Copyrighted material – 978–1–137–43172–1

Selection, introduction and editorial content © Paul Baker and Tony McEnery 2015
Individual chapters © Respective authors 2015

All rights reserved. No reproduction, copy or transmission of this publication may be made without written permission.

No portion of this publication may be reproduced, copied or transmitted save with written permission or in accordance with the provisions of the Copyright, Designs and Patents Act 1988, or under the terms of any licence permitting limited copying issued by the Copyright Licensing Agency, Saffron House, 6–10 Kirby Street, London EC1N 8TS.

Any person who does any unauthorized act in relation to this publication may be liable to criminal prosecution and civil claims for damages.

The authors have asserted their rights to be identified as the authors of this work in accordance with the Copyright, Designs and Patents Act 1988.

First published 2015 by
PALGRAVE MACMILLAN

Palgrave Macmillan in the UK is an imprint of Macmillan Publishers Limited, registered in England, company number 785998, of Houndmills, Basingstoke, Hampshire RG21 6XS.

Palgrave Macmillan in the US is a division of St Martin's Press LLC, 175 Fifth Avenue, New York, NY 10010.

Palgrave Macmillan is the global academic imprint of the above companies and has companies and representatives throughout the world.

Palgrave® and Macmillan® are registered trademarks in the United States, the United Kingdom, Europe and other countries.

ISBN 978–1–137–43172–1

This book is printed on paper suitable for recycling and made from fully managed and sustained forest sources. Logging, pulping and manufacturing processes are expected to conform to the environmental regulations of the country of origin.

A catalogue record for this book is available from the British Library.

Library of Congress Cataloging-in-Publication Data
Corpora and discourse studies : integrating discourse and corpora / edited by Paul Baker, Lancaster University, UK and Tony McEnery, University of Lancaster, UK.
pages cm

Summary: "The growing availability of large collections of language texts has expanded our horizons for language analysis, enabling the swift analysis of millions of words of data, aided by computational methods. This edited collection contains examples of such contemporary research which uses corpus linguistics to carry out discourse analysis. The book takes an inclusive view of the meaning of discourse, covering different text-types or modes of language, including discourse as both social practice and as ideology or representation. Authors examine a range of spoken, written, multimodal and electronic corpora covering themes which include health, academic writing, social class, ethnicity, gender, television narrative, news, Early Modern English and political speech. The chapters showcase the variety of qualitative and quantitative tools and methods that this new generation of discourse analysts are combining together, offering a set of compelling models for future corpus-based research in discourse"— Provided by publisher.
ISBN 978–1–137–43172–1 (hardback)
1. Discourse analysis. 2. Corpora (Linguistics) I. Baker, Paul, 1972- editor.
II. McEnery, Tony, 1964- editor.
P302.C66 2015
401'.41—dc23 2015012348

Typeset by MPS Limited, Chennai, India.

Copyrighted material – 978–1–137–43172–1

Copyrighted material – 978–1–137–43172–1

Contents

Copyrighted material – 978–1–137–43172–1

Copyrighted material – 978–1–137–43172–1

Copyrighted material – 978–1–137–43172–1

Copyrighted material – 978–1–137–43172–1

1
Introduction

Paul Baker and Tony McEnery

This book houses a collection of 13 independent studies which use the corpus linguistics methodology in order to carry out discourse analysis. In this introductory chapter we first introduce the two main concepts of the book, *corpus linguistics* and *discourse analysis*, and cover the advantages of combining the two approaches. After discussing the existing key research and debates in this relatively new field we then outline the remainder of the book's three-part structure with a brief description of each chapter.

Corpus linguistics

Corpus linguistics is a powerful methodology – a way of using computers to assist the analysis of language so that regularities among many millions of words can be quickly and accurately identified. Coming from Latin, a corpus is a body, so we may say that corpus linguistics is simply the study of a body of language – in many cases a very large body indeed. Such a body may consist of hundreds or thousands of texts (or excerpts of texts) that have been carefully sampled and balanced in order to be representative of a specific variety of language (e.g. nineteenth-century women's fiction, British newspaper articles about poverty, political speeches, teenager's text messages, Indian English, essays by Chinese students learning English). In order to facilitate more complex forms of analysis, many corpora are 'tagged', i.e. have explicit linguistic analyses introduced into them, usually in the form of mnemonic codes. This is often done automatically via computer software (for example, Amanda Potts in Chapter 14 uses a corpus of news articles tagged by a computer program called the USAS English tagger), although we note that in this volume Dan McIntyre and Brian Walker (Chapter 9) hand-tagged their corpus for different categories of discourse presentation as software was not able to make the distinctions they required. Automatic tagging performs well (although not at 100% accuracy) at grammatical or semantic tagging. For example, all of the words in a corpus may

Copyrighted material – 978–1–137–43172–1

Copyrighted material – 978–1–137–43172–1

be automatically assigned codes which indicate their grammatical part of speech (noun, verb, adjective etc.) or which semantic group they are from (living things, conflict, economics etc.). Tagging can also occur at the level of the text itself, for example, all texts may be tagged according to the gender of the author, allowing us to easily separate out and compare language according to this variable.

