Fleeing, Sneaking, Flooding

A Corpus Analysis of Discursive Constructions of Refugees and Asylum Seekers in the UK Press, 1996-2005

Costas Gabrielatos
Lancaster University, UK
Paul Baker
Lancaster University, UK

This paper examines the discursive construction of refugees and asylum seekers (and to a lesser extent immigrants and migrants) in a 140-million-word corpus of UK press articles published between 1996 and 2005. Taking a corpus-based approach, the data were analyzed not only as a whole, but also with regard to synchronic variation, by carrying out concordance analyses of keywords which occurred within tabloid and broadsheet newspapers, and diachronic change, albeit mainly approached from an unusual angle, by investigating consistent collocates and frequencies of specific terms over time. The analyses point to a number of (mainly negative) categories of representation, the existence and development of nonsensical terms (e.g., *illegal refugee*), and media confusion and conflation of definitions of the four terms under examination. The paper concludes by critically discussing the extent to which a corpus-based methodological stance can inform critical discourse analysis.[1]

Keywords: asylum; collocates; corpus; discourse; keywords; newspapers; refugees

Introduction

There are two glaringly different worlds of thought in Britain today. In the fantasies of the politically correct dictators of liberal fashion, there is no problem. They will never concede that most asylum-seekers are economic migrants, rather than people fleeing persecution. And those who take a different view are obviously xenophobic. (*Daily Mail*, November 8, 2002)

Government ministers make disparaging references to 'economic migrants'. Although not fleeing the threat of war or persecution, many economic migrants are attempting to escape from extreme poverty. Denying asylum to economic migrants means sending people back to face disease, squalor and starvation. We as a rich nation should be far more generous to such asylum-seekers. (*Guardian*, February 11, 1999)

Refugees, asylum seekers, immigrants, and migrants (henceforth RASIM) coming to the UK have attracted increased press attention in the last fifty years (Greenslade 2005). In general, the presentation of RASIM in the press is negative (van Dijk 1988, 2005), having even been termed "negative misinformation" (Greenslade 2005, 12). Most linguistic research on issues of asylum and immigration has taken a critical discourse analysis (CDA) stance, which traditionally carries out a close analysis of a small sample of texts, focusing on aspects such as positive self-presentation and negative other-presentation (e.g., Reisigl and Wodak 2001) and argumentative and linguistic strategies employed for predication, labeling, argumentation, perspectivation, and intensification/mitigation.

However, corpus-based approaches have been applied to the examination of discourses or ideologies in political texts (Flowerdew 1997; Fairclough 2000; Piper 2000; Partington 2003), scientific writing (Atkinson 1999), and newspaper articles (van Dijk 1991; Morrison and Love 1996; Caldas-Coulthard and Moon 1999; Charteris-Black 2004). Such studies have shown how corpus analysis can uncover ideologies and evidence for disadvantage, for example Hunston's (2002) study of constructions of the deaf, Baker's (2005) examination of gay men, and Mautner's (2007) study of the elderly. Hardt-Mautner (1995) and Baker (2006) give overviews of the ways in which corpus-based analysis can aid critical discourse analysis. In terms of corpus-based studies that examine refugees and related identities, Baker and McEnery (2005) carried out an analysis of a relatively small (130,000 words) corpus of British newspaper texts published in 2003, finding quantitative evidence of linguistic patterns being repeatedly used in negative constructions of refugees. A corpus-based approach also helps in addressing criticisms of CDA methodology related to the texts analyzed. More specifically, CDA studies have been criticized for arbitrary selection of texts, which is seen to cast doubts on their representativeness, and the analysis of a small number of texts or text fragments, which cannot be expected to reveal helpful patterns or insights into their frequency or distribution (see for example Stubbs 1994, 204; 1997, 103-107; Koller and Mautner 2004, 218).[2] In this study, the rationale for the source of texts (UK newspapers) was made explicit in the project aims, and the corpus was derived through a query arrived at via an algorithm developed during the initial phase of the project (see Gabrielatos 2007), which arguably ensured that the corpus was representative—that is, it comprised articles relevant to (issues related to) RASIM. In addition, the size and coverage of the corpus (140 million words, containing 175,000 full newspaper articles spanning ten years) ensured that relevant patterns would emerge in the analysis.

Corpus linguistics (CL) can pinpoint areas of interest for further/closer analysis (see also Mautner 2007). That is, emerging patterns (e.g., keywords, collocations) lead to the examination of their (expanded) concordances, or, when needed, the examination of whole texts. The examination will also point to patterns of meaning, use, or attitude. If necessary, downsampling can create a more manageable number of texts for the CDA analysis. The downsampling may be random, or informed by

frequency patterns: for example, periods of significantly higher frequency of articles on a particular topic, or of occurrences of interesting collocations (e.g., *illegal refugees*). Keyword analyses can reveal statistically significantly more frequent terms in different newspaper types, individual newspapers, or text types and genres within or across newspapers. Corpus linguistics methodology allows for a higher degree of objectivity—that is, it enables the researcher to approach the texts (relatively) free from any preconceived notions regarding their linguistic or semantic/pragmatic content. When the starting point is keyword analysis, the analyst is presented with a list of words/clusters which will then be examined in (expanded) concordances for their patterning and contextual use. Similarly, when the focus of the study is clearly defined entities, states, or processes, the analyst can examine the collocates of the lexis denoting them (e.g., *refugees*) through concordances. In either case, the words for which concordances are examined are not subjectively selected, but are those that fulfill certain statistical criteria (e.g., they score above a predetermined threshold in a statistical measure). This approach can be used in the categorization of key lexis and collocates, albeit to a lesser extent: the analyst will need to use his/her background knowledge (with all the attendant explicit or implicit preconceptions), which will render the categorization less than atheoretical. However, even given that caveat, this approach can be expected to yield a richer categorization than when the analyst has set out only to populate categories recognized in an existing theory. Of course, the initial (i.e., non-theory-specific) categories can and should be re-examined in the light of existing CDA categories.

The benefits of this combination are manifold:

- The results gain from the explanatory power of the CDA theoretical framework.
- The non-theory-specific categories emerging from the large-scale data analysis can inform the adaptation/expansion of existing CDA categories.
- The approximate quantification usually used in CDA studies (e.g., through the use of frequency adverbs) can be made more specific through (relative) frequency counts and statistical measures.

Corpus linguistics methodology can also be used to triangulate the findings of CDA studies (usually carried out on a relatively small number of texts). Quantification can reveal the degree of generality of, or confidence in, the study's findings/conclusions. CL can trace diachronic developments through the grouping and quantification of linguistic elements, as well as establish differences between (types of) newspapers, or text types/genres within the same newspaper, or between individual newspapers. (For a detailed account of the synergy between CL and CDA in the Refugees and Asylum Seekers (RAS) project see Baker et al., forthcoming).

This paper expands on Baker and McEnery's research by examining a much larger set of data (140 million words), over a longer period of time (1996-2005), from a range of different newspapers. The main research questions addressed are:

- In what ways are RASIM linguistically defined and constructed?
- What are the frequent topics of, or issues discussed in, news articles relating to RASIM?
- What attitudes toward RASIM emerge from the body of UK newspapers seen as a whole?
- Are conventional distinctions between broadsheets and tabloids reflected in their stance toward (issues relating to) RASIM?

Our approach consisted of two separate methodological strands, one which involved a corpus-based analysis, using computer software to investigate wide-scale linguistic patterns and trends in the data, while the other involved a more traditional critical discourse analysis (CDA), which carried out a close analysis of a small sample of texts in the corpus, identifying discursive strategies such as labeling social actors, forms of argumentation, and strategies of intensification or mitigation. Reassuringly, both approaches yielded a large number of shared findings, although each approach was also able to make research discoveries that the other could not. This paper focuses mainly on the corpus-based methodology; however, we are in the process of publishing papers that detail the CDA analysis and the marriage between the two methodologies (see Baker et al., forthcoming).

The British Press

British newspapers can be generally classified along a range of distinctions such as frequency of publication (dailies vs. Sunday editions), coverage (nationals vs. regionals), political stance (conservative vs. liberal), and style (broadsheet vs. tabloid). Most British newspapers make no attempt to be unbiased, but instead reveal their stance on issues in a range of explicit and subtle ways (e.g., via editorials, selection of letters to be published, or language choice such as collocations or grammatical structure).

The press can influence the views of the public in general, and, more pertinently for this project, the public's stance toward minority social/ethnic groups. Duffy and Rowden (2005, 6, cited in Greenslade 2005, 7) report that, regarding public attitudes on issues of race and immigration, there are "strong indications … that newspapers have the greatest impact." The Refugee Council (UK) reports that the "reporting and commentary about asylum seekers and refugees is often hostile, unbalanced and factually incorrect."[3] This influence can be seen as a form of "social power," that is, "the *control* exercised by one group or organisation (or its members) over the *actions* and/or the *minds* of (the members of) another group, thus limiting the freedom of action of the others, or influencing their knowledge, attitudes or ideologies" (van Dijk 1996, 84; italics in the original). Social power may be manifested through the prerogative of the press, seen collectively, to decide on the amount of coverage to be given to particular incidents.

Newspapers select, highlight, or reject content and decide on the extent and frequency of coverage according to their editorial policy or agenda (Gordon and Rosenberg 1989, 4; Statham 2002, 395; Greenslade 2005, 3). Accordingly, RASIM are not expected to have any substantial influence on what is published in the press (particularly in the national newspapers) in relation to their nature, intentions, and actions. Additionally, newspapers can better afford the space required for an extended presentation/discussion of issues compared to television, and the same newspaper can publish a number of articles not necessarily expressing the same views. Also, the increasing online availability of newspapers potentially extends the influence described above to young people (aged 16-24), who tend to prefer the Internet over newspapers and television as their source of news (Coleman et al. 2002, 23-24).[4]

However, the relation between the press and its readers is bidirectional and dynamic. Individual newspapers have a vested financial interest in reporting on issues within their readers' concerns, as well as reflecting their views and attitudes, as newspaper readers tend to read those newspapers that are generally in accord with their own perceptions and approaches (Crawley and Sriskandarajah 2005, 3). The reciprocity of influence between readers and newspapers, and, more importantly, the power of newspapers over the selection, extent, frequency, and nature of their reporting, coupled with their availability for corpus compilation, make newspapers an excellent source of data for the examination of the construction of refugees and asylum seekers.

The Corpus

The data were collected through an online interface of newspaper and periodicals (LexisNexis) by way of the following search query:[5]

> refugee* OR asylum* OR deport* OR immigr* OR emigr* OR migrant* OR illegal
> alien* or illegal entry OR leave to remain AND NOT deportivo AND NOT deportment

Data were collected from nineteen UK newspapers, including six daily tabloids (*Sun*, *Daily Star*, *People*, *Daily Mirror*, *Daily Express*, *Daily Mail*) and their Sunday editions (*Sunday Express*, *Mail on Sunday*, *Sunday Mirror*, *Sunday Star*), five daily broadsheets (*Business*, *Guardian*, *Herald*, *Independent, Telegraph*), two Sunday broadsheets (*Observer*, *Independent on Sunday*), and two regional newspapers (*Evening Standard*, *Liverpool Echo*). Data were obtained for most of the newspapers from January 1996 through October 2005, although in a few cases data were not available until 1999 (*Business*), 2000 (*Sun*, *Daily Star*, *Sunday Star*), or 2001 (*Liverpool Echo*).[6]

The data were stored in separate files containing articles spanning one calendar month from each newspaper. The resulting corpus comprised 140 million words, consisting of 175,139 articles. Table 1 shows the breakdown for national broadsheets, tabloids, and regional newspapers.

