Appendix to Chapter 6 of *Emotion Talk Across Corpora* (Appendix 6)

Contents:

A 6.1 Description of corpus (BRC baby)

The BRC baby consists of about 20,000 words from the registers of conversation, news reportage, fiction and academic discourse. Sampling procedures are described in Bednarek (2008: Chapter 5). More detailed information (including titles of books, type of sample, circulation size, level of difficulty, author gender, speaker age and occupation etc) is listed below:

Conversation (22,613 words)

Samples from 7 files:
KBC, part 1, 1-337
KBD, part 6, 1916-2243
KBP, part 5, 1233-2004
KD0, part 3, 125-346
KD8, part 15, 7154-7413
KE2, part 21, 3027-3281
KE4, part 2, 102-491

Further information on files from which samples were chosen:

KBC:
3019 words (AB) from 14 conversations recorded by 'Audrey' (PS1A9) between 2 and 9 April 1992 with 9 interlocutors, totalling 6341 s-units, 31,337 words, and 3 hours 38 minutes 41 seconds of recordings.

Speakers:

PS1A9 Audrey, 61, housewife, Lancashire

PS1AA Gordon, 61, teacher, Lancashire

PS1AB Margaret, 45, nurse, Lancashire

PS1AC Joan, 50+, clerk, Central Northern England

PS1AD Kevin, 29, computer engineer, Northern England

PS1AE Carl, 31, pharmacist, Northern England

PS1AF None ? ? ?

PS1AG Elaine, 28, housewife, Northern England

PS1AH Iris, 60, housewife, Lancashire

KBD

2690 words (AB) from 24 conversations recorded by 'Barry' (PS03W) between 1 and 6 February 1992 with 10 interlocutors, totalling 9021 s-units, 58,087 words, and 5 hours 12 minutes 10 seconds of recordings.

Speakers:

PS03W Barry, 41, entertainments consultant, Central Northern England

PS03X Terri, 35, bar staff, Home Counties

PS03Y Hugh, 30, bar staff, Irish

PS040 Alan, 38, security, Lancashire

PS041 None ? ? ?

PS042 Mark, 30, dj, London

PS043 Ken, 30, security, Lancashire

PS044 None, 35, housewife, Lancashire

PS045 Sergio, 9, student (state primary), Lancashire

KE 4:

2683 words (C1) from 22 conversations recorded by 'Valerie' (PS0WN) between 30 January and ?? ?? 1992 with 7 interlocutors, totalling 3280 s-units, 15,170 words, and 1 hour 55 minutes 21 seconds of recordings.

Speakers:

PS0WN Valerie, 36, staff nurse (pt), Scottish

PS0WP Peter, 34, sales representative, Scottish

PS0WR Jackie, 8, student (state primary), Scottish

PS0WS David, 10, student (state primary), Scottish

PS0WT Dawn, 11, student, Scottish

PS0WU None, 40+, dentist, Scottish

PS0WW None, 50+, telephone engineer, Scottish

PS0WX Dougie, 37, sales representative, Scottish

KD0:

2951 words (C1) from 106 conversations recorded by 'Kevin' (PS0HM) between 29 November and 5 December 1991 with 14 interlocutors, totalling 13,948 s-units, 77,692 words, and 10 hours 39 minutes 22 seconds of recordings.

Speakers:

PS0HM Kevin, 41, draughtsman, London

PS0HN Paul, 12, student (state secondary), London

PS0HP Ruth, 40, teacher, ?

PS0HR Michelle, 29, local government officer, European (French)

PS0HS Eric, 74, retired, London

PS0HT Adrian, 40, salesman, London

PS0HU Karen, 41, secretary, London

PS0HV Andrew, 33, local government officer, London

PS0HW Lisa, 13, student, London

PS0HX Babs (aka mutty), 70+, retired, Lower South-west England

PS0HY Joy, 70+, retired, London

PS0J0 Michael, 15, student, London

PS1KN None ? ? ?

KE 2:

2623 words (C2) from 153 conversations recorded by 'Terence' (PS0W2) between 20 and 27 February 1992 with 10 interlocutors, totalling 10,080 s-units, 77,961 words, and over 12 hours 49 minutes 22 seconds of recordings.

Speakers:

PS0W2 Terence, 70, retired (headteacher), East Anglia

PS0W3 Richard, 44, fireman, Lower South-west England

PS0W4 Margaret, 70, retired, Irish

PS0W5 Lucy, 13, student, Lower South-west England

PS0W6 Holly, 13, student, Lower South-west England

PS0W7 Adrian, 13, student, Lower South-west England

PS0W8 Danielle, 13, student, Lower South-west England

PS0W9 Christine, 40, housewife, Lower South-west England

PS0WA Mima, 50, housewife, Lower South-west England

KD8:

2654 words (C2) from 31 conversations recorded by 'Martine' (PS0LK) between 12 and 20 March 1992 with 10 interlocutors, totalling 10,787 s-units, 76,445 words, and over 7 hours 15 minutes 1 second of recordings.

Speakers:

PS0LK Martine, 25, senior technician, Welsh

PS0LL Mike, 28, construction worker, Welsh

PS0LM Merielle, 55, housewife, Welsh

PS0LN None, 45, pub landlord, Home Counties

PS0LP Harold, 58, engineer, Welsh

PS0LR Nora, 76, housewife, Welsh

PS0LS Will, 45, civil engineer, Merseyside

PS0LT Michael, 40, technical director, Home Counties

PS0LU Jim, 27, technician, Home Counties

KBP:

5993 words (DE) from 15 conversations recorded by 'Clarence' (PS065) between 13 and 19 March 1992 with 4 interlocutors, totalling 5039 s-units, 27,179 words, and 2 hours 23 minutes 42 seconds of recordings.

