A *Fresh* View of the Structure of Hard News Stories

Michaela Mahlberg and Matthew Brook O'Donnell, University of Liverpool

Using a corpus of newspaper articles divided into two categories based on textual position: TISC (text-initial sentences) and NTISC (all other sentences), the collocates and local patterns of a text-initial key word, *fresh*, are analyzed. Patterns of *fresh* + CONTROVERSY NOUN (*row, controversy, blow, embarrassment*) are particularly distinctive in first sentences including the elements FACE/SPARK ABOUT TIME AFTER, e.g. *faced fresh embarrassment over… yesterday when…* Using corpus annotation the extent and variation of this pattern is revealed and related to discourse functions, drawing in particular on White's (1997) characterisation of hard news articles and the notion of the NUCLEUS to describe text beginnings.

KEYWORDS: newspaper story structure, textual functions, text-initial key words, corpus annotation

1   INTRODUCTION

Frequencies of lexical features vary across different texts and groups of texts. Patterns of variation across different registers are described, for instance, through the multi-dimensional approach put forward by Biber (1988) and detailed comparative frequencies of lexicogrammatical features can be found in lexical and grammatical reference works (e.g. Biber et al. 1999). It has also long been acknowledged that even within texts from a single register lexical items will have different frequencies within different sections of text (Sinclair 1991; Stubbs 1996; Hoey 2004, 2005). The notion of 'key words' (Scott 1997) facilitates the identification of words that characterise one text in comparison with a reference corpus (see applications, for example, in Scott and Tribble 2006). Linguistic features identified on the basis of quantitative data open up various options for functional interpretation. Flowerdew (2008), for instance, shows how the analysis of key words can be linked to studying problem-solution patterns. A greater focus on the qualitative analysis of texts, however, can pose limits to the amount of data that can be processed. Henry and Roseberry (2001) illustrate a study of the language of move registers based on a small corpus, where each text was analysed individually to identify moves that then were collected in subcorpora. Biber et al. (2007) also present a corpus approach to move analysis that requires the hand-coding of moves in the texts of the corpus. This top-down approach to discourse studies is completed by bottom-up analyses focusing on Vocabulary-Based Discourse Units (VBDUs). VBDUs are sections of discourse whose boundaries are defined by a shift in vocabulary. This shift is identified automatically by a segmentation tool (cf. Biber et al.

2007: 161ff.). Biber et al. (2007) argue that sequences of VBDUs provide useful insights into discourse organisation.

The aim of the present study is also to explore aspects of both the quantitative and qualitative analysis of texts. We focus on newspaper texts and the present study is part of a larger project on the TEXTUAL PRIMING (Hoey 2005) of hard news stories.[1] In this article we look at what we call NUCLEUS PATTERNS, i.e. patterns of lexical items found in the first sentence of hard news stories. We aim to describe these patterns in view of their textual functions to begin news stories. The functions that we identify can be characterised as LOCAL TEXTUAL FUNCTIONS (cf. also Mahlberg 2005, 2007). They are TEXTUAL in that they describe lexical items in relation to the textual structure and the way in which information is presented in the texts; the functions are LOCAL because the description only applies to a specific set of items in a specific type of texts. The patterns examined have been identified from an analysis of the word *fresh* that has a statistically significant tendency to occur in the first sentence of hard news stories more frequently than would be expected on the basis of its overall frequency in a newspaper corpus. The paper begins with a description of this corpus and how it has been segmented to facilitate the identification of TEXT-INITIAL ITEMS using the Key Words procedure. The steps in the analysis of such items are illustrated through a study of the patterns of collocation of the word *fresh* in both text-initial and non text-initial position. From concordance lines a pattern emerges which includes semantic sets for the nouns modified by *fresh* and for the verbs to which these *fresh* NOUN combinations are linked, as well as connected functional elements. The pattern is used as the basis for an annotation scheme applied to all the occurrences of *fresh row*, *fresh embarrassment*, *fresh controversy* and *fresh blow* in the first sentences of articles in the corpus. The annotation provides the basis for both a quantitative summary and qualitative description of the textual functions of the pattern in newspaper texts.

2 *FRESH* AS A TEXT-INITIAL KEY WORD IN NEWSPAPER ARTICLES

In the Textual Priming Project a corpus of newspaper articles from the Home News section of the Guardian 1998-2004 has been compiled. The corpus contains approximately 52 million words with sentence tokenization applied. Subcorpora can be created consisting of sentences drawn from particular positions in text, such as the first sentence of every paragraph, sentences that do not begin paragraphs, and so on. The two subcorpora used in the present study are one that contains all the text-initial sentences (TISC = Text Initial Sentence Corpus) and another that contains the remainder of the article without the first sentences (NTISC = Non-Text Initial Sentence Corpus). Headlines are excluded from the quantitative component of this study. Table 1 presents an overview of the two subcorpora utilized here.

Table 1. Subcorpora of the Home News section of the Guardian 1998-2004

	TISC	NTISC	Total
Words	3,122,037	49,922,632	53,044,669
Sentences	113,288	2,227,259	2,340,547

Using the Key Words procedure in WordSmith Tools (Scott 2008) we can compare relative frequencies of the items in these two subcorpora. The method was first proposed by Mike Scott. It represents a logical and significant extension of the use of the Key Words procedure for comparing corpora containing full texts. The method aids the identification of local textual functions and allows lexical associations at the textual level (such as Hoey's (2005) TEXTUAL COLLIGATION) to be investigated. Comparing word lists from TISC and NTISC produces 3134 key words: 1905 positively key and 1229 negatively key.[2] So nearly 2000 words have a marked statistical tendency to be found in the first sentence of Home News articles from *The Guardian* in the time period 1998-2004 rather than in sentences anywhere else in an article. Examples of some of the positively key text-initial words are: *yesterday*, *after*, *according*, *emerged*, *jailed*, *police*, *accused*, *woman*, *death* and *controversial*. Examples of some of the negatively key text-initial words are: *Mr*, *said*, *that*, *would*, *she*, *do*, *although*, *spokesman*, *same* and *really*. The word *fresh* turns up as a positive key word in TISC. Table 2 shows the frequencies of *fresh* in the two subcorpora, and normalized frequencies: per 1 million words and per thousand sentences.