Using specially designed software in conjunction with a corpus, analysts are given a unique view of language within which frequency information becomes highly salient. Hence it is no surprise that the concept of frequency drives many of the techniques associated with corpus linguistics, giving the field a quantitative flavour. Many of the chapters in this book employ two frequency-based techniques in particular – keywords and collocates. Keywords are words which are more frequent than expected in one corpus, when compared against a second corpus which often stands as a 'reference', usually being representative of a notional 'standard language'. Keywords reveal words which may not be hugely frequent but are definitely statistically salient in some way. Collocation involves the identification of words which tend to occur near or next to each other a great deal, much more than would be expected if all the words in a corpus were ordered in a random jumble. Native speakers of a language have thousands of collocates stored in their memories and hearing or reading one word may often prime another, due to all of our previous experiences of hearing that word in a particular context. From an ideological point of view, collocates are extremely interesting, as if two words are repetitiously associated with each other, then their relationship can become reified and unquestioned (Stubbs, 1996: 195).

While the earliest stages of a corpus analysis tend to be quantitative, relying on techniques like keywords and collocates in order to give the research a focus, as a research project progresses, the analysis gradually becomes more qualitative and context-led, relying less on computer software. Once quantitative patterns have been identified, they need to be interpreted and this usually involves a second stage of analysis where the software acts as an aid to the researcher by allowing the linguistic data to be quickly surveyed.

For example, we may be interested in how many texts a word or feature occurs in, or whether it tends to occur at the beginning, middle or end of a text. Corpus tools often allow measures of dispersion to be taken into account, sometimes using a visual representation of a file, which can resemble a bar code, with each horizontal line indicating an occurrence of a particular word. Knowing if a word or feature is well-distributed across a corpus, or simply frequent because it occurs very frequently in a few texts, can be one way of understanding the context in which it is used. As well as position, it is essential to ascertain the way that the feature is used in the context of every utterance, sentence or paragraph it occurs in. A concordance table is simply a table of all of the occurrences of a word, phrase or other linguistic feature (e.g. grammatical or semantic tag) in a corpus, occurring with a few

Copyrighted material – 978–1–137–43172–1

Copyrighted material – 978–1–137–43172–1

words of context either side. Concordance tables can be sorted, for example by ordering the table alphabetically according to the word immediately to the right or left of the word we are analysing. This helps to group together incidences of a word that occur in similar contexts so interpretations can be more easily made. In cases where a word may occur thousands of times, we may only want to examine a smaller sample of concordance lines, so again the software will randomly reduce or 'thin' the number of lines to a more manageable amount.

Such tools enable more qualitative forms of analysis to be carried out on corpora, although we argue that a third stage of analysis – explanation – involves positioning our descriptive and interpretative findings within a wider social context. This can mean engaging with many other forms of information. For example, analyses of twentieth-century English writing from many genres might show that over time people appear to be using second person pronouns more often.[1] Such a finding could be shown via analysis of frequency and keyword lists. Dispersion analyses may indicate that such pronouns are reasonably well dispersed over different registers of writing, although seem to have especially become more frequent over time in informational and official texts. Further analysis of context via reading concordance lines may indicate that they seem to be used to indicate a personal relationship between author and reader. However, such findings would need to be positioned in relationship to social context – what do we know about social developments in the twentieth century? Can phenomena like a move towards relaxed and more informal social conventions, a tendency to denote a less hierarchical style of address, a desire to make language more accessible or even increased use of persuasive language due to the capitalist imperative to position everyone as a consumer help to explain our finding about pronouns? If the aim of our research is to be critical or inspire social change, then a fourth stage may be more evaluative, pointing out the consequences of such uses of language (asking 'who benefits?' or who is potentially disempowered), perhaps making recommendations for good practice.

Corpus analysis does not need to critically evaluate its findings, and we argue that 'curiosity'-based (as opposed to 'action'-based) research has an important role to play in linguistics. Despite all of the chapters in this collection of corpus studies being positioned as research on discourse, and all of them engaging with description and interpretation stages, some move into explanation and critical evaluation too. This is due to the fact that there is more than one way of doing discourse analysis, as the following section will show.