Table 1
Number of Words and Articles in the Corpus and Sub-corpora

	Articles		Words	
	Number	% of corpus	Number	% of corpus
Broadsheets	100,242	57.24%	87,001,072	62.36%
Tabloids	50,476	28.82%	29,883,001	21.42%
Regionals	24,421	13.94%	22,625,964	16.22%
Total	175,139	100%	139,510,037	100%

Theoretical and Methodological Framework

As corpus linguistics is more a methodological than a theoretical approach,[7] there is significant overlap between its informing theoretical concepts and its methodological tools. The corpus-based aspect of the project was mainly informed by the notions of *keyness* and *collocation*, and the latter's related notions of *semantic preference* and *semantic/discourse prosody*. However, a number of CDA notions were also utilized when grouping collocates and keywords on the basis of the semantic prosody/preference that they expressed. These were the notions of *topos* and *topic*, specific metaphors commonly known to be employed in racist discourse, as well as *referential/nomination* and *predicational* strategies.

Keyness, in corpus linguistics methodology, refers to the significantly higher/lower frequencies of particular words (termed *keywords*) in one corpus when compared to their frequency in another corpus (Scott 1999). The level of the significance of this frequency difference is statistically calculated via log-likelihood tests (Dunning 1993). Keyword analysis was used to determine words that were significantly more frequent in broadsheets and tabloids. By grouping keywords relating to specific topics, attitudes, or argumentative *topoi* (Sedlak 2000, 157-158; Reisigl and Wodak 2001, 74-76, 156-157), and exploring their use in context, it was possible to identify discourses of RASIM in different types of newspapers. As the two broadsheet and tabloid subcorpora were quite large (87 and 30 million words, respectively), a very large number of keywords was expected. In order to limit the number of keywords to be examined, the threshold for keyness was set at an extremely low p value ($p \leq 10^{-14}$; see also McEnery 2006, 233, n32)—that is, the maximum probability that the keyness of a word was due to chance was one in a hundred trillion. However, even such a strict threshold of statistical significance returned over 1,500 keywords in each comparison.

The collocates of a word contribute to its meaning (e.g., Nattinger and DeCarrico 1992, 181-182), and their examination can provide "a semantic analysis of a word" (Sinclair 1991, 115-116). Collocations give information about the most frequent or salient ideas associated with a word; for example, if *illegal* and *immigrant* are often

paired as collocates, we may be primed to think of one concept even when the other is not present (Stubbs 1996). Collocation is statistically determined, and refers to the co-occurrence of two words within a prespecified span, when the frequency of the co-occurrence is above chance, taking into account the frequencies of the "node" (the word in focus), its collocates, and the collocation itself. In this project, the span was set at five words to the left and right of the node, and collocability was determined through the combination of two measures: mutual information (MI) and log-likelihood (LL). A common criticism of MI is that it tends to favor low-frequency words (Baker 2006, 102). However, this was not perceived as a problem for our purposes—rather, the contrary: low-frequency words are usually content words (nouns, adjectives, verbs), which are the word classes that can more clearly indicate semantic/discourse prosodies or topics/*topoi*. Furthermore, the collocates were not of interest as individual words, but as groups of words used to employ particular *topoi* or signifying particular topics (see also Baker 2004).[8] One shortcoming of MI, however, is that it is not a measure of statistical significance (McEnery 2006, 22), whereas it was deemed important that the results of the quantitative analysis were statistically significant, as this would strengthen their interpretation. To this end, the log-likelihood statistic was also used, in order to filter out those words with a fairly high probability of their co-occurrence with the nodes being due to chance. In sum, the MI score was used to indicate the strength of the collocation, whereas the LL score was used to determine the statistical significance of the co-occurence.[9]

At this point, we need to introduce a technique developed within this project, namely the calculation of "consistent collocates" (henceforth, "c-collocates"), and explain the motivation for it. As, in terms of number of articles/words, the corpus leans more toward the second part of the period it spans (i.e., 2000-2005), the possibility could not be discarded that collocations calculated on the whole corpus may be more representative of the discourse/attitudes of that period, particularly as most peaks in the number of articles are in the second part of the time period in focus (see Figure 1). Furthermore, some words may be very frequent collocates in some years, and non-collocates in others, but still feature as collocates across the whole corpus. Therefore, we sought to filter out any "seasonal" collocates, in order to focus on those that would point toward consistent and, arguably, core aspects of the linguistic means used to construct the representation of RASIM in the corpus. The examination of c-collocates also contributed to the diachronic aspect of the study. Normally, diachronic studies focus on changes over time. However, it can be argued that the examination of what has remained constant over time is also a valid and helpful component of the diachronic element of a study. To this end, collocations of *RASIM* were calculated for every annual subcorpus and tabulated alphabetically per year. In order for a collocate to be deemed consistent, we stipulated that it had to be a collocate in at least seven out of the ten annual subcorpora. Only a small percentage of the collocates of *RASIM* were consistent (5.4 percent, on average), which seems to confirm that the vast majority of collocates were seasonal—that is, related

to specific events. C-collocates are expected to be more accurate in drawing a picture of the core semantic attributes of *RASIM*, and the central semantic/discourse prosodies associated with them, as well as pointing toward the *topoi* consistently utilized in the corpus texts. Because of their spread over the entire temporal span of the corpus, c-collocates can arguably represent aspects of what Baker (2006, 13) terms "the *incremental* effect of discourse." A further reason why c-collocates are more dependable than collocates derived from the whole corpus is that their calculation posits a much higher minimum frequency for the collocates. The minimum frequency, in each annual subcorpus, for a collocation to be taken into account was set to five for *refugees, seekers,* and *immigrants*, and three for *migrants*. The latter lower threshold was dictated by the fact that, because of the relatively much lower frequency of *migrants*, a minimum of five occurrences filtered out a good number of interesting collocates. However, since c-collocates needed to be present in at least seven annual subcorpora, the actual minimum frequencies for c-collocations were 35 and 21 respectively—much higher than the usual minimum of five used in corpus studies. It must also be noted that in the case of collocates present in seven or eight subcorpora, there were no instances of the collocate being absent in consecutive years.

Related to collocation are the (overlapping) concepts of semantic preference and semantic/discourse prosody. Whereas collocation is a purely lexical (i.e., formal) relation, semantic preference can be seen as its semantic extension, as it is the relation "between a lemma or word form and a set of semantically related words" (Stubbs 2001, 65). The notions of semantic/discourse prosody further expand the possibilities of a collocational analysis by allowing for the examination of the attitude expressed or revealed, by particular collocations. Semantic prosody is defined by Louw (1993, 157) as the "consistent aura of meaning with which a form is imbued by its collocates." The choice of "imbue" has been criticized, mainly because its use can be understood to presuppose that the node is devoid of meaning (Whitsitt 2005, 288-293). It is pertinent to clarify, then, that the metaphor can equally well be understood in the context of the collocates adding (elements of) their meaning to the node. Stubbs (2001, 66) prefers the term "discourse prosody" in order to "maintain a standard distinction between aspects of meaning which are independent of speakers (semantics) and aspects which concern speaker attitude (pragmatics)." Also, discourse prosody "extends over more than one unit in a linear string" (Stubbs 2001, 66). However, the notion of semantic prosody, too, allows for that extension, as semantic prosodies are "established through the proximity of a consistent series of collocates" (Louw 2000, 57). For example, the two-word string *mug of* holds a semantic preference for hot drinks (*tea*, *coffee*, *hot chocolate*), whereas the string *sat through* attributes a negative semantic/discourse prosody to its collocates, as it often occurs in constructions that describe situations where people are made to endure long and boring events.

We now need to clarify the role of collocation, semantic preference, and semantic/discourse prosody in the examination of the representation of RASIM in UK

newspapers. The assumption under which this study operates is that, although some collocations of *RASIM* in the corpus may be explained by their definitions (e.g., collocates denoting their country of origin), this cannot be expanded to the use of evaluative words (e.g., *scroungers*). Furthermore, even words that at first glance may appear to be simply descriptive, such as those referring to a person's/group's race, religion, or nationality, can effectively become pejorative if their use is gratuitous or irrelevant. Clause 12 of the Code of Practice issued by the Press Complaints Commission in the United Kingdom specifies the following: i) The press must avoid prejudicial or pejorative reference to an individual's race, color, religion, gender, sexual orientation, or to any physical or mental illness or disability. ii) Details of an individual's race, color, religion, sexual orientation, or physical or mental illness or disability must be avoided unless genuinely relevant to the story.[10] When collocates are not warranted by the factual definitions of *RASIM* (see Table 2) or by the information needs in a particular context, then it seems plausible to assume that the function of collocates is anything but purely descriptive. Rather, it seems warranted to infer that their function in discourse is to imbue RASIM with particular semantic/discourse prosodies—that is, to attach to RASIM the attributive or evaluative meaning they embody, be it at a semantic or pragmatic level. In this light, the notion of semantic preference needs to be treated with caution, as it may be understood to imply that there are elements of the nature of RASIM that attract, and therefore warrant, the use of particular (sets of) collocates and the negative evaluative meaning they may embody or suggest. Therefore, the overlapping notions of semantic/discourse prosody are deemed more suitable for this analysis, as they make explicit that the frequent use of particular collocates may result in particular meaning attributes being arbitrarily associated with RASIM. However, the notion of semantic preference is still useful, in that it allows for the categorization of collocates in cohesive semantic sets, as well as for the collective treatment of all forms of a term (e.g., *refugee, refugee's, refugees, refugees'*).

Seen from a different, but compatible, perspective, it can be argued that the collocational profile of lexis denoting the groups whose construction in the press is investigated here provides insights into the different levels of "macrostructure" (Kintch and van Dijk 1978, 365) of the discourse constructing these groups, as "a macrostructure must be implied by the (explicit) microstructure from which it is derived" (Kintch and van Dijk 1978, 366; see also Phillips 1989).[11] Following Foucault, Stubbs (1996, 158) provides a definition of *discourse* (in the sense exemplified in expressions such as *sexist/racist discourse*) that offers further support for the use of collocational analysis to investigate discourses around RASIM:

> Recurrent phrases and conventional ways of talking, which circulate in the social world, and which form a constellation of repeated meanings . . . Vocabulary and grammar provide us with the potential and resources to say different things. But often this potential is used in regular ways, in a large number of texts, whose patterns therefore

embody particular social values and views in the world. Such discourse patterns tell us which meanings are repeatedly expressed in a discourse community.