Speakers:

PS065 Clarence, 65, retired, Lancashire

PS066 Nina, 67, retired, Lancashire

PS067 Nev, 72, retired, North-east Midlands

PS068 Lil, 70, retired, Lancashire

News reportage (18,164 words)

30 articles from seven files (no tabloids):

A1E 1-145

A1G 1-158

A1M 1-138

A1N 1-160

A7 S 1-159

AL5 (all)

AJG 1 (all)

AL5 and AJG (Social): 3004 words

(1) 'Struggling dentists' pull more teeth
(2) Misconduct case GP to appeal
(3) 'The NHS is not for sale'
(4) Marital strife comes out in the wash
(5) CALLED TO ORDER
(6) Sex education for priests urged by Pope
(7) Homosexual prayer book to go on sale

A7S (Arts): 3003 words

(8) Forget Kylie — here come the crisp bags
(9) What's going on
(10) Jackboots beneath the serge
(11) OPERA: Edward Greenfield finds the Medea at Covent Garden relentlessly loud Cool, but not Callas enough.
(12) For third South arts QEH

A1N (Sports): 3058 words

(13) Cricket First Test: Azharuddin's daring defiance

(14) Motor Racing: Fighting Senna refuses to succumb

(15) Golf: Calcavecchia digs in for consolation prize

(16) Rugby Union: The dye is cast for injured Richards

(17) Football: FA investigates crowd trouble

A1M (science): 2747 words

(18) The revolution the countryside needs: Christian Wolmar says Britain is lagging behind in setting up 'telecottages'

(19) The busy sex life of the nice male

(20) Cool solutions for hot climates: David Spark looks at tropical temperature controls for vaccines

A1G (report): 3145 words

(21) Tigrayans advance on a helpless Addis Ababa

(22) Tribalism and liberation meet at Transkei burial

(23) Swapo remains favourite to win the United Nations-supervised elections next month

(24) Out of India: The snow-wreaths melt away in the heat of battle

(25) Russians send Kabul 2000 supply trucks

A1E (Commerce): 3207 words

(26) Latest corporate unbundler reveals laid-back approach: Roland Franklin, who is leading a 697m pound break-up bid for DRG, talks to Frank Kane

(27) Square Mile: Stock Exchange takes the bull by the horns

(28) View from Manhattan: Airlines fly into regulatory storm

(29) USM: Remedial action at Lincat has results

(30) MB and Caradon stay quiet on bid rumours

Fiction (20,563 words)

Beginning samples from 10 files:

AB9 1-145

AC2 1-105

BMW 1-133

C8T 1-100

CB5 1-144

CFY 1-152

FAJ 1-142

G0S 1-101

H9C 2-168

HR9 1-175

AB9 (2058 words)

Death of a Partner, Neel, Janet, Constable Company Ltd, London (1991)

= beginning sample, medium circulation size, female author, medium level of difficulty

AC2 (2159 words)

Man at the Sharp End, Kilby, M, The Book Guild Ltd, Lewes, East Sussex (1991)

= end sample, medium circulation size, male author, medium level of difficulty

BMW (1998 words)

Folly's Child, Tanner, Janet, Century Hutchinson, London (1991)

= beginning sample, high circulation size, female author, medium level of difficulty

C8T (2222 words)

Devices and Desires, James, P.D., Faber Ltd, London (1989)

= beginning sample, high circulation size, female author, medium level of difficulty

CB5 (2048 words)

Ruth Appleby, Rhodes, Elvi, Corgi Books, London (1992)

= middle sample, high circulation size, female author, medium level of difficulty

CFY (1999 words)

My Beloved Son, Cookson, C., Corgi Books, London (1992)

= middle sample, high circulation size, female author, medium level of difficulty

FAJ (2047 words)

Masai Dreaming, Cartwright, J., Macmillan Publishers Ltd, Basingstoke (1993)

= middle sample, medium circulation size, male author, high level of difficulty

G0S (2111 words)

Indigo, Warner, Marina, Chatto Windus Ltd, London (1992)

= middle sample, medium circulation size, female author, high level of difficulty

H9C (1933 words)

The Prince of Darkness, Doherty, P.C., Headline Book Publishing plc, London (1992)

= middle sample, medium circulation size, male author, medium level of difficulty

HR9 (1988 words)

They Came from SW19, Williams, Nigel, Faber Ltd, London (1992)

= end sample, high circulation size, male author, medium level of difficulty

Academic discourse (23,781 words)

Beginning samples from 10 files:

A6U 1-109

ACJ 1-101

ALP 1-101

AS6 1-80

EA7 1-93

EWW 40-139

FC1 1-83

FEF 1-121

FPG 1-101

HWV 1-93

A6U: 2411 words from:

'Being Drawn to an Image', Guy Brett, *Oxford Art Journal* (1991)

sample type unknown, from periodical, multiple authors, high difficulty

ACJ: 2666 words from:

Principles of Criminal Law, Andrew Ashworth, OUP, Oxford (1991)

= middle sample, from book, male author, high difficulty

ALP: 2285 words from:

'A Non-punitive Paradigm of Probation Practice: Some Sobering Thoughts', L.R. Singer, *British Journal of Social Work* (1991)

= middle sample, from periodical, multiple authors, high difficulty

AS6: 2073 words from:

Tackling the Inner Cities, Ben Pimlott and Susanne MacGregor, OUP, Oxford (1991)

= beginning sample, from book, multiple authors, high difficulty

EA7: 2511 words from

France in the Making, 843-1180, Jean Dunbabin, OUP, Oxford (1991)

= middle sample, from book, female author, medium difficulty

EWW (without foreword: tribute): 2299 words from:

Matrices and Engineering Dynamics, A. Simpson and A.R. Collar, Ellis Horwood Ltd, Chichester (1987)

= beginning sample, from book, multiple authors, medium difficulty

FC1: 2333 words from:

'In re A DEBTOR (NO. 784 OF 1991) 1992 April 13', J. Hoffmann, *The Weekly Law Reports*, vol 3 (1991)

= sample type unknown, from periodical, author details unknown, high difficulty

FEF: 2124 words from:

Lectures on Electromagnetic Theory, L. Solymar, OUP, Oxford (1984)

= beginning sample, from book, male author, high difficulty

FPG: 2293 words from:

Design of Computer Data Files, O. Hanson, Pitman Publishing, London (1989)

= middle sample, from book, male author, high difficulty

HWV: 2786 words from:
'Immunogenicity of a Supplemental Dose of Oral Versus Inactivated Poliovirus Vaccine'
The Lancet, London (1993)
= unknown sample type, from periodical, multiple authors, medium difficulty

A 6.2 Description of methodology

The data was analyzed and coded with the help of *Altova XMLSpy 2007* (www.altova.com), an XML editor software, which allows you to tag data with a number of attributes (Bednarek 2008: Chapter 5). Each emotion term was coded on nine linguistic variables:

1) Affect type
2) Affect trigger
3) Covert[1]-overt affect
4) Emoter
5) Hypotheticality
6) Negation
7) Part of speech[2]
8) Valence
9) Speech act

Remarks on affect type, covert and overt affect and valence are made in Bednarek (2008: Chapter 5), so that this section focuses on hypotheticality, negation, and speech act, with the analysis of the remaining variables being relatively straight-forward and in no need of further elaboration.