Discourse analysis

> 'Bid me discourse, I will enchant thine ear'
>
> (Shakespeare, Venus and Adonis)

Copyrighted material – 978–1–137–43172–1

Copyrighted material – 978–1–137–43172–1

Somewhere between Shakespeare's uplifting use of the word, and today, the word *discourse* has suffered something of an identity crisis. While the term *language* is largely understood to non-linguists, *discourse* can be an excluding shibboleth which does little to make academic research accessible or relevant to people who do not work or study in the social sciences. Part of the problem is that even among social scientists the term has a wide set of overlapping meanings. Compare the claim by Stubbs (1983:1) that discourse is 'language above the sentence or above the clause' with Fairclough (1992: 8) 'Discourse constitutes the social … Discourse is shaped by relations of power, and invested with ideologies.'

And within this edited collection, an examination of some of the collocational patterns of *discourse* is revealing of its multiplicity of meanings. Sally Hunt (Chapter 13) refers to *gendered discourses* and *discourse prosody*, Jack Hardy (Chapter 8) uses *discourse community* (as do we in Chapter 12), Karin Aijmer (Chapter 5) analyses *discourse markers*, Dan McIntyre and Brian Walker (Chapter 9) refer to *discourse presentation*, while Daniel Hunt and Kevin Harvey (Chapter 7) mention *medicalising discourse*. As many of the chapters utilise somewhat different understandings of discourse, it is pertinent to ask what they have in common. One answer is that they broadly undertake to examine 'language in use' (Brown and Yule, 1983), a concept which is ideally suited to the corpus linguistic undertaking to base analysis on large collections of naturally-occurring language. In its highest sense then, all of corpus linguistics is discourse analysis.

Therefore, the chapters in this book were chosen because they demonstrate the range of different conceptualisations of discourse that corpus linguists have utilised, indeed Daryl Hocking (Chapter 10) works with two definitions of *discourse*, one following Candlin (1997) as relating to the semiotic resources used by people to carry out practices that shape their professional, institutional and social worlds, the other based on resources used to represent practices or objects.

In Chapter 3 Svenja Adolphs, Dawn Knight and Ronald Carter view discourse in the sense of being all forms of 'language in use' while others more closely associate discourse with genres or registers of language use – so this could be used to refer to *spoken discourse* (Karin Aijmer in Chapter 5) or *digital discourse* (Dawn Knight in Chapter 2). Linked to this notion of discourse are more specific subdivisions, such as *American presidential discourse*, which Cinzia Bevitori (Chapter 6) characterises as a sub-category of *political discourse*. American presidential discourse would cover language used by American presidents, presumably in public settings (e.g. speeches, press releases, interviews). Bevitori also refers to *environmental discourse*, which could be viewed as language around the topic of the environment, and such a topic could potentially occur across a range of different genres or registers of language. However, other chapters, particularly those towards the end of this collection, conceptualise discourse from a more Foucauldian

Copyrighted material – 978–1–137–43172–1

Copyrighted material – 978–1–137–43172–1

perspective, where discourses are seen as ways of looking at the world, of constructing objects and concepts in certain ways, of representing reality in other words, with attendant consequences for power relations e.g. involving gender (Sally Hunt in Chapter 13), ethnicity (Alan Partington in Chapter 11, Amanda Potts in Chapter 14) or social class (Paul Baker and Tony McEnery in Chapter 12). Three of these four chapters follow a critical discourse analysis framework in that research has been carried out in order to highlight inequalities around the ways that certain groups are represented.

An issue with traditional methods of critical discourse analysis relates to the ways that texts and features are chosen for analysis, with Widdowson (2004) warning that 'cherry-picking' could be used to prove a preconceived point, while swathes of inconvenient data might be overlooked. The principles of representativeness, sampling and balance which underline corpus building help to guard against cherry-picking, while corpus-driven techniques like keywords help us to avoid over-focussing on atypical aspects of our texts. Corpus techniques can thus reassure readers that our analysts are actually presenting a systematic analysis, rather than writing a covert polemic.

However, an advantage of corpus-driven approaches means that techniques intended for objectively uncovering the existence of bias or manipulation in language can also be carried out from a discourse analysis perspective where the aim is not necessarily to highlight such problems. Alan Partington's chapter, for example, examines representations of Arabs in press articles but the investigation is not based on an expectation that problematic representations are necessarily 'out there' to be uncovered. Partington instead takes a more prospecting approach, bearing in mind that in terms of news values, negative reporting is to be expected so a distinction needs to be made between *negative* and *prejudiced* representation. Corpus techniques can help us to distinguish between the two, particularly if we make comparisons between different groups or different press outlets. While Potts, Hunt, and Baker and McEnery all position their research as coming from a critical discourse analysis perspective, Partington defines his research as CADS (Corpus-Assisted Discourse Studies) – note the absence of the word *critical*.

