Discourse & Society

A useful methodological synergy? Combining critical discourse analysis and corpus linguistics to examine discourses of refugees and asylum seekers in the UK press

Paul Baker, Costas Gabrielatos, Majid KhosraviNik, Michal Krzyzanowski, Tony McEnery and Ruth Wodak

>> Version of Record - May 15, 2008

What is This?

A useful methodological synergy? Combining critical discourse analysis and corpus linguistics to examine discourses of refugees and asylum seekers in the UK press

PAUL BAKER, COSTAS GABRIELATOS,
MAJID KHOSRAVINIK, MICHAŁ
KRZYŻANOWSKI, TONY MCENERY
AND RUTH WODAK
LANCASTER UNIVERSITY, UK

ABSTRACT   This article discusses the extent to which methods normally associated with corpus linguistics can be effectively used by critical discourse analysts. Our research is based on the analysis of a 140-million-word corpus of British news articles about refugees, asylum seekers, immigrants and migrants (collectively RASIM). We discuss how processes such as collocation and concordance analysis were able to identify common categories of representation of RASIM as well as directing analysts to representative texts in order to carry out qualitative analysis. The article suggests a framework for adopting corpus approaches in critical discourse analysis.

KEY WORDS: *asylum, critical discourse analysis, corpus, discourse historical approach, discrimination, method, migrants*

1. Introduction

This article describes and assesses the methodology used in the ESRC-funded project *Discourses of Refugees and Asylum Seekers in the UK Press 1996–2006* (henceforth, the RAS project), namely a novel, integrative combination of methodologies traditionally associated with corpus linguistics (CL) and critical discourse analysis (CDA).[1] We understand CDA to be an academic movement, a way of doing discourse analysis from a critical perspective, which often focuses on theoretical concepts such as power, ideology and domination. We do not view CDA as being a method nor are specific methods solely associated with it. Instead, it adopts any method that is adequate to realize the aims of specific CDA-inspired research. In general, however, many CDA practitioners have tended to use qualitative techniques, as well as taking into account analysis of the social, political, historical and intertextual contexts, which go beyond analysis

of the language within texts. Nor do we view CL as a single method, rather it utilizes a collection of different methods which are related by the fact that they are performed on large collections of electronically stored, naturally occurring texts. Many CL methods are quantitative and/or make use of statistical tests, which are performed by computer software. However, most CL methods require considerable human input, which often includes qualitative analysis (such as examining concordance lines). We have therefore tried to avoid describing CDA and CL as different 'methods' (but instead sometimes refer to methods traditionally adopted by CDA practitioners or by corpus linguists).

Because both CL and CDA are informed by distinct theoretical frameworks, their respective approaches to analysis are influenced by their informing theoretical concepts.[2] The RAS project aimed to render explicit the interaction between the various theories. Although this article focuses on the research synergy of CL and CDA (and more specifically, on the discourse–historical approach; DHA), it will, perhaps unavoidably, also comment on the more general use of CL techniques in what has been termed corpus-assisted discourse studies (CADS; Partington, 2004, 2006). In examining the combination of methods normally used by CDA and CL, we undertake to show that neither CDA nor CL need be subservient to the other (as the word 'assisted' in CADS implies), but that each contributes equally and distinctly to a methodological synergy.[3] More precisely, we address the following interrelated questions.

1. What are the respective merits and limitations of methods of analysis traditionally used by CL and CDA when the focus is on issues that CDA traditionally examines?
2. What should be the nature of such a methodological synergy?
3. How can the combination in research projects, and their potential theoretical and methodological cross-pollination, benefit CDA and CL?
4. How helpful and/or justified is the distinction between what have traditionally been termed *quantitative* and *qualitative* approaches in linguistics?

In focusing on the combination of CL and CDA techniques, it is not the intention of this article to provide a detailed account of our research findings relating to the construction of refugees and asylum seekers in the UK press, although we do describe some of our findings (Sections 6.4 and 6.5) in order to make illustrative points as they relate to wider methodological issues. Section 3 gives a short description of the research project, while the quantitative findings are more extensively discussed in Gabrielatos and Baker (2008); the qualitative results are reported in KhosraviNik (forthcoming).

2. The use of corpora and CL techniques in (critical) discourse studies

The use of methods associated with CL in order to carry out CDA is not a novel practice (Krishnamurthy, 1996; Stubbs, 1994), particularly given that both CL and CDA are relatively new movements in linguistics. Overall, the number of such studies in proportion to the number of studies in CL or CDA is extremely

small. However, more recently, it seems that use of CL techniques is becoming increasingly popular in critical approaches to discourse analysis. A case in point is a recent relevant edited collection (Fairclough et al., 2007), in which almost one in five articles is informed by corpus analysis.

Although the utility of using CL approaches in CDA and related fields has already been demonstrated (Baker, 2004a, 2006; Hardt-Mautner, 1995; Koller and Mautner, 2004; Mautner, 2000; O'Halloran and Coffin, 2004), it must also be noted that, in most such studies, the use of methods and theoretical frameworks traditionally associated with CDA and CL has not been balanced. Corpus-based studies may adopt a critical approach, but may not be explicitly informed by CDA theory and/or its traditional methods, or may not aim to contribute to a particular discourse-oriented theory (Krishnamurthy, 1996; Stubbs, 1994). Similarly, studies aiming to contribute to CDA may not be readily identifiable by corpus linguists as being corpus-based/driven[4] (Fairclough, 2000; Kovács and Wodak, 2003; Wodak et al., 1990), except for the seminal research by Gerlinde Mautner in the 1990s. Overall, the latter type of study tends to make limited or casual use of a corpus or corpus-based techniques. Sometimes, the corpus is used as a repository of examples (Flowerdew, 1997), as opposed to the analysis adhering to the 'principle of total accountability' (Leech, 1992: 112), that is, accounting for all the corpus instances of the linguistic phenomena under investigation.[5] CDA studies making use of corpora have, in general, tended to avoid carrying out quantitative analyses (see also Stubbs, 1997), preferring to employ concordance analysis (Magalhaes, 2006).[6]

When collocations (see Section 4) are examined within CDA research, they are not usually statistically calculated, but established manually through sorted concordances, and information regarding their statistical significance, the collocation span, or any frequency thresholds, is not usually provided (Piper, 2000; Sotillo and Wang-Gempp, 2004). Such approaches may miss or disregard strong non-adjacent collocates, or include non-significant collocates in the analysis. In some cases, the corpus used is very small (e.g., 25,000 words; Clark, 2007), that is, it is at the lower end of the range defining small specialized corpora (depending on the definition of 'small corpus').[7] This may be due to concerns that in a large corpus 'important features of the context of production may be lost when using such [i.e. CL] techniques' (Clark, 2007: 124), whereas a small corpus can 'be analysed manually, or is processed by the computer in a preliminary fashion . . .; thereafter the evidence is interpreted by the scholar directly' (Sinclair, 2001: xi). However, small corpora may lack some of the features in focus, or contain them in too small frequencies for results to be reliable, particularly when issues of statistical significance are not addressed. Ooi (2001: 179) suggests that 'the optimal size [of a corpus] can be reached only when the collection of more texts does not shed any more light on its lexicogrammatical or discourse patterning'; however, in the studies surveyed, there was no indication of such a concern in the corpus-building process. Finally, the corpus compilation may be flawed, in that the resulting corpus may not be representative (Meinhof and Richardson, 1994, cited in Stubbs, 1997), or, in extreme cases, the corpus may be biased (Magalhaes, 2006).[8]

However, there is a developing body of work which not only draws on both CDA and CL, but also aims to do justice to both, such as the studies by Baker and McEnery (2005) and Orpin (2005), as well as studies balancing CL and other discourse-oriented theories/methodologies, such as conversational analysis (Partington, 2003), moral panic theory (McEnery, 2006), sociolinguistics (Hardt-Mautner, 1995; Mautner, 2000, 2007), evaluation/appraisal (Bondi, 2007), stylistics (Semino and Short, 2004) and language and sexuality (Baker, 2004a).[9] The RAS project aimed to contribute to this paradigm. Ideally, the researcher(s) involved would be both corpus linguists and (critical) discourse analysts. The RAS project, arguably, adopted the next best solution: the collaboration of two teams working within the discourse–historical approach in CDA (DHA) and CL respectively.[10]

3. Description of the RAS project

3.1 FOCUS, AIMS AND RESEARCH QUESTIONS

The project aims were related to both subject matter and methodology. In terms of the former, the project set out to examine the discursive presentation of refugees and asylum seekers, as well as immigrants and migrants in the British press over a 10-year period (1996–2005). For reasons of economy, refugees and asylum seekers will be referred to by the acronym RAS, and immigrants and migrants by the acronym IM, whereas all four groups together will be referred to as RASIM. The analysis was concerned with both synchronic and diachronic aspects, while also contrasting the discourse used by broadsheets versus tabloids and national versus regional newspapers.[11] The main research questions addressed were:

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

As described earlier, it was sought to evaluate the utility of combining methods normally associated with CDA, with those normally used by CL, in order to ascertain the extent to which these approaches are complementary. A parallel aim was to demonstrate that the terms 'quantitative' and 'qualitative' may be more helpfully regarded as notional methodological extremes.[12] To that end, and in order to ensure that the results of the two research strands would be comparable, for the most part, the CL and CDA analyses were carried out separately, although there were points where both researchers contributed towards the analysis of each other, as described in Section 6.

3.2 DATA

Both strands used data from a corpus of 140 million words, compiled specifically for the project, which comprised articles related to RASIM and issues of asylum

and immigration, taken from twelve national and three regional newspapers, as well as their Sunday editions, between 1996 and 2005.[13] To aid the comparative and diachronic aspects of the project, the corpus was also divided into a number of sub-corpora, in terms of type of newspaper (broadsheets/tabloids, national/regional) and year of publication (10 annual sub-corpora). Although the CL analysis[14] made use of the whole corpus, given time and money constraints, a similar approach was not feasible for the CDA analysis. The CDA analysis thus was carried out on a sample of texts from the corpus, chosen in order to facilitate comparability of the results of the two strands (Section 6.2 describes how a sample of texts was selected for the CDA, through a novel sampling (downsizing) methodology created by the CDA researchers of the team).

Before giving illustrative examples of the different types of findings that the CL and CDA analyses uncovered, it is worth first outlining their respective theoretical and methodological profiles (Sections 4 and 5).

4. Theoretical and methodological profile of CL

It could be argued that CL methods offer the researcher a reasonably high degree of objectivity; that is, they enable the researcher to approach the texts (or text surface) (relatively) free from any preconceived or existing notions regarding their linguistic or semantic/pragmatic content. However, corpus-based analysis does not merely involve getting a computer to objectively count and sort linguistic patterns along with applying statistical algorithms onto textual data. Subjective researcher input is, of course, normally involved at almost every stage of the analysis. The analyst, informed by the quantitative aspects mentioned earlier, has to decide what texts should go in the corpus, and what is to be analysed. He/she then needs to determine which corpus-based processes are to be applied to the data, and what the 'cut-off' points of statistical significance should be. In corpus-assisted discourse analysis the researcher is normally required to analyse hundreds of lines of concordance data by hand, in order to identify wider themes or patterns in the corpus which are not so easily spotted via collocation, key word or frequency analysis. The analyst then has to make sense of the linguistic patterns thrown up via the corpus-based processes, usually with reference to one or more theoretical frameworks.

