for	Verb	gleefulness	Noun
feeling	Noun	gloom	Noun
felicity	Noun	gloomily	adverb
fervent	Adjective	gloominess	Noun
fervid	Adjective	gloomy	Adjective
fervour	Noun	glory in	Verb
fever	Noun	glum	Adjective
fevered	Adjective	glumness	Noun
feverish	Adjective	gnaw (at so)	Verb
fired-up	Adjective	gob-smacked	Adjective
flabbergast	Verb	good humour	Noun
flabbergasted	Adjective	good temper	Noun
flap (in a)	Noun	goodwill	Noun
floor	Verb	grate	Verb
flummox	Verb	grateful	Adjective
flummoxed	Adjective	gratefully	adverb
fluster	Verb	gratefulness	Noun
flustered	Adjective	gratification	Noun
flutter	Noun	gratified	Adjective
fondness	Noun	gratify	Verb
forlorn	Adjective	gratitude	Noun
forlornly	adverb	green-eyed	Adjective
fractious	Adjective	grief	Noun
frantic	Adjective	grieve	Verb
fraught	Adjective	grind down	Verb
frenzied	Adjective	grouchy	Adjective
frenzy	Noun	grudge	Noun
fret	Verb	grudge	Verb
fright	Noun	grumpiness	Noun
frighten	Verb	grumpy	Adjective
frightened	Adjective	guilt	Noun
frown on	Verb	guiltiness	Noun
frustrate	Verb	guilt-ridden	Adjective
frustrated	Adjective	guilty	Adjective
frustration	Noun	guilty conscience	Noun
fulfilled	Adjective	gusto	Noun
fuming	Adjective	guttled	Adjective
furious	Adjective	hacked off	Adjective
furiously	adverb	hanker/hanker after	Verb
furiousness	Noun	hankering	Noun
fury	Noun	happily	adverb
fuss	Verb	happiness	Noun
gaiety	Noun	happy	Adjective
gaily	adverb	harrassed	Adjective
gall	Verb	hate	Noun
gasping for	Adjective	hate	Verb
get on sb's nerves	Verb	hatred	Noun
get sb's back up	Verb	haunt	Verb

have it in for	Verb	impressed	Adjective
head over heels in love	Adjective	impulsion	Noun
heartache	Noun	in a funk	Adjective
heartbreak	Noun	in a huff	Noun
heartbroken	Adjective	in a state	Adjective
heavy-hearted	Adjective	in disbelief	Adjective
heavy-laden	Adjective	in good spirits	Adjective
heebie-jeebies	Noun	in high dudgeon	Adjective
heedful	Adjective	in seventh heaven	Adjective
hero-worship	Verb	in the dumps [down in ...]	Adjective
het up	Adjective	incense	Verb
high spirits	Noun	incensed	Adjective
hold dear	Verb	inclination	Noun
homesick	Adjective	incomprehension	Noun
homesickness	Noun	inconsolable	Adjective
honour	Verb	incredulity	Noun
honoured	Adjective	incredulous	Adjective
hope	Noun	indebted	Adjective
hopefulness	Noun	indignant	Adjective
hopeless	Adjective	indignation	Noun
hopelessly	adverb	infatuated	Adjective
hopelessness	Noun	infatuation	Noun
horrified	Adjective	inflame	Verb
horrify	Verb	infuriate	Verb
horror	Noun	infuriated	Adjective
horror-stricken	Adjective	inhibition	Noun
horror-struck	Adjective	insecurely	adverb
hostility	Noun	insecurity	Noun
hot under the collar	Adjective	intent	Adjective
huffy	Adjective	intimidate	Verb
humble	Verb	intimidated	Adjective
humbled	Adjective	irate	Adjective
humiliate	Verb	ire	Noun
humiliated	Adjective	ireful	Adjective
humiliation	Noun	irk	Verb
hung up	Adjective	irked	Adjective
hunger	Noun	irritability	Noun
hungry (for)	Adjective	irritable	Adjective
hurt	Noun	irritate	Verb
hurt	Adjective	irritated	Adjective
hurt	Verb	irritation	Noun
hurting	Adjective	itch	Noun
idealise	Verb	itch	Verb
identification (with)	Noun	itching	Adjective
identify with	Verb	jar	Verb
idolise	Verb	jauntiness	Noun
ill at ease	Adjective	jealous	Adjective
ill humour	Noun	jealousy	Noun
ill will	Noun	jitteriness	Noun
ill-disposed	Adjective	jitters	Noun
impatience	Noun	jittery	Adjective
impatience	Noun	joie de vivre	Noun
impatient	Adjective	jolliness	Noun
impress	Verb	jollity	Noun