As was mentioned above, the collocations of a word indicate the "recurrent phrases" (Stubbs 1996, 158) within which this word is used in a particular domain or genre (in our case, UK newspapers). We also need to consider that "all languages provide resources which always allow things to be expressed in different ways" (Stubbs 1996, 235). This may be a truism, but is, nevertheless, highly significant, because it highlights the choice of linguistic expressions available to journalists/ columnists. It does not seem controversial to suggest that the choice of words to be used in relation to RASIM can be used as a clear indication of the stance of the writer/newspaper toward these groups—particularly when the phraseology used is either incompatible to, or unwarranted by, the definitions of these terms, as are, for example, the expression *illegal asylum seeker* or the use of *scroungers* to refer to RAS. More importantly, as word meaning is dynamic, the frequent and consistent use of such collocations can, through the assignation of particular semantic/discourse prosodies, eventually change, expand, or contract the meaning of the terms in focus. Therefore, a collocational analysis can help render explicit the argumentation regarding RASIM pervading the texts under examination, and, in turn, point to the discoursal presentation of the groups in focus, as well as reveal elements of ongoing changes in the meaning of *RASIM*. For example, the examination of frequent collocations and the semantic/discourse prosodies they embody can reveal the linguistic means used in employing the referential/nomination and predicational strategies. In terms of methodology, collocation analyses can provide evidence that supports, refutes, or modifies conclusions based on small-scale qualitative analyses or formulated through introspection (i.e., post hoc reflection on informal observations).

These meaning changes may be enforced in a top-down manner by utilizing the power of the press to influence the use of the public, rather than emerging bottom-up through repeated use by individuals and consequently being adopted by the press. It is not our aim to resolve whether the semantic/discourse prosodies identified in this study have been assigned by the newspapers in the corpus, or whether they have simply been reproduced in them from the public domain. However, it does seem questionable to suggest that there are inherent elements in the nature of RASIM that, for instance, warrant frequent collocations with words such as *swarm*, *flood*, or *gang*—that is, words that create clear negative prosodies. Also, even if the prosodies applied to *RASIM* are reproduced rather than created by the newspapers, this does not negate the fact that, through their frequent use, these prosodies are strengthened, if not validated. The influence of the press in that respect can be described by using the following analogy by Stubbs (originally intended for language change): "each day's weather affects the climate, however infinitesimally, either maintaining the status quo or helping to tip the balance toward climatic change" (1996, 45). Simply put, through their collocational and, consequently, prosodic choices, newspapers make and communicate sociopolitical choices.

Concordance analysis is where corpus-based and CDA methodological approaches overlap. A concordance is a list of a given word or word cluster with its co-text on either side. Concordances can be sorted alphabetically on the left/right side of the word/cluster under investigation, and can thus facilitate the examination of different patterns or semantic/discourse prosodies. For example, if the focus is on adjectives predicating a word, then left-sorting would be more appropriate. Also, the concordance lines returned can be constrained to contain only particular words/clusters within a user-specified span from the node—a technique that can supplement and expand the collocational analysis. A further helpful feature of concordances is that the researcher can regulate the amount of co-text provided: from a few words on either side of the node to the whole text. Concordance analysis, then, can be used in ways akin to "qualitative" analysis. However, concordance analysis is more targeted, in that the analyst does not have to look for instances of the pattern under examination, but can specify it either in terms of a word/cluster or a sequence of grammatical categories (if the corpus has been tagged).

We turn now to the definition of the term *topos* (plural, *topoi*), and the justification of the collation of *topoi*, topics, and metaphors when grouping collocates and keywords within the corpus-based analysis, given that they are treated as different entities in CDA. *Topoi* are "conclusion rules that connect the argument with the conclusion" (Reisigl and Wodak 2001, 74-76), or, simply put, they represent "the common-sense reasoning typical for specific issues" (van Dijk 2000, 97-98), whereas "topics" simply refer to the subject matter of the discussion (Sedlak 2000, 129-130). However, *topoi* can be reasonably expected to be framed within discourse units of a compatible topic. Similarly, it is not uncommon for *topoi* to be embodied in metaphors. For example, van der Valk (2000, 234) comments that the metaphor of "water" "symbolizes the loss of control over immigration. Too many immigrants enter the country. We lost control over the process." Statements employing this metaphor (e.g., *immigrants are flooding the country*) can very well utilize a *topos* of Number.

The corpus-based analysis, therefore, involved several stages. First, the keywords were derived (via a frequency-based comparison of tabloid and broadsheet texts), with a subsequent focus on words that directly reference *RASIM* (*refugees*, *asylum seekers*, *immigrants, migrants*). The *RASIM* words and keywords were then qualitatively examined via detailed line-by-line concordance analyses, in order to identify semantic/discourse prosodies. Common collocates of these words were also investigated and subjected to further concordance analyses. Additionally, time was considered as a factor, so that frequencies of various terms could be investigated diachronically in order to explore changing discourses. The concordance program WordSmith Tools was used in order to facilitate the corpus analysis (Scott 1999). WordSmith allows analysts to view and sort concordances, as well as calculating lexical and phrasal frequencies, keywords, and collocations. Subjective researcher input was involved at almost every stage of the analysis. The analyst had to decide what was to be analyzed (e.g., whether to contrast conservative and liberal articles or to

Table 2
Contrasting Definitions of *RASIM* Terms

	Longman Dictionary of Contemporary English on CD-ROM (2003)	Oxford English Dictionary Online (2007)	Refugee Council[12]
Refugee	Someone who has been forced to leave their country, especially during a war, or for political or religious reasons.	One who, owing to religious persecution or political troubles, seeks refuge in a foreign country. A runaway; a fugitive from justice, etc. rare. Someone driven from his home by war or the fear of attack or persecution; a displaced person.	Someone whose asylum application has been successful and who is allowed to stay in another country having proved they would face persecution back home.
Asylum seeker	Someone who leaves their own country because they are in danger, especially for political reasons, and who asks the government of another country to allow them to live there.	A person seeking refuge, esp. political asylum, in a nation other than his or her own.	Someone who has fled persecution in their homeland, has arrived in another country, made themselves known to the authorities and exercised the legal right to apply for asylum.
Immigrant	Someone who enters another country to live there permanently.[13]	One who or that which immigrates; a person who migrates into a country as a settler. An animal or plant that has migrated into a given area, esp. one now living there; also, an animal (esp. a bird) that regularly or occasionally migrates into a given area.	—
Illegal immigrant	Someone who comes to live in another country without official permission.	—	Someone who has arrived in another country, intentionally has not made themselves known to the authorities and has no legal basis for being there.
Migrant	Someone who goes to live in another area or country, especially in order to find work.	That migrates; characterized by migration. Also (occas.): wandering, nomadic. A person who moves temporarily or seasonally from place to place A person who moves permanently to live in a new country, town, etc., esp. to look for work, or to take up a post, etc.;	[economic migrant] Someone who has moved to another country to work.
Emigrant	Someone who leaves their own country to live in another.	One who removes from his own land to settle (permanently) in another.	—

focus on broadsheet and tabloid articles), as well as determine which corpus-based processes were to be applied to the data and what the "cut-off" points of statistical relevance should be. In terms of the analysis of semantic/discourse prosodies, the researcher was required to analyze hundreds of lines of concordance data by hand in order to identify wider themes that were not so easily spotted via collocation, keyword, or frequency analysis. The examination of expanded concordances of keywords and collocates also informed the initial categorization of keywords and collocates, as the contextual clues provided by the expanded co-text helped clarify their use (see Brown and Yule 1982, 47). Subsequently, the *topoi*, topics, and metaphors identified by CDA research were used to fine-tune the categorization.

Definitions of *RASIM*

Before embarking on the examination of the overlapping use of *RASIM*, it seems pertinent to consider existing "official" definitions of the four terms, as well as the related term *emigrant* (see also Krishnamurthy 1996; Kjellmer 2003). Table 2 contrasts dictionary definitions of the terms under examination with those of the Refugee Council.

An important distinction to note is that the dictionary definitions present an asylum seeker as a refugee who has applied for asylum and so imply the temporal sequence *refugee → asylum seeker*, whereas the definitions of the Refugee Council clearly imply the opposite sequence. If newspapers operate according to the dictionary definitions, then *asylum seekers*, seen as seeking permanent or long-term residence, would perhaps be expected to share a large proportion of collocates with *immigrants/migrants*. Conversely, *refugees*, seen as transient, would be expected to show little overlap in these categories of c-collocates with *immigrants/migrants*.

Results

Due to issues of space, we cannot report all of the results of our analysis in this paper, but instead have summarized some of the most interesting findings under the categories of diachronic patterns, collocates and *topoi*, confusion/conflation of *RASIM* terms, broadsheet vs. tabloids, and nonsensical terms.

Diachronic Change

A frequency count of the number of articles per month in the corpus (Figure 1) reveals two key trends: first, a clear upward trend between 1996 and 2005,[14] and second, a number of sharp rises ("spikes") and falls, which may be indicative of occasional/seasonal interest in RASIM in the media.

Figure 1
Number of Articles per Month in the Corpus (1996-2005)
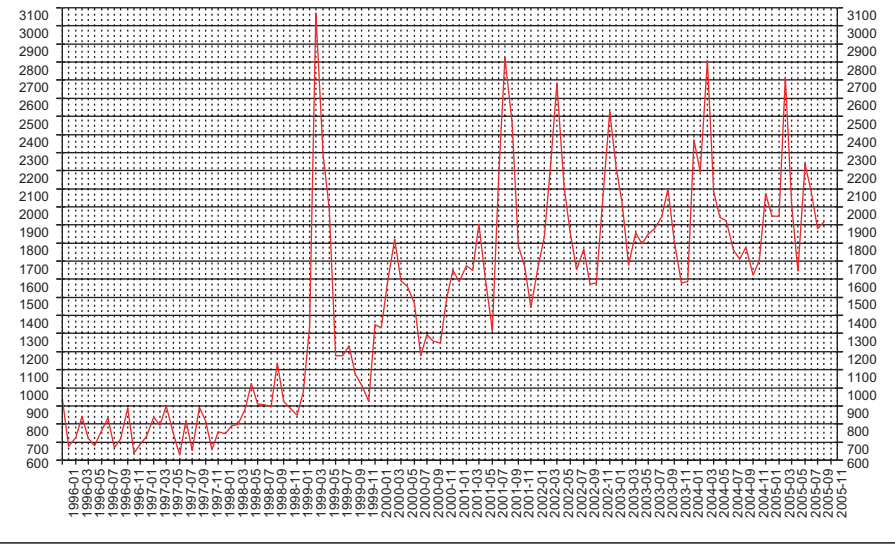
Aside from any sociopolitical agendas pursued by newspapers, actual events, too, can be reasonably expected to play a role in shaping public sentiment, as well as influencing the amount and bias of press coverage relating to RASIM (see Greenslade 2005, 5-6). The existence of a causal link between events and press attention to RASIM is supported by the examination of the events around the major peaks in Figure 1, summarized in Table 3.

