

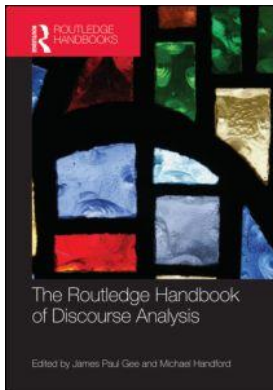
This article was downloaded by: 10.3.97.143

On: 30 Sep 2023

Access details: *subscription number*

Publisher: *Routledge*

Informa Ltd Registered in England and Wales Registered Number: 1072954 Registered office: 5 Howick Place, London SW1P 1WG, UK



## The Routledge Handbook of Discourse Analysis

James Paul Gee, Michael Handford

### Corpus-based discourse analysis

Publication details

<https://www.routledgehandbooks.com/doi/10.4324/9780203809068.ch13>

Lynne Flowerdew

**Published online on: 16 Nov 2011**

**How to cite :-** Lynne Flowerdew. 16 Nov 2011, *Corpus-based discourse analysis from: The Routledge Handbook of Discourse Analysis* Routledge

Accessed on: 30 Sep 2023

<https://www.routledgehandbooks.com/doi/10.4324/9780203809068.ch13>

**PLEASE SCROLL DOWN FOR DOCUMENT**

Full terms and conditions of use: <https://www.routledgehandbooks.com/legal-notices/terms>

This Document PDF may be used for research, teaching and private study purposes. Any substantial or systematic reproductions, re-distribution, re-selling, loan or sub-licensing, systematic supply or distribution in any form to anyone is expressly forbidden.

The publisher does not give any warranty express or implied or make any representation that the contents will be complete or accurate or up to date. The publisher shall not be liable for an loss, actions, claims, proceedings, demand or costs or damages whatsoever or howsoever caused arising directly or indirectly in connection with or arising out of the use of this material.

# Corpus-based discourse analysis

Lynne Flowerdew

---

## Introduction

Discourse analysis covers a vast range of areas and is also one of the least clearly defined fields in applied linguistics (Stubbs, 1983; Aijmer and Stenström, 2004). Blommaert (2005: 2) notes that, traditionally, discourse has been treated in linguistic terms as ‘language-in-use’, informing areas such as pragmatics and speech act theory. However, for Blommaert discourse has a wider interpretation as ‘language-in-action’, i.e. ‘meaningful symbolic behaviour’. Jucker *et al.* (2009b: 5) define this wider use of the term discourse as ‘the totality of linguistic practices that pertain to a particular domain or that create a particular object’. A useful distinction is made by Gee (2001), who defines the ‘language-in-use’ aspect as ‘discourse’ (with a little ‘d’) and the more ‘language-in-action’ orientation as ‘Discourse’ (with a capital D), involving not only linguistic practices but other semiotic elements. Discourses are created through recognition work of ‘ways with words, actions, beliefs, emotions, values, interactions, people, objects, tools and technologies’ that constitute a way of being a member of a particular discourse community (*ibid.*, p. 20).

Corpus linguistics is a field of enquiry whose essential nature, like that of discourse analysis, has also come under scrutiny. The main contention revolves around ‘corpus-driven’ vs. ‘corpus-based’ linguistics and whether corpus linguistics is a theory or a methodology. The field of corpus linguistics in the ‘corpus-driven’ sense is underpinned by a phraseological, syntagmatic approach to language data (see Sinclair, 2004), consisting of five categories of co-selection with the core lexical item and semantic prosody as obligatory elements, and collocation, colligation and semantic preference as optional categories. Proponents in the ‘corpus-driven’ camp regard corpus linguistics as essentially a theory with corpus enquiries revealing hitherto unknown aspects of language, thus challenging the ‘underlying assumptions behind many well established theoretical positions’ (Tognini Bonelli, 2001: 48). For this reason they oppose any a priori mark-up of the data, arguing that it would obscure the syntagmatic, lexico-grammatical patternings associated with the phraseological approach. However, most corpus linguists take a less extreme view, tending towards a more ‘corpus-based’ approach. For example, Aarts (2002) views corpus linguistics as a methodology for validating existing descriptions of language on which to make changes in the description where corpus data does not fit. While McEnery *et al.* (2006) conceive of corpus linguistics as a new philosophical approach to linguistic enquiry; they do not consider it to have the status of a theory (see also Biber *et al.*, 1998; Conrad, 2002). In spite of these different theoretical positions, corpus linguistics is generally regarded as a methodology, and ‘corpus-based’ is used as an umbrella term for a range of corpus enquiries, which is the sense adopted in this chapter.

Although discourse analysis and corpus linguistics both make use of naturally occurring attested data, they have intrinsically ontological and epistemological differences, as noted by Virtanen

(2009). Doing corpus analysis is not the same as doing discourse analysis (DA). Leech (2000: 678–680, cited in McEnery *et al.*, 2006) observes that there is a ‘cultural divide’ between the two: ‘while DA emphasizes the integrity of the text, corpus linguistics tends to use representative samples; while DA is primarily qualitative, corpus linguistics is essentially quantitative; while DA focuses on the contents expressed by language, corpus linguistics is interested in language *per se*’ (p. 111). Moreover, Biber *et al.* (1998) have noted that the software tools such as concordance packages for corpus analysis do not lend themselves to focusing on language characteristics extending across clause boundaries, or to semantic analysis, and are therefore not suitable for discourse analyses. The main epistemological differences, though, between the two fields lie in the fact that corpus analyses, by virtue of their methodological status, treat the text as product rather than as an unfolding discourse as process and social action: ‘the computer can only cope with the material products of what people do when they use language. It can only analyse the textual traces of the processes whereby meaning is achieved’ (Widdowson, 2000: 4).

As far back as 1998, I drew attention to the potential of corpus linguistics for ‘doing’ discourse analysis (Flowerdew, 1998). McEnery *et al.* (2006: 111) state that the aforementioned cultural divide ‘is now diminishing’, and Partington (2004) proposes that corpus and discourse methods are complementary. This chapter seeks to examine to what extent corpus and discourse approaches have now established a common meeting point, given their inherent differences in epistemologies and methodologies.

Studies of corpus-based discourse analyses will be discussed from the following three main approaches, adapted from categories proposed by Hyland (2009: 20). These have areas of overlap as, in reality, each approach also draws on aspects of the other approaches.