As mentioned in the Introduction, CL methodologies are not uniform. However, the techniques used in the RAS project are widespread in CL studies. In many respects, the approach used was compatible with the 'corpus-driven' paradigm of CL research (Tognini-Bonelli, 2001). That is, the CL analysis started with the examination of relative frequencies and emerging statistically significant lexical patterns in the corpus and sub-corpora mainly involving the four terms in focus: refugee(s), asylum seeker(s), immigrant(s), migrant(s),[15] and the close examination of their concordances. In fact, concordance analysis was used to supplement all other methodological tools. Two theoretical notions, and their attendant analytical tools, were central in the analysis: keyness and collocation.

Keyness is defined as the statistically significantly higher frequency of particular words or clusters in the corpus under analysis in comparison with another corpus, either a general reference corpus, or a comparable specialized corpus. Its purpose is to point towards the 'aboutness' of a text or homogeneous corpus (Scott, 1999), that is, its topic and the central elements of its content. In the RAS project, a key word analysis was carried out to examine differences between tabloids and broadsheets. As the topic of the corpus texts was known (RASIM and/or issues of asylum and migration), the examination of the strongest key words and clusters[16] in the two sub-corpora, combined with concordance analysis, provided helpful indications of the respective stance towards RASIM of the two types of newspaper. However, it may also be beneficial to examine the keyness not only of word-forms, but also of lemmas, word families,[17] and, more pertinently for this project, semantically/functionally related words (Baker, 2004b, 2006). By grouping together key words relating to specific topics, metaphors or topoi (as ascertained through concordance analysis), it was possible to create a general impression of the presentation of RASIM in the broadsheets and tabloids.

The definition of *collocation* adopted in the RAS project is the above-chance frequent co-occurrence of two words within a pre-determined span, usually five words on either side of the word under investigation (the *node*) (see Sinclair, 1991). The statistical calculation of collocation is based on three measures: the frequency of the node, the frequency of the collocates, and the frequency of the collocation. Because the collocates of a node contribute to its meaning (Nattinger and DeCarrico, 1992), they can provide 'a semantic analysis of a word' (Sinclair, 1991), but can also 'convey messages implicitly' (Hunston, 2002). On one level, collocation is a lexical relation better discernable in the analysis of large amounts of data, and, therefore, it is less accessible to introspection or the manual analysis of a small number of texts (Hunston, 2002). On another level, the meaning attributes of a node's collocates can provide a helpful sketch of the meaning/function of the node within the particular discourse. At this point, we need to introduce the concepts of semantic preference, and semantic/discourse prosody (terms which are sometimes used inconsistently or interchangeably), as they can be seen as the semantic extension of collocation. Semantic preference refers to semantic, rather than evaluative, aspects; it is the relation 'between a lemma or word form and a set of semantically related words' (Stubbs, 2001: 65). For example, the two-word cluster *glass of* shows a semantic preference for the set of words to do with cold drinks (water, milk, lemonade, etc.) Semantic prosody is evaluative, in that it often reveals the speaker's/writer's stance; it is the 'consistent aura of meaning with which a form is imbued by its collocates' (Louw, 1993: 157). Discourse prosody, also evaluative, 'extends over more than one unit in a linear string' (Stubbs, 2001: 65); Stubbs provides the example of the lemma CAUSE, which 'occurs overwhelmingly often with words for unpleasant events' (Stubbs, 2001). The notion of discourse prosody makes it explicit that collocates need not be adjacent to the node for their meaning to influence that of the node.

The analysis of emerging significant lexis and lexical patterns was supplemented throughout with the examination of their concordances. A concordance presents the analyst with instances of a word or cluster in its immediate co-text. The number of words on either side of the word/cluster in focus can be usually set to fit the researcher's needs, and concordance lines can be expanded up to the whole text. Also, concordance lines can be sorted in various ways to help the analyst examine different patterns of the same word/cluster. Concordance analysis affords the examination of language features in co-text, while taking into account the context that the analyst is aware of and can infer from the co-text. It is no wonder, therefore, that it has proven to be the single CL tool that discourse analysts seem to feel comfortable using (see Section 2). In turn, this indicates that CL is no stranger to 'qualitative' analysis (see also Section 6.3). Furthermore, as concordance analysis looks at a known number of concordance lines, the findings can be grouped (e.g., topoi related to a specific word or cluster) and quantified in absolute and relative terms for possible patterns to be identified (e.g., the tendency of words/clusters to be employed in the utilization of particular topoi – see the following section).

A frequent criticism of CL is that it tends to disregard context (Mautner, 2007; Widdowson, 2000). Mautner (2007) argues that 'what large-scale data are not well suited for . . . is making direct, text-by-text links between the linguistic evidence and the contextual framework it is embedded in'. These criticisms seem to stem from restricted conceptions of CL, and would apply more accurately to CL studies that limit themselves to the automatic analysis of corpora, and are of a descriptive rather than an interpretative nature. The examination of expanded concordances (or whole texts when needed) can help the analyst infer contextual elements in order to sufficiently recreate the context (Brown and Yule, 1982). During language communication, addressees do not need to take the full context into account, as according to the principle of local interpretation, addressees need not construct a context more complex than that needed for interpretation (Brown and Yule, 1982). In turn, the co-text provided by the (expanded) concordances helps in 'limiting the interpretation' to what is contextually appropriate or plausible (Brown and Yule, 1982: 59).

Having outlined the approach taken in the CL strand of the project, we now turn to consider the theoretical and methodological stance taken by the CDA component of the research in the discourse historical approach.

5. Theoretical and methodological profile of CDA

CDA provides a general framework for problem-oriented social research.[18] Every 'text' (e.g., an interview, focus group discussion, TV debate, press report, or visual symbol) is conceived as a semiotic entity, embedded in an immediate, text-internal co-text as well as intertextual and sociopolitical context (the 'four-level-model' of context in the DHA; Wodak, 2000, 2001). The DHA thus takes into account the intertextual[19] and interdiscursive[20] relationships

between utterances, texts, genres and discourses, as well as extra-linguistic social/sociological variables, the history and 'archaeology' of an organization, institutional frames of a specific context of situation and processes of text production, text-reception and text consumption.

Van Dijk (2008) emphasizes that 'the "core" of CDA remains the systematic and explicit analysis of the various structures and strategies of different levels of text and talk'. Thus, CDA must draw on specific approaches or concepts of anthropology, history, rhetoric, stylistics, conversation analysis, literary studies, cultural studies, semantics, pragmatics, philosophy and sociolinguistics when approaching or investigating complex social phenomena.

Furthermore, CDA is informed by social theory and views discursive and linguistic data as a social practice, both reflecting and producing ideologies in society (of course, all scientific endeavour is socially committed, as Habermas, 1967, clearly illustrated for the social and natural sciences). In this way, all CDA approaches have to be regarded not only as 'tools', but also as discourse theories (Van Dijk, 2008; Wodak and Chilton, 2007).

CDA researchers are fundamentally interested in analysing opaque as well as transparent structural relationships of dominance, discrimination, power and control, as they are manifested in language. For CDA, language is not powerful on its own – it gains power by the use people make of it and by the people who have access to language means and public fora. In agreement with its critical theory predecessors, CDA emphasizes the need for interdisciplinary work in order to gain a proper understanding of how language functions in constituting and transmitting knowledge, in organizing social institutions or in exercising power in different domains/fields in our societies (Wodak, 2004a).

Very few linguistic forms have not, at some stage, been pressed into the service of the expression of power, for example, by a process of syntactic or textual metaphor. CDA analyses the ways in which such linguistic forms are used in various expressions and manipulations of power and control (Chilton, 2004). Power is signalled not only by grammatical forms within a text, but also by a person's control of a social occasion, by means of the genre of a text, or by access to certain public spheres. It is often exactly within the genres associated with given social occasions that power is exercised or also challenged (see, for example, the investigation of organizational discourses in their hierarchical structures and implied inclusion/exclusion patterns) (Blommaert, 2005; Iedema, 2003; Krzyżanowski and Oberhuber, 2007; Muntigl et al., 2000; Wodak, 1996, 2007a).

Those groups who are in control of most influential public discourses, that is symbolic elites such as politicians, journalists, scholars, teachers and writers, play a special role in the reproduction of dominant knowledge and ideologies in society (Van Dijk, 2005). Because prejudices are not innate, but socially acquired, and because such acquisition is predominantly discursive, the public discourses of the symbolic elites are the primary source of shared ethnic prejudices and ideologies (Van Dijk, 1993).

CDA theories argue that the theorization of context is constitutive for the text analysis (see earlier; Fairclough and Wodak, 1997). In this way, 'context' cannot be reduced to exploring the seemingly 'objective' dimensions of the broader locution of utterances (time, space, speakers, etc.); context has to be perceived and interpreted so that speakers produce utterances they regard as adequate and hearers interpret them due to their *perceptions* of context and their schematic knowledge (Van Dijk, 2005). Hence, Van Dijk claims that we need to assume 'context models' which allow (subjective) understanding of what is said and meant in the interaction.

By contrast, a 'critical' analysis would not only be interested in accounting for what linguistic elements and processes exist in a text or set of texts, but would also need to explain why and under what circumstances and consequences the producers of the text have made specific linguistic choices among several other options that a given language may provide. That is, a critical analysis takes into account *absences as well as presences* in the data (Kress and Van Leeuwen, 2001). This justifies the use of CDA rather than purely descriptive, data-driven approaches which are epistemologically inadequate in accounting for the complex linguistic choices made during the processes of text production.

The CDA component of our project was based on categories of analysis taken from the discourse–historical approach in CDA (DHA). Created by Ruth Wodak and collaborators at the University of Vienna, DHA combines theoretical discourse studies with ethnographic fieldwork and interdisciplinarity. This approach was first developed in order to trace the constitution of an anti-Semitic stereotyped image, or '*Feindbild*' as it emerged in public discourse (particularly press reporting) in the 1986 Austrian presidential campaign of Kurt Waldheim (Wodak, 2004b; Wodak et al., 1990).

Thus, the CDA approach we adopted focused on macro-structural categories (such as the specific genre) and on text-inherent categories developed in the DHA approach of CDA for the analysis of positive self-presentation and negative other-presentation (Reisigl and Wodak, 2001). These dimensions include *inter alia* strategies employed for predication, labelling, argumentation, perspectivation and intensification/mitigation (Table 1). Each of these strategies is manifested textually through a number of linguistic indicators, such as specific lexical items to construct in-groups and out-groups, along with adjectives, attributes, metaphors and the selection of verbs. In addition, argumentative devices which legitimize constructions of RASIM were examined. The recontextualization of specific topoi[21] in the press could be made explicit as well as the various perspectives of reporting (direct/indirect speech, meta-pragmatic verbs, etc).