Kevin Harvey and Daniel Hunt (Chapter 7) also offer an interesting perspective on corpus approaches to critical discourse analysis. Their chapter examines the online language of people who suffer from eating disorders – but this is not a traditional CDA study that aims to highlight how a powerful text producer unfairly treats a less powerful group. Instead the analysis shows that some people personalise their disorder as 'talking' to them. Harvey and Hunt discuss how such a representation can both help to mitigate the stigma around the illness and provide support to others but it may also constrain understandings that afford more control to the person with the illness. However, in positioning their research as critical, they cite Toolan (2002), who argues that a critically motivated analysis can focus on discourses that are simultaneously enabling and disempowering. The point

Copyrighted material – 978–1–137–43172–1

Copyrighted material – 978–1–137–43172–1

we wish to make here is that corpus linguistics is extremely well-placed to enable discourse analytical research to be carried out from a range of different 'starting positions', depending on the meaning(s) of *discourse* we wish to work with.

The development of a synergy

The relationship between corpus linguistics and discourse analysis has been in development for a quarter of a century, focussed on different groupings over time. The paragraphs below give a vaguely sequential summary of some of the main proponents of what has been referred to more recently as a 'synergy', although it is admittedly brief and thus incomplete; apologies are made in advance to anyone who is missed.

The early work in the field tended to use untagged corpora and was often highly reliant on concordance analyses. Pioneering work was connected to the University of Birmingham in the early 1990s, coming out of early research in corpus linguistics by John Sinclair and taken up by Michael Stubbs, Susan Hunston, Bill Louw, Ramesh Krishnamurthy, Wolfgang Teubert and Carmen Caldas-Coulthard, among others. While corpus research at Birmingham had initially been focussed at the lexical and grammatical levels, an early theoretical concept was that of prosodies. Sinclair (1991) showed how the verb phrase *set in* had a negative prosody, tending to co-occur or collocate with negative associations like *rot*. While *set in* has no intrinsically negative meaning in itself, it is hypothesised that people unconsciously remember the contexts that they have heard it in the past and then will use it themselves in similarly negative ways. Louw referred to this phenomenon as semantic prosody, defining it as the 'consistent aura of meaning with which a form is imbued by its collocates' (Louw, 1993: 157). Semantic prosodies could be exploited for ironic effect, and Louw gives a droll example from a novel by David Lodge where people attending a conference are described as 'austerely bent on self-improvement'. While *self-improvement* usually has positive associations, the phrase *bent on* is regularly used to describe negative behaviours, allowing the author to signal a somewhat different attitude towards the people he is writing about. Semantic prosodies, identified through corpus techniques, are thus an effective way of indicating a text producer's underlying stance – the concept was further developed for analysis of ideologies by Stubbs (1996, 2001), who coined a related term *discourse prosody*.

Other early work at Birmingham was more concerned with representation of different identity groups in corpora. For example, Caldas-Coulthard (1995) carried out a study of gender representation in news stories, indicating a gender bias that was heavily skewed in favour of men, while Krishnamurthy (1996) studied the contexts of identity words like *tribal*, *race* and *racial*. He concluded that 'tribal clearly has pejorative connotations,

Copyrighted material – 978–1–137–43172–1

Copyrighted material – 978–1–137–43172–1

and if we continue to use it, and apply it only to certain groups of human beings, we are merely recycling the prejudices that the English-speaking culture has developed with regard to those groups.' (ibid: 197). Rather than examining a particular social group, Teubert (2001) examined the discourse around Euro-scepticism in Britain, focussing on how subliminal messages were created through the repetition of 'stigma' and 'banner' keywords, the former including *bureaucrat*, *corruption* and *federal*, while the latter involved *independence*, *peace* and *prosperity*. Finally, Hunston's corpus-based work on evaluation (2004, 2011), emerging from her research on pattern grammar (Hunston and Francis, 1999), has also been influential in laying the groundwork for later discourse-based research.

While holding a visiting position at Lancaster University, Gerlinde Hardt-Mautner published the first paper that aimed to describe the potential of combining a corpus linguistics approach with critical discourse analysis, a form of (mostly qualitative) analysis that was popularised by Norman Fairclough at Lancaster (see Hardt-Mautner, 1995). Tony McEnery and Paul Baker were also influenced by Fairclough's approach, developing methods for corpus-based discourse analysis which relied on analyses of keywords and tagged data, see for example, Baker (2005) on the representation of gay men in different sets of publically available texts and McEnery's (2005) corpus analysis of swearing, both in terms of its use and attitudes around it. Baker (2006) published a second monograph illustrating how corpus techniques could be employed in discourse analysis of holiday brochures, parliamentary debates and newspaper articles.