Table 2. Occurrences of *fresh*

	TISC	NTISC	Keyness (Log likelihood)
Occurrences	923	3984	944.01
Per million words	295.6	79.8	
Per 1000 sentences	8.1	6.2	

The difference in frequency between the subcorpora does not necessarily point to differences in the functional behaviour of *fresh*. It simply establishes a preference for the word to be used in the first sentence. To investigate the textual functions of *fresh* and thereby ascertain whether the word is in fact used differently when it occurs in the first sentence, we take two steps. First, in this section, we look at the collocates of *fresh* in TISC and NTISC. Then in section 3 we focus on the patterns of *fresh* in the text-initial sentences, i.e. we go beyond the information from fixed-width concordances and aim to relate the patterns to the functions of *fresh* in the news stories by examining the whole of the first sentence, which is a key element of the NUCLEUS (White 1997).

Table 3 shows the top 21 collocates for *fresh* in both TISC and NTISC. This number was selected simply because there are 21 collocates in TISC that occur at least 10 times one word to the right (R1 position) of *fresh*. The table shows that some of the

---

[2] A p value of 0.000001 was used in the WordSmith Key Words program to calculate the key word lists in this study. It was set to retrieve all the key words (both positive and negative) for TISC compared to NTISC.

top collocates of *fresh* are shared between the subcorpora: *evidence, round, start, air, appeal* and *from* occur among the top 21 collocates in each list (capitalized in Table 3). However, the table already gives an indication of different meanings of *fresh* and how these are distributed according to text position. In the NTISC list the words in italics refer to food and water and can be summarized under the heading of CONSUMABLES: *fruit*, *food* and *water*. We could also include *air* in this group. In contrast, words from the group of CONSUMABLES (except *air* if included) do not occur as the top collocates of *fresh* in TISC. In the same way, we find another group in TISC that does not figure among the most frequent collocates of *fresh* in NTISC. The TISC list contains a number of nouns with negative connotations: *controversy, row, embarrassment, blow, allegations, pressure, crisis, attack, doubts* and *criticism*. In the NTISC list, apart from *allegations* and perhaps *questions*, there are not any nouns that are clearly negative. A word that may appear to have negative connotations is *faced* and below we illustrate uses of the verb FACE in a negative context. The occurrences of *faced* in the NTISC R1 collocate list, however, are part of the compound *fresh-faced*. Of the ten words with apparently negative connotations in the TISC list (*controversy*, *row*, etc.), eight are also key words of TISC compared with NTISC, only *doubts* and *criticism* are not text-initial key words. For the key words, their frequency in the first sentence of the corpus of hard news stories is statistically significant when compared with their relative frequency in sentences from the remainder of the articles. Similarly, all nouns in the CONSUMABLES group of NTISC are not key in TISC.

Table 3. The top 21 collocates of *fresh* at position R1

	TISC				NTISC			
Rank	R1	TK?	freq.	per 1000 occs	R1	NK?	freq.	per 1000 occs
1	EVIDENCE	N	55	59.6	START	N	232	58.2
2	**controversy**	Y	41	44.4	EVIDENCE	Y	171	42.9
3	**row**	Y	41	44.4	AIR	N	134	33.6
4	**embarrassment**	Y	27	29.3	FROM	N	106	26.6
5	**blow**	Y	25	27.1	*water*	N	85	21.3
6	ALLEGATIONS	Y	20	21.7	and	Y	81	20.3
7	pressure	Y	20	21.7	*fruit*	N	80	20.1
8	ATTEMPT	Y	19	20.6	questions	Y	54	13.6
9	ROUND	N	17	18.4	investigation	N	46	11.5
10	crisis	Y	16	17.3	ALLEGATIONS	N	44	11.0
11	START	N	16	17.3	UN	N	40	10.0
12	AIR	Y	15	16.3	elections	N	39	9.8
13	attack	Y	13	14.1	inquiry	N	35	8.8
14	doubts	N	12	13.0	ATTEMPT	N	33	8.3
15	APPEAL	Y	10	10.8	look	Y	33	8.3
16	calls	N	10	10.8	ROUND	N	31	7.8
17	criticism	N	10	10.8	APPEAL	N	29	7.3
18	FROM	Y	10	10.8	faced	N	29	7.3
19	INVESTIGATION	Y	10	10.8	*food*	N	28	7.0
20	political	Y	10	10.8	resolution	N	27	6.8
21	wave	Y	10	10.8	information	Y	25	6.3
		15 of 21 key				5 of 21 key		

TK? = Is the collocate a key word in TISC against NTISC word list?
NK? = Is the collocate a key word in NTISC against TISC word list?

Table 3 contains two columns (headed 'TK?' and 'NK?') indicating whether the collocates are found on the appropriate key word list: TISC compared with NTISC for the most frequent R1 collocates of *fresh* in TISC (left hand list) and NTISC compared with TISC for the top 21 NTISC collocates. At this point, the initial overview already highlights a tendency for *fresh* to behave differently in the first sentence of an article compared to non-initial sentences. In the first sentence of an article *fresh* is found in the company of words that also have a tendency to occur in first sentences, and that appear to have negative connotations. With these observations we have to bear in mind that we have focused on only those collocates of *fresh* in TISC that occur at least 10 times. These collocates have been analysed in terms of rank and whether they are key words for either subcorpus independently of their co-occurrence with *fresh*.

A more targeted and systematic analysis of the differences in the collocational behaviour of *fresh* between TISC and NTISC can be carried out using the concept of Key Collocates. The key collocate procedure (O'Donnell, forthcoming) compares the frequency of each of the collocates of a node item in one corpus or set of concordance lines with their frequency as a collocate of the same node in a reference corpus/concordance line set. For example, within a 10-word span of *fresh* in its 923 occurrences in TISC, *controversy* is found 48 times (52.0 per 1000 instances). In contrast, in the 3984 occurrences of *fresh* in NTISC and within the same span, *controversy* occurs only 23 times (5.8 per 1000 instances). Statistical measures, either Chi-Square or Log Likelihood, can be applied to these figures in the same way as to word frequencies in two corpora when calculating key words (see Rayson, Berridge & Francis 2004). For *controversy* the log likelihood value is 80.55 which is above the critical value of 15.13 at the 0.01% (p < 0.0001). So we can say that *controversy* is a key collocate of *fresh* in text-initial sentences in comparison to non-initial sentences. Key collocates can be calculated for all occurrences within a specified span of the node word, focused on just one side of the node or on a single position (such as R1 position). Table 4 shows the right key collocates of *fresh* (both positive and negative) in TISC with reference to the collocates of *fresh* in NTISC.