Clearly, major wars, natural disasters, and terrorist attacks resulted in an increased focus on RASIM. However, it is also interesting that two of the "spikes" in our data co-occurred with political events in the United Kingdom: the Asylum Bill (March-April 2004) and the UK general elections (March-May 2005). During these periods, the construction of RASIM worked as part of (usually negative) media comment on government policies (what could be termed a "political rivalry" discourse). At these times RASIM were thus functionalized as part of a struggle for political hegemony, being discursively constructed as a people who merely constitute the topic of political debate, somewhat dehumanized as an "issue." Also, informal observations during the concordance analysis indicate that the corpus newspapers tend to adopt a positive or neutral stance when referring to RAS entering (or being likely to enter), temporarily residing, or applying for asylum to countries other than the United Kingdom.

Table 3
Correspondence between Events and Spikes in Press Attention to RASIM
Issues

Period	Events
March-May 1999	War in Kosovo
	Fighting between separatist guerrillas and paramilitary forces in East Timor
September-October 2001	Twin towers attack
	U.S. attack on Afghanistan
	Australian "boat people" incident
April-May 2002	Siege of the Church of the Nativity for 38 days.[15]
	War in Afghanistan
	East Timor independence
December 2002-February 2003	Chechen suicide truck-bomb attack
	Iraq disarmament crisis
March-April 2004	Second round of French presidential elections.
	The Asylum Bill in the United Kingdom.
	EU expansion-related immigration checks scandal.
	Madrid explosions
	Palestinian suicide bombers
	Violence in Kosovo
	Darfur ceasefire
	Assassination of Pim Fortyun.
March-May 2005	UK general elections
	Earthquake in Sumatra

A further possible contributing factor toward the rise in news stories over time may be that, over the last two decades, there has been a significant overall increase in the number of those seeking asylum in the United Kingdom (Figure 2).

As can be seen in Figure 2, asylum applications in the United Kingdom fell dramatically between 2002 and 2005. The rising amount of press attention toward RAS could be seen as reflecting, to some degree, the rise of RAS in previous years, rather than being due to the actual numbers of RASIM in 2002-2005. A less charitable interpretation would be that the UK press, seen as a whole, chose to disregard the falling number of asylum applications. A further plausible contributing factor is the fact that the United Kingdom attracts the second-highest percentage of asylum applications in the European Union (EU), after Germany. To be precise, more than one-fifth (21.2 percent) of new asylum applications within the EU are made in the United Kingdom. Additionally, despite the fall in asylum applications since 2002, net migration in the United Kingdom, in line with that in the EU, has kept increasing since the mid-1980s, regardless of whether the new residents have come as asylum seekers or immigrants. This development needs to be seen in the context of the ongoing EU expansion, which is usually followed by internal mobility. Fears of large numbers of citizens from new EU states moving to the United Kingdom seem to have merged with, and been

Figure 2
Annual Number of Asylum Applications (Including Dependents)
in the UK, 1982-2005
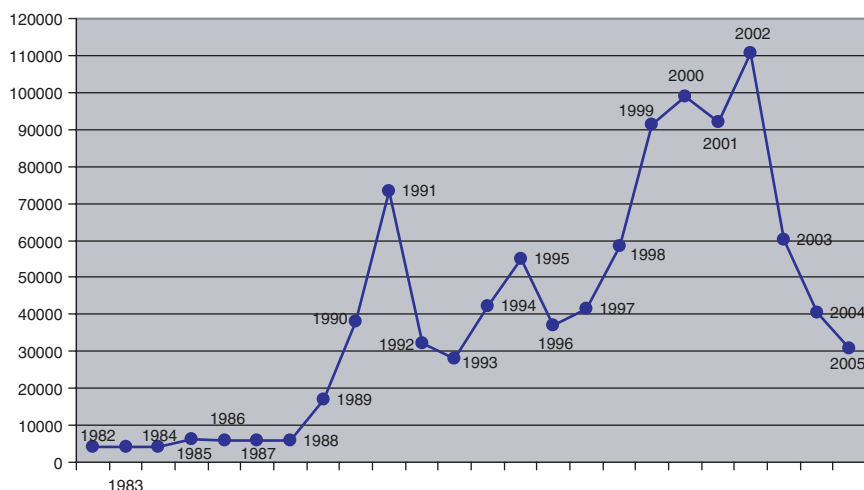
reinforced by, the increase in asylum seekers coming to the United Kingdom between the mid-1980s and mid-1990s. Van Leeuwen and Wodak (1999, 83-84) report a similar response in Austria, and also in the EU in general.

Examining C-collocates

The c-collocates of *refugees*, *asylum seekers*, *immigrants*, and *migrants* provide strong indications of the stance adopted in their representation. A first impression is that the discourses of RASIM in UK newspapers revolve around a small number of topics and employ a limited number of *topoi*, most of which denote a negative stance. The vast majority (86 percent) of the content c-collocates could populate eight categories (see Table 4 for an outline). Although the categories are informed by CDA theory,[16] they were not imposed on the c-collocates; rather, they emerged from the examination of their use in context (through concordance analysis).

From a psycholinguistic perspective, we also need to consider that, when reading, we do not normally store a verbatim version of a text in long-term memory; rather, what is stored in long-term memory is our understanding or, more precisely, our interpretation of the propositions put forward in the text (e.g., Clark and Clark 1977, 135–141). Therefore, it seems plausible that a central factor influencing what readers understand and remember (i.e., their interpretations) is the frequency of specific collocations and the semantic/discourse prosodies they communicate. More specifically,

Table 4
CDA-informed Categories Used in Grouping Collocates of *RASIM*

Category	Definition and Examples
Provenance/ destination/transit	Used to refer to all forms of *RASIM*. Words referring to the country, region, or continent that RASIM come from (provenance) e.g., Iran, Lebanon, Pakistan, Turkey, China; or the country, region, or continent they want to go to (destination) e.g., UK, or the place that they temporarily stay, or are held, while waiting to be allowed go to their place of destination, or while waiting for their case to be decided (transit) e.g., France. As the same word may refer, in different texts, to more than one sub-category (e.g., *France* is both a country of destination and transit), these collocates were grouped together.
Number	Used mainly to refer to refugees, these are words denoting their large number (e.g., *flooding, pouring, streaming*).
Entry	Used mainly to refer to refugees, asylum seekers, and migrants, these are verbs denoting entry to the country of destination/transit (e.g., *arrive, come, enter*), verbs denoting their journey (e.g., *cross*), verbs denoting their (illegal) attempts to enter (e.g., *trying*), verbs denoting their being in transit (e.g., *waiting*), or the place of entry (e.g., *borders, Dover*).
Economic problem	Mainly used for asylum seekers, immigrants, and migrants, these are words denoting either that they are a financial burden because they receive state benefits (e.g., *benefits, claiming, receive*), or a financial threat because they compete for jobs with existing citizens (e.g., *jobs, working*).
Residence	Mainly used for refugees and asylum seekers, these are words denoting their residence during transit (e.g., *camp, shelter, temporary*), or in the destination country (e.g., *housed, settled*), or verbs used to express opposition to the latter (e.g., *allowed, granted*).
Return/repatriation	Mainly used for refugees and asylum seekers, these are words referring to RASIM returning, or being made to return, to their country (e.g., *back, refused, return, sent*).
Legality	Mainly used for asylum seekers and immigrants, these are words concerning the legal status of RASIM. They are either direct—that is, attributative (e.g., *bogus, genuine, illegal*), or indirect—that is, they imply their illegality (e.g., *caught, detained, smuggled*).
Plight	Mainly used to refer to refugees, these are words referring to the situations that necessitated their leaving their country (e.g., *fear, forced, persecution*), the manner of their leaving (e.g., *escape, fleeing*), their current/recent state (e.g., *displaced, homeless*), or their current needs (e.g., *aid*).

what readers are expected to remember is not so much the verbatim collocations, but the prosodies—particularly as the same prosody can be embodied in a range of collocation patterns (e.g., *bogus/fake/illegal asylum seekers, bogus/illegal immigrants*). This makes the frequency of semantic/discourse prosodies much higher than that of the individual collocation patterns that give rise to them. Baker (2006,

114) concludes that "collocates can … act as triggers, suggesting unconscious associations which are ways that discourse can be maintained" (see also Fairclough 1989).[17] In this light, the semantic prosodies generated by frequent collocates can help create, reinforce, or exploit a *topos* without the need for any explicit argumentation, in that the reader is more likely to attend to the connotational rather than the denotational level of particular frequent collocations (e.g., *swarms of refugees*). From a CDA perspective, readers may accept the metaphor, and the premise-conclusion shortcut (*topos*) triggered by the metaphor, rather than engage critically with either of them, particularly as the same *topos* can be embodied in a range of collocations (and the attendant prosodies). For example, the "*topos* of burdening/weighing down" (Reisigl and Wodak 2001, 78)[18] can be utilized or reinforced by the strategic use of "quantity" or "group" collocations, particularly those expressed through emotionally charged metaphors (e.g., *flood/river/tide/wave of refugees*; *hordes/gangs of refugees*), which give rise to negative semantic prosodies related to the alleged inordinate number, and perceived attendant threat, of refugees. In fact, this seems to be compatible with the "*topos* of definition" (or "*topos* of name-interpretation"): "if an action, a thing or a person (group of persons) is named/designated (as) X, the action, thing or person (group of persons) carries or should carry the qualities/traits/attributes contained in the (literal) meaning of X" (Reisigl and Wodak 2001, 76).

A frequent element of the reporting/discussion on issues relating to RASIM is their number. The frequent quantification of RAS has been established through a quantitative analysis of a sample of 16,000 concordance lines of each term, revealing that one in five uses of *refugees/asylum seekers* is accompanied by some form of quantification. Three lemmas are c-collocates (FLOOD, POUR, STREAM), all of which embody water metaphors. Also, there is a clear imbalance in their distribution, with *refugees* having by far the most quantity c-collocates. The difference between *refugees* and *asylum seekers* could be explained with reference to their dictionary definitions, which seem to be the ones that newspapers operate with. Refugees are treated as potential or imminent entrants, and, therefore, newspapers feel they need to report on the "problem." Asylum seekers, on the other hand, are treated as already being in the country. This is also supported by the fact that, apart from one form (*poured*), these lemmas are usually used in habitual or progressive contexts. Finally, *migrants* does not register any quantity c-collocates, which may be seen as an indication of its relatively positive (or less negative) treatment in the corpus.

It is not surprising that "entry" and "residence" are two of the three most populated categories (the other is "provenance/destination/transit"), as all four groups are fundamentally defined by their desire to enter and reside in a country different from their own. However, this also indicates a preoccupation on the part of the UK press (seen as a whole) with RASIM entering, and staying in, the United Kingdom. Similarly, it is interesting to note the absence of any return/repatriation c-collocates for *IM*. Logically, if newspapers were concerned with the repatriation of RAS, they would be expected to be so for IM as well—particularly in light of the largely interchangeable

use of these terms (see below). However, the lack of such c-collocates for *IM* can be resolved if we consider that these terms are mostly used in the corpus to refer to citizens from the new EU countries who, by law, cannot be made to leave. This interpretation seems to be supported by the strong overlap of "entry" c-collocates in general, and the c-collocate *trying* in particular, in that, as these people cannot be made to leave, their arrival/entry becomes all the more an issue.