In 2008, we collaborated with a team of critical discourse analysts led by Ruth Wodak, who had co-developed the Discourse Historical Approach to CDA, in order to carry out a large-scale study into the representation of refugees in the British press (see Baker et al., 2008; Gabrielatos and Baker, 2008). This was followed by a similar study examining how the press presented Muslims and Islam (Baker et al., 2013). The research coming out of Lancaster has thus tended to be more closely connected to schools of critical discourse analysis, as well as attempting to incorporate analysis of different types of social, historical and political context in order to explain findings.

Developing concurrently is a third approach, headed by Alan Partington at the University of Bologna, called Corpus Assisted Discourse Studies (CADS), as noted. Focussing on political and press registers, CADS takes a less overtly critical stance to analysis (see above), and has involved analysis of seemingly innocuous abstract concepts like *science* (Taylor, 2010) and *moral* (Marchi, 2010). CADS was also an early pioneer of diachronic corpus research, studying changes in representations in different years of newspapers, for example (Clark, 2010), while another methodological innovation involved the development of a technique called concordance keywords (Taylor, 2010). A handbook-length book devoted to CADS methods was published in 2013 by Partington et al.

Copyrighted material – 978–1–137–43172–1

Copyrighted material – 978–1–137–43172–1

Currently, further advances in the combination of discourse analysis and corpus linguistics are being made at the University of Nottingham, with research centring around multi-modality (see Adolphs and Carter, 2013; Adolphs, Knight and Carter, this volume), and online interaction (see Harvey, 2012; Hunt and Harvey, this volume). There are also numerous researchers working around the world, either independently or in groups looking at corpora and discourse analysis, some of whom are represented in this collection.

Three debates

As with any emerging field, and particularly one which combines elements of different fields together, the use of corpus methods to carry out discourse analysis has instigated a number of challenges and ongoing discussions among its practioners, of which our thoughts on three are discussed below: covering bias, 'so what' findings and ethics/copyright.

The issue of bias is contentious, being linked to researcher values and orientation. It can be argued that a corpus approach allows research to be carried out from a 'naïve' perspective, so the pre-existing views and prejudices of the analyst do not interfere with the early stages of analysis. It is overly hopeful to expect any analyst to approach any topic objectively, and post-structuralists would point to the myth of the neutral researcher, even within the sciences (see Burr, 1995). Accepting this limitation, we would hope that a corpus approach would at least limit the extent of some of our biases – a keyword list produced by a computer is not biased in the way that humans are. It simply works by identifying an ordered list of statistically salient words that the human analyst then needs to account for, many of which would not have been foreseen as key by a human in advance. So far so good. However, the keywords procedure itself suffers from other forms of less conspicuous bias. For one, it is a method that focusses on difference – word *x* in corpus *a* is relatively more frequent than word *x* in corpus *b*. However, if we only focus on a few words that have strong differences in frequencies between two corpora, we may fail to acknowledge words that are reasonably similar in frequency for both corpora – indeed, a more interesting analysis might be one which remarks on both similarities *and* differences. There are workarounds to take similarity into account (such as using a third reference corpus as a form of triangulation), but viewing a keyword technique as unbiased is only true to a certain extent. As well as being intrinsically biased towards identifying differences, the process itself is not immune from (biased) researcher interference. Humans need to decide when and how to employ it, and as there are numerous settings that can be altered (such as the statistical test used to calculate keyness, the minimum number of times a word must appear in a corpus in order for it to be candidate as key, or the cut-off for statistical significance), it is highly likely that two independent

Copyrighted material – 978–1–137–43172–1

Copyrighted material – 978–1–137–43172–1

researchers working on the same corpus will produce different keyword lists. Even then, especially when working with large corpora of millions of words, it is likely that hundreds of words will emerge as key, probably far too many for an analyst to do justice to. Therefore, the researcher needs to make choices about which keywords should receive the most attention – a process that can end up being somewhat subjective as certain keywords may 'jump out' as being more interesting than others. Similar accusations can be levelled at other techniques like collocates, wordlists or even concordance analyses. Concordance analyses can be particularly subjective, and underline the fact that much corpus-based discourse analysis is actually qualitative in nature. Two researchers may draw very different conclusions from looking at the same set of concordance lines. Ultimately, rather than viewing corpus linguistics as *problematically* biased, it is more helpful to accept that there is no such thing as unbiased human research (and that such a goal may not necessarily be attractive in any case), but instead aim for wider transparency about methodological decisions and a more nuanced set of stated claims about the benefits of using computational methods.