Two related criticisms of CDA concern the selection of texts to be analysed, and their representativeness (Koller and Mautner, 2004; Stubbs, 1997).

> The hidden danger is that the reason why the texts concerned are singled out for analysis in the first place is that they are not typical, but in fact quite unusual instances which have aroused the analyst's attention. (Koller and Mautner, 2004: 218)

TABLE 1. *Strategies of positive self-presentation and negative other-presentation (adapted from Wodak, 2001)*

Strategies	Objective	Devices	Examples from the news corpus
Referential/Nomination	Construction of in-groups and out-groups	Membership categorization. Biological, naturalizing and depersonalizing metaphors and metonymies. Synecdoches	'... the **pitiful convoy**'. '... an **army** of 110,000 Iraqi refugees'
Predication	Labelling social actors more or less positively or negatively, deprecatorily or appreciatively	Stereotypical, evaluative attribution of negative or positive traits. Implicit and explicit predicates	'Calais is still **crawling** with asylum seekers **trying to break into Britain**.'
Argumentation	Justification of positive or negative attributions	Topoi used to justify political inclusion or exclusion, discrimination or preferential treatment	'... if too many arrive in an uncontrolled manner, the structures of society in an already overcrowded island cannot cope'
Perspectivation, framing or discourse representation	Expressing involvement positioning speakers' point of view	Reporting, description, narration or quotation of events and utterances	'BRITAIN **was warned last night** it faces a massive benefits bill to pay for the looming influx of immigrants . . .'
Intensification, mitigation	Modifying the epistemic status of a proposition	Intensifying or mitigating the illocutionary force or (discriminatory) utterances	'... the politically correct dictators of liberal fashion . . . will **never** concede that **most** asylum-seekers are economic migrants, rather than people fleeing persecution.'

While CDA practitioners are explicit about their stance with regard to the subject of their analysis, some could be accused of selecting texts which they either incorrectly believe to be representative or have been chosen in order to 'prove a point'. Therefore, texts that present a more complex or even contradictory picture might be overlooked.[22]

CDA studies have also been criticized for analysing a small number of texts, or short texts and text fragments (Stubbs, 1994, 1997). Stubbs (1994: 204) argues that 'some patterns of language use are not directly observable, because they are realized across thousands or millions of words of running text, and because they are not categorical but probabilistic'. A small-scale analysis may not be able to identify which linguistic patterns are cumulatively frequent (and therefore likely to represent powerful discourses) and those which are less frequent (and therefore may constitute minority or resistant discourses). In reference to the media, for example, Fairclough (1989: 54) observes,

> The hidden power of media discourse and the capacity of . . . power-holders to exercise this power depend on systematic tendencies in news reporting and other media activities. A single text on its own is quite insignificant: 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.

Clearly, neither CDA nor CL has insight into the psychology of discourse processing, effects, memory, etc., so it is necessary to qualify Fairclough's statement. For example, a single important speech may have a vast impact, while other, more routine ones, repeated daily may hardly get noticed (see Baker, 2006: 19–21 for further discussion).

Therefore, although CL and CDA can both be seen to have strengths and weaknesses, it is hoped that a combination of the two would help to exploit their strong points, while eliminating potential problems. The following section describes how this complementary methodology was carried out.

6. Combining CDA and CL: description and evaluation

In this section we discuss different aspects of the combination of methodologies normally associated with CDA and CL, providing illustrative examples of cases where one of the techniques revealed elements of the representation of RASIM in the corpus that the other could not, as well as examples of cases where one technique supplemented the other.

6.1 CONTEXT-BASED RESEARCH

An initial starting point for the project (in keeping with approaches to CDA) was to investigate aspects of the wider context surrounding the issue of RASIM in the UK.

One aspect of this was to examine how the terms refugee, asylum seeker, immigrant and migrant were conceptualized by 'official' sources, to wit, dictionaries and organizations (e.g., the Refugee Council) who were directly involved with these social groups (see also Krishnamurthy, 1996). A comparison of official

definitions proved to be illuminating: for example, dictionary definitions tended to define an asylum seeker as a refugee who had applied for asylum, implying the temporal sequence refugee → asylum seeker, whereas the Refugee Council defined a refugee as someone whose application for asylum had been successful, implying the opposite sequence asylum seeker → refugee. This fundamental disagreement among official definitions proved to be useful in contextualizing the frequent confusion, conflation and inconsistency which both the CL and CDA researchers independently found in the UK press when such terms were used.

Willing to contextualize the findings of our text-based analyses, we supported them with relevant migration-related statistical information; for example, official figures on the numbers of asylum applications to the UK and the EU as a whole (since the early 1980s) along with net migration during this period. These figures indicated that asylum applications fell sharply after 2002 (which was interesting in light of the fact that the corpus analysis later found that articles about RASIM steadily increased in number over time, even after 2002). However, we also found that the UK had one of the highest rates of asylum applications in the EU, and that net migration to the UK has been increasing since the 1980s, regardless of whether new residents had come as asylum seekers or immigrants. Therefore, analysis of official definitions and government statistics was useful in helping to frame the more linguistic-based research findings (both from the CL and CDA analyses) within a wider context.

We also carried out research on readership figures and demographics for the different newspapers that we included in our corpus. While binary categories like tabloid versus broadsheet or conservative versus liberal sometimes proved to be difficult to maintain, an analysis of previous research on the British press (Conboy, 2006; Richardson, 2004) proved to be useful in giving us ideas about how to categorize different sets of newspapers in order to carry out comparisons. We decided that the broadsheet versus tabloid distinction would be worth investigating. Thus, an initial examination of context proved to be useful in 'setting the scene' for further analysis. Armed with this contextual information we were able to form the research questions outlined in Section 3.1, as well as to decide how to compare different sections of the corpus against each other. We therefore agree with Hardt-Mautner (1995), who argues that researchers ought to carry out background research and form hypotheses in advance of doing corpus-assisted analysis, rather than approaching the corpus from a naive position.[23]

6.2 DATA SELECTION AND DOWNSAMPLING

In this section, we examine how CL techniques provided a 'map' of the corpus, pinpointing areas of interest for a subsequent close analysis (see also Mautner, 2007). Emerging lexical patterns (e.g., key words/clusters, collocates) led to the examination of their (expanded) concordances, or, when needed, the examination of whole texts. This approach is supported by Stubbs (1994: 212), who stresses 'the need to combine the analysis of large-scale patterns across long texts with the detailed study of concordance lines'.

In the RAS project, the corpus was designed in accordance with the project's focus and aims, and is representative in terms of the text source (UK newspapers), text topic (RASIM, and more generally, asylum and immigration issues) and time span (1996–2005) (Gabrielatos, 2007). The texts to be analysed using CDA techniques were selected from a pool of articles published in periods of increased references to RASIM, as indicated by a quantitative analysis. More precisely, the number of corpus articles per month was plotted in order to establish the diachronic development in newspaper coverage of issues pertaining to RASIM. As the plot showed clear frequency 'spikes', corresponding months were examined for local or international events related to RASIM, the wide reporting of which may have caused the significant increase in articles.[24] Texts were selected, through downsampling,[25] from the articles within these periods of increased reporting, with the additional restriction that they reported on the specific events.

The application of downsampling in the project revealed an interesting 'blind spot' for CL, which seems to support arguments for erring on the side of building a larger rather than a smaller corpus. It was initially decided that a sub-corpus containing articles published up to one week before the incidents deemed to have contributed to the spikes, would be constructed and analysed by using both CDA and CL techniques. However, the 'spikes' sub-corpus proved to be too large for the CDA analysis (which necessitated the use of further downsampling). This was not surprising, as CDA in-depth analysis is very labour-intensive. What was less expected was that the sub-corpus would prove to be too small for any significant or helpful patterns to emerge from the collocation and key word analyses.

6.3 CDA AND CL: INTERACTION AND SYNERGY

Partington (2003: 12) presents a scalar view of the uses of CL methodology which points towards a rationale for using CL-related methods to carry out CDA.

> At the simplest level, corpus technology helps find other examples of a phenomenon one has already noted. At the other extreme, it reveals patterns of use previously unthought of. In between, it can reinforce, refute or revise a researcher's intuition and show them why and how much their suspicions were grounded.

As noted earlier, theories of language use underpinning CDA result in a focus on grammatical features (e.g., agentivity, passivization, metaphors). The synergy with the particular approach to CL adopted here adds a focus on lexical patterns. Also, CL processes can help quantify discoursal phenomena already recognized in CDA; that is, establish their absolute and relative frequencies in the corpus, through the examination of the different linguistic means utilized to express them. Even when the CL analysis does not set out to examine existing CDA notions, it can utilize a CDA theoretical framework in the interpretation of the findings. For example, a number of central CDA notions were utilized when grouping collocates and key words on the basis of the semantic preference or semantic/discourse prosody that they communicated. These were the notions of topos and topic, specific metaphors commonly employed in racist discourse, as well as

the referential (or nomination) and predicational strategies (for definitions and discussion see Section 5, and also KhosraviNik, forthcoming). So, as well as providing a framework from argumentation theory as employed in DHA (the concept of topoi), which the CL researcher could use in order to organize the emerging linguistic/discourse patterns found, the CL researcher was able to compare his/her findings against existing immigration-related topoi (provided by a substantial amount of previous CDA research; see Reisigl and Wodak, 2001; Wodak and Van Dijk, 2000).

It can be argued that the CDA notions described earlier enabled the assignation of more explicit and finer semantic/discourse prosody values than merely assigning a general positive/negative bias. At the same time, the corpus size and coverage, coupled with the quantitative aspect of CL, not only provided support for the prominence of central topoi, topics and metaphors already identified in CDA studies, but also indicated their relative frequency.

6.4 COLLOCATION ANALYSIS

Owing to the diachronic nature of the RASIM corpus, and in conjunction with its large size, extracting collocates from the whole corpus can be reasonably expected to include a large number of 'seasonal collocates', that is, collocates that are very frequent in a small number of years. The presence of such collocates was confirmed by the collocational analysis of annual sub-corpora, which revealed that, on average, 92 percent of the collocates of RASIM were only present in no more than five of the ten year-long sub-sections of the corpus.[26] In order to filter out these seasonal collocates, and focus on those collocates that are both salient and central to the representation of RASIM in the corpus newspapers, the notion of consistent collocates (henceforth c-collocates) was introduced.[27] C-collocates were deemed those present in at least seven of the ten annual sub-corpora. C-collocates were then categorized according to the characteristic they applied to RASIM. The categorization was supported by concordance analysis, and was then refined taking into account the CDA notions of topos and topic, as well as metaphors recognized in CDA (Reisigl and Wodak, 2001; Sedlak, 2000; Van Leeuwen, 1996; Wodak and Van Dijk, 2000).