It is also worth pointing out that the word *trying* is a c-collocate of *refugees* and *asylum seekers*. Although *trying* (when referring to attempts to enter the destination country) is understandable as a c-collocate of *immigrants*, as they may not be legal, it is, nevertheless, puzzling as one of *RAS*, as it indicates the existence, or wish for, barriers to their entry, which, in turn, indicates a negative stance toward RAS. Let us now examine some representative examples in which the shared c-collocations play an important, if not central, role in constructing the identities of RASIM.

(1) MORE than 150 illegal immigrants are trying to sneak into the UK through the Channel Tunnel. Bosses say it's now time for an urgent crackdown by the British and French Governments. The daredevil refugees are risking their lives—and disrupting services for thousands of legitimate travellers. Eurotunnel finance director Richard Shireffs said hundreds of illegal immigrants were trying to creep through the tunnel or stow away on trucks. "It has been really getting quite serious for the last year. The immigrants are causing massive disruption to services. They are putting staff at risk and are a safety worry to themselves and to passengers. They try to walk through the tunnel, which has electrified rails. If we catch them they just turn round and try again." (*Daily Star*, February 20, 2001)

In example (1), *trying* is not incongruent with *illegal immigrants*, as, arguably, legal immigrants would not try to "sneak into the UK." However, the choice of the co-referring expression reveals the overlap, if not conflation, of terms discussed in more detail below: *illegal immigrants* in the first sentence of the example is co-referred to by *daredevil refugees* in the second. More clues to the newspaper's anti-immigration and anti-asylum agenda come from the second and last sentences in the example (both reported views used as support, or, at best, left unchallenged), in which the failures of the current ("soft") system are "exposed" and tougher measures are called for. Similar discoursal strategies, this time the c-collocation between *trying* and *asylum seekers*, are demonstrated in example (2):

(2) Backbenchers noted last week that post-September 11, Britain remains unprepared for terrorist attack. This is something else I noted first. A year ago I warned that in the event of a dreadful human epidemic, which then I'd assumed would be natural, the regime would be hopeless. Now it's official. Again, I told you so. Calais is still crawling with asylum seekers trying to break into Britain. Why do they come? Perhaps they still regard Blair as marginally more benign than Sadam Hussein. The key word is marginally. But who will engineer a regime change for the British? (*Sunday Times*, July 28, 2002)

This time, however, *trying* (as well as *break into*) is incongruous with *asylum seekers*, as an asylum seeker is, by any definition, someone who has registered, or intends to register, with the authorities (see Table 1 for definitions). Furthermore, the report of the attempt to "break into Britain" is framed by references to terrorism, the use of a negative metaphor ("dreadful human epidemic"), the large number of asylum seekers ("crawling"),[19] and thinly veiled calls, again through the use of a metaphor, for tougher measures ("regime change"). Very similar patterns are found in the example of the c-collocation between *trying* and *refugees*.

(3) FRENCH police caught 39 refugees trying to sneak into Britain yesterday hidden in the back of a German lorry. The asylum seekers had survived for three days in total darkness with just eight litres of water to drink between them, officials said. The immigrants—all Sri Lankans—were discovered in a random check at the port of Le Havre. (*Express*, January 21, 2002)

In example (3) the c-collocation involves, again, the verb *sneak*, and, in line with example 1, *refugees* is co-referred to as *immigrants*.

Although the vast majority of representations of RASIM in the UK press referenced mainly negative *topoi* and stereotypes, there were some more positive representations, particularly in broadsheets, which, however, were often framed within, or combined with, negative representations (examples 4-6; see also Gabrielatos and Baker 2006a):

(4) The country needs the talent and vibrancy an immigrant community will bring to a flagging native population base. Quieting public concern fanned by the disgraceful lack of control over entry to Britain and the failure to send home those eventually found to have no entitlement to remain is but a prelude to educating the public to recognise something more stark: Britain needs immigrants. We are not alone. The European Union saw birth rates fall below the natural replacement ratio of population as long ago as the last half of the Seventies, when Europe's demographic time bomb began ticking. Now the birth rate per woman is running annually at 1% below replacement rate. (*Business*, February 17, 2002)

(5) AN ARMY of 110,000 Iraqi refugees is heading for Britain to escape the looming war with Saddam Hussein. Ten thousand Kurds have already trekked to Turkish ports ready to cross into Greece—from where they can easily reach the UK. Greek officials say another 100,000 are set to join the exodus if war begins, dwarfing the 92,000 asylum seekers who entered Britain last year. (*Sunday Express*, December 29, 2002)

(6) HUNDREDS of illegal immigrants waiting to sneak into Britain staged a protest demo in France—complaining that it's getting too difficult. More than 250 immigrants, mainly Kurds and Afghans, took to the streets of Sangatte just outside Calais, where they are cared for in a Red Cross centre. They complained they are

"being treated like animals" when they try to escape and called on the French government to let them travel to England unhindered. The peaceful protest afternoon was allowed by French riot police who stood guard with batons. (*Daily Star*, August 4, 2001)

All three examples feature the close juxtaposition of positive and negative aspects relating to RASIM. Example (4) contains a mainly positive stance, relying on the *topos* of "usefulness" (see Sedlak 2000), albeit tempered by references to problems with the asylum/immigration system. In example (5) the plight of RAS ("to escape the looming war") is framed within references to their large numbers and their presentation as "an army," whereby arguably diminishing the effect on the reader of the mention of their plight. In example (6), "peaceful protest" sits uneasily with "illegal" and "sneak into," whereas the status of immigrants implied by the use of "cared for" seems to be contradicted by the later use of "try to escape."

The terms *Eastern* and *Europe* are c-collocates of *IM* only, which indicates regular references to citizens from the new/candidate EU countries coming to the United Kingdom. An examination of concordance lines reveals that the vast majority of *Eastern* occur within *Eastern Europe* and *Eastern European*. The attitudes expressed in these cases are more balanced, with articles arguing for the benefits of immigration from new/candidate EU countries, as in example (7), but they do not outnumber those arguing against, as in example (8).

(7) THE IMMINENT expansion of the EU has prompted fears of an influx of benefit-seeking immigrants from eastern Europe. But the reality is that many immigrants are well-qualified professionals willing to take low-paid, menial work that would probably be spurned by those educated to a similar level in Britain. (*Daily Telegraph*, February 7, 2004)

(8) BRITAIN was warned last night it faces a massive benefits bill to pay for the looming influx of immigrants, including gypsies, from eastern Europe. Tens of thousands of poor, unskilled workers seeking benefits are preparing to head for the UK by exploiting our lax border controls, ministers were told. (*Express*, February 9, 2004)

We also need to consider that, although in the EU "moves from one Member State to another are defined as mobility and not migration,"[20] *mobility* registers just one significant collocate in the corpus (*EU*), one which is only indirectly related to the European countries that had recently entered or were about to enter the EU between 1996 and 2005, whereas *immigration* has twelve such collocates.[21]

Conflation/Confusion of *RASIM* Terms

This section will examine closely the case of the overlapping use of the terms *refugees, asylum seekers, immigrants,* and *migrants* by examining their shared c-collocates. A look at the extent of the overlap of c-collocates between the four

Table 5
Overlap Ratio of Consistent Collocates

	Refugees	Asylum Seekers	Immigrants	Migrants
Refugees		34%	33%	18%
Asylum seekers	47%		46%	26%
Immigrants	34%	40%		39%
Migrants	38%	38%	79%	

terms (Table 5) provides the first indication for the overlap in their use. There are two different ratios of overlap for each pair of terms, as each term has registered different numbers of c-collocates. For ease of reference, comparisons start with the shaded column. For example, *refugees* shares 34 percent of its consistent collocates with *asylum seekers*, whereas *asylum seekers* shares 47 percent of its consistent collocates with *refugees*. The significant overlaps between *refugees* and *asylum seekers*, as well as between *immigrants* and *migrants*, are, of course, expected. What is unexpected, and at the same time indicative of the "misuse of terminology" reported by Greenslade (2005, 5) and the interchangeable use of *RAS* and *IM* reported by Baker and McEnery (2005), is the equally significant ratio of overlap between the consistent collocates of *immigrants/migrants* and *refugees/asylum seekers*. More precisely, in half of the cases, the unexpected overlap ratios are equal to the expected ones. For example, *asylum seekers* shares the same proportion of its consistent collocates with *refugees* (expected) and *immigrants* (unexpected).

The very high overlap of the content c-collocates of *immigrants* and *migrants* can be explained with reference to their almost identical definitions (Table 1). To a lesser extent, definitions can also account for the high overlap between *refugees* and *asylum seekers* in that, whichever definition is adopted, the two terms are linked temporally, or, in other words, both terms can refer to the same group of people in different stages of their journey to safety. However, a more interesting pattern is revealed by the examination of the directional overlap of c-collocates (Table 4). Whereas *migrants* shares the vast majority (79 percent) of its c-collocates with *immigrants*, the corresponding overlap of c-collocates of *immigrants* with *migrants* is only half (39 percent). This seems to indicate that the sense of *migrant* is part of the (more general) sense of *immigrant*, or that *migrant* has more specialized uses than *immigrant*—a view that appears to be consistent with dictionary definitions, which connect *migrant* with (short-term) work. To a lesser extent, the difference in the overlap between *asylum seekers-refugees* (47 percent) and *refugees-asylum seekers* (34 percent) can be interpreted as suggesting that, in UK newspapers, the sense of *refugees* is more inclusive than that of *asylum seekers*, in that an *asylum seeker* is also a *refugee*, but a *refugee* may not become an *asylum seeker*. If this observation is valid, then UK newspapers seem to approach the core sense of these terms according

to dictionary definitions, rather than the definitions of the Refugee Council. Let us now examine some representative examples of the overlapping use of RASIM.

(9) The entire set of rules governing migrants, be they asylum-seekers or any other type of immigrant, needs a total overhaul. (*Mirror*, November 28, 2001)

In example (9) the terms *migrant, immigrant*, and *asylum seeker* are conflated; in fact, *migrant* and *asylum seeker* are presented as "types of immigrant." All three groups are presented as constituting a problem that needs to be solved. This combination is observed even in positively disposed pieces, as in example (10) below.

(10) In a speech billed as refreshing thinking on race, Mr Singh pointed to a Mori opinion poll that found widespread negative perceptions of asylum and immigration: "This is indicative of a wider problem in British society of how immigration and asylum is packaged for the public. I think we have a fundamental problem of explanation and presentation." He said that he accepted there were economic migrants trying falsely to claim asylum. He also said the arrival of asylum seekers posed challenges to overstretched public services and poor communities. (*Guardian*, May 17, 2002)

Although the attitude toward RASIM reported in the article, as well as the stance of reporting itself, are clearly positive, it is, nevertheless, accepted that migrants may falsely claim asylum and that asylum seekers may pose an economic problem.

The intercollocation of *immigrants* and *asylum seekers*, seen in conjunction with the weak and one-directional collocation of *immigrants* and *refugees*, provides further indications that UK newspapers operate with definitions of *RAS* that are, in one crucial respect, diametrically different from the definitions used by the Refugee Council (see Table 2). In examples (11) and (12) we find more cases of overlap or inaccurate usage, *refugees* and *asylum seekers* being constructed as hyponyms of *immigrants*.