- **Textual:** approaches that focus on language choices, meanings and patterns in texts, including those based on the Swalesian (Swales, 2004) notion of genre and the problem–solution pattern. Also considered are various phraseological elements operating at the level of discourse.
- **Critical:** an approach that brings an attitude of criticality, such as critical discourse analysis (CDA), but also draws on other methods, e.g. systemic functional linguistics (SFL).
- **Contextual:** analyses that adopt a more sociolinguistic approach to the corpus data, where situational factors are also taken into account. This approach draws on conversation analysis, speech act theory and pragmatics.

Corpus-based discourse analyses can be viewed not only as adopting one of the main approaches above together with its attendant discourse area, e.g. the problem–solution pattern primarily associated with the textual approach, but also in terms of subject areas, i.e. workplace discourse, media discourse, academic discourse and so on, as a reflection of certain ideological positionings – discourses of racism, discourse of power – and mode – whether spoken or written or a ‘hybrid’ discourse, such as is emerging in the new technologies of blogs and chat rooms. Moreover, many of the discourse-based studies cited below implicitly subscribe to the ‘corpus-driven’ approach with their focus on the phraseological nature of language, in which the lexical item has primacy.

### Corpus-based textual approach

The corpus studies discussed in this section have mainly a text-based – that is, ‘language-in-use’ – focus. However, at the same time many also address the interpersonal nature of language, so that the analyses are also reader- and/or writer-oriented.

### *Models of discourse*

Handford's (2010a) overview article on the value of using specialized corpora for researching the discourse of particular genres reviews the operationalization of the Swalesian, new rhetoric and SFL approaches to genre across professional, academic and non-institutional genres. The Swalesian notion of genre as a goal-driven communicative event associated with particular discourse communities is the model that has been most widely applied in studies of written academic and professional texts, many previously hand-tagged for move structures. J. Flowerdew and Forest (2010) apply Swales' (1990: 141) CARS ('create a research space') model, originally posited for academic research article introductions, to PhD literature reviews, investigating the patterning of the keyword 'research' across different moves and steps (see also Gledhill, 2000; Upton and Connor, 2001; Flowerdew, 2008a for other genre-motivated corpus research). However, Bhatia *et al.*'s (2004) study of genre moves in law cases reveals the limitations of a purely corpus analytic approach: in order to make a pragmatic distinction between seemingly synonymous verbs, such as *dismiss* and *reject* in law cases, Bhatia *et al.* (*ibid.*, p. 213) state that it would be necessary to 'look for evidence from institutional practices', as corpora cannot (usually) provide such information. Different from, yet complementing these genre analyses in the spirit of the Swalesian tradition, are those studies reported in Biber *et al.* (2007a) and Csomay (2005) on vocabulary-based discourse units (VBDUs), which are identified automatically through comparison of 50-word 'windows', i.e. segments, of text. Although this is not a functional approach, nevertheless it is another way of identifying topic or move boundaries in text.

The studies cited above all commence from a lexico-grammatical, bottom-up perspective. Kanoksilapatham's (2007) research, on the other hand, takes a rhetorical top-down perspective at the outset. In her study of biochemistry research articles Kanoksilapatham first develops an analytical discourse-based framework through the identification of rhetorical move types and then uses Biber's multi-dimensional analysis to determine the linguistic characteristic of each move (see Parodi, 2009 for application of Biber's model to disambiguating discourse variation across academic and professional genres in Spanish).

Another discourse model that has been applied to corpus investigations is the problem-solution (P-S) pattern (see Hoey, 2001). Flowerdew (2008b) used the appraisal system from systemic-functional linguistics (Martin and White, 2005) for classifying keywords for the P-S pattern in technical reports into different types of evaluative lexis, followed by micro-analyses of the semantic relation of cause and effect. Ali Mohamed (2007) investigated the problem element in another text type, Malaysian and British journalistic business texts, also applying Martin's appraisal system to categorizing interpersonal and evaluative meanings. A key feature of this study is her use of the WMatrix corpus tool (Rayson, 2008) to identify different semantic fields characteristic of the problem element. Development of such tools serves to address Biber's criticism, noted earlier, that discourse studies are not served well by the existing concordance tools. A corpus study investigating the P-S pattern in terms of a more reader- and writer-oriented perspective accompanying the textual analysis is that by Alonso Belmonte (2009), which investigates two corpora: newspaper editorials and op-eds. An interactional analysis of different communicative acts (e.g. justification, exemplification) associated with different elements of the pattern was complemented by an illocutionary analysis with the corpus coded for speech acts such as assertions, shared-knowledge assertions and so on, indicating how writers conduct interaction with their readers. Alonso Belmonte's interactional analysis involving various types of discourse relations shares some features with Thompson and Mann's rhetorical structure theory (RST), which Renkema (2009: 174–177), with reference to three small-scale RST corpora, has advocated as a starting point for further corpus linguistic research.

## *Linguistic devices with discourse functions*

The focus of this section is on Hoey's theory of lexical priming as it operates at the textlinguistic level, and on three types of devices, namely lexical bundles, metadiscourse and metadiscoursal nouns. A key feature of the corpus studies reviewed below is that they are contrastive in nature, highlighting variation across different university disciplines, genres and registers.

### Lexical priming

Hoey's (2004, 2005) theory of lexical priming maps out a theoretical relationship between lexis and textlinguistics, showing how semantic associations, collocation and colligation operate at a discursive level. Hoey argues that some lexical items have a bias towards (or against) certain textual functions such as cohesion, theme choice and paragraph division and are also tied to particular genres and communities of users. For example, in the phrase *sixty years ago today*, all items, when they are the theme, have a preference for occurring in paragraph initial position (Hoey, 2004: 188).

### Lexical bundles

Biber *et al.* (2004) classify bundles (contiguous sequences of (usually) three-, four- or five-word n-grams) into three main functional categories: discourse organizers, referential expressions and stance expressions, the latter consisting of epistemic stance bundles that comment on the knowledge status of the information (e.g. *I don't know if*) and attitudinal/modality stance bundles which express speaker attitude towards certain actions (e.g. *I want you to*). Hyland's (2008) categorization of lexical bundles is similar to that of Biber's, but differs in that, like his classification of metadiscourse, they are organized around categories that reflect either the writer or the reader involvement in the text. Both Biber's and Hyland's research studies on lexical bundles have been invaluable for highlighting the functional differences between spoken and written registers (Cortes, 2004; Biber, 2006) and disciplinary variation in the academy (Hyland, 2008).