A related issue with a corpus analysis of discourse involves questions about what an analysis should actually tell us. Just as the press are governed by 'news values' (Galtung and Ruge, 1965) which include features like timeliness, eliteness, superlativeness, proximity, negativity and novelty, academics are, to an extent, influenced by research values which help to determine which types of research attract funding and also which findings may be prioritised by researchers in their reports. A research finding which causes the reader to splutter 'So what? I knew that already!', is generally seen as less valuable than one which causes a response of 'I did not know that' or 'That goes against what I expected'. In analysing so much language data at once, a common predicament faced by many researchers, especially in the early stages, is that the analysis produces numerous 'so what' findings. Alternatively, it can be easy to get so caught up in our own research topic that we do not realise that some of our findings may not be so earth-shattering to others. For example, with our research on the representation of Islam in the British press (see Baker et al., 2013), it is fairly unsurprising to report that many newspapers, especially tabloids and/or conservative papers reported in ways which appeared negatively oriented towards Muslims. Yet it is easy to identify front pages of newspapers which contain headlines and images that could be interpreted as Islamophobic, without carrying out a corpus study.[2] And it would be 'so what' to say that large-scale terrorist attacks like 9.11 and 7.7 caused huge spikes in frequencies of news stories that referred to Islam, as we found in our corpus analysis. So is there any value in such a corpus-based study? We would argue that there are several ways that such a corpus study is worth carrying out. First, it at least provides a more credible, large-scale grounding for making what looks like an obvious claim. A handful of biased front page headlines may be extremely

Copyrighted material – 978–1–137–43172–1

Copyrighted material – 978–1–137–43172–1

salient, but they may not represent the bulk of daily reporting over a period of years. Being able to draw conclusions based on extremely large samples of data adds validity to claims, even if they confirm what we suspected, while providing a quantitative summary gives substance to what may have been a suspicion. So we may not be surprised to be told that Muslims are referred to as extremist more than moderate in the British press, but how widespread is this practice? Knowing, for example, that mentions of Muslims who are extremists outnumber mentions of moderates by 100 to 1 might prompt a different response to being told that the ratio is only 2 to 1 (in fact it is 9 to 1; Baker et al., 2013: 265).

Second, a corpus analysis can indicate more subtle and insidious ways that an obvious outcome is realised, it can reveal the 'tricks of the trade' in other words. Techniques like keywords, collocates and concordances help to give a much more detailed insight into the workings of language in use. For example, in Sally Hunt's analysis of representation of gender in the popular Harry Potter series of books (Chapter 13), it is not hugely surprising to find that female characters are represented in arguably a more restricted and less empowered way than male characters. However, Hunt demonstrates how such representations are embedded within repetitive patterns around seemingly innocent body words like *feet*, *hand*, *fingers*, *arm*, *shoulder*, *head* and *face*. What the male and female characters are shown to be doing with parts of their bodies (e.g. what they carry on their shoulders or how they use their hands to silence others or indicate a desire to communicate), illustrates an important and possibly subconscious way in which agency is constructed very differently for the two sexes in these books. Such patterns are much less obvious than more overt cases of gender bias such as the use of a sexist label or a character making an on-record remark about the differences between boys and girls.

And thirdly, a corpus analysis can reveal findings that are genuinely surprising, going against our expectations. For example, Dawn Knight (Chapter 2) indicates an unexpected (to us), finding around the use of modal verbs in online language production (e-language). Emails, text messages, tweets, blogs etc. are sometimes viewed as a kind of hybrid form of writing and speech, and Knight hypothesises at the start of her analysis that modal verb frequency in e-language would be higher than written discourse but less frequent than speech, coming somewhere between the two. However, e-language actually contains *more* use of modals than both writing and speech, a finding which goes against expectations and needs to be explained by consideration of the ways that people communicate in speech, writing and e-language.

In any case, we should be cautious in dismissing any finding because we think it is 'so what'. People retain a remarkable capacity for holding variant beliefs and in any case, much 'shared knowledge' is actually very specifically associated to a particular culture and time period. There is value in

Copyrighted material – 978–1–137–43172–1

Copyrighted material – 978–1–137–43172–1

producing academic research which will serve as an historical account for future generations.

Moving on, we come to an ongoing debate surrounding copyright and ethics within corpus linguistics which is made doubly problematic due to the 'discourse' or 'critical' nature of the research in this collection. Many qualitative forms of textual research are not as hampered by concerns over copyright, due to the fact that smaller amounts of text are collected and analysed, and this can often be labelled as 'fair use'. However, corpus analyses usually require much larger datasets which can raise questions about copyright. Many early corpus building projects devoted large amounts of time and money to securing copyright clearance for every text that was included in their corpus – but such a model is not ideal and some corpus linguists would argue that it should be unnecessary to go to such lengths in order to carry out non-profit making research which would only involve the brief quotation of a few concordance lines of text in any case. A recent document by the UK Intellectual Property Office noted a new copyright exception to non-profit research which involves text and data mining. This 'allows researchers to make copies of any copyright material for the purpose of computational analysis if they already have the right to read the work (that is, work that they have "lawful access" to). They will be able to do this without having to obtain additional permission to make these copies from the rights holder' (Intellectual Property Office, 2014: 6). Such an exception only applies to research within the UK though, and copyright rules differ from country to country.