(11) Though some immigrants, including Asian refugees from African chauvinism, have contributed to this country's economic, professional, political and cultural life and deserve full credit for doing so, on balance mass immigration, in my view, has done more harm than good. (*Times*, April 2, 2001)

(12) Gradually—very gradually—a policy solution seems to be emerging. Countries start by allowing entry to job-creating entrepreneurs. Then they admit workers with skills that are in short supply. They also accept, perhaps for a limited period, large numbers of unskilled workers needed to keep hospital floors swept and construction projects going. Then they allow in some immigrants who are genuine asylum seekers. (*Times*, July 30, 2000)

Broadsheets vs. Tabloids

This section details the keyword analysis that was obtained by comparing the broadsheet and tabloid newspapers against each other. Although keyword analysis can offer very useful insights into the differences in content or style of the corpora under comparison, it is a fairly blunt instrument. The usefulness of a keyword comparison is contingent on the analysts' understanding of the nature of keyness and the make-up of the corpora compared. However, the corpora compared in this study are sufficiently large to ensure that keyness is not the product of the chance high density of a word in a relatively small number of articles. This is further ensured by the statistically calculated probability that keyness is due to chance, which is extremely low for all keywords examined in this section: no higher than 10^{-15}, or one in a quadrillion. Another consideration is that the keyness of a word does not ensure that it is directly associated with one or more of the query terms (as is the case with collocates). Rather, it indicates that the word is used in the same linguistic context as the query terms (e.g., the same text), and points to the co-existence of their semantic/pragmatic meanings in discourse. In other words, the keywords in one corpus, seen as a group or sets of groups (as in the present analysis), indicate the phraseology, and, by extension, the topics or stances usually favored by a specific (group of) newspaper(s) as compared to those of another. Of course, the results of a keyword analysis need to be triangulated against those of more fine-grained corpus techniques, such as collocation analysis. Finally, the categorization of keywords was informed by the examination of concordance lines, and, if necessary, larger parts of a text, or the whole text, from which the concordance line had been extracted. When examining the categorized keywords in this section, we will be taking into account the number and nature of (un)populated categories, as well as the number of keywords in each category. For ease of comparison, unpopulated categories have not been removed from the tables.

Table 6 shows the keywords in two groups of UK national newspapers in the corpus (tabloids, broadsheets) when compared with each other. In each case, keywords, in line with c-collocates, have been mainly grouped according to the *topoi*/topics distinguished in CDA studies (see Table 4); however, other categories had to be introduced to cater for a large number of keywords. The examination of keywords will start with those in tabloid newspapers for two reasons: tabloids return the largest number of keywords, and these keywords can be used to populate all the existing CDA categories, the vast majority of which refer to negative attitudes toward, and representation of, RASIM.

Overall, keywords indicate that tabloids dedicate their discussion of asylum and immigration to issues relating to the number of RASIM, the manner and place of entry to the destination country, their alleged abuse of the asylum/immigration system and the laws of the country (the "political rivalry" discourse referred to earlier), and the purported threat they pose to the safety and welfare of the citizens of the country. More importantly, apart from a number of keywords referring to the

Table 6
Tabloid vs. Broadsheet Keywords

Categories		Tabloid keywords	Broadsheet keywords
Query and related terms		asylum, emigrated, emigrating, immigrant, immigrants, immigration, migrants, refugee, refugees, seeker, seekers	claimant, diaspora, dissident, dissidents, émigré, émigrés, exile, exiles, foreign, gypsies
Entry	General	arrivals, arrive, arrived, arriving, passengers	
	Mode	aboard, container, eurostar, ferries, ferry, freight, lorries, lorry, train, trains, truck, trucker, truckers	
	Place/route	airport, calais, channel, chunnel, dover, eurotunnel, heathrow, tunnel	
	Legality	passport, passports, smuggle, smuggled, smuggling, sneak, sneaked, sneaking, stowaway, stowaways	
Number		flood, flooding, hundreds, influx, number, numbers, rocketed, swamped, thousands	dozens
Abuse of system		abuse, abused, abusing, bride, brides, cheat, cheating, cheats, marriage, marriages, married, marry, marrying, racket, racketeers, scam, scams, wedding, weddings	
Economic burden/threat		allowance, begging, benefit, benefits, cost, costing, expense, handout, handouts, job, jobs, jobless, rocketed, scroungers, spongers, strain, taxes, taxpayer, taxpayers	
Threat	Crime	arrested, crime, criminal, criminals, cronies, crooks, gang, gangmasters, gangs, gangsters, murder, murdered, murderer, murderers, murdering	
	Terrorism	antiterrorism, antiterrorist, anthrax, extremists, fanatic, fanatical, fanatics	
Residence		applications, citizenship, claim, claimants, claimed, claiming, claims, permit, permits	
Legality of claim		bogus, failed, fake, false, genuine, granted, illegal, illegally, illegals, legal, legally, posing, sham	authentic, legitimacy
Issues with UK asylum system		crisis, laws, lax, loophole, mockery, shambles	
Unwelcome		boot, booted, crack, crackdown, deport, deportation, deported, deporting, kick, kicked, leave	
Plight		desperate, desperately, devastated, devastating, horrific, horrified, horror, misery, tragedy, tragic, tragically, victim, victims	airstrikes, ceasefire, conflict, conflicts, coup, displaced, earthquake, genocide, holocaust, humanitarian, humanity, massacres, occupation, occupied, sanctions, survivors, torture
Positive effect			diversity
Other terms of interest		moslem, moslems	enlargement, european, integration
			christian, christians, hebrew, islamic, islamist, jewish, jews, Judaism, muslim, muslims, orthodox, religious, secular, synagogue
			ethnic, nationalism, nationalist, separatist, settlement, settlements, settlers, zionism, zionist
			fighters, freedom, guerrilla, guerrillas, hamas, hizbollah, hizbullah, insurgency, militants, rebel, rebels
			qa'eda
			military, militia, militiamen, militias

plight of refugees and asylum seekers, the majority of keywords populate categories reflecting a negative stance toward RASIM. It is also indicative that the two religion-related keywords (*Moslem*, *Moslems*) are considered to be terms of abuse.[22]

When we examine the broadsheet keywords, it is clear that there are striking differences. Apart from the category of "plight," which is equally well populated in both cases, the population and makeup of all other categories indicate a radically different treatment of, and stance toward, RASIM on the part of the two groups of newspapers. For example, broadsheets seem to prefer to use a wider variety of terms to refer to RASIM and use more terms with either positive or neutral connotations (e.g., *émigré, exile*). It is also interesting that more than half of the broadsheet keywords do not belong to CDA-related categories. Keeping in mind that these categories are mostly related to negative stance/representation, this seems to point to an overall more positive, or less negative, stance of broadsheets toward RASIM. Also, the majority of broadsheet keywords seem to suggest a more in-depth treatment of asylum/immigration issues, placed within a wider context (e.g., international, social, political, religious), with, arguably, more balanced argumentation. For example, the examination of religion-related keywords in concordances and whole texts indicates that issues of asylum and immigration are covered within discussions of a variety of topics (e.g., relations between Christianity and Islam), rather than only in articles treating perceived problems with asylum/immigration.

Nonsensical Terms

This section examines the frequency distribution of a number of expressions involving RASIM, the use of which indicates blatant negative bias: *illegal refugee(s)/asylum seeker(s)* and *bogus immigrant(s)/migrant(s)*, as well as their "positive" counterparts: *legal refugee(s)/asylum seeker(s)* and *genuine immigrant(s)/migrant(s)*.[23] Irrespective of the definition that one may operate on, none of these negative expressions makes sense. Also, their seemingly positive counterparts are equally nonsensical, and negatively biased, as they can only exist by virtue of their negative "opposites." In fact, their use prompted the Press Complaints Commission to issue a guidance note to newspaper editors informing them that "there can be no such thing in law as an 'illegal asylum seeker'" and asking them to refrain from the term's use.[24] In the light of the analysis of c-collocates, which provides compelling evidence for the overlap in use of the terms *refugees, asylum seekers, immigrants,* and *migrants,* it would be beneficial also to establish the extent to which different newspapers take this overlap to extremes by misapplying the adjectives *legal/illegal* to *RAS* and *genuine/bogus* to *IM*. In so doing, the newspapers employ inaccurate and misleading terms, which, at best, reveal ignorance, and, at worst, a negative agenda.

However, there is also a methodological interest in such an analysis, as far as corpus research is concerned. The quantitative analysis of the phraseology employed in expressing a strong stance, such as extreme negative bias, can provide a measure of

Figure 3
Frequency of "legal/illegal RAS + genuine/bogus IM" per Thousand Articles
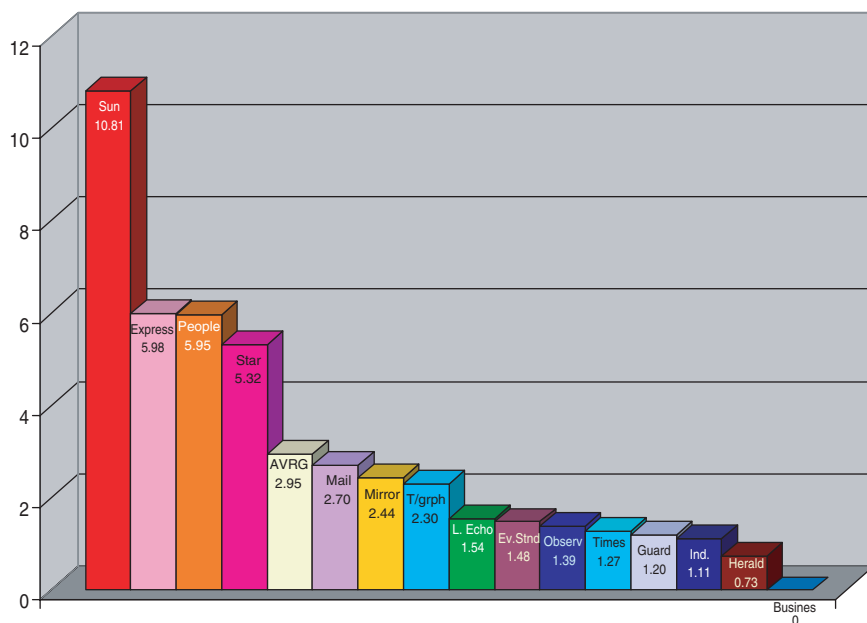
the "quality" of a newspaper—that is, the objectivity and balance of its reporting and discussion, in relation to a specific topic (see also Gabrielatos 2006; Gabrielatos and Baker 2006a, 2006b). It must be noted that articles in which the use of these expressions is challenged, or their nonsensical nature is clarified, have not been included in the analysis below. It should also be noted that these articles comprise only a small fraction of the cases, since all but one[25] were published after the Press Complaints Commission's guidelines had been issued. As the newspaper subcorpora are of different sizes, the comparison will be based on the normalized, rather than the actual, frequencies. The usual practice in corpus-based studies is to present normalized frequencies in terms of a fixed number of words, e.g., frequencies per million words. However, as tabloids tend to have shorter articles than broadsheets, the frequency per million words would tend to underestimate the relative frequency of the expressions in broadsheets and overestimate it in tabloids. For this reason, normalized frequencies were calculated in terms of articles (per thousand articles; Figure 3). A first observation is that the use of these nonsensical and negatively biased expressions is clustered around a relatively small number of newspapers: one-third of the newspapers in the corpus account for almost three-quarters of occurrences.