### Metadiscourse

Hyland's (2005) pioneering corpus research on metadiscourse is essentially interpersonal with its focus on those aspects of text that embody writer–reader interactions. On the basis of his model, which consists of an interactive (i.e. helping to guide the reader through the text) and interactional (involving the reader in the text) dimension, Hyland conducted corpus searches across various disciplines and levels of academic writing on the linguistic resources, realizing various functions subsumed under the two dimensions (see Hyland and Tse, 2004; Hyland, 2005). Bondi's (2001, 2004) quantitative and qualitative study of economics texts also examines the dialogic argumentative structure of academic text. Of interest is that Granger (1998) has found excessive the use of such signaling phrases for introducing arguments in learner academic writing. Likewise, Aijmer's (2009) study of 'I don't know' shows that this bundle is overwhelmingly used as a speech management function in learners' spoken English when compared with native speaker use.

Metadiscourse has also been investigated in the Michigan Corpus of Academic Spoken English (MICASE), most notably by Mauranen (2001, 2003) and Swales and Malczewski (2001). Mauranen's concept of metadiscourse differs somewhat from that of Hyland's, as she views this feature as fundamentally textual on account of its organization of ongoing discourse (*So let me just elaborate a little bit and then we...*). However, Mauranen argues that at the same time it is also

interpersonally motivated, as it imposes the speaker's order on the discourse situation, thus socializing students into the discourse. Likewise, Crawford Camiciottoli (2007) assigns an expert to novice communication role to metadiscursive phrases in business lectures (e.g. *That is the main point, The key thing*).

### Metadiscourse nouns

One area that has received a lot of attention is how certain nouns function at a discourse level. Drawing on Biber *et al.*'s distinction between epistemic and attitudinal markers, Charles (2003) compares the use of epistemic nouns, e.g. *assumption*, and stance nouns, e.g. *problem*, in post-graduate theses from the field of politics and materials science. Such metalinguistic nouns were found to function retrospectively, thus having an interpersonal function, as they indicate to the reader how the proposition is to be interpreted. Like Hyland, Charles (2006) also interprets her findings with reference to the different epistemologies and ideology of the discipline, noting that the higher frequency of metalinguistic nouns found in the politics corpus can be accounted for by the fact that knowledge construction in this discipline draws mostly on resources that are language-based, predominantly in written form (see J. Flowerdew, 2003 for research on nouns with discourse properties in a corpus of biology texts). A study examining the textlinguistic function of common nouns in a general corpus, i.e. the Bank of English, is that by Mahlberg (2003). Mahlberg notes the interactive nature of such nouns, identifying giving emphasis as one of their support function: 'It would doubtless be too much to expect Spurs fans to suddenly express a sweetness for Alan Sugar, *a man* who's been subjected to more abuse and hate mail than the average child molester' (p. 102).

Corpus linguistic techniques have thus proved of great value in shedding light on how various language choices and patterns operate at a textlinguistic level, either at a level above the clause or sentence, or within the framework of discourse models, very often using a multi-pronged approach, for instance combining Biber's MDA with Swalesian genre move structures.

### Corpus-based critical approach

In CDA the focus is on 'discourses' rather than on discourse per se. This notion refers to a broad range of linguistic and nonlinguistic social practices and ideological assumptions that co-construct, for example, 'discourses of power' or 'discourses of racism' – in other words, discourse with a capital 'D' (Gee, 2001). The techniques of CDA are multi-fold and vary. Text-analytic techniques draw on SFL, pragmatics and speech act analysis, and are integrated with concepts from contemporary social and cultural theory. Thus CDA is not a method as such in itself, but rather 'an academic movement', drawing on a kaleidoscope of methods increasingly those associated with corpus linguistics (Baker *et al.*, 2008). Two main approaches to CDA have developed since the 1960s. In the approach associated with Fairclough (2000) the analytical framework centres on a discursive event, an instance of language use, analysed not only as text, but also as discursive and social practice. The discourse–historical approach associated with the Viennese school (Wodak and Meyer, 2009) takes a more interdisciplinary, sociolinguistic perspective to the data, in which ethnography can also be a part of the analytical procedures.

What role do corpora play in these two approaches to CDA? Titscher *et al.* (2000: 158) state: 'Its [CDA's] point of departure is always the assumption that inequality and injustice are repeatedly reproduced in language and legitimized by it'. Corpus linguists working in CDA attempt to link recurring patterns in text with sociolinguistic features from the original contextual environment and vice versa. As Mautner (2009b: 124) points out: 'Doing so *critically* means unveiling and

challenging taken-for-granted assumptions about language and the social, as well as recognizing discourse as a potentially powerful agent in social change'. Moreover, corpus-based CDA studies make use of both quantitative techniques, that is, frequency and keyword lists – complemented by more detailed qualitative textual analysis and combined in such a way so as to uncover the *non-obvious meaning*, unavailable to conscious awareness, in the discourse under investigation. Using corpora for CDA analysis would also help to offset Widdowson's (1995, 2004) criticism of CDA that ideological significance is assigned to co-textual relations on very scant evidence, thus helping to reduce researcher bias.

In fact corpus-based CDA is a relatively new field (see Hunston, 2002, and Mautner, 2009a for a review of key studies), put on the map by the pioneering work of Stubbs (1996, 2001) and Hardt-Mautner (1995). These studies have given rise to the newly emerging interdisciplinary field of corpus-assisted discourse studies (CADS), an approach underpinned by Fairclough's concept of CDA.

### *Corpus-Assisted Discourse Studies (CADS)*

Some CADS case studies tend more towards the textual approach; such is Partington's (2003, 2007) research on the language strategies, metaphors and motifs used by journalists and spokespersons in US press conferences and on how these reflect their respective world views. Others, for instance the study by Krishnamurthy (1996) on the construction of people and race, are positioned clearly in the critical discourse camp.

Many of these studies examine the pervasive phenomenon of evaluation (see Hunston and Thompson 2000), applying Martin and White's (2005) Appraisal system in systemic functional grammar (see Bednarek, 2006 for an in-depth study of evaluation in media text). Much work in this area has also been carried out under the auspices of the CorDis project, which examines, from an interactive discourse perspective, how the conflict in Iraq was discussed and reported in the Senate and Parliament and in various media outlets (Morley and Bailey, 2009; Haarman and Lombardo, 2009). Duguid (2007), for example, examines the dialogistic positioning of Tony Blair and of his two advisors in the Hutton enquiry, noting the frequency of the collective noun *people*, which Fairclough (2000) has also noted surfacing as a keyword in his corpus of Blair speeches. In the corpus extracts below, the use of *people* serves to make the interactive, dialogistic nature of the discussions explicit, illustrating 'the continuous inter-textual concerns of the team, where a constant second-guessing goes on about how actions or texts will be perceived by those outside' (Duguid, 2007: 91):

You should not have gone to war – *people* can have a disagreement about that...

to, as it were, offer the name, but on the other hand, not to mislead *people*

but *people* would say, 'when did you know?'