Yet even if some governments are making it easier for 'text mining' research to be carried out, there are ethical problems which involve gaining the respect of the academic community you work in that could present a challenge to one's individual moral values and could result in unintended consequences if not heeded. It is important to secure permission to analyse 'private' uses of language (e.g. personal conversations, privately sent letters or emails, etc.) but how do we treat online data, of the type which the author makes public: tweets, blogs, newsgroups, comments sections of newspapers etc.? On the one hand it could be argued that someone who posts a personal opinion, narrative or attack is 'fair game' to be included in a corpus because anyone can potentially read their post – it is to all intents and purposes, public. However, researchers working in the social sciences often want to ensure that they treat the people they study with sensitivity. Reid (1996) notes that it is doubtful that people who post online messages were intending to have their writing appear in a different public domain, and therefore a blogger may not have considered that the language in their blog posts would come under scrutiny in an academic journal, read by a different sort of audience than they had envisaged. Someone posting an offensive tweet might be a child and drawing attention to them may have unintended consequences for that poster. But asking permission of the hundreds of people who have

Copyrighted material – 978–1–137–43172–1

Copyrighted material – 978–1–137–43172–1

unwittingly contributed towards your corpus may be very time-consuming and difficult (many of them may have out-of-date contact information). It could also result in a skewed corpus if we have to remove all the people who do not give consent. To an extent, if an analysis focusses on decontextualised language use and we stick to reporting frequencies of words, then this is less problematic. But quoting (even snippets) of text can result in a conflict between copyright and ethics. Do we attribute 'authorship' of a post or do we anonymise the identity of the poster to protect their identity, knowing that in some cases an online search of the text we quote in our analysis may reveal the poster's identity anyway?

We do not believe this is a debate that can be easily resolved, and there are many factors which mean that a single solution cannot be applied to all cases. In our own chapter, which focusses on an analysis of tweets, we made the decision not to quote usernames of tweeters or to quote any tweets which advocated violence, but had we used a different corpus, we may have decided on a different way of dealing with ethical issues. Debate is an important part of any developing field, and it is likely that the issues we have raised in this section will continue to provide discussion with regard to best practice in the coming years.

Outline of this volume

The remaining chapters in this book are broadly divided into three parts. Part 1 (Chapters 2–5) considers discourse as related to *modes* or text types (CMC, multimodal texts, mediated texts and spoken texts). We have tried to focus these early chapters on newer and/or under-researched forms of text, going beyond the analysis of the written word in order to discuss the new challenges that such texts bring with them for discourse analysis, and how such challenges can begin to be met with corpus approaches. Part 2 (Chapters 6–10) more broadly considers discourse as related to *social practice*, and consequently this group of chapters deals with environmental discourse, health discourse, academic discourse and news discourse (from an historical perspective). Part 3 (Chapters 11–14) relates discourse more closely to *ideology* or attitude, and here we are concerned with discourse as a means of representing a social identity or concept through language use. Partington's study covers the field of Corpus Assisted Discourse Studies while the other three chapters more broadly follow a critical discourse analysis perspective. We should note though, that many of the chapters in this collection draw on multiple notions of discourse, so our categorisation system could have been carried out differently and is not as neat as first appears.

Beginning with Part 1, in Chapter 2, Dawn Knight examines use of modal verbs in the Cambridge and Nottingham e-Language Corpus (a relatively new corpus containing text from blogs, discussion boards, emails, text messages and tweets). Comparing individual and joint modal verb frequencies

Copyrighted material – 978–1–137–43172–1

Copyrighted material – 978–1–137–43172–1

both within the different e-language registers and against other spoken and written corpora, she finds differences at every level of comparison – modals are particularly unpopular in tweets but very common in emails and text messages, and as discussed above, surprisingly, modals were more frequent in e-language than both writing and speech. Knight relates the presence of modals to the extent to which the intention is to communicate to a wider (often unknown) audience or whether the audience is small and specific. In addition she notes that compared to speech, e-language has a crucial inadequacy – the lack of gestural, paralinguistic and extra-linguistic cues that are used to communicate meaning. A higher reliance on modality in e-language may be for compensatory reasons.