The first important observation is that the vast majority (86%) of content c-collocates could be classified under only eight categories of reference, namely: (a) Provenance/transit/destination, (b) Number, (c) Entry, (d) Economic problems, (e) Residence, (f) Return/repatriation, (g) Legality and (h) Plight. These categories are regularly used in ways which negatively reference RASIM, particularly those concerned with Entry, Economic problems and Legality (c-collocates are shown in bold below):

> BRITAIN was warned last night it faces a massive benefits bill to pay for the looming **influx** of immigrants, including gypsies, from eastern Europe.
>
> (*The Express*, 9 February 2004)

> Calais is still crawling with asylum seekers **trying** to break into Britain.
>
> (*Sunday Times*, 28 July 2002)

The high proportion of references to RAS in the categories of Entry, Residence and Provenance/transit/destination suggests that there may be a preoccupation in the UK press with RASIM entering and staying in the UK (which was subsequently confirmed by concordance analyses). Unexpectedly, the terms immigrant(s) and migrant(s) were found to strongly collocate with *fled* and *fleeing* – these are unlikely collocates as immigration, unlike the seeking of asylum, is a planned process. Also, the concordance analysis indicated that about one in five references to refugees and asylum seekers are accompanied by quantification (the Number category). A common strategy was to quantify RAS in terms of water metaphors (POUR, FLOOD, STREAM), which tend to dehumanize RAS, constructing them as an out-of-control, agentless, unwanted natural disaster.

Interestingly, both the CL and CDA researchers independently found numerous examples of negative categories of references (the corpus linguist examined collocates, while the CDA researcher carried out a close analysis of individual texts). The CL researcher also found a small number of examples of positive categories of reference used in some of the broadsheets. These stressed the advantages of diversity, which, due to the smaller amount of data used in the CDA research, was concluded to be almost non-existent, for example,

> The country needs the talent and vibrancy an immigrant community will bring to a flagging native population base. (*Business*, 17 February 2002)

The small number of categories suggests that the interest of the newspapers in RASIM is focused rather than comprehensive; their nature indicates that the attitude towards RASIM is negative rather than positive. It is equally interesting that the four terms in focus share a good number of c-collocates, as shown in Table 2, which points towards overlap in use.

The significant overlap of c-collocates between refugees–asylum seekers (40.5%), and immigrants–migrants (59%) is perhaps to be expected, as the terms in each pair share a lot of characteristics. However, the overlap between refugees and asylum seekers, on the one hand, and immigrants and migrants, on the other hand, is unexpected. These findings lend further support to the conclusion that the four terms are used as near synonyms in the corpus. Equally telling is the overlap in the categories which emerged from the concordance analysis of

TABLE 2. *Overlap ratio of consistent collocates*

Pairs of terms		% of shared c-collocates
Immigrants	Migrants	59.0
Asylum seekers	Immigrants	43.0
Refugees	Asylum seekers	40.5
Refugees	Immigrants	33.5
Asylum seekers	Migrants	32.0
Refugees	Migrants	28.0

TABLE 3. *Overlap in categories*

	Refugees	Asylum seekers	Immigrants	Migrants
Refugees		ENTRY NUMBER ECON. BURDEN RETURN	ENTRY RESIDENCE	ENTRY
Asylum seekers	ENTRY PLIGHT NUMBER RETURN		ENTRY LEGALITY PDT* RESIDENCE	ENTRY
Immigrants	ENTRY RESIDENCE PLIGHT NUMBER	ENTRY PLIGHT RESIDENCE LEGALITY		ENTRY ECON. THREAT
Migrants	ENTRY RESIDENCE PLIGHT PDT*	PLIGHT	PDT* ENTRY RESIDENCE ECON. THREAT LEGALITY	

*PDT = Provenance/Destination/Transit.

the c-collocates of RASIM. Table 3 gives the categories shared by each pair of terms. The first column indicates the term which is the focus of the comparison in each case.[28]

   A first impression is that the discourses of RASIM in UK newspapers revolve around a small number of topics/categories and employ a limited number of topoi, most of which denote a negative stance. A large number of topoi/topics/ categories is shared by *refugees*, *asylum seekers* and *immigrants*, but less so by *migrants*. This seems to tie in with the findings of the CDA analysis, which identified an overall positive use of the term *migrants*. A second observation is that the overlap in terms of categories goes beyond what could be predicted on the basis of the common definitions of the terms. For example, although the overlap in the category of Entry can be predicted by the definitions (all four groups come into the destination country), the overlap in the category of Plight seems to run contrary to the definitions of immigrants and migrants (migration is usually a planned process). This points to the interpretation that the extensive overlap is not merely the result of overlapping senses or conflicting definitions (explicit or implicit). Rather, the observed overlap, in conjunction with the nature of the categories, seems to be indicative of a wider approach towards issues of asylum and immigration, one which has been identified in a number of CDA studies, namely, that RASIM are less than welcome. A further interesting finding to emerge from the combination of collocation and concordance analysis is that the same c-collocate may index a number of different topics or topoi in the discourse – sometimes more than one at the same time. Table 4 provides examples with the c-collocate *allowed*.

TABLE 4. *Examples of the c-collocate* allowed *indexing topoi/topics related to* RASIM

Entry/ Legality	FRANCE finally closed a loophole yesterday which has *allowed* thousands of illegal immigrants to sneak into Britain on Eurostar trains. (*Daily Star*, 5 February 2002)
Residence	Jack Straw faced embarrassment last night as it emerged that almost one-third of new asylum seekers have been *allowed* to remain in Britain. (*Daily Mail*, 26 November 1998)
Economic burden	But illegal immigration and bogus asylum seekers have been *allowed* to pour into Britain to live off our taxes. (*The Sun*, 21 April 2005)
Economic threat	Job firms in the South have been inundated with pleas for jobs from illegal migrants who are *allowed* to work. (*Sunday Mirror*, 16 April 2000)
Return/ Plight	Only a small number of refugees have been *allowed* to return home, and there has been muted progress in merging Serb-controlled areas with those of the Muslim–Croat federation. (*The Times*, 25 September 1997)

Similarly, c-collocates also index the interchangeable use of RASIM, as the examples with the c-collocate *trying* indicate (our emphasis).

> English, Welsh and Scottish Railways (EWS) warned last night that it cannot, and will not, endure another six months of lost revenue – (pounds) 10m so far – arising from security problems caused by **asylum seekers** in France. EWS services to and from Europe have fallen by 60% since November because so many **refugees** are *trying* to board trains bound for Britain via the Channel tunnel. (*The Herald*, 7 May 2002)

> CHANNEL Tunnel security came under scrutiny last night after 44 **illegal immigrants** were intercepted *trying* to reach Dover. The desperate **asylum seekers** walked seven miles in complete darkness before being caught. (*The Mirror*, 31 August 2001)

> All the passengers were **illegal immigrants** *trying* to make their way to Greece. Survivors identified them as Pakistanis, Moroccans and Bangladeshis. Mr Dokuzoglu said Indian and Afghan **refugees** were also believed to be on board. (*The Guardian*, 2 January 2001)

6.5 QUANTIFYING BIAS: A CORPUS-BASED ANALYSIS OF *POSE AS*

The approach adopted in the RAS project provided evidence that 'qualitative' techniques can be employed, even when the corpus is extremely large,[29] while also retaining the 'quantitative' aspect, which provides further evidence that CL methodology is 'much more than bean counting' (Biber and Conrad, 2001). Let us take the example of the concordance analysis of the multiword unit *POSE as* (which comprises the forms *pose as*, *posed as*, *posing as*, *poses as*) in the tabloid and broadsheet sub-corpora.

Initial analysis showed that tabloids use *POSE as* almost three times more often than broadsheets in general, and eight times as often in reference to RASIM (both differences are statistically significant – see note 30). Comparisons in this section only refer to the uses of *POSE as* in relation to RASIM. The general picture of *POSE as*

is that tabloids employ a negative stance towards RASIM more often than the broadsheets, which, however, by no means implies the traditional stereotype that broadsheets employ a more neutral/positive stance than tabloids (cf. also Van Dijk, 1991). The picture becomes more complex when we examine the statistical significance of the differences:[30] although tabloids almost always (98.1%) adopt a negative stance, broadsheets also do so in a considerable proportion of cases (75.6%), therefore the difference is not statistically significant. In contrast, broadsheets use *POSE as* in positive contexts almost 12 times more frequently than tabloids, and the difference is statistically significant. An example of *POSE as* being used in a non-negative context is given below:

> There might have been an immigration crisis 20 years ago, although that is debatable, but the so-called 'system' of immigration that Britain is stuck with serves to deprive our industry of essential talent (much of which has then migrated elsewhere) while effectively obliging some individuals to **pose as** 'bogus' asylum-seekers in an attempt to enter the country. (*The Times*, 11 September 2005)

In addition, it was observed that the negative stance towards RASIM denoted through the use of *POSE as* could be either accepted or challenged in the articles. Close analysis of expanded concordances (or whole texts, when necessary) revealed a very small number of contexts in which *POSE as* is used when referring to RASIM (and issues of asylum and immigration). Overall, *POSE as* was used in eight frames of the following type: 'Actor(s) POSE as X to achieve Y' (see Tables 5 and 6). When examining the stance of broadsheets and tabloids in these different frames, it became clear that not all differences were statistically significant. This approach clarified that neither tabloids nor broadsheets are consistent in their stance towards RASIM (as shown in their use of *POSE as* in relation to them). More importantly, the analysis established that agentivity (i.e., whether RASIM or others were the agents of *POSE as*) did not correspond one-to-one with positive or negative stance. The statistically significant differences can be interpreted as indicating that broadsheets are more likely than tabloids to challenge negative presentations of RASIM or criticize calls for a stricter immigration and asylum system (Table 5). However, in other frames (Table 6) tabloids and broadsheets show broad similarities in their stance towards RASIM – indicating that their perceived differences in reporting are not clear cut. In broad terms, the differences between tabloids and broadsheets in their reporting on (issues related to) RASIM seem to revolve around tabloids adopting a predominantly negative stance, whereas broadsheets demonstrate a more balanced stance (e.g., combining positive and negative arguments). The latter finding offers a different perspective on intuitive views of broadsheets as being consistently neutral or only positive towards RASIM. At the same time, the findings suggest that UK national newspapers generally termed 'tabloid' form a more homogeneous group than those termed 'broadsheet' (see also Gabrielatos, 2006; Gabrielatos and Baker, 2006a, 2008).