Coffin and O'Halloran (2005) also make use of the Appraisal system, specifically of judgement, graduation and affect, first to carry out a detailed qualitative analysis of a report from *The Sun*. For example, in the sentence below, bold indicates graduation, underlining – judgement and italics – affect (adapted from Coffin and O'Halloran, 2005: 149):

**Two million jobs** will be lost if Tony Blair signs the EU treaty (negative indirect judgement of Blair), it was *feared* last night.

They then used a 45-million word newspaper corpus, made up of *The Sun* and its Sunday version, *The News of the World*, to check any potential over-interpretation of their Appraisal analysis. Their

concordances of *United States of Europe* reveal many of the local lexico-grammatical environments to indicate a negative evaluation.

towards their ambition of a	United States of Europe,	stretching from Shetland
could pave the way for a	United States of Europe.	British people have made
leader's bleak plan for a	United States of Europe	came as a hammer blow to
the road towards a Federal	United States of Europe.	Hague has never tried to
forming into a giant	United States of Europe	–with the same tax and
for a hopeless dream of a	United States of Europe.	He is certain to pay the

Concordance lines for *United States of Europe* from the *sunnow* sub-corpus of the Bank of English (adapted from Coffin and O'Halloran, 2005: 157).

Because of the negative prosody of the *United States of Europe*, Coffin and O'Halloran argue that Sun readers will be potentially predisposed to evaluate related expressions negatively even when they occur in a seemingly neutral statement, as in the case of the last sentence (*Mr Blair will be expected to sign up to the constitution blueprint by the end of June*) in their text chosen for qualitative analysis (see also Coffin and O'Halloran, 2006; O'Halloran, 2009).

### *Corpus-informed critical discourse studies*

Another perspective on corpus-based approaches to CDA, derived from the discourse-historical approach, is offered by the team of linguists working at Lancaster University on the project *Discourses of Refugees and Asylum Seekers in the UK press 1996–2006*. Their research is based on the analysis of a 140-million-word corpus of British news articles about refugees, asylum seekers, immigrants and migrants (collectively referred to as RASIM).

A key difference between CADS and this discourse-historical inspired study is that in the RASIM project a wide spectrum of background information on the social, political, historical and cultural context of the corpus data was used both to formulate hypotheses on which to base research questions and to inform interpretation of the corpus data. Key terms like 'refugee' were examined to see how they were conceptualized by 'official' sources such as dictionaries and organizations directly involved with these groups. Text-based analyses were also supported by official statistical information on the number of asylum applications.

The RASIM research has some affinity with the CADS approach through its focus on the identification of key words and collocation patterns, and of their underlying semantic preference and discourse prosodies (Baker and McEnery, 2005; Baker *et al.*, 2008), but it has less focus on SFL categories for linguistic analysis. It also differs from CADS in that these patterns were then mapped onto the discourse-historical CDA notions of *topos* and *topic* and also on the metaphors commonly employed in racist discourse, as a means of revealing elements of the underlying discourses relating to RASIM. For example, one of the common metaphors found to frame refugees was that of 'water', symbolizing the loss of control over immigration. Gabrielatos and Baker (2008: 15) point out: 'Statements employing this metaphor (e.g., *immigrants are flooding the country*) can very well utilize a *topos* of Number.' In assigning evaluative significance to various framing discourses for refugees and immigrants, Baker *et al.* (2008) also consulted the British National Corpus (BNC) to ascertain normative patterns of language use against which to compare the findings from the newspaper corpus (see Stubbs, 1996).

Corpus studies underpinned by the discourse-historical approach are few and far between, no doubt one reason being the intricate nature of the analyses drawing on a web of contextual strands



at various stages of such a study as the one above. Contextual approaches are also the focus of the following section dedicated to various types of spoken corpora.

### Corpus-based contextual approach

The analysis of spoken discourse through corpus linguistic techniques is largely concerned with how various rhetorical and pragmatic devices are operationalized by participants in specific social situations. This approach can be traced back to pioneering research carried out by Aijmer (1986) and others on the 500,000-word London-Lund Corpus (LLC) of spoken English of casual conversation. Stenström's (1994) application, on the basis of the Sinclair/Coulthard model, of the exchange sequence of questions and answers to this corpus is also a landmark in analyses of spoken discourse. Altenberg's (1998) study of the LLC was one of the first to explore phraseology in spoken corpora. His research shows that functional sequences often have a core with optional extensions (e.g., [oh/ yes] *I see*) and that sequences may overlap, being sometimes interrupted by non-formulaic language, in accordance with Sinclair's (1991) 'open-choice' principle.

Since the 1990s other studies making use of the LLC have adopted a more finely grained, integrative, multi-layered approach, also paying attention to prosodic elements of the discourse for meaning-making. Aijmer (2002) analysed the discourse particle *actually* by following Brazil's (1995) prosodic model of proclaiming and referring tones and by noting correlations between its prosody and discourse functions in different positions in the utterance. Other studies linking prosody with discourse function include Wichmann's (2004) study of *please* requests in the ICE-GB Corpus (the British contribution to the International Corpus of English) and Cheng and Warren's (2008) study mapping speakers' discourse intonation choices onto word association patterns in a corpus of public discourse, one of the four sub-corpora in the two-million-word Hong Kong Corpus of Spoken English (HKCSE). A CDA perspective on this corpus is given in Cheng (2004), who shows how prosodic choices can be exploited by speakers so that a politician might assert common ground where none exists (see also Warren, 2004).

The 5-million-word CANCODE (Cambridge and Nottingham Corpus of Discourse in English) marked a watershed in the era of computerized spoken corpora for its design and range of analytical procedures (McCarthy, 1998; Carter and McCarthy, 2004). As well as being marked up for demographic data, it has as its main organizing principle five genre contexts (transactional, professional, pedagogical, socializing and intimate) and focuses on three types of goal-oriented exchanges. These three interaction types cover information provision (e.g. commentary by museum guide), collaborative task (e.g. choosing and buying a television), and collaborative idea (e.g. chatting with hairdresser) (see Koester, 2006 whose analysis of workplace discourse, a part of which is based on CANCODE, adopts similar investigative procedures, with a focus on the analysis of transactional and relational goals). The multi-faceted sociolinguistic, analytical approach of the CANCODE data draws on praxis theory (meanings are negotiated face to face and emerge from the unfolding discourse), which itself implicates and extends, as context is taken into account, the notion of adjacency pairs, turn-taking, turn boundaries and sequencing associated with conversation analysis. Hughes and McCarthy (1998) also posit the notion of an *interpersonal grammar*, exemplifying how certain grammatical features, e.g. tags and amplificatory noun-phrases occupying the tail slot of a sentence – 'It's very nice *that road up through Skipton to the Dales*' – signal relationships between participants and their stance or attitudinal positioning towards the emergent discourse (see Leech, 2000 for a detailed discussion on discourse grammar).