joviality	Noun	melancholy	Noun
joy	Noun	merriment	Noun
joyful	Adjective	merry	Adjective
joyfully	adverb	miff	Verb
joyfulness	Noun	miffed	Adjective
joylessness	Noun	mind	Verb
joyous	Adjective	mirthful	Adjective
joyousness	Noun	miserable	Adjective
jubilant	Adjective	miserably	adverb
jubilation	Noun	misery	Noun
jumpily	adverb	misgiving(s)	Noun
jumpiness	Noun	miss	Verb
jumpy	Adjective	mistrust	Verb
keen	Adjective	mistrustful	Adjective
keenness	Noun	moonstruck	Adjective
keyed-up	Adjective	morose	Adjective
knock your socks off	Verb	moroseness	Noun
knocked out	Adjective	mortification	Noun
lap up	Verb	mortified	Adjective
lather (in a)	Noun	mortify	Verb
leaning	Noun	mournful	Adjective
let down	Verb	mournfulness	Noun
light-hearted	Adjective	mystification	Noun
lightheartedly	adverb	mystified	Adjective
like a cat on a hot tin roof	Adjective	mystify	Verb
like a cat on hot bricks	Adjective	nag	Verb
liking	Noun	narked	Adjective
livid	Adjective	needle	Verb
lividly	adverb	nerves	Noun
loathe	Verb	nerviness	Noun
loathing	Noun	nervous	Adjective
long	Verb	nervousness	Noun
longing	Noun	nervy	Adjective
look up to	Verb	nettle	Verb
love	Noun	nettled	Adjective
love	Verb	niggle (at so)	Verb
lovesick	Adjective	nonplus	Verb
love-struck	Adjective	nonplussed	Adjective
low	Adjective	nostalgia	Noun
low spirits	Adjective	obeisance	Noun
lugubriousness	Noun	obligated	Adjective
lust	Noun	odium	Noun
lust	Verb	offence/offense	Noun
luxuriate	Verb	offend	Verb
mad	Adjective	offended	Adjective
madden	Verb	on cloud nine	Adjective
maddened	Adjective	on edge	Adjective
make sb's blood boil	Verb	on tenterhooks	Adjective
make sb's hackles rise	Verb	on the warpath	Adjective
malaise	Noun	oppress	Verb
malcontent	Adjective	optimism	Noun
malice	Noun	out of sorts	Adjective
manic	Adjective	outrage	Noun
melancholic	Adjective	outrage	Verb

outraged	Adjective	pride yourself on	Verb
over the moon	Adjective	prize	Verb
overawe	Verb	protectiveness	Noun
overawed	Adjective	proud	Adjective
overexcite	Verb	psych	Verb
over-excited	Adjective	psyched up	Adjective
overjoyed	Adjective	punch-drunk	Adjective
overstrung	Adjective	put off	Verb
overwhelmed	Adjective	put sb's back up	Verb
overwrought	Adjective	puzzle	Verb
pain	Noun	puzzled	Adjective
pain	Verb	puzzlement	Noun
panic	Noun	qualm	Noun
panic	Verb	queasy	Adjective
panicked	Adjective	rabid	Adjective
panicky	Adjective	rabidly	adverb
panic-stricken	Adjective	rage	Noun
paranoid	Adjective	rancorous	Adjective
partial to	Adjective	rancour	Noun
partiality (for)	Noun	rancorousness	Noun
passion	Noun	rankle	Verb
passionate	Adjective	rapture	Noun
passionately	adverb	rattily	adverb
peevd	Adjective	rattle	Verb
peevish	Adjective	ratty	Adjective
peevishness	Noun	regard	Noun
penchant for	Noun	regret	Noun
penitent	Adjective	regretful	Adjective
perplex	Verb	regretfully	adverb
perplexed	Adjective	relish	Noun
perplexity	Noun	relish	Verb
perturb	Verb	reluctance	Noun
perturbation	Noun	remorse	Noun
perturbed	Adjective	remorseful	Adjective
pessimism	Noun	repel	Verb
petrified	Adjective	repentance	Noun
petrify	Verb	repentant	Adjective
petulance	Noun	repugnance	Noun
petulant	Adjective	repulsion	Noun
pine	Verb	resent	Verb
pique	Noun	resentfulness	Noun
pique	Verb	resentment	Noun
piqued	Adjective	resignation	Noun
pity	Noun	respect	Noun
pity (so)	Verb	respect	Verb
please	Verb	respectfulness	Noun
pleased	Adjective	revel in	Verb
pleasure	Noun	revere	Verb
possessiveness	Noun	reverence	Noun
predilection	Noun	revolt	Verb
prepared	Adjective	revulsion	Noun
pressured	Adjective	rile	Verb
prey (on)	Verb	riled	Adjective
pride	Noun	romanticise	Verb