Hypotheticality

Under the heading of *hypotheticality* the analysis focused on whether an emotion was described as being experienced (in the past or present) in reality, or whether its experience was predicted (in the future) or just hypothesized (in a possible world). Table A.21 shows typical analyses of emotions as 'hypothetical', 'will' (future) or 'real':

Table A.21: Analysis of hypotheticality

Coding	**Typical analyses**
'Hypothetical'	deontic and dynamic modality when referring to non-real, non-actualized emotion; hypotheticality; intention etc: *would/'d*, *could* (past + future), *should*, *can*, (*in order*) *to*, *the purpose was to*, *have to*, *must be*, *might*, *in any desired order*, *ought to*, *impossible to*, *if*, *as if*, *as often as she wished*, *had been about to*, *unless*, *wanting* someone *to feel*, *for* ('in order to'), *try to look like a man who enjoyed*, *designed to*, *test whether*, *whether or not x is around to be impressed*, *the opportunity to*, *would work for*, *I want to see…*, *inclinations towards*, *prevent*, *whatever you want*
'Will'	*won't*, *shall*, *'ll*, *will*, … *ahead*
'Real'	everything else, including evidentiality and epistemic modality, reported emotions, e.g.: *seem*, *evidence of*, *I bet*, *on the face of it*, *predict*, *obviously*, *I understand* (that), *a message that*, *you don't say you love*, *accuse of*, *perhaps*, *I thought you liked it*, *apparently* etc[3]

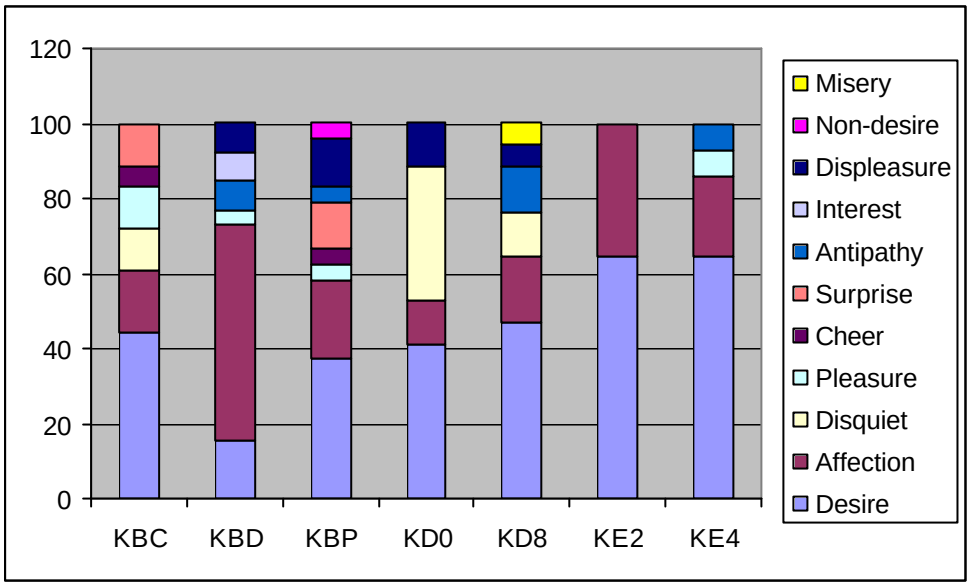
Negation

The categorization of negation relates to whether or not an emotion is negated. As noted in Bednarek (2008: Section 5.3.2.2), there is a distinction in appraisal theory between negative emotions (*sad*) and negated positive emotions (*not happy*), with the latter coded as 'neg + hap' rather than '-hap'. Thus, 'negated' relates to grammatical negation (*not*, *no*, *never* etc) whereas 'non-negated' includes morphological (negative prefix (*un-*, *in-*, *dis-*) or suffix (*-less*)), and lexical negation (such as *disinclined*, *refuse*, *reluctant*, *dislike*, *doubtfully*). On emotion words and negation see also Nöth (1992).

Speech act

Finally, the category that I have labelled 'speech act' here purely relates to whether the emotion is questioned (e.g. *do you **want** x*, *do you **love** me*?) or asserted (e.g. *I am **furious***). This does not in fact correspond to whether the speech act as such is a question or not; for instance examples such as *Why does it interest you? What do you want*? are questions but contain asserted emotions (in contrast to *Does it **interest** you? Does she **want to** collect you*?). Only if the experience of the emotion is questioned to some extent was this coded as 'question' rather than 'statement' (tag questions were not counted as questions either, relating to mitigation/hedging). 'Question' also includes reported questions, for example *consider whether they really **wanted** to*.

A 6.3 Affect sub-types in the conversation sub-corpus of the BRC baby



	Desire	Affection	Disquiet	Pleasure	Cheer	Surprise	Antipathy	Interest	Displeasure	Non-desire	Misery
KBC	44.4	16.6	11.1	11.1	5.5	11.1	0	0	0	0	0
KBD	15.4	57.7	0	3.8	0	0	7.7	7.7	7.7	0	0
KBP	37.5	20.8	0	4.2	4.2	12.5	4.2	0	12.5	4.2	0
KD0	41.2	11.8	35.3	0	0	0	0	0	11.8	0	0
KD8	47.1	17.6	11.8	0	0	0	11.8	0	5.9	0	5.9
KE2	64.7	35.3	0	0	0	0	0	0	0	0	0
KE4	64.3	21.4	0	7.1	0	0	7.1	0	0	0	0

(Figures refer to percentages of affect sub-type with respect to all emotion terms in given file; for instance, 44.4% of all emotion terms in KBC realize Desire)

A 6.4 Applications and implications

In Bednarek (2008: Chapter 6) I noted that the findings might have some implications for:

- the application of appraisal theory;
- the modelling of probabilistic (intra- and inter-) register variation;
- natural language processing (automated register recognition, parsing of affect);
- language teaching and lexicography.

The following sections provide a more detailed discussion of these aspects.