The approach also draws on rational action (language as strategic motivated choices), encompassing Brown and Levinson's (1987) politeness theory. McCarthy (1998) reports that the CANCODE data show speech acts to be far more indirect and subtle in their unfolding than

invented examples; the verb form *disagree* occurs mostly in contexts reporting disagreement with someone, or disagreeing with propositions rather than people. Adolphs' (2008) research on speech acts and pragmatic functions in CANCODE is primarily motivated by this 'language as rational action' perspective.

While corpus linguistics and discourse analysis have benefited from each other's strengths, there still remain some weaknesses to address. Widdowson (2000: 4) has remarked that corpus-based methods focus on the text as product and 'cannot account for the complex interplay of linguistic and contextual factors whereby discourse is enacted'. In a similar vein, Virtanen (2009: 62) remarks that 'the main problem on the road from discourse to corpora and back again remains the lack of contextual dynamism'. Furthermore, Lee (2008) notes the lack of non-verbal aspects of communication accompanying spoken corpora. The following section reviews some very recent endeavours that serve to address these issues, thereby investing corpus-based discourse analysis with more of a 'language in action' orientation. Yet at the same time the changing face of 'discourse' raises new challenges for corpus linguists.

### Recent developments and new challenges

As Widdowson (2000: 4) has noted: 'It [the computer] cannot produce ethnographic descriptions of language use.' This fact, together with 'the lack of immediacy of the discourses vis-à-vis the analyst', which 'may be a hindrance for types of discourse analysis that rely on intimate knowledge of the data, participants and context' (Lee, 2008: 95) no longer remain such stumbling blocks as evidenced by Handford's (2010b) research of business meetings from CANBEC (Cambridge and Nottingham Corpus of Business English). Handford maps out a sophisticated set of practices (professional, social, discursive), including ethnographic data in the form of field notes and interviews, in order to understand professional practices of the genre. Through this multi-perspective lens Handford is able not only to capture the routinized aspects of the genre but also to tap into the 'contextual dynamism' of the communication to shed light on discourse features.

Until very recently corpora were analysed from a monomodal perspective, either written or spoken. Multimodal corpora are now being compiled with a view to aligning non-linguistic and linguistic aspects of spoken discourse. Pioneers in this field working in the Hallidayan tradition (Baldry and Thibault, 2006, and forthcoming) take a systemic-functional orientation to the discourse to determine how different semiotic resources (language, gaze, gesture, movement etc.) interact to create meaning. They propose the notion of 'visual collocation' to refer to the probability of constellations of visual items in a particular setting; for example, in car advertisements featuring test drivers, the car is often found as the phenomenon in a gaze transitivity frame, collocating with difficult testing terrains such as deserts. Another key initiative is that reported in Carter and Adolphs (2008) on a corpus of video-taped MA and PhD supervision sessions: the authors advocate the importance of taking a discourse-level perspective on the integration of verbal and visual elements of the corpus data, noting a correlation between different types of non-verbal backchanneling (head nods) and information structure and function. A third pioneering endeavour is that undertaken by Gu (2006), from a situated discourse perspective in which video streams and synchronized sounds take precedence over the orthographic transcription in the analysis. This type of multimodal analysis moves from the analytic unit of a situated discourse through several layers to a prosodic unit of illocutionary force. Gu's work, with its emphasis on language as a social phenomenon used in meaning-making, has some affinity with Halliday's SFL, and also draws on Kress' (2001) work on multimodality in its study of social action over time.

However, there is now a ‘new modal order’ emerging in this era of digital literacies, specifically computer-mediated communication involving e-mail, discussion groups, Internet relay chats (IRC) and weblogs (Beißwenger and Storrer, 2008; King, 2009; Ooi, 2009). Although some analyses have been carried out into weblogs (Ooi *et al.*, 2007), this is still a fledgling area as far as corpus-based discourse analysis is concerned, and one that poses enormous challenges. How can all the semiotic elements in a corpus of weblogs, with its manifold modalities (text, video, pictures, audio files, hypertextual links to other blogs,) be accommodated within a discourse-analytic framework? Even the written text alone will require new software for analysing discourse features of internet communication such as emoticons, which can have evaluative, expressive or regulative functions, and other conventions (such as upper case) for simulating prosodic features. Moreover, King (2009) notes the challenge for corpus linguists in analysing turn-taking in chat rooms, in which one turn can often be split into many in order to keep up with the real-time unfolding of conversation.

This review chapter of corpus-based discourse analysis has exemplified how the field has moved from single-pronged to more multi-pronged approaches, from a language in use to a more complex language in action perspective, and from monomodal to multimodal analyses. This complex synergy of methods, approaches and tools has enabled a rapprochement of the two fields, corpus linguistics no longer hovering on the periphery of discourse analysis but now assuming a central role. However, new forms of discourse are evolving that have thrown up new challenges for corpus linguistics, in the never-ending quest to get at the heart of what ‘discourse’ really entails.

## Further reading

Aijmer, K. and Stenström, A.-B. (eds.) (2004) *Discourse Patterns in Spoken and Written Corpora*. Amsterdam: John Benjamins.

This edited collection contains a range of corpus-based studies on different aspects of discourse: cohesion and coherence, metadiscourse and discourse markers, and text and information structure.

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

This book provides a very reader-friendly introduction to how key words, frequency and dispersion, and collocational networks can inform discourse analysis, especially those studies of a CDA nature.

Charles, M., Pecorari, D., and Hunston, S. (eds.) (2010) *Academic Writing: At the Interface of Corpus and Discourse*. London: Continuum.

This volume contains a collection of articles that focus on different types of discourse: genre and disciplinary, interpersonal and learner discourse.

Paltridge, B. (2006) *Discourse Analysis*. London: Continuum.

Chapter 7 in this volume, ‘Corpus Approaches to Discourse Analysis’, presents a very useful overview of the relationship between corpus analysis and discourse analysis.