Table 4. Right Key Collocate Comparison between TISC and NTISC instances of *fresh*

collocate	First Sentence Freq. (TISC)	Rest of article Freq. (NTISC)	Keyness value (Log Likelihood)
yesterday	89	34	166.54
night	73	26	140.77
last	83	43	133.52
over	99	93	103.60
when	71	49	95.37
row	50	25	82.02
controversy	46	23	75.46
after	58	47	69.00
blow	25	14	38.45

embarrassment	29	21	37.63
today	18	6	35.66
start	16	236	-32.64
and	52	544	-47.50

The nouns controversy, row, embarrassment, and blow that we identified in Table 3 as belonging to a group of nouns with negative connotations are key collocates, and in the following sections we refer to them as the controversy group (they are highlighted in bold in Table 3). So far we have looked at these nouns in view of frequency information. A concordance analysis can add detail to the patterns of the nouns. Concordance 1 below shows all 40 occurrences of fresh controversy in the subcorpus oft text-initial sentences (TISC) sorted by the first and second word on the left.

```
1    r in waiting, has become embroiled in a fresh controversy over claims that he as
2                                           Fresh controversy over American bombing
3    Perrier comedy award yesterday courted fresh controversy by unveiling an all ma
4    108ft tower in Oxford are set to create fresh controversy.
5    Branson's bid to run the Lottery faced fresh controversy last night after his c
6    f the Holocaust next weekend are facing fresh controversy after the Government s
7               PETER Mandelson is facing fresh controversy after the disclosure y
8    ime Minister, John Prescott, was facing fresh controversy yesterday when it was
9                MPs are braced for fresh controversy over their parliamenta
10                                          FRESH controversy about Alastair Campbel
11   le Sir Ronnie Flanagan was embroiled in fresh controversy yesterday when it was
12   Tony Blair was last night embroiled in fresh controversy over the future size a
13   government was last night embroiled in fresh controversy over a series of initi
14   Government was last night embroiled in fresh controversy over the premature lea
15          Barclays was embroiled in fresh controversy last night with the re
16   tax and Europe - left him embroiled in fresh controversy last night.
17   olicy on asylum seekers was engulfed in fresh controversy last night after a Fre
18    Tony Blair was last night enmeshed in fresh controversy over Bernie Ecclestone
19   d cricket tour of Zimbabwe was mired in fresh controversy last night after Rober
20   as a United States senator was mired in fresh controversy yesterday when a crimi
21   Reform of the Lords will erupt into fresh controversy today when a minority
22   m of the House of Lords will erupt into fresh controversy today when a minority
23   TONY Blair was last night plunged into fresh controversy about the sex lives of
24      BRITISH athletics was plunged into fresh controversy over drugs last night
25   inister Lord Sainsbury was plunged into fresh controversy last night after it em
26          The RUC was plunged into fresh controversy last night when it bel
27    Tory party treasurer, was plunged into fresh controversy last night amid report
28   opment Secretary, Clare Short, ran into fresh controversy last night after defen
29                                          Fresh controversy about Michael Ashcroft
30   ir Richard Branson was at the centre of fresh controversy yesterday when it emer
31   s, in remarks which last night prompted fresh controversy.
32   , Ron Davies, moved last night to quash fresh controversy over voting in last ye
33   mond for tea in Downing Street, risking fresh controversy over the Labour party'
34   ottery grant in a move certain to spark fresh controversy over use of `good caus
35   nown living entities looks set to spark fresh controversy about the nature and o
36    attack on A-level standards but stoked fresh controversy by admitting he had no
37                                          FRESH controversy yesterday surrounded t
38              MO MOWLAM triggered fresh controversy yesterday by admitting
39              JACK STRAW triggered fresh controversy yesterday after callin
40                     There was fresh controversy over the issue of so-c
```

Concordance 1: 40 occurrences of *fresh controversy* in TISC

The left-hand context shows two main verb patterns. The first and most common pattern expresses that someone finds themselves thrown into a problematic situation. This pattern is associated with *face, embroil in, plunge into* as the most frequent verbs. For

the verb *face* this meaning is linked to the active voice; *embroiled in*, *plunged into*, *engulfed in* occur in a passive constructions. Other less frequent verb forms in this group with similar meanings are *engulfed in*, *enmeshed in*, *mired in* and *braced for* as well as *ran into*. Another construction with similar meaning is *be at the centre of*. The second main verb pattern illustrates that someone or something *sparks*, *triggers* or *prompts controversy*. A less frequent pattern shows *controversy* at the beginning of the sentence and the five lines for this pattern are given in Concordance 2 below extending the context to the right.

```
2    Fresh controversy over American bombing flared last night after …
10  FRESH controversy about Alastair Campbell's role in the Kosovo conflict erupted
yesterday after …
29  Fresh controversy about Michael Ashcroft's pounds 3m donation to the Conservative
party erupted last night after …
38  FRESH controversy yesterday surrounded the deal to bring radioactive nuclear waste
from Georgia for reprocessing at Dounreay after …
41 There was fresh controversy over the issue of so-called designer babies last night
after …
```

Concordance 2: *controversy* at the beginning of sentences

The verb *erupt* occurs twice in these examples, which fits into the pattern of the other verbs, most of which express a sense of forcefulness or dramatic confrontation. Additionally the extended contexts above all illustrate occurrences of *after*. The patterns on the right of *controversy* in Concordance 1 also hint at functions of *after* in the first sentence of newspaper articles. Patterns on the right of *controversy* include time references with *last night, today, yesterday* as well as *after, when* and combinations of time references such as *last night after, yesterday when*. Other frequent patterns include *over* and *about*.