Appraisal theory

In terms of appraisal theory, there is a need for discourse analysts to try out the new classificatory system of affect (including the nine variables identified above), in order to test its applicability, its advantages and disadvantages. More (theoretical) research is also called for in the areas of:

- Surprise and counter-expectation: should this be established as an evaluative (appraisal) system in its own right?
- Types of covert affect: what is the usage and patterning of, and difference between nouns such as *worries*, *disappointments*, adjectives such as *worrying*, and adverbs such as *sadly*? Are there also verbs that indicate covert affect?
- Authorial vs. non-authorial appraisal: does non-authorial appraisal involve intersubjectivity (engagement)?[4]
- Appraisal and grammatical metaphor: when do appraisal expressions construe modality, when affect/engagement (see Martin 1992, 2000b, Martin & White 2005: 54-56)?
- What is the role of talk about *not* experiencing an emotion, rather than talk about experiencing an emotion? Galasinski notes that in his data there is a 'narrative awareness of the fact that certain events in people's lives […] might be associated with certain emotional experiences and they explicitly acknowledge this by denying having these experiences.' (Galasinski 2004: 83). How is this related to appraisal? It might also be interesting to examine talk about hypothetical emotions (Galasinski 2004: 84). Since the BRC baby was coded for hypotheticality (see above), this aspect can easily be investigated in the future.
- Affect and intensity (graduation): how are degrees of intensity conveyed in emotion talk? This is likely to be a huge research project: 'Studying language intensity in emotion talk will be challenging because modifiers of emotion terms, inflectional and intonational changes, and even exclamations will likely need to be considered' (Anderson & Leaper 1998: 443). For already existing studies on intensity/involvement see Bednarek (2008: Section 1.3).
- Appraisal and *polyphony* (Downes 2000): what is the interplay between different evaluative (appraisal) systems and can a typological description capture this?
- Appraisal and textual structure: how do the prosodic structure of interpersonal meaning, and the periodic structure of textual meaning interact? How is evaluative meaning

> distributed in texts (elaborating on research by Martin 1992, 1997, 2002a, 2004, Macken-Horarik 2003: 317, Martin & White 2005: 85-89)?

Current research addresses these issues in more detail (e.g. Bednarek 2007). It must also be pointed out that the focus of Bednarek (2008) was on *emotion* talk (the use of emotion terms) rather than *emotional* talk (compare Bednarek 2008: Chapter 1), though there is a whole range of resources for emotional talk without the use of emotion terms: 'An explicit emotion vocabulary is not necessary for powerful displays of emotion with language in its full pragmatic environment' (Goodwin & Goodwin 2000: 254).

Modelling register variation

Findings about frequencies in text are useful for a variety of reasons, providing us with information about the semiotic system itself (Halliday 2005: 45) and the modelling of register variation. The establishment of probability profiles (in Bednarek 2008: emotion profiles) has implications for 'at least five areas of theoretical enquiry: developmental, diatypic, systemic, historical and metatheoretic.' (Halliday 2005: 73). Knowing about the frequency of words is also important in various areas of language teaching, for example the design of curricula, the writing of materials and the testing of language proficiency (Leech et al 2001: ix), and in the fields of natural language processing, linguistics, psychology, and cultural studies (Leech et al 2001: x). Leech et al also point out that

> for the various uses of frequency information mentioned earlier, particularly in the educational arena, we need to reckon on different frequency profiles for different varieties of the language. The idea that one monolithic frequency list for the whole language can satisfy all needs is, of course, unrealistic. (Leech et al 2001: xi).

As we have seen in Bednarek (2008), there is some variation in the emotional profiles of the four registers investigated here. If we want to include such register variation in a description of affect, how can we model it? SFL lends itself quite well to probabilistic modelling, since it recognizes that there are probabilistic tendencies in language, and recently several studies have addressed the issue of modelling co-selection or intersections of systems (e.g. Matthiessen 2006, Tucker 2006). It is suggested that system networks can be used to represent probabilistic tendencies:

> the system network can represent not only the dependency of one system on another or others, but also any probabilistic correlation between any two features anywhere in the network. We are thus able to represent the relationship between choices in the various systems, such as transitivity, tense, mood etc., which are manifested through syntagmatic co-occurrence. […] Importantly, the notion of **pre-selection** is introduced into the grammar. The selection of any feature, or combination of features, may lead to the pre-selection of subsequent features, either in terms of absolute pre-selection or of the setting of probabilities on the features in the system(s) in question. (Tucker 2006: 92; emphasis in original).

The modelling of probabilistic variation is arguably an integral part of any description of the lexico-grammatical resources of a language (Matthiessen 2006). Additionally, it may provide input for any future computational linguistic modelling. Even though networks are specifically used in SFL, they also represent a user-friendly device to taxonomize probabilistic tendencies of corpus-linguistic (CL) evidence regardless of the theory researchers adhere to. For examples of probabilistic modelling in SFL see Tucker (2006), and Matthiessen (2006). For a discussion of SFL vs. CL principles see Hunston & Thompson (2006), especially Hunston (2006). However, more research is necessary on intra-register variation (see Bednarek 2008: Chapter 6 for preliminary comments). The crucial question is whether there are more similarities between texts across registers than between texts within a certain register. This should be determined with the help of sophisticated statistical measures (Biber 1989, Kilgarriff 2001).

Natural language processing

Automated register recognition

Can the findings about the frequency of emotion terms in different registers help automated register recognition? A good overview of previous research and criteria on differentiating between text types is given by Stubbs & Barth (2003: 79). It seems likely that in as far as emotion terms are content words, they are at least partly dependent on topic, and 'may therefore be frequent in an individual text, but absent in another text from the same text-type.' (Stubbs & Barth 2003: 67). Further, the kind of analysis that I have undertaken was meaning-sensitive, and cannot yet be automated. More promising are word-chains (Stubbs & Barth 2003: 62) or lexico-grammatical patterns/local grammars as described in Chapters 3 and 4.

Parsing affect

As Hunston & Sinclair propose, parsers can be developed on the basis of local grammars that can automatically extract information from texts (e.g. the automatic retrieval of definitions from texts) (Hunston & Sinclair 2000: 82). A parser that could be developed on the basis of the description in Bednarek (2008: Chapter 3),[5] i.e. that is programmed with patterns that allow it to automatically produce a simple table or list of emoters, emotions and triggers from texts, might be useful enough (though it is important that negation be included to capture the distinction between *love* and *no/not love* – but this is easy when pre-processed corpora are used). Although the relations between emoter, emotion and trigger may vary according to the particular emotion involved, the human analyst can decide what these relations are, on the basis of the output of the parser and his/her linguistic and non-linguistic knowledge. If the parser output is emotion = *love*, trigger = *you*, the relation is one of direction; whereas if the parser output is trigger = *the results*, emotion = *surprise*, the relation is one of cause.
Ultimately, however, more detail (though not all) that can be found in FrameNet could be added to parsers. For example, if the degree element could automatically be parsed by the software, it would be possible to produce a list of emoters, emotions, triggers *and* the degree of emotion involved. In terms of appraisal theory, the results of this parsing would show both affect and graduation – two of the sub-systems of appraisal (Bednarek 2008: Section 1.4). More and more details could gradually be included (from FrameNet and other corpus research) to make the parser more and more sophisticated (to enable it to deal with variations of patterns, e.g. pseudo-clefts, changed word-order and so on. Compare Francis et al 1996: 611-615).