## References

- Aarts, J. (2002) ‘Review of *Corpus Linguistics at Work*’, *International Journal of Corpus Linguistics*, 7 (1): 118–123.
- Ädel, A. and Reppen, R. (eds.) (2008) *Corpora and Discourse: The Challenges of Different Settings*. Amsterdam: John Benjamins.
- Adolphs, S. (2008) *Corpus and Context: Investigating Pragmatic Functions in Spoken Discourse*. Amsterdam: John Benjamins.
- Aijmer, K. (1986) ‘Discourse variation and hedging’, in B. Aarts and W. Meijs (eds.) *Corpus Linguistics II: New Studies in the Analysis and Exploitation of Computer Corpora*. Amsterdam: Rodopi, pp. 1–18.
- Aijmer, K. (2002) *English Discourse Particles; Evidence From a Corpus*. Amsterdam: John Benjamins.
- Aijmer, K. (2009) ‘“So er I just sort I dunno I think it’s just because...”’: a corpus study of *I don’t know* and *dunno* in learners’ spoken English’, in A. Jucker *et al.* (eds.) *Corpora: Pragmatics and Discourse*. Amsterdam: Rodopi, pp. 151–168.

- Ali Mohamed, A. (2007) 'Semantic fields of problem in business English: Malaysian and British journalistic business texts', *Corpora*, 2 (2): 211–239.
- Alonso Belmonte, I. (2009) 'Towards a genre-based characterization of the problem-solution textual pattern in English newspaper editorials and op-eds', *Text and Talk*, 29 (4): 393–414.
- Altenberg, B. (1998) 'On the phraseology of spoken English: the evidence of recurrent word combinations', in A. Cowie (ed.) *Phraseology: Theory, Analysis and Applications*. Oxford: Clarendon Press, pp. 101–122.
- Baker, P. (ed.) (2009) *Contemporary Corpus Linguistics*. 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., KhosraviNik, M., Krzyżanowski, M., Mcenery, T., and Wodak, R. (2008) 'A useful methodological synergy? Combining critical discourse analysis and corpus linguistics to examine discourses of refugees and asylum seekers in the UK Press', *Discourse and Society*, 19 (3): 273–306.
- Baldry, A. and Thibault, P. (2006) 'Multimodal corpus linguistics', in G. Thompson and S. Hunston (eds.) *System and Corpus; Exploring Connections*. London: Equinox, pp. 164–183.
- Baldry, A. and Thibault, P. (forthcoming) *Multimodal Corpus Linguistics*. London: Routledge.
- Bednarek, M. (2006) *Evaluation in Media Discourse*. London: Continuum.
- Beißwenger, M. and Storrer, A. (2008) 'Corpora of computer-mediated communication', in A. Lüdeling and M. Kytö (eds.) *Corpus Linguistics: An International Handbook. Vol. 1*. Amsterdam: Mouton de Gruyter, pp. 292–309.
- Bhatia, V., Langton, N., and Lung, J. (2004) 'Legal discourse: opportunities and threats for corpus linguistics', in U. Connor and T. Upton (eds.), *Discourse in the Professions: Perspectives from Corpus Linguistics*. Amsterdam: John Benjamins, pp. 203–231.
- Bhatia, V., Flowerdew, J., and Jones R. (eds.) (2008) *Advances in Discourse Studies*. London: Routledge.
- Biber, D. (2006) *University Language: A Corpus-Based Study of Spoken and Written Registers*. Amsterdam: John Benjamins.
- Biber, D., Connor, C., and Upton, T. (2007a) *Discourse on the Move: Using Corpus Analysis to Describe Discourse Structure*. Amsterdam: John Benjamins.
- Biber, D., Conrad, S., and Cortes, V. (2004) 'If you look at...: lexical bundles in university teaching and textbooks', *Applied Linguistics*, 25 (3): 371–405.
- Biber, D., Conrad, S., and Reppen, R. (1998) *Corpus Linguistics: Investigating Language Structure and Use*. Cambridge: Cambridge University Press.
- Biber, D., Csomay, E., Jones, J., and Keck, C. (2007b) 'Introduction to the analysis and identification of vocabulary-based discourse units', in D. Biber et al. (eds.), *Discourse on the Move*. Amsterdam: John Benjamins. pp. 155–173.
- Blommaert, J. (2005) *Discourse*. Cambridge: Cambridge University Press.
- Bondi, M. (2001) 'Small corpora and language variation: reflexivity across genres', in M. Ghadessy, A. Henry, and R. Roseberry (eds.) *Small Corpus Studies and ELT*. Amsterdam: John Benjamins.
- Bondi, M. (2004) 'The discourse function of contrastive connectors in academic abstracts', in K. Aijmer and A.-B. Stenström (eds.) *Discourse Patterns in Spoken and Written Corpora*. Amsterdam: John Benjamins, pp.139–156.
- Brazil, D. (1995) *A Grammar of Speech*. Oxford: Oxford University Press.
- Brown, P. and Levinson, S. (1987) *Politeness*. Cambridge: Cambridge University Press.
- Carter, R. and Adolphs, S. (2008) 'Linking the verbal and the visual: new directions for corpus linguistics', *Language and Computers*, 64: 275–291.
- Carter, R. and McCarthy, M. (2004) 'Talking, creating: interactional language, creativity and context', *Applied Linguistics*, 25 (1): 62–88.
- Charles, M. (2003) '"This mystery...": a corpus-based study of the use of nouns to construct stance in theses from two contrasting disciplines', *Journal of English for Academic Purposes*, 2: 313–326.
- Charles, M. (2006) 'The construction of stance in reporting clauses: a cross-disciplinary study of theses', *Applied Linguistics*, 27: 492–518.
- Cheng, W. (2004) '//→ FRIENDS//↗ Ladies and GENTlemen//: Some preliminary findings from a corpus of spoken public discourses in Hong Kong', in U. Connor and T. Upton (eds.), *Discourse in the Professions: Perspectives from Corpus Linguistics*. Amsterdam: John Benjamins, pp. 35–50.
- Cheng, W. and Warren, M. (2008) '//→ ONE country two SYStems//: The discourse intonation patterns of word associations', in A. Ädel and R. Reppen (eds.) *Corpora and Discourse: The Challenges of Different Settings*. Amsterdam: John Benjamins, pp. 135–153.