A corpus-assisted approach, which looks for specific linguistic patterns and carries out tests of statistical significance is therefore able to quantify notions like 'bias'. However, it should be noted that corpus-assisted discourse analysis is

TABLE 5. *Statistically significant differences between tabloids and broadsheets in the use of* POSE *as in relation to* RASIM

Actors	Action	Context	Stance	Freq. tabloids (%)	Freq. broadsheets (%)
Aliens, Terrorists, Fanatics, Militants, Criminals, Crooks, Gangs, Spies, Beggars	(BE instructed/helped to) POSE *as RAS* in order to gain illegal entry	Reported as the view of third party (e.g. politician, public) and challenged	Negative stance towards 'tough measures' Neutral/positive stance towards *RASIM*	—	11.5
Reporters	POSE *as* (illegal) *RASIM* to investigate current asylum system	Examination of the asylum and immigration system in the UK	Negative stance towards perceived strictness of system Negative stance towards attempts to show up perceived laxity of system Positive stance towards *RASIM* (and their plight)	—	11.5
RASIM	POSE *as* . . . doctors, nurses, students, taxi drivers, workers, scientists, athletes, sports fans, artists, EU citizens, tourists . . . to gain entry/find work/receive benefits	The problem with *RASIM*	Negative stance towards *RASIM*	23.4	—

TABLE 6. *Statistically non-significant differences between tabloids and broadsheets in the use of* POSE *as in relation to RASIM*

Actors	Action	Context	Stance	Freq. tabloids (%)	Freq. broadsheets (%)
Aliens, etc. Terrorists Fanatics/Militants Criminals/Crooks/ Gangs Spies Beggars	(BE instructed/ helped to) POSE as RAS to gain illegal entry	Reported as fact, or taken for granted	Positive stance towards tougher measures. Indirectly negative towards *RASIM*	51.4	58.9
Police Reporters	POSE as RASIM to expose problems with current asylum system	Examination of the asylum and immigration system in the UK	Positive stance towards practice Negative stance towards perceived laxity/ inadequacy of current system Negative stance towards *RASIM*	22.9	16.7
Individuals	POSE as RASIM to gain illegal entry	Current system forces immigrants to pose as RAS	Neutral/positive stance towards *RASIM*	—	1.3
Soldiers Criminals	POSE as aid workers/ peacekeepers to exploit/harm RASIM	Plight of RAS	Positive stance towards RASIM (and their plight)	1.9	—
Criminals	POSE as aid workers to illegally help RASIM	Strategies used to get RASIM illegally into the country	Negative stance towards RASIM	0.9	—

rarely able to eschew the analysis of context; in the earlier analysis, expanded concordance lines needed to be examined in order to ascertain stance (positive, negative or neutral) towards actors associated with *POSE as*.

We now turn to assess the benefits of the DHA, where context played an even more important role in the analysis.

6.6 CDA TOOLS: STRATEGIC DISCOURSE, PERSPECTIVIZATION AND NARRATIVIZATION

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 normally take into account the social, political, historical and cultural context of the data (see Section 6.1).

To give an example of how a close analysis of context (both in terms of how asylum seeker is used within the context of a full article, and in terms of how the article itself relates to events in the outside world), we wish to focus on an excerpt from *The Guardian*'s report of a speech by the (then) leader of the (opposition) Conservative Party, Michael Howard:

> Mr Howard's third charge was Mr Blair's failure to get a grip on asylum and 'pussyfoot around' on immigration. This completely ignores the coercive controls that Labour has introduced, cutting asylum applications by two-thirds since October 2002, with 1,000 cases a day being denied entry by 2003. (*The Guardian*, 11 April 2005)

Analysis of the wider context of this article revealed that Howard's speech occurred during the Conservative Party's 2005 election campaign, which focused on issues surrounding RASIM. In the article in *The Guardian* article, RASIM were thus referred to as the subject matter of a 'political rivalry discourse', often used by newspapers to criticize (or occasionally support) the current government, particularly at points of political rivalry (e.g., during an election).

The *Guardian* article is critical of Howard's speech. However, despite this, the newspaper discursively represents immigrants and asylum seekers in a way which echoes the representations by the right-wing press and the Conservative Party. So overall, the article represents immigrants and asylum seekers as an 'issue' which is being debated between two political parties. Immigrants and asylum seekers are not portrayed as being a heterogeneous set of people or as doing or saying anything. Instead they are objectified and backgrounded, being referred to in terms of 'applications' alongside quantification (1000 cases a day). Despite the article being in a liberal broadsheet newspaper, the way of representing asylum seekers reproduces an ideology that has been established by conservatives (and which we found was particularly dominant in the right-wing press).

Hence, although the article might be perceived as writing against Conservative anti-immigration rhetoric, it contributes towards the political rivalry discourse and at times even confirms the negative representation of immigrants and asylum seekers by trying to convince its readership that the Labour Party has already been tough in reducing the numbers of immigrants and asylum seekers (references to the 'coercive controls' that Labour has introduced). This is also 'strategic discourse' in the sense that the article is not necessarily focused on the

ethicality or appropriateness of what is being debated; rather it is focused on how the debate between the two main British political parties could be won.[31]

In an excerpt from another article about Howard's campaign (in the right-wing newspaper *The Mail*), we can see how DHA is effective in identifying strategies of perspectivation, positive self-presentation and negative other-presentation:

> MICHAEL HOWARD stood by his views on immigration last night in the face of a hostile TV ambush. The Tory leader repeated his calls for a cap on migration and health checks on immigrants at ports when he appeared on ITV1's *Ask The Leaders* programme. He defended his policy throughout in the face of aggressive questioning from presenter Jonathan Dimbleby. Tory immigration policy was highlighted from the first question in what appeared to be a co-ordinated attack. But, despite the hostile approach, Mr Howard maintained his claim that urgent action was needed to restore public confidence in immigration controls . . . The programme gave every impression of being meticulously planned with a high degree of co-ordination between questioners and presenter. At one point student Dean Delani, 18, shouted at Mr Howard: 'You are inciting xenophobia and hatred in our country. You don't realise what it's like for me.' But the Tory leader replied: 'It doesn't take the debate much further to pin labels on me or abuse me and insult me in the way you have just done.' He added: 'I profoundly disagree with you. What I say to people who hold the view you hold is that if you disagree with these proposals tell us what you would do.' (*The Mail*, 19 April 2005)

Here *The Mail* incorporates a number of discursive techniques to support the in-group social actor (Howard) and negatively perspectivize the out-group (the presenter and members of the audience who did not agree with him). Howard is described as being successful, for example, 'stood by his views', 'defended his policy throughout' and 'maintained his claim' (all positively connotated political jargon), whereas those who disagree with him are described in terms of a 'hostile TV ambush', 'engaging in aggressive questioning', a 'hostile approach' and a 'co-ordinated attack', hence employing negatively connotated war-metaphors and war-jargon. Howard is also directly quoted as claiming that a member of the audience is labelling, abusing and insulting him. These pre-dicational strategies therefore suggest that the out-group are seen to be engaging in negative or unfair actions.

Our analysis also focused on who is written about, how much space they are given and whether they are directly or indirectly quoted. For example, in *The Mail* article above, Howard's words are directly quoted, as compared with the presenter, whose words tended to be summarized (not shown in the excerpt). Howard is also given more space to express his views than members of the constructed out-group, and is reported as being calm (as a serious politician is expected to behave). However, members of the out-group are represented as being threatening, for example, 'At one point student Dean Delani, 18, shouted at Mr Howard: "You are inciting xenophobia and hatred in our country . . ."'. This (and similar cases) are suggestive of a pattern where the newspaper gives more space and direct citation to an in-group member, while citation to

out-group members is provided when they are (or can be represented as being) inarticulate, extremist, illogical, aggressive or threatening. This analysis therefore shows how newspapers position themselves in relationship to the issues they report, in this case, *The Mail* aligns itself with Howard's stance towards immigration.

DHA is therefore helpful in terms of identifying strategies based around language usage (which can be overlooked by the sorts of frequency-based lexical analyses implemented in CL). DHA's strength, however, is not in locating and analysing referential strategies per se. It builds on a network of referential, predicational and argumentative strategies along with analysis of metaphors, presuppositions, mitigation and hyperboles, etc. in deconstructing a text, all of which require a close analysis of context. Additionally, journalistic features, for example, the order of the information, agenda setting and space allocation, in general, and quotation patterns, in particular, play an important role in implementing particular perspectives, and hence, ideologies. In this way, the CDA (DHA) analysis also provides explanatory power to the descriptive results of the CL analysis.

7. Conclusions and recommendations

Overall, each approach can be used to help triangulate the findings of the other, taking into account the coherence, or lack of it, of the findings and the theoretical frameworks informing CDA and CL. Both approaches can be used as entry points, creating a virtuous research cycle (in Table 7, it could be argued that all the 'stages' listed are potential entry or starting points). As shown earlier, CL can provide a general 'pattern map' of the data, mainly in terms of frequencies

TABLE 7 . *Possible stages in corpus-assisted critical discourse analysis*

1	Context-based analysis of topic via history/politics/culture/etymology. Identify existing topoi/discourses/strategies via wider reading, reference to other CDA studies
2	Establish research questions/corpus building procedures
3	Corpus analysis of frequencies, clusters, keywords, dispersion, etc. – identify potential sites of interest in the corpus along with possible discourses/topoi/ strategies, relate to those existing in the literature
4	Qualitative or CDA analysis of a smaller, representative set of data (e.g., concordances of certain lexical items or of a particular text or set of texts within the corpus) – identify discourses/topoi/strategies (DH approach)
5	Formulation of new hypotheses or research questions
6	Further corpus analysis based on new hypotheses, identify further discourses/ topoi/strategies, etc.
7	Analysis of intertextuality or interdiscursivity based on findings from corpus analysis
8	New hypotheses
9	Further corpus analysis, identify additional discourses/topoi/strategies, etc.

(e.g., number of texts per period and/or newspaper, number of words in sub-corpora, type/token ratios, mean sentence length), key words/clusters and collocations, as well as their diachronic development (the latter contributing to the historical perspective in DHA). This helps pinpoint specific periods for text selection (e.g., through downsampling) or sites of interest. Similarly, the CDA analysis (or reference to previous studies) can point towards patterns to be examined through the CL lens for triangulation (e.g., use of nonsensical terms like *bogus asylum seeker*, or particular *topoi*).

CL can also examine frequencies, or, at least, provide strong indicators of the frequency, of specific phenomena recognized in CDA (e.g., topoi, topics, metaphors), by examining lexical patterns, and can add a quantitative dimension to CDA. The approximate quantification usually used in CDA studies (e.g., through the use of frequency adverbs, *usually, normally, frequently*) can be made more specific through (relative) frequency counts and statistical measures. The corpus-based approach also uncovered a small number of articles where 'positive' topoi of RASIM were employed in the corpus. This was different to the CDA analysis, which, focusing on a smaller number of articles, concluded that positive topoi were almost non-existent. The corpus analysis tended to focus around lexical patterns and collocations. This approach is mostly 'lexical' and is most productive when accounting for what DHA calls 'referential' strategies (less so for predicational strategies). The DHA analysis therefore at times facilitated a more detailed analysis, taking into account larger amounts of textual context as well as the structure and characteristics of the employed genres.