    The patterns of *controversy* are very similar to the patterns of the other key collocates of *fresh* in TISC. An examination of the 41 lines for *row* in R1 position reveals that most are covered by verb constructions on the left or right that overlap with the patterns of *controversy* illustrated by Concordance 1. Below are examples for *row* illustrating the remaining new verb patterns.

```
8            Tony Blair yesterday ignited a fresh row over the issue of privatisatio
9   British Museum has become involved in a fresh row with the Greeks and a leading
31  r, Ariel Sharon, threatens to provoke a fresh row with the Palestinians today by
32   results released today have provoked a fresh row about the credibility of the "
33          Ken Livingstone has provoked a fresh row with the motoring lobby by cal
35                             A fresh row broke out yesterday between a
```

Concordance 3: Verb patterns for *row* that do not figure in Concordance 1 for *controversy*

The patterns, however, are still similar in meaning: *ignite* and *provoke* are similar to *spark* and *trigger*, *break out* fits with *erupt* and *involved in* relates to *embroiled in*. The patterns on the right are also similar to those of *controversy* as we find time references as well as *over* and *about*. Additionally, a repeated pattern for *row* is *a fresh row with*, in 7 of 41 lines (17%).

```
7            Ken Livingstone has provoked a fresh row with the motoring lobby by
11               Downing Street risked a fresh row with the media last night when
17    guilty of bullying, is embroiled in a fresh row with the beleaguered professional
20      Ariel Sharon, threatens to provoke a fresh row with the Palestinians today by
28  British Museum has become involved in a fresh row with the Greeks and a leading
30      John Prescott was at the centre of a fresh row with the Tories last night over
33          The government was embroiled in a fresh row with the BBC yesterday after
```

Concordance 4: Examples of *a fresh row with*

We try to illustrate the similarities between the four nouns of the CONTROVERSY group in Table 5 below. The way to read this table is that *fresh* collocating with one of the *controversy* nouns co-occurs with four other elements. One of these elements is the verb that indicates that someone or something faces or sparks a difficult situation or that problems arise (for the verb *erupt*). The verbs *face* and *spark* are taken to represent this meaning and the table does not imply that all of the examples listed co-occur with each of the nouns. The verb *deal*, for instance, only occurs with *blow*, as in *schools will be dealt a fresh blow today*. The verb *face* is the verb with the greatest overlap across the nouns, but still shows tendencies of different frequencies: for *embarrassment* 18 of the 27 examples are accounted for by *face* (i.e. about 67%), for *row* in contrast, only four out of 41 lines, figure examples of *face* (i.e. about 10%).

Table 5: Pattern of the CONTROVERSY group

FACE/SPARK	fresh	NOUN	ABOUT	TIME	AFTER
be embroiled in		*controversy*	*over*	*yesterday*	*when*
be at the centre of	*fresh*	*row*	*about*	*last night*	*as*
be dealt		*embarrassment*	*to*	*today*	
ignite		*blow*			
heap on					
erupt					

Support for this pattern comes from the list of key collocates in Table 4. Words from the NOUN, TIME, ABOUT and AFTER categories appear as key collocates of *fresh* in TISC compared to *fresh* in NTISC. Looking at the left key collocates of *fresh* gives many of the verbs in the FACE/SPARK group as well as further instances of words belonging to time.[3]

The time reference with a point of time such as *yesterday* also occurs with all nouns, and all nouns illustrate examples with relative time relationships that can be illustrated for instance by *after*. The ABOUT element of the pattern is realised by forms that are more specific to each noun, e.g. *about* can occur with *controversy* and *row* but there are no examples for *embarrassment* and *blow* in the concordance lines. What the words in this category share is that they are prepositions giving some indication as to

---

[3] Key left collocates (span L1 to L5) of *fresh* in TISC when compared to instances in NTSIC are: *night*, *yesterday*, *last*, *facing*, *into*, *plunged*, *faced*, *suffered*, *was*, *embroiled*, *Tony*, *today*, *centre*, *a* and *after*.

what the problem is about, as in example (1) of *blow* where the authority of Iain Duncan Smith is at stake. The example also shows that words of the AFTER category can occur together in a sentence, here we find both *when* and *after*.

(1) *Iain Duncan Smith <u>faced</u> a <u>fresh</u> <u>blow</u> <u>to</u> his authority <u>last night</u> <u>when</u> a key aide was ousted from Tory Central Office after a revolt by MPs and senior members of the party's governing board.*

Table 5 so far is a summary of the patterns of the CONTROVERSY group that are visible on the basis of concordance lines. The pattern is discussed in more detail in the following sections.

3   THE CONTROVERSY GROUP AND THE NUCLEUS PATTERN

What we have identified in section 2 is a group of nouns that have a tendency to collocate with *fresh* in the first sentence of hard news articles. In this section we want to look at the textual functions of the *fresh* patterns by focusing on features of newspaper articles. Characteristics of the *fresh* pattern appear to relate to the specific way in which information is presented in news stories. White (1997) illustrates how the generic organization of hard news reports serves to achieve social and ideological objectives. He claims that a significant number of hard news reports have a structure that accounts for two parts of an article: the nucleus and the main body. The nucleus is "most typically constituted by the combination of its headline and its opening sentence" (White 1997: 111) and the body of the article is what follows the nucleus. White points out that in most cases there is great overlap between the information in the headline and the first sentence with the headline being shorter and primarily functioning to "signpost key meanings". In the present paper we do not include the headline in the study but focus on the first sentence alone. So strictly speaking we are only looking at part of the nucleus. An important element of White's argument is that the structure of the body of the article is characterised by a series of specifications referring back to the nucleus. Our focus is on the first sentence without investigating details of different types of specifications. For White (1997: 104) hard news reports deal with threats to the social order. The nucleus (headline plus first sentence) "acts to launch the reader immediately into the heart of the social-order disruption about which the report is organised" (White 1997: 112). As the nucleus directly goes to the "peak or climax of social-order disruption" (White 1997: 112), it is also "an interpersonal peak" (White 1997: 115), with the greatest density of intensifying interpersonal meanings (White 1997: 114). White (1997) looks at lexis as one means of intensification and points out that in reports dealing with communicative events, intensification is often found with the reporting of verbal communication: "politicians slam their opponents rather than criticizing or disagreeing with them, political parties find themselves plunged into a heated row rather than engaging in debate" (White 1997: 109).