rotten	Adjective	sorrow	Verb
rub up the wrong way	Verb	sorrowful	Adjective
rueful	Adjective	sorrowfully	adverb
ruffle	Verb	sorrowfulness	Noun
sad	Adjective	sorrowing	Adjective
sadden	Verb	sour grapes	Noun
saddened	Adjective	sourness	Noun
sadly	adverb	spite	Noun
sadness	Noun	spitefulness	Noun
satisfaction	Noun	spleen	Noun
satisfied	Adjective	spook	Verb
satisfy	Verb	squirmy	Adjective
savour	Verb	stagger	Verb
scandalise	Verb	staggered	Adjective
scandalised	Adjective	starry-eyed	Adjective
scare	Noun	startle	Verb
scare	Verb	startled	Adjective
scared	Adjective	stirred up	Adjective
scourge	Verb	straining at the leash	Adjective
seething	Adjective	stress	Noun
self-contempt	Noun	stress out	Verb
self-disgust	Noun	stressed	Adjective
self-dislike	Noun	strop	Noun
self-doubt	Noun	stroppiness	Noun
self-hatred	Noun	strung up	Adjective
self-loathing	Noun	stumped	Adjective
self-pity	Noun	stun	Verb
self-pitying	Adjective	stung	Adjective
self-reproach	Noun	stunned	Adjective
send over the edge	Verb	stupefaction	Noun
shake up	Verb	stupefied	Adjective
shaken	Adjective	stupefied	Adjective
shame	Noun	stupefy	Verb
shame	Verb	suicidal	Adjective
shamefaced	Adjective	sulk	Noun
shattered	Adjective	sulkiness	Noun
sheepish	Adjective	sulky	Adjective
sheepishness	Noun	surprise	Noun
shock	Noun	surprise	Verb
shock	Verb	surprised	Adjective
shocked	Adjective	suspense	Noun
show up	Verb	swear by sth	Verb
sick	Adjective	sweep so off their feet	Verb
sicken	Verb	sympathy	Noun
sickened	Adjective	take a dim view of	Verb
smitten	Adjective	take a fancy to	Verb
soft spot	Noun	take a shine to	Verb
sold on	Adjective	take aback	Verb
solicitude	Noun	take exception	Verb
sombre	Adjective	taken aback	Adjective
sombrely	adverb	taken with	Adjective
sombreness	Noun	tear apart	Verb
sore	Adjective	teed off	Adjective
sorrow	Noun	temptation	Noun