Importantly, the project demonstrated the fuzzy boundaries between 'quantitative' and 'qualitative' approaches. More specifically, it showed that 'qualitative' findings can be quantified, and that 'quantitative' findings need to be interpreted in the light of existing theories, and lead to their adaptation, or the formulation of new ones. The non-theory-specific categories emerging from the large-scale data analysis helped inform the adaptation/expansion of existing DHA categories. Indeed, all categories which are quantified are first established in a qualitative, subjective way; they are qualitative categories which are then quantified.

Moreover, the corpus-based analysis tends to focus on what *has* been explicitly written, rather than what *could have been* written but was not, or what is implied, inferred, insinuated or latently hinted at. As shown earlier, DHA allows the analyst to step outside the corpus in order to consult other types of information (such as dictionary definitions, policy documents or government correspondence to newspapers). For example, the corpus analysis may be able to identify which newspapers use a nonsensical term like *bogus asylum seeker*, but a fuller understanding of the term's significance is only available if we consider sources outside the corpus. Such sources would also give examples of other possible ways of expressing the concept, for example, *failed asylum seeker* (which may or may not appear in the corpus, or may not be frequent or significant enough to be included in the analysis). Moreover, pragmatic devices and subtle, coded strategies or concepts can not be readily analysed through corpus linguistic means (Wodak, 2007b).

The RASIM corpus allowed for the comparison of patterns in particular newspapers, and between widely accepted newspaper types (tabloids versus broadsheets). In addition, corpus analysis, combining both quantitative and qualitative techniques, enabled the quantification of the 'quality' of reporting, as far as RASIM are concerned, and, thus, supported (and was supported by) as well as refined the current categorization of UK national newspapers in broadsheets and tabloids (Gabrielatos and Baker, 2006b, 2008). However, the analysis did not differentiate between different subgenres (Mautner, 2007). The comparative capabilities of corpus-based research can also help trace diachronic developments, both in terms of changes (as in the case of nonsensical collocations) and consistency (as in the case of c-collocates).

The combination of methodologies traditionally associated with CDA (DHA) and CL in research projects, and their potential theoretical and methodological cross-pollination, seem to benefit both CDA and CL. Combining methods strengthens the theoretical basis of both DHA and CL (e.g., expressing semantic/discourse prosodies in terms of DHA topoi/topics). CL, in general, and concordance analysis, in particular, can be positively influenced by exposure and familiarity with CDA analytical techniques, and the theoretical notions and categories of DHA can inform the quantitative CL analysis. Also, CL needs to be supplemented by the close analysis of selected texts using CDA theory and methodology. CDA, in turn, can benefit from incorporating more objective, quantitative CL approaches, as quantification can reveal the degree of generality of, or confidence in, the study findings and conclusions, thus guarding against over- or under-interpretation (O'Halloran and Coffin, 2004). As the project indicated that the CL analysis can overlap with that of CDA (DHA), it would be desirable to further examine the extent to which a CL approach is able, on its own, to contribute to critical approaches to discourse analysis. At the same time, it would seem unreasonable for CL to ignore the findings of the considerable body of work in CDA and related fields, or the theoretical notions informing, and deriving from, relevant non-CL research.

NOTES

1. The use of the plural (*methodologies*) indicates that neither CL nor CDA are uniform in terms of their traditional methodological tools and approaches (for CL, see McEnery and Gabrielatos, 2006; for CDA, see Wodak, 2004a; Wodak and Meyer, 2001 for extensive overviews).

2. For example, some CDA approaches (Fairclough, 1989) are informed inter alia by systemic–functional grammar (SFL; Halliday and Matthiessen, 2004), which can be seen to inform the research focus on language aspects such as passivization or agentivity. At the same time, the analysis and interpretation of results are carried out with constant reference to the sociocultural context (Halliday, 1978); other approaches in CDA avoid SFL and focus on argumentation theory and draw primarily on text linguistics (*Discourse–Historical Approach* by Reisigl and Wodak, 2001; *Socio-cognitive Approach* by Van Dijk, 1993, 1998; Jäger, 2001, focuses on metaphors and collective symbols and locates his approach primarily in structural grammar;

and so forth). In turn, the CL methodological approach used in the project is informed by lexical grammar (Sinclair, 2004) in general, and, in particular, by the related notions of collocation (Sinclair, 1991), semantic preference (Stubbs, 2001) and semantic/discourse prosody (Louw, 1993; Stubbs, 2001).

3. For example, the collocational profiles, and the significant intercollocation, of the terms *refugees*, *asylum seekers*, *immigrants* and *migrants*, as well as the clear grouping of collocates in terms of CDA and argumentation categories (i.e., topoi, topics, metaphors), analysed within the context of semantic preference (Stubbs, 2001) and semantic prosody (Louw, 1993), provided strong indications of the discoursal construction of these groups in the UK press (Gabrielatos and Baker, 2006a, 2006b, 2008).

4. For a discussion of the distinction between the corpus-based and corpus-driven approaches see McEnery and Gabrielatos (2006).

5. This principle also informs the quantitative nature of corpus-based studies.

6. Wodak (1986) and Wodak and Schulz (1986) used large corpora and analysed these both in quantitative and qualitative ways, drawing on more sociolinguistic approaches and combining these with CDA (see Titscher et al., 2000; Wodak, 1996 for overviews). Moreover, the 'French School' in CDA has always relied on corpus-driven methodologies (see Fairclough and Wodak, 1997; Wodak and De Cillia, 2006, for overviews). However, all these approaches never reflected the full capacity of CL.

7. For example, Aston (1997) gives a range of 20,000–200,000 words, whereas Gavioli (2005) sets the range at 50,000–100,000, while also citing corpora as small as 6854 words.

8. Magalhaes' study aimed to examine the use of the words *negra(s), negro(s), pretos* (black/blacks) and *pardos* (brown/browns) in a corpus drawn from a newspaper, in order to investigate discourses of racism surrounding these terms. However, the corpus articles were derived by way of a query comprising the terms race, racism and racist. This query would be expected to return texts in which racism is overtly/explicitly mentioned or discussed, and, consequently, the terms in focus would, by necessity, be examined within the context of racism – something that would almost certainly bias the results.

9. For an extensive examination of the use of CL techniques in discourse analysis see Baker (2006).

10. The CDA team consisted of Ruth Wodak, Michal Krzyżanowski and Majid KhosraviNik. The CL team consisted of Costas Gabrielatos, Paul Baker and Tony McEnery.

11. For a more detailed account of the CL part of the research, see Gabrielatos and Baker (2008).

12. For a discussion of the tension between 'quantitative' and 'qualitative' approaches to discourse studies see Baker (2004a).

13. For details on the corpus compilation and make-up, see Gabrielatos (2007) and Gabrielatos and Baker (2008).

14. The software used to carry out the corpus analysis was WordSmith Tools 3.0 and 4.0 (Scott, 1999, 2007).

15. The decision to examine these four terms (and not others such as *aliens*) was also made subjectively, based on our research questions.

16. A cluster (also termed an *n-gram* or *lexical bundle*) is a sequence of two or more words, not necessarily a grammatical or meaningful unit. A key word analysis can also be applied to clusters.

17. 'A word family consists of a base word and all its derived and inflected forms. . . . [T]he meaning of the base in the derived word must be closely related to the meaning of the base when it stands alone or occurs in other derived forms, for example, hard and hardly would not be members of the same word family' (Bauer and Nation, 1993: 253).

18. The history of the field is summarized in detail in Renkema (2004) and Van Dijk (2008).

19. Intertextuality refers to the fact that all texts are linked to other texts, via synchronic and diachronic dimensions. Such links can be established in different ways: through continued reference to a topic or main actors; through reference to similar events; or by the transfer of main arguments from one text into the next (i.e., recontextualization). (See Wodak, 2008 for extensive definitions.)

20. Interdiscursivity indicates that discourses are linked to each other in various ways. If we define discourse as primarily topic-related, i.e., a discourse on X, then a discourse on un/employment often refers, for example, to topics or subtopics of other discourses, such as gender or racism: arguments on systematically lower salaries for women or migrants might be included in discourses on employment.

21. *Topoi* are 'conclusion rules that connect the argument with the conclusion' (Reisigl and Wodak, 2001), or, simply put, they represent 'the common-sense reasoning typical for specific issues' (Van Dijk, 2000), whereas *topics* simply refer to the subject matter of the discussion (Sedlak, 2000). 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 utilizing this metaphor (e.g., immigrants are flooding the country) can very well employ a topos of Number (Gabrielatos and Baker, 2008). Reisigl and Wodak (2001) list the whole range of metaphors typically used in debates on immigration.

22. In many CDA studies which have investigated large data samples, such restricted and biased data selection did not occur (see, e.g., Blackledge, 2005; Jäger, 2001; Kovács and Wodak, 2003; Krzyżanowski and Oberhuber, 2007; Richardson, 2004; Wodak, 1986; Wodak and Van Dijk, 2000; Wodak et al., 1990, 1999; Wodak and Schulz, 1986). Criticism directed towards CDA in general often focuses exclusively on Anglo-American research (Wodak, 2006). Hence, some criticisms are biased and even false if generalized to the whole paradigm of CDA.

23. The sequence, however, can be flexible. The results of the first analysis can then be interpreted in the light of subsequent research in the relevant sociopolitical context, which, in turn, will lead to more directed and fine-tuned analysis.

24. These events were: the war in Kosovo (March–May 1999), the 9/11 terrorist attacks on the USA (September–October 2001), the war in Afghanistan (April–May 2002), the Iraq disarmament crisis (December 2002–Feburary 2003), the UK asylum bill (March–April 2004), and the UK general elections (March–May 2005).

25. This form of downsampling was suggested in an EU research project on 'voices of migrants' (Krzyżanowski and Wodak, 2007; Project XENOPHOB, 2003–05).

26. For more details on the collocation analysis, see Baker et al. (2007, 2008) and Gabrielatos and Baker (2006b, 2008).

27. The notion of *consistency* has also been utilized by Scott (1999), who uses it in relation to word lists and key words. A consistency analysis shows the number of texts or sub-corpora that a key word is found in.

28. The extent of the overlap of c-collocates for each pair is directional; that is, it depends on the number of c-collocates in each category that each term has registered, and the ratio of overlap in each case. Please note that only categories with at least one-third overlap are included in Table 3.

29. Compared with the usual size of specialized corpora (the RASIM corpus is 140 million words).

30. Calculations were based on the total instances of *POSE as* when referring to RASIM, and were carried out manually using Paul Rayson's online log-likelihood calculator (http://ucrel.lancs.ac.uk/llwizard.html). The probability that differences were due to chance was extremely low: no higher than one in a quadrillion (1,000,000,000,000,000).

31. To be fair to the CL analysis, it did uncover clues to the existence of such discourses, as suggested by the analysis demonstrated in the previous section.

REFERENCES

Aston, G. (1997) 'Small and Large Corpora in Language Learning', in B. Lewandowska-Tomaszczyk and P. Melia (eds) *Practical Applications in Language Corpora*, pp. 51–62. Lodz, Poland: University Press.

Baker, P. (2004a) *Public Discourses of Gay Men*. London: Routledge.