The patterns that we identified in section 2 already highlight some aspects of the nucleus as "the text's anchor point" (White 1997: 116). The CONTROVERSY nouns (*controversy*, *row*, *embarrassment*, *blow*) all label some difficult or unpleasant situation, the verbs illustrate the highthened impact (White's example of *plunged into a heated row*

specifically relates to the patterns of *row* identified above), and the adjective *fresh* itself—that was the starting-point for our analysis—indicates how the significance of the content of the news article is emphasised. The evaluative meaning of *fresh* also relates to the news value of RECENCY (cf. Bell 1991) and to the time references that are important to news reporting. The pattern of the CONTROVERSY group:

FACE/SPARK *fresh* NOUN ABOUT TIME AFTER

can be interpreted as a basic structural pattern to introduce the content of the news report. The CONTROVERSY nouns carry evaluative meaning and highlight the newsworthiness but at the same time they are general enough to cover a variety of subject matters. In this sense the nouns are what Francis (1994) calls LABELS. The function of the nouns can also be seen as a type of minimal speech presentation, when they summarise contrasting opinions that can then be followed by more specific reports as the newspaper article develops (Semino and Short 2004: 44f.).

The individual elements of the pattern function to express relationships between different aspects that relate to the story. To investigate the pattern in more detail we have manually annotated the first sentence of each of the articles of the four CONTROVERSY nouns, marking occurrences of the above elements of the pattern.[4] For each noun we selected the articles where the noun occurs as a collocate of *fresh*. We allowed for intervening premodifiers, such as *fresh political controversy*, along with cases where the nouns were in R1 position as in the concordances above. The pattern initially summarises the similarities between the patterns of the four nouns as shown in their concordances. An analysis of the full first sentences also accounts for the elements of the pattern when they occur outside the limited span of the concordance. Table 6 shows the results of the analysis for *row*. The table does not list the elements *fresh* and *row* as they occur in all of the examples. Of the initial 47 occurrences of *fresh row* we included 43 articles in the counts, because of duplication of articles[5], and in one case the pattern occurred in a non-finite clause, which does not reflect the main tendency of the pattern - the verb of the pattern is the main verb of the first sentence. The analysis for *row* shows that all four elements of the CONTROVERSY pattern occur. Additionally there are seven examples with further premodifiers of *row* that follow the adjective *fresh*. In order to make the automatic counting of the manual annotation easier we did not include the auxiliaries, but only the lexical verbs, and also counted *at the centre of* as verbal element (see example below).

---

[4] Annotation was carried out using the linguistic mode in the Fab4 browser (http://bodoni.lib.liv.ac.uk/fab4/) and exported to XML. A simple script was then used to count the frequencies of each tag and collate the lexical items contained in them.

[5] The corpus is derived from the archive of the Guardian newspaper from 1998-2004. We have found that there are quite often two versions of a single article that represent either pre-publication and publication versions or early and late edition versions. Sometimes it appears that new material has been added to the end of the article, whereas in other instances more detailed editing has taken place and a new headline and revised first sentence can be observed. In this study we have removed duplicate lines from concordances and annotated only one of duplicate/related articles.

Table 6: The nucleus pattern for *fresh row*. 43 of 47 articles annotated

FACE/SPARK 43 (100%)		PRE-MODIFIER 7 (16%)		ABOUT 29 (67%)		TIME 33 (77%)		AFTER 31 (72%)	
at the centre of	7	*political*	3	*over*	22	*yesterday*	16	*after*	11
embroiled in	6	*education*	1	*about*	6	*last night*	15	*when*	11
sparked	4	*race*	1	*as*	1	*today*	1	*as*	3
provoked	3	*leadership*	1			*just days after*	1	*by*	3
		sleaze	1						
(x2) *facing, ran into, broke out, triggered, faced, erupted, plunged into* (x1) *blew up, involved in, embroiled, threatens to provoke, ignited, engulfed in, erupted in, risked, fighting*								(x2) *with*, (x1) *for*	

N.B. (x2) & (x1) indicate that the word occurs twice and once respectively in element

The percentages in Table 6 reflect the frequency of each of the elements of the CONTROVERSY pattern in the *fresh row* sentences annotated. Tables 8 to 10 in the Appendix show the same results from the annotation of sentences containing *fresh embarrassment*, *fresh controversy* and *fresh row* respectively. Table 6 shows that the TIME element is the most frequent in the nucleus pattern occurring in *77%* of the sentences containing *fresh row*.

In 11 of the 43 texts all four elements of the pattern occur together: FACE/SPARK *fresh row* ABOUT TIME AFTER occurs in 26 % of the articles and is the most frequent combination of elements. In example (2) the four elements of the pattern around *fresh row* are underlined. The time element is realised by *yesterday,* we count *at the centre of* as verb element, *over* indicates what the row is about, and *after* introduces what led up to the row.

(2) *The government was <u>yesterday</u> <u>at the centre of a</u>* **fresh row** <u>over</u> *plans for a shake-up in the way history is taught in schools, <u>after</u> the Tories joined forces with traditionalist history experts to accuse Labour of undermining Britain's national identity.*

The elements listed in the AFTER category in Table 6 do not necessarily indicate that something happened actually before a row broke out, in this sense indicating a temporal relationship. More important to the AFTER category is the indication of aspects of the subject matter of the article. While the verbs that belong to the CONTROVERSY pattern are similar in that their main meaning in the first sentence is to express that there is a row, the other elements of the pattern fill in the information of what happened to bring about a row. Example (3) illustrates the AFTER pattern being realised by *as* and the sentence presents Blair's action of branding his critics "small 'c' convervatives" being interpreted as igniting a row. In example (4) the non-finite clause with *by* provides the information about what is described as provoking the row.