Since I am neither a computational linguist nor an AI researcher, I cannot authoritatively discuss how easy or difficult the development of such a parser could be (for a discussion of some issues regarding a local grammar of evaluative adjectives see Hunston & Sinclair 2000: 82, and Hunston 2002: 180). However, the following factors could cause some problems for such an application:

- Patterns come in different forms and are changed by processes such as clefting, fronting, passivization etc (see Francis et al 1996: 611-615, Hunston & Francis 2000).
- The difference between undirected and directed affect patterns is superficial: triggers of presumably undirected affect patterns may be explicitly stated in the context or inferable by readers/hearers.

- Patterns can have different mappings depending on the lexical item or meaning group involved (see also Hunston 2003: 7, Hunston & Sinclair 2000: 88), e.g.:

(n) V n:

(i) emoter emotion trigger (*I admired them like I admire Tom Wolfe,* BNC, CHA 2382)

(ii) trigger emotion emoter (*Her reaction surprised me*, BNC, H0D 2326)

ADJ *for* n:

(i) emotion *for* trigger (*Regan, anxious for an alternative theoretical platform from which to put into orbit his conviction*, BRC, CM8 491; *enthusiastic for young athletes to do well*, BRC, CH6 370; *desperate for money*, BRC C8E 198, *anxious for her sons*, BRC, CCD 2410; *willing for more fun,* BRC, FPF 2661)

(ii) emotion *for* empathy target (*I am happy for him*, BNC, H8G 790; *I am delighted for you*, BRC, AE0 2465; *I'm pleased for you*, BRC, KCX 644; *I'm frightened for him*, BRC, CH3 3192; *I am very disappointed for Jimmy*, BNC, AHC 662).

ADJ n

(i) emotion emoter (*anxious faces*, BRC, H8A 1451)

(ii) emotion trigger (*sad sound*, BNC, G0Y 2062)

Thus, the order of the elements emoter, emotion, and trigger varies with directed affect patterns (depending on the verb), and the pattern ADJ *for* can express both directed (with trigger) and undirected affect (with empathy target). In such cases, the parser has to be given lists of lexical items that occur with each mapping. With the pattern ADJ n, the matter is more complex, since it seems to depend on the noun whether directed or undirected affect is concerned. Research would be necessary to determine which kinds of nouns typically function as emoters and which as triggers.[6]

- The local grammar description involves patterns (e.g. passives **PV *by* n**, basic patterns such as **ADJ n**) that also occur in areas outside affect:

PV *by* n	He … was respected and admired by all of his colleagues in the forces (FrN) (affect)
PV *by* n	After their divorce Jane worked on, helped by students at the growing research station. (BNC, A7D 448)
ADJ n	They were admitted by a surprised servant (BRC, CD2 1395) (affect)
ADJ n	The old man and the girl are listening attentively. (BNC, A04 1362)

Again, the parser would have to be given lists of lexical items inscribing affect to distinguish such patterns.

- In some cases, it depends on the context whether *when*-clauses and *because*-clauses can be mapped as trigger or not. In examples 1-2 below the clauses seem to express circumstances of time with respect to an emotion directed by an emoter at a trigger, rather than triggers of emotions themselves (as in examples 3-5):

(1) **When she was quite small** [circumstance] she had been surprised [emotion] to see the woman carrying a folded table and a Gladstone bag [trigger]. (BRC, AEA 176)

(2) **When he looked up** [circumstance], he seemed surprised [emotion] to find Lacuna still staring at him [trigger]. (BRC, F9X 3448)

(3) **When presently the young person came in** [trigger], he was definitely surprised [emotion].

(4) She was asked to study, so she did, and seemed surprised [emotion] **when her application produced good marks** [trigger]. (BRC, HWE 2039)

(5) He looked even more surprised [emotion] **when Elaine walked into the lounge with two men** [trigger]. (BRC, FAB 2451)

Similarly, *because*-clauses can either refer to the triggers of an emotion (you are disappointed because something happened), as in examples 6-7 or they can be used to elaborate *why* someone feels disappointed at a trigger (examples 8-9):

(6) I'm really disappointed [emotion] **cos I don't think my mum will let me stay now!** [trigger] (BRC, KCE 251)

(7) I was really disappointed [emotion] **cos erm I didn't get that one from Mencap** [trigger] (BRC, KCP 5756)

(8) […] last time on a, I was really disappointed [emotion] I'd been refused the course [trigger] **cos I thought oh, it'll be right up my street that** (BRC, KBW 15192)

(9) In fact, we're disappointed [emotion] with it [trigger] **cos usually we have vegetables out of our own garden all the year round** (BRC, KC0 5317)

Since it seems that *when*- and *because*-clauses are only triggers when no other triggers are identifiable in the sentence (examples 3-7), the parser must be trained to include *when*- and *because*-clauses as triggers **only** if no other triggers have already been parsed.

However, if a way can be found of solving these (and perhaps other) problems (maybe through combining a thesaurus-based seed list of emotion terms with information on patterns), a parser could be developed that automatically retrieves inscribed affect from texts, listing emoters, emotions and triggers. As John Patrick (p.c.) has noted, sentiment classification is of

growing interest in computational linguistics, with an increase in publications about detecting and classifying sentiment and subjectivity in discourse (compare also Taboada & Grieve 2004, Whitelaw et al 2005, Bloom et al 2006), and many important conferences have been held on this topic. The analysis of appraisal has a wide-range of applications, for example 'data and web mining, market research, and customer relationship management' (Whitelaw et al 2005: 1). Some of these approaches already use appraisal theory and include different appraisal *attributes* such as *attitude type*, *orientation*, *force*, *polarity* and *target type* (Taboada & Grieve 2004, Whitelaw et al 2005, Bloom et al 2006, focussing on adjectives).