Baker, P. (2004b) 'Querying Keywords: Questions of Difference, Frequency, and Sense in Keywords Analysis', *Journal of English Linguistics* 32(4): 346–59.

Baker, P. (2006) *Using Corpora in Discourse Analysis*. London: Continuum.

Baker, P. and McEnery, T. (2005) 'A Corpus-based Approach to Discourses of Refugees and Asylum Seekers in UN and Newspaper Texts', *Journal of Language and Politics* 4(2): 197–226.

Baker, P., Gabrielatos, C. and McEnery, T. (2008) 'Using Collocational Profiling to Investigate the Construction of Refugees, Asylum Seekers and Immigrants in the UK Press', paper given at the 7th conference of the American Association for Corpus Linguistics, Brigham Young University, Provo, UT, 13–15 March 2008.

Baker, P., McEnery, T. and Gabrielatos, C. (2007) 'Using Collocation Analysis to Reveal the Construction of Minority Groups: The Case of Refugees, Asylum Seekers and Immigrants in the UK Press', paper given at Corpus Linguistics 2007, University of Birmingham, 28–30 July 2007. Available at: http://eprints.lancs.ac.uk/602/.

Bauer, L. and Nation, P. (1993) 'Word Families', *International Journal of Lexicography* 6(4): 253–79.

Biber, D. and Conrad, S. (2001) 'Quantitative Corpus-based Research: Much More than Bean Counting', *TESOL Quarterly* 35(2): 331–6.

Blackledge, A. (2005) *Discourse and Power in a Multilingual World*. Amsterdam: Benjamins.

Blommaert, J. (2005) *Discourse. A Critical Introduction*. Cambridge: Cambridge University Press.

Bondi, M. (2007) 'Key-words and Emotions: A Case Study of the Bloody Sunday Enquiry', in N. Fairclough, G. Cortese and P. Ardizzone (eds) *Discourse and Contemporary Social Change*, pp. 407–32. Bern: Peter Lang.

Brown, G. and Yule, G. (1982) *Discourse Analysis*. Cambridge: Cambridge University Press.

Chilton, P. (2004) *Analysing Political Discourse*. London: Routledge.

Clark, C. (2007) 'A War of Words: A Linguistic Analysis of BBC Embedded Reports During the Iraq Conflict', in N. Fairclough, G. Cortese and P. Ardizzone (eds) *Discourse and Contemporary Social Change*, pp. 119–40. Bern: Peter Lang.

Conboy, M. (2006) *Tabloid Britain: Constructing a Community Through Language*. London: Routledge.

Fairclough, N. (1989) *Language and Power*. London: Longman.

Fairclough, N. (2000) *New Labour, New Language?* London: Routledge.

Fairclough, N. and Wodak, R. (1997) 'Critical Discourse Analysis', in T.A. van Dijk (ed.) *Discourse as Social Interaction*, pp. 258–84. London: SAGE.

Fairclough, N., Cortese, G. and Ardizzone, P. (eds) (2007) *Discourse and Contemporary Social Change*. Bern: Peter Lang.

Fairclough, N. and Wodak, R. (2008) 'Recontextualizing the Bologna Strategy in Higher Education', in N. Fairclough, R. Jessop and R. Wodak (eds) *CDA and the Knowledge-based Economy in Europe*. London: Falmer Press.

Flowerdew, J. (1997) 'The Discourse of Colonial Withdrawal: A Case Study in the Creation of Mythic Discourse', *Discourse & Society* 8(4): 453–77.

Gabrielatos, C. (2006) 'Towards Quantifying "Quality" in the Press: Comparing the Stance of UK Broadsheets and Tabloids Towards Refugees, Asylum Seekers and Immigrants', 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 2006. Available: http://eprints.lancs.ac.uk/231.

Gabrielatos, C. (2007) 'Selecting Query Terms to Build a Specialised Corpus from a Restricted-access Database', *ICAME Journal* 31: 5–43.

Gabrielatos, C. and Baker, P. (2006a) 'Representation of Refugees and Asylum Seekers in UK Newspapers: Towards a Corpus-based Comparison of the Stance of Tabloids and Broadsheets', paper presented at the first international conference on Critical Approaches to Discourse Analysis Across Disciplines, University of East Anglia, UK, 29–30 June 2006. Available: http://eprints.lancs.ac.uk/250/.

Gabrielatos, C. and Baker, P. (2006b) 'Representation of Refugees and Asylum Seekers in UK Newspapers: Towards a Corpus-based Analysis', paper given at the joint annual meeting of the British Association for Applied Linguistics and the Irish Association for Applied Linguistics: From Applied Linguistics to Linguistics Applied: Issues, Practices, Trends, University College, Ireland, 7–9 September 2006. Available: http://eprints.lancs.ac.uk/265/.

Gabrielatos, C. and Baker, P. (2008) 'Fleeing, Sneaking, Flooding: A Corpus Analysis of Discursive Constructions of Refugees and Asylum Seekers in the UK Press 1996–2005', *Journal of English Linguistics* 36(1): 5–38.

Gavioli, L. (2005) *Exploring Corpora for ESP Learning*. Amsterdam: Benjamins.

Habermas, J. (1967) *Theorie und Praxis*. Neuwied am Rhein: Luchterland.

Halliday, M.A.K. (1978) *Language as Social Semiotic: The Social Interpretation of Language and Meaning*. London: Arnold.

Halliday, M.A.K. and Matthiessen, M.I.M. (2004) *An Introduction to Functional Grammar*, 3rd edn. Oxford: Oxford University Press.

Hardt-Mautner, G. (1995) *Only Connect. Critical Discourse Analysis and Corpus Linguistics*. UCREL Technical Paper 6. Lancaster, UK: Lancaster University. Available at: http://ucrel.lancs.ac.uk/papers/techpaper/vol6.pdf

Hunston, S. (2002) *Corpora in Applied Linguistics*. Cambridge: Cambridge University Press.

Iedema, R. (2003) *The Discourses of Post-bureaucratic Organization*. Amsterdam: Benjamins.

Jäger, S. (2001) *Kritische Diskursanalyse*. Duisberg: DISS.

KhosraviNik, M. (forthcoming) 'The Representation of Refugees, Asylum Seekers and Immigrants in British Newspapers: A Critical Discourse Analysis', *Journal of Language and Politics*.

Koller, V. and Mautner, G. (2004) 'Computer Applications in Critical Discourse Analysis', in C. Coffin, A. Hewings and K. O'Halloran (eds) *Applying English Grammar: Corpus and Functional Approaches*, pp. 216–28. London: Arnold.

Kovács, A. and Wodak, R. (eds) (2003) *NATO, Neutrality and National Identity*. Vienna: Böhlau.

Kress, G. and Van Leeuwen, T. (2001) *Multimodal Discourse: The Modes and Media of Contemporary Communication*. London: Arnold.

Krishnamurthy, R. (1996) 'Ethnic, Racial and Tribal: The Language of Racism?', in C.R. Caldas-Coulthard and M. Coulthard (eds) *Texts and Practices: Readings in Critical Discourse Analysis*, pp. 129–49. London: Routledge.

Krzyżanowski, M. and Oberhuber, F. (2007) *(Un)Doing Europe. Discourses and Practices of Negotiating the EU Constitution*. Brussels: P.I.E.-Peter Lang.

Krzyżanowski, M. and Wodak, R. (2007) 'Multiple Identities, Migration and Belonging: "Voices of Migrants"', in C.R. Caldas-Coulthard and R. Iedema (eds) *Critical Discourse and Contested Identities*, pp. 95–119. Basingstoke, UK: Palgrave Macmillan.

Leech, G. (1992) 'Corpora and Theories of Linguistic Performance', in J. Svartvik (ed.) *Directions in Corpus Linguistics: Proceedings of the Nobel Symposium 82, Stockholm, 4–8 August 1991*, pp. 105–22. Berlin: Mouton de Gruyter.

Louw, B. (1993) 'Irony in the Text or Insincerity in the Writer? The Diagnostic Potential of Semantic Prosodies', in M. Baker, G. Francis and E. Tognini-Bonelli (eds) *Text and Technology: In Honour of John Sinclair*, pp. 157–76. Philadelphia/Amsterdam: John Benjamins.

McEnery, T. (2005) *Swearing in English: Bad Language, Purity and Power from 1586 to the Present*. London: Routledge.

McEnery, T. and Gabrielatos, C. (2006) 'English Corpus Linguistics', in B. Aarts and A. McMahon (eds) *The Handbook of English Linguistics*, pp. 33–71. Oxford: Blackwell.

Magalhaes, C.M. (2006) 'A Critical Discourse Analysis Approach to News Discourses and Social Practices on Race in Brazil', *DELTA* 22(2): 275–301.

Mautner, G. (2000) *Der britische Europa-Diskurs: Methodenreflexion und Fallstudien zur Berichterstattung in der Tagespresse*. Vienna: Passagen-Verlag.

Mautner, G. (2007) 'Mining Large Corpora for Social Information: The Case of *Elderly*', *Language in Society* 36(1): 51–72.

Meinhof, U.H. and Richardson, K. (eds) (1994) *Text, Discourse and Context: Representations of Poverty in Britain*. London: Longman.

Muntigl, P., Weiss, G. and Wodak, R. (2000) *European Union Discourses on Unemployment*. Amsterdam: John Benjamins.

Nattinger, J.R. and DeCarrico, J.S. (1992) *Lexical Phrases and Language Teaching*. Oxford: Oxford University Press.

O'Halloran, K. and Coffin, C. (2004) 'Checking Overinterpretation and Underinterpretation: Help from Corpora in Critical Linguistics', in C. Coffin, A. Hewings and K. O'Halloran (eds) *Applying English Grammar: Corpus and Functional Approaches*, pp. 275–97. London: Arnold.

Ooi, V. (2001) 'Investigating and Teaching Genres Using the World Wide Web', in M. Ghadessy, A. Henry and R.L. Roseberry (eds) *Small Corpus Studies and ELT*, pp. 175–203. Amsterdam: Benjamins.

Orpin, D. (2005) 'Corpus Linguistics and Critical Discourse Analysis: Examining the Ideology of Sleaze', *International Journal of Corpus Linguistics* 10(1): 37–61.

Partington, A. (2003) *The Linguistics of Political Argumentation: The Spin-doctor and the Wolf-pack at the White House*. London: Routledge.

Partington, A. (2004) 'Corpora and Discourse, a Most Congruous Beast', in A. Partington, J. Morley and L. Haarman (eds) *Corpora and Discourse*, pp. 11–20. Bern: Peter Lang.

Partington, A. (2006) 'Metaphors, Motifs and Similes Across Discourse Types: Corpus-assisted Discourse Studies (CADS) at Work', in A. Stefanowitsch and S. Gries (eds) *Corpus-based Approaches to Metaphor and Metonymy*, pp. 267–304. Berlin: Mouton de Gruyter.