(3) *Tony Blair <u>yesterday</u> <u>ignited</u> a* **fresh row** <u>*over*</u> *the issue of privatisation <u>as</u> he branded his fiercest union critics of public service changes "small 'c' conservatives", akin to those who had been unwilling to abandon Clause 4, or reject far left militants in the 1980s.*

(4) *Tony Blair <u>yesterday</u> <u>provoked</u> a* **fresh political row** <u>*over*</u> *the role of the intelligence agencies <u>by</u> appointing John Scarlett, the official responsible for the widely disputed Iraqi weapons dossier, as the new head of MI6.*

Examples (2) to (4) all include the four elements that most frequently co-occur; example (4) additionally illustrates a premodifier that is combined with *fresh*. The variation in the co-occurrence of the elements of the noun pattern can be described in relation to the evaluative meanings in the nucleus. As the examples above have shown the FACE/SPARK element of the pattern together with the *fresh* NOUN part mainly seem to have an interpersonal function. They emphasize the interpretation of the subject matter of the article as negative in view of the social order—in the sense of White (1997). The ABOUT and AFTER elements of the pattern provide more information and introduce the content of the article. The elements FACE/SPARK and NOUN are the obligatory parts of the patterns, the adjective *fresh* occurs in all cases because that was the text-initial key word that led us to identify the pattern. A future study will have to examine instances of the nouns in the CONTROVERSY group that are not premodified by *fresh*. Initial analysis suggests that the pattern identified around *fresh* + CONTROVERSY noun also occurs when *fresh* is not present. The way in which ABOUT and AFTER can introduce aspects of the content of the article make it possible that both occur together, as in examples (2) to (4). Particularly in (3) and (4) the AFTER elements (*as, by*) give a more general view on the action described in the ABOUT elements of the pattern. Example (5) shows *after* can also be sufficient on its own to outline the topic.

(5) *Labour was <u>engulfed in</u> a* **fresh row** <u>*last night*</u> <u>*after*</u> *a panel headed by the party chairman, Charles Clarke, excluded the prominent leftwinger Mark Seddon from a byelection shortlist.*

Similarly, there are examples where *over* occurs on its own. In contrast to examples (2) to (4) above, in (6) the *over* element is now more detailed.

(6) *Michael Meacher, the environment minister, is <u>facing</u> a* **fresh row** <u>*over*</u> *a London incinerator's toxic ash which has been distributed widely to the building trade for use in breeze blocks and roads.*

A similar function to *about* can also be achieved by the additional modifier, as *race* in (7) illustrates.

(7) *A* **fresh race row** <u>*erupted in*</u> *the Conservative party <u>last night</u> <u>after</u> it emerged that Lord Tebbit is to share a platform with a key adviser to the French National Front leader, Jean-Marie Le Pen.*

The importance of either an ABOUT or AFTER element is further highlighted when we consider the minimal pattern for *row*. Of the 43 examples of *fresh row* annotated, example (8) is the closest to a minimal pattern with the typical ABOUT or AFTER elements. In addition to *fresh row* the first sentence contains the obligatory verb *embroiled in* and the sentence adverbial, introduced with *Just days after*, is to some extent similar to the other time elements found. Although there is an *after* in the adverbial, the action of the clause is not what is seen as the row. It is the second sentence that introduces the subject matter of the article: the imminent court battle. In this case the nucleus contains two sentences.

(8) *(1) JUST DAYS <u>after</u> disastrous viewing figures forced ITV to scale back its pounds 10 million flop Survivor, the show is <u>embroiled in</u> a* **fresh row**. *(2) A court battle is looming in America that threatens to spark the biggest scandal in reality television.*

The minimal noun patterns also points to patterns more specific to the individual nouns and not shared across all patterns. Example (9) illustrate that *row with* is another pattern of the noun. The description of the two parties involved in the row is sufficient as preview for the article. In the example the *after being found guilty of bullying* is not classified as *after* pattern because it is part of the relative clause of the subject.

(9) *Kamlesh Bahl, the former Law Society vice-president who was forced out in March 2000 after being found guilty of bullying, is <u>embroiled in</u> a* **fresh row** <u>with</u> *the beleaguered professional body.*

Example (9) also illustrates another point that we have not touched upon so far. The patterns can also be analysed in further detail by describing the role relationships of the participants. However, this aspect is excluded from the present study.

The final element of the noun pattern that we want to look at is the TIME category. Table 6 shows that the TIME category is the most frequent element. It does not relate as closely to the other elements of the pattern as it is a more general feature of text-initial sentences of hard news report. Recency is a news value that relates to the news cycle and therefore words belonging to the TIME category are common in an overwhelming proportion of first sentences. *Yesterday*, *today*, *last night*, *last week*, *weekend* and *tomorrow* are all positive text-initial keywords. *Yesterday* occurs 34,646 times in the 113,288 sentences of TISC. That is about 1 in every 3 sentences. In NTISC it occurs 57,278 times in 2,227,259 sentences, which is around 1 in every 39 sentences.[6] *Yesterday* is so common in TISC that it has no significant collocates itself (i.e. no downward collocates). This explains why the TIME category appears to be a more loosely linked or less central component of the pattern we have been discussing. The frequency of TIME words in text-initial sentences accounts for why it would play a role in the pattern.

After the detailed illustration of the CONTROVERSY group pattern for *row,* Table 7 now presents a comparison across the four nouns of the group. The Appendix

---

[6] If we consider just non-initial sentences (i.e. not the first sentence of the text or any paragraph in it [NISC]), the figures are even more pronounced: 13,363 occurrences of *yesterday* in 1,064,493 sentences, which 1 in every 80 sentences.

provides individual tables for the other CONTROVERSY nouns: *embarrassment*, *blow* and *controversy*.

Table 7. Comparison of occurrences of components of CONTROVERSY pattern for all four nouns

fresh NOUN	FACE /SPARK	PRE- MODIFIER	ABOUT	TIME	AFTER	*Most frequent pattern*
controversy (40 articles)	100%	5%	43%	75%	75%	AFTER TIME FACE/SPARK (38%)
row (43 articles)	100%	16%	67%	77%	72%	ABOUT AFTER TIME FACE/SPARK (28%)
embarrassment (28 articles)	100%	7%	46%	93%	96%	AFTER TIME FACE/SPARK (46%)
blow (21 articles)	100%	0%	5%	81%	81%	AFTER TIME FACE/SPARK (71%)

Table 7 shows that all 4 elements FACE/SPARK TIME ABOUT AFTER co-occur with all four nouns. The final column of the table gives the most frequent combination of elements for each noun. For instance, the FACE/SPARK AFTER TIME combination occurs in 71% (15) of the articles containing *fresh blow* in the first sentence. The table highlights that the FACE/SPARK verb is an obligatory element of the noun pattern. This information reflects that the noun pattern characterises the main clause of the first sentence. The three flexible elements also occur with all nouns. However, we see differences in what the main combination of elements is for each noun. Whereas the most frequent pattern for *row* contains all three flexible elements TIME ABOUT AFTER, the nouns *controversy, embarrassment* and *blow* tend to occur most frequently with the pattern TIME AFTER. Below are examples of this pattern for each of the three nouns.