Figure 4
Annual Development of Frequency per Thousand Articles of
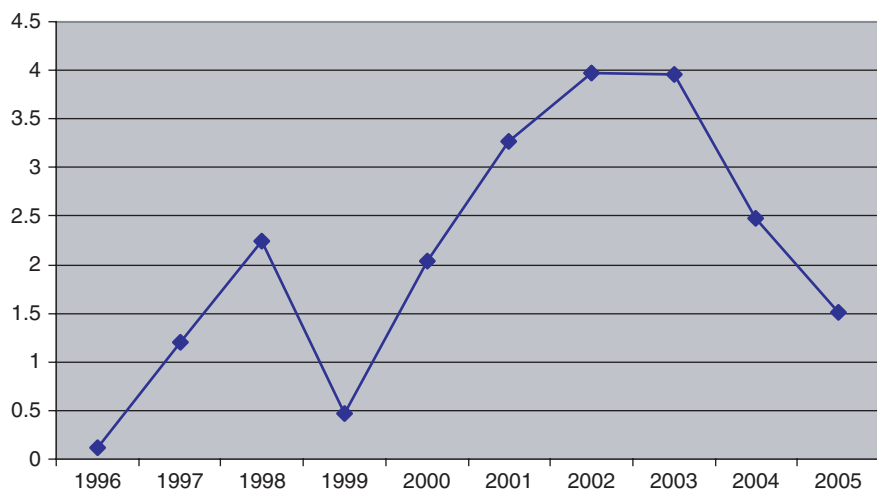**"legal/illegal RAS + genuine/bogus IM"**

Figure 3 gives an interesting indication of how "tabloid" a newspaper is, supporting our introspective categorization of UK national newspapers. Some newspapers are clearly at the "tabloid" end of the spectrum (*Express, People, Star, Sun*), while others are toward the "broadsheet" end (*Business, Evening Standard, Guardian, Herald, Independent, Liverpool Echo, Observer, Times*), with a group occupying the middle (*Mail, Mirror, Telegraph*). Those at the "tabloid" end show frequencies well above the average normalized frequency, those at the "broadsheet" end have frequencies below, and including, the median (*Liverpool Echo*), whereas those around the middle fall between the average and the median.

In line with the diachronic aspect of the project, the frequency development of these nonsensical expressions over time was also examined (see Figure 4).

Figure 4 reveals an interesting pattern. The fall in references in 1999 coincides with an overall "spike" in terms of media reporting on RASIM (see Figure 1). However, this spike was due to the conflicts in Kosovo and East Timor, which suggests that, in this case, the media were referring to refugees whose status could not be disputed. In addition, these refugees neither resided in, nor were likely to enter or seek asylum in, the United Kingdom. The rise in these terms during 2001-2003 could be the result of a time lag in terms of a media moral panic resulting from the peak in asylum applications from 1999-2002 (see Figure 2) as well as more general concerns about "foreigners" stemming from the 9/11 attacks on the United States. However, the sharp decline after 2003 could be mainly attributed to the guidelines to newspaper editors issued by the Press Complaints Commission on October 21, 2003.

Conclusion

The analysis has been exceptionally revealing in terms of enabling an understanding of the complex and often ambiguous media representations of RASIM in the United Kingdom. The continued use of conflated and confused meanings of *RASIM* words, along with the deployment of nonsensical terms and collocates indexing negative *topoi* or embodying negative metaphors, suggests that the conservative and tabloid British press are responsible for creating and maintaining a moral panic around RASIM, which has increasingly become the dominant discourse in the UK press.

Unlike many other CDA-based approaches, our corpus-based analysis was inductive: the corpus analysis directed the researcher toward salient or frequent linguistic patterns. The use of WordSmith Tools allowed for such patterns to be accurately and quickly identified. It must be reiterated that not all of the corpus-based analysis involved objectively running statistical algorithms over text data without human intervention. The analyst had to make sense of the linguistic patterns thrown up via the corpus-based processes. To this end, the CDA-based concept of *topoi* was useful in helping to provide a) a theoretical framework to organize the emerging linguistic/discourse patterns around, and b) existing *topoi* (provided by a substantial amount of previous CDA research) regarding asylum and immigration, against which the corpus researcher could compare his findings. At the same time, the non-theory-specific categories emerging from the large-scale data analysis can inform the adaptation/expansion of existing CDA categories.

However, a traditional corpus-based analysis is not sufficient to explain or interpret the reasons why certain linguistic patterns were found (or not found). Corpus analysis does not usually take into account the social, political, historical, and cultural context of the data. For this reason, a multidimensional CDA analysis that also goes beyond the "linguistic" elements of the text is instrumental in allowing researchers to consider issues such as:

- processes of text production and reception of the news data under analysis;
- the social context of the news industry in the United Kingdom (e.g., the competitive news market);
- (changing) political policy in the United Kingdom and elsewhere surrounding RASIM;
- statistics regarding immigration and asylum applications;
- social attitudes toward RASIM;
- meta-data, e.g., reports or talk about the newspaper texts under examination;
- macro-textual structures;
- text-inherent structures (coherence and cohesion devices).

While the corpus-based approach was therefore useful in identifying large-scale trends as well as minority cases, we would recommend that such a methodology only form part of the analysis, informing and being informed by a critical discourse

analysis approach. Clearly, there is more work that can be carried out on the corpus. It would be interesting to determine whether keywords were evenly dispersed over time, or whether certain words were key at various points. Additionally, an examination of where keywords appear within individual articles would be worth carrying out. Do certain keywords mainly occur at the beginning, middle, or end of articles? This would be a useful way of investigating whether keywords are linked to specific rhetorical effects and places in terms of argumentation strategies.

How much of the "bias" found in individual newspapers and texts is inherent to the text proper and how much is inherited from newswire sources that newspaper stories are often developed from? This would require acquiring Press Association data, as well as carrying out clustering procedures that could automatically compare the similarity coefficients of related news stories, identifying which parts are different within each. A related procedure would be to compare editorials or letters to the editor (which are not based on newswire material) with other newspaper text types and newswire sources.

Further work could be done in terms of relating discourses surrounding RASIM to methods of production and reception regarding individual newspapers—for example, examining how issues such as readership figures and demographics relate to the types of constructions that are found. And finally, we need to consider how theories of collocation relate to audience reception. Current theories within corpus linguistics claim that people are unconsciously primed to infer various meanings due to the cumulative effect of all of their previous encounters with a word, collocation, or phrase (e.g., Hoey 2005). But is that actually the case? And is the effect always the same, for everybody? What factors may have an impact on it? Cognitive or psycholinguistic techniques need to be carried out in order to investigate this priming effect more closely. These are issues we hope to address as part of a continuing analysis of this fascinating subject.

Notes

1. This project was made possible due to ESRC funding (grant number RES-000-22-1381).

2. For counter-arguments to these criticisms, see Baker et al. (forthcoming); for detailed suggestions on, and examples of, the use of CL techniques in (C)DA see Stubbs (1994, 216-218; 1997, 110-114).

3. For examples, see: http://www.refugeecouncil.org.uk/gettinginvolved/campaign/campaigners_pack/press_myths.htm.

4. All newspapers in the corpus are available online.

5. See Gabrielatos (2007) for a detailed discussion of how the query terms were derived.

6. The smaller amount of data at the start of the corpus collection point was taken into account when working out increases in references to particular words or phrases over time.

7. For a summary see McEnery and Gabrielatos (2006, 34-35).

8. However, function words can also indicate areas of interest, due to their cohesive function in discourse (see McEnery [2006] for a discussion of the function of *and* in moral panic discourses).

9. A word was accepted as a collocate if MI ≥ 3 and LL ≤ 6.63 (p ≤ 0.01).

10. http://www.pcc.org.uk/cop/practice.html.

11. "The microstructure is the local level of the discourse, that is, the structure of the individual propositions and their relations" (Kintch and van Dijk 1978, 365).

12. http://www.refugeecouncil.org.uk/practice/basics/truth.htm.

13. It is interesting to notice that both examples given in the Longman dictionary (which is corpus-based) carry negative prosodies: "an illegal immigrant," "a new wave of immigrants from the Middle East."

14. The frequencies from 1996 to 1999 (inclusive) are underrepresented, as some newspapers were not available on the database during this period. Of course, this makes the 1999 spike all the more striking.

15. The "Nativity Siege" incident is not straightforward. According to some accounts, armed/militant Palestinians trying to flee invading Israeli troops (as part of Operation Defensive Shield) took the clergy and civilians inside the church hostage. However, according to other accounts, a number of Palestinians (some of them armed and/or wanted by the Israeli army) took refuge in the church; the priests and nuns/monks stayed voluntarily in order to protect the church and those seeking refuge.

16. The categories are informed by central topics, metaphors, and *topoi* recognized in CDA.

17. Fairclough (1989, 54) states that "the effects of media power are cumulative, working through the repetition of particular ways of handling causality and agency, particular ways of positioning the reader, and so forth."

18. "If a person, an institution or a 'country' is burdened by specific problems, one should act in order to diminish these burdens" (Reisigl and Wodak 2001, 78).

19. Interestingly, the lemma CRAWL strongly collocates in the 100-million-word British National Corpus with the words *beetle, worms, ants, insects,* and *flies*, suggesting a metaphorical construction of asylum seekers as insect pests.

20. Source: http://ec.europa.eu/employment_social/news/2003/oct/eie2003_chap6_en.pdf, p. 190.

21. Bulgaria, Cypriot, Cyprus, Czech, Eastern [Europe], Estonia, Hungary, Poland, Portuguese, Romanians, Slovak, Slovenia.

22. The *Daily Mail* and *Daily Express* are the two newspapers that used these spellings most frequently. The Media Committee of the Muslim Council of Britain wrote to the editors of both newspapers on July 16, 2002, asking them to standardize their spelling of common Arabic words so as not to cause unnecessary offense to British Muslims. A possible reason for the unpopularity of *Moslem* could be that it can result in the pronunciation "mawzlem," which sounds similar to the Arabic word for 'oppressor.' The *Express* then started to favor the spelling *Muslim* in July 2002, whereas the *Mail* started using the spelling *Muslim* more often than *Moslem* around July 2003.

23. We are grateful to Ruth Wodak for drawing our attention to the nonsensical collocations and suggesting their closer examination.

24. http://www.pcc.org.uk/news/index.html?artiCle=OTE=.

25. This was in an article by Roy Greenslade, reporting on a ruling by the Press Complaints Commission, which "persuaded the *London Evening Standard* to apologize to the Refugee Council for incorrectly using the phrase 'illegal asylum seeker'" (*The Guardian*, Guardian Media Pages, p. 6, March 19, 2003).

References

Atkinson, Dwight. 1999. *Scientific Discourse in Sociohistorical Context*. Mahwah, NJ: Erlbaum.

Baker, Paul. 2004. Querying Keywords: Questions of Difference, Frequency, and Sense in Keywords Analysis. *Journal of English Linguistics* 324:346-59.