- Coffin, C. and O'Halloran, K. (2005) 'FINDING THE GLOBAL GROOVE: Theorising and analyzing dynamic reader positioning using APPRAISAL, corpus, and a concordancer', *Critical Discourse Studies*, 2 (2): 143–163.
- Coffin, C. and O'Halloran, K. (2006) 'The role of appraisal and corpora in detecting covert evaluation', *Functions of Language*, 13 (1): 77–110.
- Connor, U. and Upton, T. (eds.) (2004a) *Applied Corpus Linguistics: A Multi-Dimensional Perspective*. Amsterdam: Rodopi.
- Connor, U. and Upton, T. (eds.) (2004b) *Discourse in the Professions: Perspectives From Corpus Linguistics*. Amsterdam: John Benjamins.
- Conrad, S. (2002) 'Corpus linguistic approaches for discourse analysis', *Annual Review of Applied Linguistics*, 22: 75–95.
- Cortes, V. (2004) 'Lexical bundles in published student disciplinary writing: examples from history and biology', *English for Specific Purposes*, 23: 397–423.
- Cowie, A. (ed.) (1998) *Phraseology: Theory, Analysis and Applications*. Oxford: Clarendon Press.
- Crawford Camiciottoli, B. (2007) *The Language of Business Studies Lectures*. Amsterdam: John Benjamins.
- Csomas, E. (2005) 'Linguistic variation within university classroom talk: a corpus-based perspective', *Linguistics and Education*, 15 (3): 243–274.
- Duguid, A. (2007) 'Soundbiters bit. Contracted dialogistic space and the textual relations of the No. 10 team analysed through corpus-assisted discourse studies', in N. Fairclough, G. Cortese, and P. Ardizzone (eds.) *Discourse and Contemporary Social Change*. Frankfurt am Main: Peter Lang, pp. 73–94.
- Fairclough, N. (2000) *New Labour, New Language?* London: Routledge.
- Flowerdew, J. (2003) 'Signalling nouns in discourse', *English for Specific Purposes*, 22: 329–346.
- Flowerdew, J. and Forest, R. (2010) 'Schematic structure and lexico-grammatical realization in corpus-based genre analysis: the case of research in the PhD literature review', in M. Charles, D. Pecorari, and S. Hunston (eds.) *Academic Writing at the Interface of Corpus and Discourse*. London: Continuum, pp. 15–36.
- Flowerdew, L. (1998) 'Corpus linguistic techniques applied to textlinguistics', *System*, 26 (4): 541–552.
- Flowerdew, L. (2008a) 'Determining discourse-based moves in professional reports', in A. Ädel and R. Reppen (eds.) *Corpora and Discourse: The Challenges of Different Settings*. Amsterdam: John Benjamins, pp. 117–131.
- Flowerdew, L. (2008b) *Corpus-Based Analyses of the Problem-Solution Pattern*. Amsterdam: John Benjamins.
- 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.
- Gee, J. P. (2001) *An Introduction to Discourse Analysis*. London: Routledge.
- Gledhill, C. (2000) 'The discourse function of collocation in research article introductions', *English for Specific Purposes*, 19: 115–135.
- Granger, S. (1998) 'Prefabricated patterns in advanced EFL writing: collocations and formulae', in A. Cowie, (ed.) *Phraseology: Theory, Analysis and Applications*. Oxford: Clarendon Press, pp. 145–160.
- Gu, Y. (2006) 'Multimodal text analysis: a corpus linguistic approach to situated discourse', *Text and Talk*, 26 (2): 127–167.
- Haarman, L. and Lombardo, L. (eds.) (2009) *Evaluation and Stance in War News*. London: Continuum.
- Handford, M. (2010a) 'What can a corpus tell us about specialist genres?', in M. McCarthy and A. O'Keeffe (eds.) *The Routledge Handbook of Corpus Linguistics*. London: Routledge, pp. 255–269.
- Handford, M. (2010b) *The Language of Business Meetings*. Cambridge: Cambridge University Press.
- Hardt-Mautner, G. (1995) *Only Connect: Critical Discourse Analysis and Corpus Linguistics*. UCREL Technical Paper 6. Lancaster: University of Lancaster.
- Hoey, M. (2001) *Textual Interaction: An Introduction to Written Discourse Analysis*. London: Routledge.
- Hoey, M. (2004) 'Textual colligation: a special kind of lexical priming', in K. Aijmer and B. Altenberg (eds.) *Advances in Corpus Linguistics*, Amsterdam: Rodopi, pp. 171–194.
- Hoey, M. (2005) *Lexical Priming: A New Theory of Words and Language*. London: Routledge.
- Hughes, R. and McCarthy, M. (1998) 'From sentence to grammar: discourse grammar and English language teaching', *TESOL Quarterly*, 32 (2): 263–287.
- Hunston, S. (2002) *Corpora in Applied Linguistics*. Cambridge: Cambridge University Press.
- Hunston, S. and Thompson, G. (eds) (2000) *Evaluation in Text: Authorial Stance and the Construction of Discourse*. Oxford: Oxford University Press.
- Hyland, K. (2005) *Metadiscourse*. London: Continuum.
- Hyland, K. (2008) 'As can be seen: lexical bundles and disciplinary variation', *English for Specific Purposes*, 27 (1): 4–21.