Language teaching and lexicography

Language teaching

Pattern-based descriptions of local grammars are also potentially interesting for language teaching purposes. Patterns, as Hunston (2002: 173-174) notes, can be taught to students to improve both fluency and accuracy. The particular advantage of a local grammar approach to patterns lies on the one hand in their transparent, very 'user-friendly' category labels, i.e. semantic labels such as *emoter*, *emotion*, and *trigger* that directly reflect the discourse function of sentence elements (which the traditional categories of subject, verb etc do not). On the other hand, the organization of local grammars according to meaning means that exercises can be developed that cover a particular area of meaning – one that might be important to the particular students involved. Hunston (2002) discusses some examples of teaching methods appropriate to pattern teaching (without a specific discussion of local grammars).

Lexicography

A number of questions for lexicography seem to arise from the findings outlined in Bednarek (2008). The first concerns the variability of lexico-grammatical patterns investigated in Bednarek (2008: Chapter 4). While some of these terms show a very strong preference for **inter-register stability**, others demonstrate **inter-register variation**. In other words, with some

terms there is not much variation in terms of the most common L1 or R1 collocate, whereas others do vary across registers.

Table A.22 below sums up the most frequent L1 collocates of the analyzed emotion terms in the four corpora. As we can see, *enthusiastic*, *hate* (V) and *delighted* (and perhaps also *admire*, *surprised*, and *surprise* (V)) exhibit relatively stable L1 collocates across at least three corpora (red), whereas other emotion terms are more varied (e.g. *surprise* (N), *anxious*, *disappointed*, *impressed*, *pleased*, *willing*, *affection*, *frightened*). That is, we can make a distinction between lexico-grammatically stable (**LG-stable**) and lexico-grammatically volatile (**LG-volatile**) emotion terms.

Table A.22 L1 collocates of emotion terms across corpora

	L1			
	A	C	N	F
surprised	be	**'m**, (not)	be, (was)	was, (be)
surprise (N)	**no**	a	a	**in**
surprise (V)	(not)	n't, (it)	not	**to**
admire (V)	to	NN (I, and)	I, (to)	to
anxious (A)	was	NN (was)	**is**	was
disappointed	be	**(a) bit**	be, **very**	**was**
enthusiastic	an	NN (very)	an	an
impressed	**be**	**very**	**so**	**was**
pleased	**be**	very	very	**was**
willing	be	**are/re**	be	**was**
affection	and	NN (had, no)	with, and, 's	**of**, with
hate (V)	NN (I)	I	I, (he)	I, (she, he)
delighted	NN (be)	was	was	was
frightened	NN (too, BECOME)	was	**are**	was
hate (N)	NN	NN	NN	**of**

Considering only LG-volatile terms, news reportage shares collocates with conversation (*a surprise*, *very pleased*) (pink), with academic discourse (*be disappointed*, *be willing*, *and affection*) (green) and fiction (*with affection*) (blue). Conversation shares two L1 collocates with news reportage (as mentioned), and one with fiction (*was frightened*) (grey), but none with academic discourse. Fiction shares collocates with news reportage and conversation (as noted), and with academic discourse (*was anxious*) (yellow).

Moving on to R1 collocation, Table A.23 (on the next page) sums up their distribution across the four corpora.

Table A.23 R1 collocates across corpora

	R1			
	A	C	N	F
surprised	**that**, **by**, (to)	if	to, (**at**, if)	to
surprise (N)	to, (that)	**for**	to, (that)	to, (**and**)
surprise (V)	**us**	me	**the**	me, **him**, (**her**)
admire (V)	the	NN (it)	the	the
anxious (A)	to	NN (to)	to	to
disappointed	by, with, that	with	by	that, (**when**)
enthusiastic	about	NN (about)	about	about
impressed	by	**with**	by	by
pleased	to	**with**	to	to
willing	to	to	to	to
affection	**and**	NN	for	for
hate (V)	NN (by, the, their, to)	it, to	the, (to)	the, (to, it)
delighted	NN (with)	to	to, (with)	to
frightened	NN (to)	of	of	of
hate (N)	NN	NN	NN	**for**

As becomes apparent here, three emotion terms are very much LG-stable (collocates shared by all registers): *anxious* (*to*), *enthusiastic* (*about*) and *willing* (*to*). An additional eight emotion terms are relatively LG-stable (sharing R1 collocates among three registers: either all written registers or all except academic discourse): *surprised*, *surprise* (N), *admire* (V), *impressed*, *pleased*, *hate* (V), *delighted* and *frightened*. Other than that, we find LG-volatile terms such as *surprise* (V) and *disappointed*. Concerning these, news reportage shares one collocate with conversation (*surprised if*), and two with academic discourse (*surprise that*, *disappointed by*) and fiction (*affection for*, *hate the*). Conversation (apart from sharing with news reportage) also shares two collocates with fiction (*surprise me*, *hate it*), and one with academic discourse (*disappointed with*). Fiction shares with conversation and news reportage (as noted), and also has one collocate in common with academic discourse (*disappointed that*). Summing up all shared collocates, the picture looks as Table A.24 (on the next page) visualizes. We can see that academic discourse and news reportage are most similar (five shared collocates). Conversation and news reportage, conversation and fiction, and news reportage and fiction are also quite similar, sharing three collocates, respectively. This is perhaps best explicable by the fact that both news reportage and fiction contain dialogue (quotations by news actors and character dialogue) which mimic conversation, whereas both

also involve descriptions of emoters (news actors or characters). Furthermore, the news story is, after all, also a particular type of story/narrative. Finally, fiction and academic discourse share two collocates, and conversation and academic discourse share only one,[7] being relatively dissimilar (compare also the findings of the large-scale quantitative analyses in Chapter 2).[8]

Table A.24 Shared L1 and R1 collocates across corpora

N° of shared L1 and R1 collocates (where variation occurs)			
Academic	*Conversation*	*News*	*Fiction*
✔ ✔ ✔ (L1) ✔ ✔ (R1)		✔ ✔ ✔ (L1) ✔ ✔ (R1)	
	✔ ✔ (L1) ✔ (R1)	✔ ✔ (L1) ✔ (R1)	
	✔ (L1) ✔ ✔ (R1)		✔ (L1) ✔ ✔ (R1)
		✔ (L1) ✔ ✔ (R1)	✔ (L1) ✔ ✔ (R1)
✔ (L1) ✔ (R1)			✔ (L1) ✔ (R1)
✔ (R1)	✔ (R1)		