(10) *The RUC was* <u>plunged into</u> **fresh controversy** <u>last night</u> <u>when</u> *it belatedly admitted that an unidentified man had been charged for possessing a dossier of personal details on 150 leading nationalists likely to be of use to terrorists.*

(11) *THE BBC was* <u>facing</u> **fresh embarrassment** <u>last night</u> <u>when</u> *one of its most senior news presenters attacked managers for undermining television news coverage.*

(12) *LABOUR's attempts to fend off a pensions revolt* <u>suffered</u> **a fresh blow** <u>yesterday</u> <u>as</u> *a poll revealed overwhelming backing for a rise for all OAPs.*

Although it is not the most frequent pattern for them, both *controversy* and *embarrassment* can also occur with the three elements TIME ABOUT AFTER. However, the patterns for *blow* contain only one example (13) of ABOUT and the example of ABOUT does not contain an AFTER. Five (20%) of the instances of *fresh blow* in TISC have *to* in R1 position. An initial examination of all instances of *blow* in TISC indicates a collocation with *to* in R1 position, suggesting that the noun *blow* might bring this association with it to the *fresh* NOUN pattern which could block or interfere with the ABOUT element. However, the lack of the ABOUT element in sentences containing *fresh blow* in TISC requires further analysis and explanation.

(13) *Stephen Byers, the transport secretary, <u>faced</u> a **fresh blow** <u>to</u> his credibility <u>yesterday</u> <u>over</u> allegations that he refused to speak to his most senior civil servant at the height of the latest Jo Moore row.*

The main function of the nouns in the pattern is to express that there is a problematic situation. This evaluative emphasis is highlighted when we look at different meanings of the nouns. A particularly striking example is *embarrassment*. The *Collins Cobuild Dictionary for Advanced Learners* (2001), based on corpus data, gives the following definition for the most frequent meaning of *embarrassment*: "Embarrassment is the feeling you have when you are embarrassed." The second meaning that is listed explains: "An embarrassment is an action, event or situation which causes problems for a politician, political party, government, or other public group." An example for the second meaning is "The poverty figures were undoubtedly an embarrassment to the president." The CONTROVERSY group patterns mainly draw on this second meaning of 'political embarrassment'. However, the noun is not used as a countable noun–*an embarrassment*–but as a more abstract characterisation of a problematic situation, as we have seen in the examples above. The meaning of *embarrassment* as a feeling that one has is less frequent and seems to be realised in the newspaper articles mainly in the sense of someone doing something that is cringe-making or embarrassing for other people as in example (14).

(14) *THE Duchess of York heaped **fresh embarrassment** on the royal family yesterday by revealing that she and her former husband do not bring lovers back to their shared Berkshire home, writes Luke Harding.*

4   CONCLUSIONS

With the information from Table 7 and the functions that we have identified for the elements of the noun pattern we can summarise the characteristics of what we have outlined as a text-initial CONTROVERSY pattern. The nouns in the CONTROVERSY group (*controversy, row, embarrassment, blow*) resemble one another in that they all have a tendency to occur in the first sentence of hard news article where they collocate with the text-initial key word *fresh*. The verbs that the CONTROVERSY nouns collocate with are characterised by an evaluative meaning expressing heightened intensity in the main clause of the text-initial sentence. The combination FACE/SPARK + *fresh* NOUN is an evaluative pattern that interprets the subject matter of the article as newsworthy. The verb in the main clause of the sentence is mainly evaluative and therefore provides an interpretation of the action that characterises the subject matter. This subject matter can be introduced by different realisations and combinations of the flexible elements of the pattern ABOUT and AFTER. It is interesting to note that *after, when* and *over* are positive key word for TISC compared with NTISC, that is, they occur in a significantly greater number of TISC sentences than would be expected on the basis of their occurrences in NTISC. In contrast, *about* is negatively key in TISC making its role in the identified pattern worthy of further examination. Additionally, the noun pattern can have further elements that are more specific to each noun, we have seen examples of *row with* or *blow to*. A more detailed description of these patterns could be achieved by

analysing the participant roles in the clause. The TIME element in the pattern is also a flexible element, but is less closely linked to the noun pattern. Its frequency is mainly due to the frequency of time words in the first sentence. It highlights the recency of the subject matter and stresses that news are reported within a news cycle of 24 hours (often less). The verbs in the CONTROVERSY pattern carry evaluative meaning, but also the nouns have an evaluative function. Their informative meaning is very general and they can function as labels for the subject matter of the article. It seems that the meaning of the nouns is shaped by their occurrence in the pattern. Their functions relate to the structure of the text in the sense that the nucleus of the article introduces the subject matter and evaluates it as a newsworthy event. For the description of the NUCLEUS pattern it is also important to note that we started the analysis with the text-initial key word *fresh*. The adjective *fresh* led us to the identification of the CONTROVERSY nouns and the description of their local textual functions associated with the NUCLEUS pattern. Further studies will be necessary to the characterise the functions of the nouns when they do not appear as collocates of *fresh* but still appear in the first sentence of newspaper articles.

REFERENCES

Bell, Allan. 1991. *The language of news media*. Oxford: Blackwell.
Biber, Douglas. 1988. *Variation across speech and writing*. Cambridge: Cambridge University Press.
Biber, Douglas, Johansson Sitg, Leech Geoffrey and Conrad, Susan. 1999. *Longman grammar of spoken and written English*. London: Longman.
Biber, Douglas, Connor Ulla, and Upton Thomas A. 2007. *Discourse on the Move. Using Corpus Analysis to Describe Discourse Structure*. Amsterdam/Philadelphia: John Benjamins.
Carter, Ronald. 1987. *Vocabulary: Applied linguistic perspectives*. London: Routledge.
Collins Cobuild Dictionary for Advanced Learners. 2001. 3rd ed. Glasgow: HarperCollins.
Flowerdew, Lynn. 2008. *Corpus-based analyses of the problem-solution pattern: A phraseological approach*. Amsterdam/Philadelphia: John Benjamins.
Francis, Gill. 1994. Labelling discourse: An aspect of nominal-group lexical cohesion. In Malcolm Coulthard (ed.), *Advances in written text analysis*, 83-101. London: Routledge.
Henry, A. & Roseberry, R. 2001. Using a small corpus to obtain data for teaching a genre. In M. Ghadessy, M., A. Henry, and R. Roseberry, (eds) (2001) *Small Corpus studies and ELT: Theory and practice*, 93-133. Amsterdam/Philadelphia: Benjamins.
Hoey, Michael. 2004. Textual Colligation: a special kind of Lexical Priming. In K. Aijmer and B. Altenberg (eds) *Advances in Corpus Linguistics: Papers from ICAME 23, Götenborg 22-26 May 2002 (The Theory and Use of Corpora)*. Amsterdam: Rodopi.
Hoey, Michael. 2005. *Lexical Priming: A new theory of words and language*. London: Routledge.
Mahlberg, Michaela. 2007. "Lexical items in discourse: identifying local textual functions of *sustainable development*". In Hoey, M., Mahlberg, M., Stubbs, M.,