- Hyland, K. (2009) *Academic Discourse*. London: Continuum.
- Hyland, K. and Tse, P. (2004) 'Metadiscourse in academic writing: a reappraisal', *Applied Linguistics*, 25 (2): 156–177.
- Jucker, A., Schreier, D., and Hundt, M. (eds.) (2009a) *Corpora: Pragmatics and Discourse*. Amsterdam: Rodopi.
- Jucker, A., Schreier, D., and Hundt, M. (2009b) 'Corpus linguistics, pragmatics and discourse'. In A. Jucker *et al.* (2009a), pp. 3–9.
- Kanoksilapatham, B. (2007) 'Rhetorical moves in biochemistry research articles', in D. Biber *et al.* (eds.), *Discourse on the Move*. Amsterdam: John Benjamins, pp. 73–119.
- King, B. (2009) 'Building and analyzing corpora of computer-mediated communication', in P. Baker (ed.) *Contemporary Corpus Linguistics*. London: Continuum, pp. 301–320.
- Koester, A. (2006) *Investigating Workplace Discourse*. London: Routledge.
- Kress, G. (2001) *Multimodal Discourse: The Modes and Media of Contemporary Communication*. London: Arnold.
- Krishnamurthy, R. (1996) Ethnic, racial and tribal: the language of racism? In R. Caldas-Coulthard, and M. Coulthard (eds.) *Texts and Practices: Readings in Critical Discourse Analysis*, pp. 129–149. London: Routledge.
- Lee, D. (2008) 'Corpora and discourse analysis: new ways of doing old things', in V. Bhatia, J. Flowerdew, and R. Jones (eds) *Advances in Discourse Studies*. London: Routledge, pp. 86–99.
- Leech, G. (2000) 'Grammar of spoken English: new outcomes of corpus-oriented research'. *Language Learning*, 50 (4): 675–724.
- Mahlberg, M. (2003) 'The textlinguistic dimension of corpus linguistics', *International Journal of Corpus Linguistics*, 8 (1): 97–108.
- Martin, J. and White, P. (2005) *The Language of Evaluation: Appraisal in English*. Basingstoke: Palgrave/Macmillan.
- Mauranen, A. (2001) 'Reflexive academic talk: observations from MICASE', in R. Simpson and J. Swales (eds.) *Corpus Linguistics in North America*. Ann Arbor, MI: University of Michigan Press, pp. 165–178.
- Mauranen, A. (2003) "'But here's a flawed argument": socialization in and through metadiscourse', in P. Leistyna and C. Meyer (eds.) *Corpus Analysis: Language Structure and Language Use*. Amsterdam: Rodopi, pp. 19–34.
- Mautner, G. (2009a) 'Corpora and critical discourse analysis', in P. Baker (ed.) *Contemporary Corpus Linguistics*. London: Continuum, pp. 32–46.
- Mautner, G. (2009b) 'Checks and balances: how corpus linguistics can contribute to CDA', in R. Wodak and M. Meyer (eds.) *Methods of Critical Discourse Analysis*. London: Sage, pp. 122–143.
- McCarthy, M. (1998) *Spoken Languages and Applied Linguistics*. Cambridge: Cambridge University Press.
- McEnery, T., Xiao, R., and Tono, Y. (2006) *Corpus-Based Language Studies*. London: Routledge.
- Morley, J. and Bailey, P. (eds.) (2009) *Corpus-Assisted Discourse Studies on the Iraq Conflict*. London: Routledge.
- O'Halloran, K. (2009) 'Inferencing and cultural reproduction: a corpus-based critical discourse analysis', *Text and Talk*, 29 (1): 21–51.
- Ooi, V. (2009) 'Computer-mediated language and corpus linguistics', in Y. Kawaguchi, M. Minegishi, and J. Durand (eds.) *Corpus Analysis and Variation in Linguistics*. Amsterdam: John Benjamins, pp. 103–120.
- Ooi, V., Tan, P., and Chiang, A. (2007) 'Analysing personal weblogs in Singapore English: the Wmatrix approach', in *EVariEng (Journal of the Research Unit for Variation, Contacts, and Change in English)*. Volume 2: *Towards Multimedia in Corpus Studies*. Finland: University of Helsinki.
- Parodi, G. (2009) 'Academic and professional written genres in disciplinary communication', in J. Renkema (ed.) *Discourse, of Course: An Overview of Discourse in Research Studies*. Amsterdam: John Benjamins, pp. 93–110.
- Partington, A. (2003) *The Linguistics of Political Argument*. London: Routledge.
- Partington, A. (2004) 'Corpora and discourse, a most congruous beast', in A. Partington, J. Morley, and L. Haarmann (eds.) *Corpora and Discourse*, Bern: Peter Lang, pp. 11–20.
- Partington, A. (2007) 'Irony and reversal of evaluation'. *Journal of Pragmatics*, 39: 1547–1569.
- Rayson, P. (2008) 'From key words to key semantic domains'. *International Journal of Corpus Linguistics*, 13 (4): 519–549.
- Renkema, J. (2009) *The Texture of Discourse*. Amsterdam: John Benjamins.
- Simpson, R. and Swales, J. (eds.) (2001) *Corpus Linguistics in North America*. Ann Arbor, MI: University of Michigan Press.
- Sinclair, J. McH. (1991) *Corpus, Concordance, Collocation*. Oxford: Oxford University Press.
- Sinclair, J. McH. (2004) *Trust the Text*. London: Routledge.
- Stenström, A.-B. (1994) *An Introduction to Spoken Interaction*. London: Longman.

- Stubbs, M. (1983) *Discourse Analysis*. Chicago, IL: University of Chicago Press.
- Stubbs, M. (1996) *Text and Corpus Analysis*. Oxford: Blackwell.
- Stubbs, M. (2001) *Words and Phrases: Corpus Studies of Lexical Semantics*. Oxford: Blackwell.
- Swales, J. (1990) *Genre Analysis: English in Academic and Research Genres*. Cambridge: Cambridge University Press.
- Swales, J. (2004) *Research Genres*. Cambridge: Cambridge University Press.
- Swales, J. and Malczewski, B. (2001) 'Discourse management and new-episode flags in MICASE', in R. Simpson and J. Swales (eds.) *Corpus Linguistics in North America*. Ann Arbor, MI: University of Michigan Press, pp. 145–164.
- Titscher, S., Meyer, M., and Vetter, E. (2000) *Methods of Text and Discourse Analysis*. London: Sage Publications.
- Tognini Bonelli, E. (2001) *Corpus Linguistics at Work*. Amsterdam: John Benjamins.
- Upton, T. and Connor, U. (2001) 'Using computerized corpus analysis to investigate the textlinguistic discourse moves of a genre', *English for Specific Purposes*, 20: 313–329.
- Virtanen, T. (2009) 'Discourse linguistics meets corpus linguistics: theoretical and methodological issues in the troubled relationship', in A. Renouf and A. Kehoe (eds.) *Corpus Linguistics: Refinements and Reassessments*, Amsterdam: Rodopi, pp. 49–65.
- Warren, M. (2004) 'A corpus-driven analysis of the use of intonation to assert dominance and control', in U. Connor and T. Upton (eds.), *Applied Corpus Linguistics. A Multi-Dimensional Perspective*. Amsterdam: Rodopi, pp. 21–33.
- Wichmann, A. (2004) 'The intonation of *please*-requests: a corpus-based study', *Journal of Pragmatics*, 36: 1521–1549.
- Widdowson, H. G. (1995) 'Discourse analysis: a critical view', *Language and Literature* 4 (3): 157–172.
- Widdowson, H. G. (2000) 'The limitations of linguistics applied', *Applied Linguistics*, 21(1): 3–25.
- Widdowson, H. G. (2004) *Text, Context, Pretext: Critical Issues in Discourse Analysis*. Oxford: Blackwell.
- Wodak, R. and Meyer, M. (2009a) 'Critical discourse analysis: history, agenda, theory and methodology', in R. Wodak and M. Meyer (eds.) *Methods of Critical Discourse Analysis*. London: Sage, pp. 1–33.
- R. Wodak, M. Meyer, (eds.) (2009b) *Methods of Critical Discourse Analysis*. London: Sage.