What are the implications of these findings for lexicography? If senses of words differ in various corpora (Kilgarriff 1997b: 108), and if lexico-grammatical patterns are register-sensitive, we could develop dictionaries that reflect this. This could take the form of an electronic dictionary (with a query system similar to that described by Teubert 2004) that includes five different dictionaries: one 'general' dictionary that lists meanings and lexico-grammatical patterns which are common and stable across registers, and four 'register' dictionaries (one each for conversation, news reportage, fiction, and academic discourse) that list meanings and lexico-grammatical patterns that are common in and characteristic for the respective register. The general dictionary should ideally include frequency bands for each register (note that the *Longman Dictionary of Contemporary English* (LDOCE) already makes a difference between written and spoken language frequency). Importantly, all 'sub-dictionaries' should include detailed information for meanings, patterns and pragmatic functions as well as authentic examples in the given register. Current dictionaries and other learner sources include a lot of corpus-based information about frequency (Neale 2006: 149-155) and collocation but this is not always referred to and can then remain relatively worthless to the user (Sinclair 2004c: 143). And some corpora that are the sources of dictionaries consist largely of media language such as the Bank of English (Mahlberg 2004: 119), which has some implications on the resulting description of English.

In other words, what I propose is that the approach that Biber et al (1999) have taken with respect to grammar can and should be extended to dictionaries, especially in view of the more or less unlimited space available when they are produced in electronic rather than in printed form. Thus, when learners read a newspaper or a novel, they can choose the 'news

dictionary' or the 'fiction dictionary'; when they want to focus on spoken conversation, they use the 'conversation dictionary'; and academic learners will need mostly the 'academic dictionary'.[9]

Other potential applications

Other potential applications of some of the findings in Bednarek (2008) include the development of guidelines or exercises for fiction writing in English, with respect to characterization and the portrayal of character emotion. It would also be possible to use the affect classification in Chapter 5 for a character analysis à la Toolan (2001). In analogy to Toolan's proposal for a character-trait inventory, we could make use of an emotional response inventory. One of Toolan's character-trait categories is +/- emotional, and this could be elaborated to incorporate the kind of emotionality that is present or absent (Table A.25 below). In so doing, we can build up an exact emotional picture of a character, contributing to the analysis of characterization in works of fiction.

Table A.25

✓ = affect type present
× = affect type absent
+ = positive affect sub-type present
- = negative affect sub-type present

	Character 1	*Character 2*	*Character 3*	*Character 4*	*Character 5*
un/happiness	✓	×	✓	×	✓
affection/antipathy	+ (affection)	×	- (antipathy)	×	- (antipathy)
cheer/misery	- (misery)	×	- (misery)	×	+ (cheer)
in/security	×	✓	×	✓	✓
quiet/disquiet	×	-	×	-	+
trust/distrust	×	-	×	+	-
dis/satisfaction	✓	×	✓	×	×
interest/ennui	-	×	-	×	×
pleasure/displeasure	-	×	+	×	×
dis/inclination	✓	×	✓		✓
desire	+	×	-		-
non-desire	-	×	-		+
surprise	✓	-	✓	✓	-

Finally, another application outside linguistics concerns the frequency findings of emotion terms in Bednarek (2008: Chapter 2). Rather than basing their analyses on elicited or free-listed emotion terms, psychologists and anthropologists could work with those emotion terms that are most frequently used in conversation.

Notes

1 With respect to covert affect note also that not all terms that are derivationally related to emotion terms should be included as covert affect. With some, the emotional meaning has been bleached to a large extent; e.g. *pity* in *It is a pity that…* should not be coded as covert affect but rather as judgement even though it shares patterns with covert affect. Examples of terms that were excluded from the analysis of covert/overt affect are (Table N.1):

Table N.1: Excluded terms

Excluded from (covert/overt) affect	*dazzling, nae bother, was a real stunner, jarring, trouble, aspirant members,* x *is a pain, jollification, buoyant markets, striking, strikingly, x was found wanting, pleasant, pleasure, pleasurable, awesome, preoccupation, delightful, their own interests*

2 When analyzing part of speech, a rough distinction was made between the POS categories adjective – adverb – verb – noun – other (relating to idioms). In order to facilitate the analysis, *-ed* and *-ing* verb participles (such as *impressed by, admiring, dazzling, cheered, brightening, much-loved, fed-up with, troubled by, tempted to)* were automatically coded as adjectives (compare Bednarek 2008: Chapter 3, but see Osmond 1997: 112).

3 *Would you like* was analyzed as 'real' desire rather than 'hypothetical' affection.

4 The terms *authorial* (1st person) and *non-authorial* (2nd & 3rd person) affect are suggested by White (2001a). It becomes clear from White's discussion that the distinction has to do with sourcing and the claiming or transferral of responsibility:

- With authorial affect, 'the writer is the source of the emotion by which the evaluation is conveyed and hence takes some responsibility for that evaluation' (White 2001a: 6);
- With non-authorial affect, '[t]he writer presents herself as merely **reporting** on the emotional reactions [of others]' (White 2001a: 6, emphasis mine); 'It is, in a sense, an **attributed** evaluation, responsibility of which has been transferred to an external source' (White 2001a: 6, emphasis mine).

For a discussion that relates non-authorial affect to engagement (attribute: acknowledge) see Bednarek (2007).

5 The approach used in Bednarek (2008: Chapter 3) for the description of affect patterns was more detailed than local grammar (and less 'strict' about the notion of patterns), but less detailed than FrameNet. It used the local grammar approach to functionally organize and describe one area of meaning (affect), and the FrameNet approach to include more detail for the parsing of the affect elements:

Inspired by local grammar	**Inspired by FrameNet**
description of one area of meaning (a unified description)	include non-patterns
functional orientation	include patterns of items that co-occur with emotion term
meaning groups	include some collocates that are part of specific (and important) patterns, e.g. expressor

Along the lines of suggestions by Hunston (2003: 16), the functional local grammar approach was used to 'tidy up' and 'organize' the details emerging from FrameNet and corpus analyses, and some non-patterns were labelled as local grammar elements (e.g. **V n *in* n**: *We admire this characteristic in others,* BNC, B19 1079), whereas phrases that are part of noun phrases/adjective phrases (e.g. *Rory hated his mother in a hat*, BRC, G0A 829) were not listed, e.g. as **V n *in* n** (though sometimes the difference may be difficult to capture). The resulting analysis improved the delicacy of the pattern approach, which 'given the impressive scale of this work, […] is often rather coarse' (Stubbs 2001: 460), and was more concise than the FrameNet approach, which can be overwhelming in its detail.