Teubert, W. *Text, Discourse and Corpora. Theory and Analysis,* 191-218. London: Continuum.

Mahlberg, Michaela. 2005. *English General Nouns: A Corpus Theoretical Approach.* Amsterdam: John Benjamins.

O'Donnell, Matthew Brook. Forthcoming. Some extensions and applications of the 'key word' procedure.

Rayson P., Berridge D. and Francis B. 2004. Extending the Cochran rule for the comparison of word frequencies between corpora. In Volume II of G. Purnelle, C. Fairon, A. Dister (eds.) *Le poids des mots: Proceedings of the 7th International Conference on Statistical analysis of textual data (JADT 2004), Louvain-la-Neuve, Belgium, March 10-12, 2004*, Presses universitaires de Louvain, 926-936.

Scott, Mike, 2008. WordSmith Tools Version 5.0.

Scott, Mike. 1997. PC Analysis of key words – and key key words. *System,* (25), 1-13.

Scott, Mike & Tribble, Chris. 2006. *Textual patterns: Key words and corpus analysis in language education.* Amsterdam/Philadelphia: John Benjamins.

Semino, Elena & Short, Mick. 2004. *Corpus Stylistics. Speech, Writing and Thought Presentation in a Corpus of English Writing.* London/New York: Routledge.

Sinclair, John. 1991. *Corpus, concordance, collocation.* Oxford. Oxford University Press.

Stubbs, Michael. 1996. *Text and corpus analysis: Computer-assisted studies of language and culture.* London. Blackwell.

White, Peter. 1997. Death, disruption and the moral order: the narrative impulse in mass-media 'hard news' reporting. In F. Christie and J. R. Martin (eds.) Genre *and institutions. Social processes in the workplace and school,* 101-133. London: Continuum.

Examples quoted from the *Guardian*:

(1) "Tories oust Duncan Smith aide", 8 May 2003, Nicholas Watt and David Hencke
(2) "Tories protest at history shift", 5 August 1999, Rebecca Smithers
(3) "Blair attacks 'wreckers' on left and right", 4 February 2002, Patrick Wintour
(4) "Row as author of Iraq dossier is made head of MI6", 7 May 2004
(5) "Labour faces control freak taunt over leftwinger", 9 January 2002
(6) "Meacher faces row over toxic ash sold to builders", 5 July 2001
(7) "Tebbit in Le Pen row for Tories", 17 October 2002, Nicholas Watt
(8) "Knives out for Survivor show as contestant goes to court", 10 June 2001
(9) "Bahl makes fresh victimisation claims against Law Society", 5 December 2002
(10) "RUC failed to tell top nationalists of dossier", 19 May 1999, John Mullin
(11) "Sissons sneers at BBC 'beauty contest' news policy", 31 August 1998
(12) "Pensions poll blow shocks Labour", 24 September 2000
(13) "Byers 'wouldn't talk to top civil servant'", 4 March 2002
(14) "Fergie flaunts family affairs as post-divorce 'role model'", 29 April 1998

Table 8: The nucleus pattern for *fresh embarrassment*: 28 of 28 articles annotated

FACE/SPARK	PRE-MODIFIER	ABOUT	TIME	AFTER
28 (100%)	2 (7%)	13 (46%)	26 (93%)	27 (96%)
faced 14 *facing* 3 *heaped* 2	*political* 1 *pre-election* 1	*over* 12 *for* 1	*last night* 13 *yesterday* 11	*when* 14 *after* 5 *by* 3 *as* 3
(x1) *face, struggling to contain, moved to head off, suffered, cause, risks, save from, was heaped, was facing*			(x1) *are set to, today*	(x1) *with, following*

Table 9: The nucleus pattern for *fresh controversy*: 40 of 43 articles annotated

FACE/SPARK	PRE-MODIFIER	ABOUT	TIME	AFTER
40 (100%)	2 (5%)	17 (43%)	30 (75%)	30 (75%)
embroiled in 6 *plunged into* 5 *facing* 3 *erupted* 2 *triggered* 2	*political funding* 1 *political* 1	*over* 13 *about* 4	*last night* 18 *yesterday* 10 *today* 1 *this week* 1	*after* 14 *when* 11 *as* 2
(x1) *set to create, ran into, moved last night to quash, surrounded, prompting, looks set to spark, spark, waded into, at the centre of, mired in, enmeshed in, there was, courted, flared, engulfed by, risking, erupt into, seek to avoid, braced for, prompted, faced, engulfed in*				*with, amid, by*

Table 10: The nucleus pattern for *fresh blow*: 21 of 25 articles annotated

FACE/SPARK	PRE-MODIFIER	ABOUT	TIME	AFTER
21 (100%)	0 (0%)	1 (5%)	17 (81%)	17 (81%)
suffered 6 *will be dealt* 3 *faced* 2		*over* 1	*last night* 6 *yesterday* 6 *today* 5	*when* 12 *after* 3 *with* 1

				as 1
(x1) *hit by, dealing, was facing, has taken, were dealt, deal, have been dealt, received, has been dealt, will suffer*				

Michaela Mahlberg & Matthew Brook O'Donnell
School of English, Modern Language Building
University of Liverpool
Liverpool
L69 7ZR, UK

m.mahlberg@liverpool.ac.uk
m.b.odonnell@liverpool.ac.uk