Piper, A. (2000) 'Some Have Credit Cards and Others Have Giro Cheques: "Individuals" and "People" as Lifelong Learners in Late Modernity', *Discourse & Society* 11(3): 515–42.

Reisigl, M. and Wodak, R. (2001) *Discourse and Discrimination: Rhetorics of Racism and Anti-Semitism*. London: Routledge.

Renkema, J. (2004) *Introduction to Discourse Studies.* Amsterdam: Benjamins.

Richardson, J. (2004) *(Mis)representing Islam: The Racism and Rhetoric of British Broadsheet Newspapers*. Amsterdam: Benjamins.

Scott, M. (1999) *WordSmith Tools Help Manual, Version 3.0*. Oxford: Mike Scott/Oxford University Press.

Scott, M. (2007) *Oxford WordSmith Tools Version 4.0*. Available: http://www.lexically.net/downloads/version4/wordsmith.pdf.

Sedlak, M. (2000) 'You Really Do Make an Unrespectable Foreign Policy', in R. Wodak and T.A. van Dijk (eds) *Racism at the Top: Parliamentary Discourses on Ethnic Issues in Six European States*, pp. 107–68. Klagenfurt, Austria: Drava-Verlag.

Semino, E. and Short, M.H. (2004) *Corpus Stylistics*. London: Routledge.

Sinclair, J. (1991) *Corpus, Concordance, Collocation*. Oxford: Oxford University Press.

Sinclair, J. (2001) 'Preface', in M. Ghadessy, A. Henry and R.L. Roseberry (eds) *Small Corpus Studies and ELT*, pp. vii–xv. Amsterdam: Benjamins.

Sinclair, J. (2004) *Trust the Text: Language, Corpus and Discourse*. London: Routledge.

Sotillo, S.M. and Wang-Gempp, J. (2004) 'Using Corpus Linguistics to Investigate Class, Ideology, and Discursive Practices in Online Political Discussions: Pedagogical Applications of Corpora', in U. Connor and T.A. Upton (eds) *Applied Corpus Linguistics*, pp. 91–122. Amsterdam: Rodopi.

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

Stubbs, M. (1997) 'Whorf's Children: Critical Comments on Critical Discourse Analysis', in A. Wray and A. Ryan (eds) *Evolving Models of Language*, pp. 100–16. Clevedon: Multilingual Matters.

Stubbs, M. (2001) *Words and Phrases: Corpus Studies of Lexical Semantics*. Oxford: Blackwell.

Titscher, S., Meyer, M., Wodak, R. and Vetter, E. (2000) *Methods of Text and Discourse Analysis*. London: SAGE.

Tognini-Bonelli, E. (2001) *Corpus Linguistics at Work*. Amsterdam: Benjamins.

Van der Valk, I. (2000) 'Parliamentary Discourse on Immigration and Nationality in France', in R. Wodak and T.A. van Dijk (eds) *Racism at the Top: Parliamentary Discourses on Ethnic Issues in Six European States*, pp. 221–60. Klagenfurt, Austria: Drava-Verlag.

Van Dijk, T.A. (1991) *Racism and the Press*. London: Routledge.

Van Dijk, T.A. (1993) *Elite Discourse and Racism*. Thousand Oaks, CA: SAGE.

Van Dijk, T.A. (1998) *Ideology*. London: SAGE.

Van Dijk, T.A. (2000) 'Parliamentary Debates', in R. Wodak and T.A. van Dijk (eds) *Racism at the Top: Parliamentary Discourses on Ethnic Issues*, pp. 45–79. Klagenfurt: Drava.

Van Dijk, T.A. (2005) *Discourse and Racism in Spain and Latin America*. Amsterdam: Benjamins.

Van Dijk, T.A. (2008) *Discourse Reader*. London: SAGE.

Van Leeuwen, T. (1996) 'The Representation of Social Actors', in C.R. Caldas-Coulthard and M. Coulthard (eds) *Texts and Practices*: *Readings in Critical Discourse Analysis*, pp. 32–70. London: Routledge.

Weiss, G. and Wodak, R. (eds) (2003) *Critical Discourse Analysis: Theory and Interdisciplinary*. Basingstoke, UK: Palgrave.

Widdowson, H.G. (2000) 'On the Limitations of Linguistics Applied', *Applied Linguistics* 21(1): 3–25.

Wodak, R. (1986) *Language Behavior in Therapy Groups*. Los Angeles, CA: University of California Press.

Wodak, R. (1996) *Disorders of Discourse*. London: Longman.

Wodak, R. (2000) 'Recontextualization and the Transformation of Meanings: A Critical Discourse Analysis of Decision Making in EU Meetings About Employment Policies', in S. Sarangi and M. Coulthard (eds) *Discourse and Social Life*, pp. 185–206. Harlow, UK: Pearson Education.

Wodak, R. (2001) 'The Discourse Historical Approach', in R. Wodak and M. Meyer (eds) *Methods of Critical Discourse Analysis*, pp. 63–94. London: SAGE.

Wodak, R. (2004a) 'Critical Discourse Analysis', in C. Seale, G. Gobo, J.F. Gubrium and D. Silverman (eds) *Qualitative Research Practice*, pp. 197–213. London: SAGE.

Wodak, R. (2004b) 'Discourses of Silence', in L. Thiesmeyer (ed.) *Discourse and Silencing*, pp. 179–209. Amsterdam: Benjamins.

Wodak, R. (2006) 'Review Article: Boundaries in Discourse Analysis', *Journal of Language in Society* 35(4): 595–611.

Wodak, R. (2007a) 'What Now?' – Some Reflections on the European Convention and its Implications', in M. Krzyżanowski and F. Oberhuber (eds) (*Un)doing Europe. Discourses and Practices of Negotiating the EU Constitution*, pp. 203–17. Brussels: P.I.E.-Peter Lang.

Wodak, R. (2007b) 'Pragmatics and Critical Discourse Analysis: A Cross-disciplinary Inquiry', *Journal of Pragmatics and Cognition* 15(1): 203–27.

Wodak, R. (2008) 'Introduction', in R. Wodak and M. Kryżanowski (eds) *Qualitative Discourse Analysis in the Social Sciences*. Basingstoke, UK: Palgrave Macmillan.

Wodak, R. and Chilton, P. (eds) (2007) *A New Agenda in Critical Discourse Analysis*, 2nd edn. Amsterdam: Benjamins.

Wodak, R. and De Cillia, R. (2006) *Ist Österreich ein deutsches Land? Anmerkungen zur Sprachenpolitik der Zweiten Republik.* Innsbruck, Austria: Studienverlag.

Wodak, R. and Krzyżanowski, M. (eds) (2008) *Qualitative Discourse Analysis in the Social Sciences*. Basingstoke, UK: Palgrave Macmillan.

Wodak, R. and Meyer, M. (eds) (2001) *Methods of Critical Discourse Analysis*. London: SAGE.

Wodak, R. and Schulz, H. (1986) *The Language of Love and Guilt*. Amsterdam: Benjamins.

Wodak, R. and Van Dijk, T.A. (eds) (2000) *Racism at the Top. Parliamentary Discourses on Ethnic Issues in Six European States*. Klagenfurt, Austria: Drava-Verlag.

Wodak, R., De Cillia, R., Gruber, H., Mitten, R., Nowak, P. and Pelikan, J. (1990) *Wir sind alle unschuldige Täter! Diskurshistorische Studien zum Nachkriegsantisemitismus.* Frankfurt-am-Main: Suhrkamp.

Wodak, R., De Cillia, R., Reisigl, M. and Liebhart, K. (1999) *The Discursive Construction of National Identity*. Edinburgh: Edinburgh University Press.

PAUL BAKER is a Senior Lecturer in English Language and Linguistics at Lancaster University. His recent publications include *Polari, The Lost Language of Gay Men* (2002), *Using Corpora in Discourse Analysis* (2006), *A Glossary of Corpus Linguistics* (2006 with Tony McEnery and Andrew Hardie) and *Public Discourses of Gay Men* (2005). He has also published papers in *Sociolinguistics*, *Literary and Linguistic Computing* and *Journal of English Linguistics*. He is the commissioning editor of *Corpora* journal. See http://www.ling.lancs. ac.uk/profiles/48/. ADDRESS: Department of Linguistics and English Language, Lancaster University, Lancaster LA1 4YT, UK. [email: p.baker@lancaster.ac.uk]

COSTAS GABRIELATOS is a Research Associate at Lancaster University, where he is also working towards a PhD in linguistics, doing corpus research on English *if*-conditionals. His main areas of interest are the expression of modality and time in English, use of corpus-based approaches in (critical) discourse analysis, and the implications of corpus-based research insights for language teaching in general, and pedagogical grammar in particular. For details see http://www.ling.lancs.ac.uk/profiles/305/.

MAJID KHOSRAVINIK is a Research Associate and Teaching Assistant in Linguistics and English Language at Lancaster University. He is working towards a PhD on critical discourse analysis and discursive strategies of 'self' and 'other' representation in news discourse. Majid has been involved in discourse analytical studies in the context of Iran including a CDA analysis of political ideologies in Iranian newspapers. His research interests include: discourse and racism/discrimination, discourse and politics, language and identity, language and gender, and language policy. See www.ling.lancs.ac.uk/profiles/856.

MICHAŁ KRZYŻANOWSKI is research fellow at the Department of Linguistics and English Language, Lancaster University, UK, and Assistant Professor at the School of English, Adam Mickiewicz University, Poznań, Poland. His main research interests are in critical discourse analysis and the development of the CDA's 'discourse–historical' approach to the synchronic and diachronic study of social, political and institutional change. He has published extensively on discourses in EU Institutions and on discourses of politics and the media in Europe. For details about current projects and publications see: http://www.ling.lancs.ac.uk/profiles/309.

TONY MCENERY is Professor of English Language and Linguistics at Lancaster University. He has researched extensively on the use of corpora in applied and theoretical linguistics and is the author of *Corpus Linguistics* (with A. Wilson, Edinburgh University Press, 1996) and *Swearing in English* (Routledge, 2005). He has held numerous grants exploring the application of corpus linguistics in applied linguistics, including grants from the UK research councils and the European Union. He is currently exploring the use of corpus data in the diagnosis and explanation of moral panics. See also http://www.ling.lancs.ac.uk/profiles/295/.

RUTH WODAK is Distinguished Professor of Discourse Studies at Lancaster University. Besides various other prizes, she was awarded the Wittgenstein Prize for Elite Researchers in 1996 and is also head of the Wittgenstein Research Centre 'Discourse, Politics, Identity' at the University of Vienna. Her research interests focus on discourse analysis, gender studies, language and/in politics, identity politics, prejudice and discrimination, and on ethnographic methods of linguistic field work. She is a member of the editorial board of

a range of linguistic journals and co-editor of the journals *Discourse & Society*, *Critical Discourse Studies* and *Language and Politics*. She has held visiting professorships in Uppsala, Stanford University, University of Minnesota, University of East Anglia and Georgetown University. See http://www.ling.lancs.ac.uk/profiles/265/ for current research projects.