———. 2005. *Public Discourses of Gay Men.* London: Routledge.

———. 2006. *Using Corpora in Discourse Analysis*. London: Continuum.

Baker, Paul, and Tony McEnery. 2005. A Corpus-Based Approach to Discourses of Refugees and Asylum Seekers in UN and Newspaper Texts. *Journal of Language and Politics* 42:197-226.

Baker, Paul, Costas Gabrielatos, Majid Khosravinik, Michal Krzyzanowski, Tony McEnery, and Ruth Wodak. Forthcoming. A Useful Methodological Synergy? Combining Critical Discourse Analysis and

Corpus Linguistics to Examine Discourses of Refugees and Asylum Seekers in the UK Press. Under review in *Discourse and Society*.

Brown, Gillian, and George Yule. 1982. *Discourse Analysis.* Cambridge: Cambridge University Press.

Caldas-Coulthard, Carmen Rosa, and Rosamund Moon. 1999. Curvy, Hunky, Kinky: Using Corpora as Tools in Critical Analysis. Paper delivered at the Critical Discourse Analysis Meeting, University of Birmingham, UK.

Clark, Herbert H., and Eve V. Clark. 1977. *Psychology and Language*. New York: Harcourt Brace Jovanovich.

Charteris-Black, Jonathan. 2004. *Corpus Approaches to Critical Metaphor Analysis*. Basingstoke: Palgrave Macmillan.

Coleman, Stephen, Barry Griffiths, and Eleanor Simmons. 2002. *Digital Jury—The Final Verdict*. London: Hansard Society. Also available online: http://www.ofcom.org.uk/static/archive/bsc/pdfs/about/digjury.pdf.

Crawley, Heaven, and Dhananjayan Sriskandarajah. 2005. Preface. In *Seeking Scapegoats: The Coverage of Asylum in the UK Press*, edited by Roy Greenslade, 3. London: Institute for Public Policy Research.

Dunning, Ted. 1993. Accurate methods for the statistics of surprise and coincidence. *Computational Linguistics* 191:61-74.

Fairclough, Norman. 1989. *Language and Power*. London: Longman.

———. 2000. *New Labour, New Language?* New York: Routledge.

Flowerdew, John. 1997. The Discourse of Colonial Withdrawal: A Case Study in the Creation of Mythic Discourse. *Discourse and Society* 8:453-77.

Gabrielatos, Costas. 2006. Toward Quantifying 'Quality' in the Press: Comparing the Stance of UK Broadsheets and Tabloids toward Refugees, Asylum Seekers and Immigrants. Paper delivered at the joint meeting of the Corpus Research Group and the Language, Ideology and Power Research Group, Department of Linguistics and English Language, Lancaster University, 13 June. Available online: http://eprints.lancs.ac.uk/231/.

———. 2007. Selecting Query Terms to Build a Specialised Corpus from a Restricted-Access Database. *ICAME Journal* 31:5-43.

Gabrielatos, Costas, and Paul Baker. 2006a. Representation of Refugees and Asylum Seekers in UK Newspapers: Toward a Corpus-Based Comparison of the Stance of Tabloids and Broadsheets. Paper delivered at Critical Approaches to Discourse Analysis Across Disciplines (CADAAD), University of East Anglia, Norwich, UK, 29-30 June. Available online: http://eprints.lancs.ac.uk/250/.

———. 2006b. Representation of Refugees and Asylum Seekers in UK Newspapers: Toward a Corpus-based Analysis. Paper delivered at the Joint Annual Meeting of the British Association for Applied Linguistics and the Irish Association for Applied Linguistics (BAAL/IRAAL), University College, Cork, Ireland, 7-9 September. Available online: http://eprints.lancs.ac.uk/265/.

Gordon, Paul, and David Rosenberg. 1989. *Daily Racism: The Press and Black People in Britain*. London: The Runnymede Trust.

Greenslade, Roy. 2005. *Seeking Scapegoats: The Coverage of Asylum in the UK Press*. London: Institute for Public Policy Research. Also available online: www.ippr.org.uk/members/download.asp?f=/ecomm/files/wp5_scapegoats.pdf&a=skip.

Hardt-Mautner, Gerlinde. 1995. Only Connect: Critical Discourse Analysis and Corpus Linguistics. *UCREL Technical Paper 6*. Lancaster, UK: University of Lancaster.

Hoey, Michael. 2005. *Lexical Priming: A New Theory of Words and Language*. London: Routledge.

Hunston, Susan. 2002. *Corpora in Applied Linguistics*. Cambridge, UK: Cambridge University Press.

Kintsch, Walter, and Teun A. van Dijk. 1978. Toward a Model of Text Comprehension and Production. *Psychological Review* 85:363-94.

Kjellmer, Göran. 2003. Synonymy and corpus work: On *Almost* and *Nearly*. *ICAME Journal* 27:19-27.

Koller, Veronica, and Gerlinde Mautner. 2004. Computer Applications in Critical Discourse Analysis. In *Applying English Grammar: Corpus and Functional Approaches*, edited by Caroline Coffin, Ann Hewings, and Kieran O'Halloran, 216-28. London: Arnold.

Krishnamurthy, Ramesh. 1996. Ethnic, Racial and Tribal: The Language of Racism? In *Texts and Practices: Readings in Critical Discourse Analysis,* edited by Carmen Rosa Caldas-Coulthard and Malcolm Coulthard, 129-49. London: Routledge.

Longman Dictionary of Contemporary English on CD-ROM. 2003. London: Longman.

Louw, Bill. 1993. Irony in the Text or Insincerity in the Writer? The Diagnostic Potential of Semantic Prosodies. In *Text and Technology. In Honour of John Sinclair*, edited by Mona Baker, Gill Francis, and Elena Tognini-Bonelli, 157-76. Philadelphia; Amsterdam: John Benjamins.

———. 2000. Contextual Prosodic Theory: Bringing Semantic Prosodies to Life. In *Words in Context: A Tribute to John Sinclair on his Retirement*, edited by Chris Heffer and Helen Sauntson, 48-94. Discourse Analysis Monograph no. 18, CD-ROM. Birmingham, UK: University of Birmingham.

Mautner, Gerlinde. 2007. Mining Large Corpora for Social Information: The Case of 'Elderly.' *Language in Society* 361:51-72.

McEnery, Tony. 2006. Swearing in English: Bad Language, Purity and Power from 1586 to the Present. London: Routledge.

McEnery, Tony, and Costas Gabrielatos. 2006. English Corpus Linguistics. In *The Handbook of English Linguistics*, edited by Bas Aarts and April McMahon, 33-71. Oxford, UK: Blackwell.

Morrison, Andrew, and Alison Love. 1996. A Discourse of Disillusionment: Letters to the Editor in Two Zimbabwean Magazines 10 Years after Independence. *Discourse and Society* 7:39-76.

Nattinger, James R., and Jeanette S. DeCarrico. 1992. *Lexical Phrases and Language Teaching*. Oxford, UK: Oxford University Press.

Oxford English Dictionary Online 2007. Oxford, UK: Oxford University Press. http://www.oed.com/.

Partington, Alan. 2003. The Linguistics of Political Argument: The Spin-Doctor and the Wolf-Pack at the White House. London: Routledge.

Phillips, Martin. 1989. *Lexical Structure of Text*. Discourse Analysis Monograph no. 12, English Language Research, University of Birmingham, UK.

Piper, Alison. 2000. Some People have Credit Cards and Others have Giro Cheques: 'Individuals' and 'People' as Lifelong Learners in Late Modernity. *Discourse and Society* 11:515-42.

———. 2001. *Discourse and Discrimination: Rhetorics of Racism and Anti-semitism*. London and New York: Routledge.

Scott, Mike. 1999. *WordSmith Tools Help Manual*. Version 3.0. Oxford, UK: Mike Scott and Oxford University Press.

Sedlak, Maria. 2000. You Really Do Make an Unrespectable Foreign Policy. In *Racism at the Top: Parliamentary Discourses on Ethnic Issues in Six European States*, edited by Ruth Wodak and Teun A. van Dijk, 107-68. Klagenfurt: DRAVA-Verlag.

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

Statham, Paul. 2002. United Kingdom UK. In *Racism and Cultural Diversity in the Mass Media: An Overview of Research and Examples of Good Practice in the EU Member States, 1995-2000*, edited by Jessika ter Wal, 395-419. Vienna: European Research Centre on Migration and Ethnic Relations. Also available online: http://eumc.europa.eu/eumc/material/pub/media_report/MR-CH4-15-United-Kingdom.pdf.

Stubbs, Michael. 1994. Grammar, Text, and Ideology: Computer-Assisted Methods in the Linguistics of Representation. *Applied Linguistics* 15(2): 201-23.

———. 1996. *Text and Corpus Analysis: Computer-Assisted Studies of Language and Culture*. Oxford, UK: Blackwell.

———. 1997. Whorf's Children: Critical Comments on Critical Discourse Analysis. In *Evolving Models of Language*, edited by Alison Wray and Anne Ryan, 100-16. Clevedon, UK: Multilingual Matters.

———. 2001. *Words and Phrases: Corpus Studies of Lexical Semantics*. Oxford, UK: Blackwell.

van der Valk, Ineke. 2000. Parliamentary Discourse on Immigration and Nationality in France. In *Racism at the Top: Parliamentary Discourses on Ethnic Issues in Six European states*, edited by Ruth Wodak and Teun A. van Dijk, 221-60. Klagenfurt: DRAVA-Verlag.

van Dijk, Teun A. 1988. How 'They' Hit the Headlines. Ethnic Minorities in the Press. In *Discourse and Discrimination,* edited by Geneva Smitherman-Donaldson and Teun A. van Dijk, 221-62. Detroit, MI: Wayne State University Press.

———. 1991. *Racism and the Press*. London; New York: Routledge.

———. 1996. Discourse, Power and Access. In *Texts and Practices. Readings in Critical Discourse Analysis*, edited by Carmen-Rosa Caldas-Coulthard and Malcolm Coulthard, 84-104. London: Routledge.

———. 2000. On the Analysis of Parliamentary Debates on Immigration. In *The Semiotics of Racism*, edited by Martin Reisigl and Ruth Wodak, 85-104. Vienna: Passagen Verlag.

———. 2005. *Racism and Discourse in Spain and Latin America*. Amsterdam; Philadelphia: John Benjamins.

van Leeuwen, Theo, and Ruth Wodak. 1999. Legitimizing Immigration Control: A Discourse-Historical Analysis. *Discourse Studies* 1:83-118.

Whitsitt, Sam. 2005. A Critique of the Concept of Semantic Prosody. *International Journal of Corpus Linguistics* 103:283-305.

Costas Gabrielatos is a research associate and PhD student at Lancaster University, UK, doing corpus research on English *if*-conditionals. His current interests are the expression of time and modality in English, corpus-based discourse analysis, and the role of corpora in language teaching.

Paul Baker is a senior lecturer in Linguistics and English Language at Lancaster University, UK. He is an editor for the journal *Corpora*. His recent publications include *Using Corpora for Discourse Analysis* and *A Glossary of Corpus Linguistics* with Tony McEnery and Andrew Wilson.