tense	Adjective	unnerved	Adjective
tensely	adverb	unquiet	Noun
tension	Noun	unsatisfied	Adjective
terrified	Adjective	unsettle	Verb
terrify	Verb	unsettled	Adjective
territorial	Adjective	up in arms	Adjective
terror	Noun	upbeat	Adjective
terrorise	Verb	uplifted	Adjective
terrorised	Adjective	upset	Noun
terror-stricken	Adjective	upset	Verb
testy	Adjective	uptight	Adjective
tetchy	Adjective	urge	Noun
thankful	Adjective	value	Verb
thankfully	adverb	venerate	Verb
thankfulness	Noun	veneration	Noun
thirst	Noun	vengeful	Adjective
thirst	Verb	venom	Noun
thirsty	Adjective	vex	Verb
thrill	Verb	vexation	Noun
thrilled	Adjective	vexed	Adjective
thrown off balance	Adjective	vindictive	Adjective
thunderstruck	Adjective	vindictiveness	Noun
ticked off	Adjective	wallow in	Verb
tickle	Verb	wanderlust	Noun
tickled	Adjective	wariness	Noun
tired	Adjective	warm/warm to	Verb
tizzy	Noun	warmness	Noun
torment	Noun	waspish	Adjective
torment	Verb	weakness	Noun
tormented	Adjective	weigh down	Verb
torture	Noun	well-disposed	Adjective
torture	Verb	whim	Noun
touched	Adjective	wholehearted	Adjective
traumatise	Verb	wilfulness	Noun
traumatized	Adjective	will	Noun
treasure	Verb	willies	Noun
trepidation	Noun	willing	Adjective
triumph	Noun	willingness	Noun
triumphant	Adjective	wired	Adjective
trouble	Verb	wish	Noun
troubled	Adjective	wistfulness	Noun
twitchily	adverb	with bated breath	adverb
twitchiness	Noun	woe	Noun
twitchy	Adjective	woeful	Adjective
umbrage	Noun	wonder	Noun
unease	Noun	work up	Verb
uneasiness	Noun	worked-up	Adjective
uneasy	Adjective	world-weariness	Noun
unglued	Adjective	worried	Adjective
unhappily	adverb	worry	Noun
unhappiness	Noun	worry	Verb
unhappy	Adjective	worship	Noun
unhopeful	Adjective	worship	Verb
unnerve	Verb	worshipful	Adjective

wound	Verb
wounded	Adjective
wrath	Noun
wrathful	Adjective
wretched	Adjective
wretchedly	adverb
wretchedness	Noun
wrought-up	Adjective
yearn	Verb
yearning	Noun
yearningly	adverb
yen (have a for)	Noun
zeal	Noun
zest	Noun
zing	Noun

## Notes

- 1 For research on affect/emotion and gender see Lutz (1990), Gallois (1994: 306-307), Anderson & Leaper (1998), Planalp (1999: 36), Goldshmidt & Weller (2000), Galasinski (2004), Oatley et al (2006: 246-248); on affective communication and social groups see Besnier (1990), Irvine (1990), Gallois (1994).
- 2 Oakey (2002) discusses some of the advantages/disadvantages of this approach: ‘The sampling methods used in its [the BNC] construction meant that partial texts were included, and thus some areas of discourse may not be represented equally. Nonetheless, it is fair to say that a subset of this size yields sufficient occurrences for meaningful comparisons to be made between sub-genres.’ (Oakey 2002: 115-116)
- 3 Abbreviations: Hu = humanities, me = medicine, ns = natural sciences, sos = social sciences, as = applied science, ple = politics, law, education; bs = broadsheets, art = arts/cultural material, fin = finance/commerce, nw = home/foreign news, sc = science, lst = lifestyle etc, sp = sports, tbl = tabloid, uk = unknown.
- 4 Crucially, there is no assumption that one ‘emotion term’ corresponds to one ‘emotion’ as defined in psychology or biology. Rather, emotion terms are ‘ways of speaking’ (Galasinski 2004: 6) or ‘discursive phenomena’ (Edwards 1999: 279), and ‘[I]exis provides categories of affect in the form of folk taxonomies.’ (Downes 2000: 108).
- 5 ‘There is certainly a core of prototypical emotions on which subjects will largely agree that they are emotions. But there is a large border area where judgements differ.’ (translation mine)
- 6 In retrospect, some of the words in the categories involving ‘interest’ could have been added to the list, as they can be dealt with under the affect sub-categories of dis/satisfaction: interest and ennui, even though these categories ‘take us to the borders of **affect**’ (Martin & White 2005: 50, emphasis in original).
- 7 ‘A run-on entry is [...] a word morphologically derived from a dictionary headword which is not itself defined, but is printed, along with its word class and possibly an example, usually in a bold typeface, at the end of the entry for the word it is derived from’ (Kilgarriff 1997a: 153).
- 8 6353 occurrences in conversation, 5347 occurrences in fiction.
- 9 13,060 occurrences in conversation, 10,461 occurrences in fiction.
- 10 10,787 occurrences in conversation.