6 In order to investigate this phenomenon further I analyzed the right-hand noun collocates for four emotion adjectives in detail: *anxious*, *disappointed*, *delighted* and *surprised*. Of these four adjectives, three occur only with nouns that indicate an emoter. This noun is either a conscious participant (or metonymically an institution, a collective), an expressor, or a linguistic/mental activity:

Disappointed (excluding *disappointed expectations*)

- noun = conscious participant (e.g. *birdwatcher*, *competitors*, *English graduate*, *ghost*, *man*, *parents*, *pigeons*, NAME, *cleric; clubs* (metonymic))
- noun = expressor (e.g. *eyes*)
- noun = linguistic/mental phenomenon (e.g. *editorial*, *huffing and puffing*, *rage*)

Delighted

- noun = conscious participant (e.g. *grandfather*, NAME, *boy*, *fan, parents, Republicans*, *workers*)
- noun = expressor (e.g. *cackle*, *cascade* [of laughter], *countenance*, *face*, *kisses, laugh, laughter*, *squeal*, *whoop, kisses*)
- noun = linguistic/mental phenomenon (e.g. *anticipation*, *perception*, *disbelief*, *outrage*, *wonder*)

Surprised

- noun = conscious participant (e.g. *beast*, NAME, *servant*, *young men*, *bosses*, *patients, Toronto fan*)
- noun = expressor (e.g. *expression, look*, *fingers*, *voice*)
- noun = linguistic/mental phenomenon (e.g. *approval*, *reply*)

(*Surprised* also occurs with *way* modifying an action by an emoter: *The* ***surprised*** *way he had looked at her*, BRC, CEC 2373)

The parsing of such adjectives does not entail any major problems: all ADJ n patterns can be parsed as undirected affect (with some perhaps additionally involving appreciations of semiotic phenomena).

With *anxious*, the situation is more complex (even disregarding its use in sexual descriptions: *anxious* + 'genital' as well as one instance of *an anxious blue blur* in BRC, FRL 1397).

When the right-hand noun collocate points to an emoter, the noun is either a participant, an expressor or a linguistic/mental phenomenon as with the other three adjectives, or (in addition) a noun referring to an activity that is clearly related to an actor:

- noun = conscious participant (e.g. *anxious children*, *clients, individual*, *patient*, *people*, *person, lovers*, *owners, shop stewards*, *worker*, *test takers*, *wife* etc, maybe also *anxious presence*)
- noun = expressor (e.g. *expression, eye(s), face(s), features*, *neck*, *tone*, *voice*) → showing the emoter's anxiety
- noun = linguistic/mental phenomenon (e.g. *questions, reflection*, *consultations, discussion, fussing, moan*, *quaking, account*, *debate, inquiry*, *reception*, *silence, attention, attachment*, *spiritual commitment*, *[mental] state, an anxious shrinking from everything except duty*) → pointing to the emoter's (sayer's/senser's) anxiety

- noun = activity related to an actor (e.g. *galvanising the Gnomes into **anxious** industry*, BRC, G1L 79*; She put her hands on the bedspread, looked into his eyes, an **anxious** searching*, BRC, H7F 1228; *She took an **anxious** step towards him*, BRC, BMW 389) → pointing to the emoter's (actor's) anxiety

But *anxious* can also occur with nouns that point to triggers (rather than or in addition to emoters). One example was already mentioned above:

(F1) OLDHAM keeper Jon Hallworth is facing an **anxious** battle to be fit in time for Saturday's big Premier League kick-off. (BRC, CH7 869)

In this example, it is presumably the 'battle to be fit in time' that causes Jon Hallworth's anxiety but the noun also metonymically allows us to infer the emoter (as the one facing the battle). *Anxious* frequently occurs with nouns indicating a duration of time, too (*moment*(*s*), *days*, *plane journey*, *day*(*s*), *week*(*s*)). For example:

(F2) But Roy had thrown away his script, and he spent an **anxious** plane journey trying to remember his lines. (BRC, CH1 3108)

(F3) The 26-year-old midfielder, who played in all three of England's European championship games, now faces an **anxious** wait for Taylor's verdict. (BRC, CH7 2804)

(F4) JONATHAN SPEELMAN gave his supporters an **anxious** day in the sixth round of the Pilkington Glass World Chess Championship semi-finals yesterday. (BRC, A4K 729)

What these examples are saying is that these times (moments, days etc) are spent by emoters in a state of anxiety because of something that happens during these times. It is not the days themselves that cause the anxiety, but the happening during these days. For example, that Roy has thrown away his script causes his anxiety (rather than the plane journey) or the worry about what Taylor's verdict might be or Jonathan Speelman's chess play in the semi-finals on the day in question. That is, such nouns can neither be parsed as emoter nor as trigger. Even more complex is an adjective like *sad*, which can express both emotion ('being unhappy', 'showing unhappiness of emoters') with collocating nouns as emoters or 'making someone feel unhappy' with collocating nouns as triggers) and opinion (''unacceptable, deserving blame', 'boring, not fashionable', 'in poor condi-

tion'). Collocating nouns cannot easily be used to distinguish these meanings without the wider context. Thus, *a sad man* might be unhappy, or 'boring, not fashionable' and *a sad consequence* can arguably either indicate affect ('making you unhappy') or opinion ('unacceptable, deserving blame'). The collocates for *sad* would need to be examined in detail to find out how accurate they predict the meaning of *sad*.

7 This points to the fact that conversation is similar in one sense to both news reportage and fiction (e.g. in containing a certain amount of narrative), while being different in another sense from all written discourse, e.g. in being spoken and interactive (Biber et al 1999: 16). Academic discourse, on the other hand, is sometimes similar to news reportage and fiction (because like them, it is a written register), while also differing from conversation, news reportage and fiction in that it presumably does not contain as much narrative and hardly any dialogue.

8 These findings might change if the analysis of patterns is broadened to positions other than L1 and R1 and to patterns that are less frequent in each register (compare Bednarek (2008: Chapter 4) for the methodology of the analysis of lexico-grammatical patterning).

9 More detail would also be helpful where emotion adverbs are concerned, which are often only run-on entries in dictionaries on account of their infrequency, though they have a variety of different meanings (see Appendix 2 for some of the meanings of *desperately*, *happily*, *cheerfully*).