Svenja Adolphs, Dawn Knight and Ronald Carter demonstrate developments in multimodal corpus research where 'non-linguistic' data-streams are incorporated into an analysis of a small spoken corpus (Chapter 3). In this preliminary case study the location where speech takes place is used as a driving factor to examine spoken language use during a series of visits to galleries. Unlike the previous chapter, no hypothesis was developed prior to the research and instead an analysis of frequent words drives the analysis towards use of deixis. The study also highlights some of the potential difficulties encountered when building spoken corpora (e.g. not all of the participants visited every location as expected, batteries ran out and the GPS signal was lost), indicating some of the unforeseen practical complications that occur when moving beyond the analysis of written corpora. The analysis gradually focusses in on the use of evaluative language, noting that while inside the galleries people made more frequent use of *like* as an evaluative, while outside the galleries *like* was used more often as a quotative, with the evaluative *love* being more frequent instead. The analysis thus shows up unexpected directions in analysis, demonstrating neatly how context relates to language use.

In Chapter 4, Monika Bednarek examines issues that arise when corpus linguists study another type of multimodal text – film and television. Bednarek is not just concerned with the written scripts but in the ways that this audio-visual medium uses narrative, body language, visual communication, kinesics and proxemics as part of the discourse. After discussing how such corpora can be created or obtained, with particular consideration of transcription issues, Bednarek gives a case study involving the analysis of a corpus of scripts from the American serial *Nurse Jackie* that were created by a fan of the programme. A keyword analysis shows how past tense forms like *joined* indicate the transcriber's unfamiliarity with scripting conventions while incorrect pronoun use also suggests that such fan scripts may be of limited value for the investigation of multimodality in television and film. Following that, Bednarek carries out a multimodal analysis of a single scene from the series, showing how visuals and verbal text are combined in the programme in order to create meaning. Camera focus on facial expression

Copyrighted material – 978–1–137–43172–1

Copyrighted material – 978–1–137–43172–1

is often required in order to interpret elements of the script that would otherwise be mystifying or misinterpreted by a 'reader'. Additionally, use of eye movement and eye contact contribute towards characterisation, indicating one of the many ways in which a transcript of spoken words alone would result in a partial (and thus somewhat deficient) analysis of this medium.

Karin Aijmer also looks at spoken language data, although rather than considering the more stylised and scripted mode of television drama, in Chapter 5 she examines a corpus of naturally occurring speech, focussing on how discourse markers are frequently employed by speakers in order to guide the hearer to the interpretation of an utterance. Considering the multifunctional nature of discourse markers, Aijmer notes how they are of particular value for corpus linguists who are interested in speech, providing a case study of the marker *actually*, comparing it across corpora of speech from different varieties of English and examining its position and functions. While Aijmer concludes that the various meanings of *actually* can be derived from a core meaning, she notes variation in terms of language variety and position. For example, occurring in the right periphery it corrects a preceding claim while in the left periphery it introduces a shift in perspective. In Hong Kong and Singaporean English it has subjective meanings, relating to showing how an utterance should be understood or viewed as relevant, whereas for Great Britain and New Zealand English it is intersubjective, pointing backwards to an earlier claim to correct it. Aijmer's analyses of this discourse marker indicate the extent to which corpus analysis needs to be qualitative, going beyond the concordance line to examine sometimes quite lengthy extracts of dialogue in order to identify meaning.

Part 2 of the collection focusses more on discourse as social practice and begins with Cinzia Bevitori's chapter on environmental discourse (Chapter 6). Taking a diachronic approach, Bevitori examines a corpus consisting of speeches made by ten American presidents, focussing on their language around the environment. Taking the word *environment* (and its related forms) as the starting point of the analysis, Bevitori gradually shifts to focus on collocates such as *protect*, *energy* and *clean*. By aligning the results of the corpus analysis to consideration of different social and political contexts, Bevitori is able to show how meanings of *environment* are not stable but vary according to changing political priorities, events (such as the 1973 oil crisis, the state of the economy or global conflicts) and world views of those in power. In particular she demonstrates how three concepts, environmental protection, energy conservation and cleanliness have gradually become more intertwined over time.

Daniel Hunt and Kevin Harvey in Chapter 7 consider health discourses, looking at how people communicate in two online contexts when discussing eating disorders, specifically anorexia. Using keywords, they identify lexis that signpost the various ways that eating disorders are framed by participants in online discussions. In one general health forum aimed at

Copyrighted material – 978–1–137–43172–1

Copyrighted material – 978–1–137–43172–1

teenagers, keywords like *am*, *I'm*, *anorexic* and *stone* are shown to contribute towards a medicalising discourse which foregrounds disease and deviance. In a second forum which is more specifically devoted to discussion of anorexia, grammatical patterns around keywords like *ED* (*eating disorder*) are less likely to construct an eating disorder as a possession or trait of an individual but instead they tend to be constructed as independent entities, somewhat separate from their sufferers. The eating disorder is granted grammatical agency: 'it is totally ed using your voice to express itself'. By contrasting the two forums, Hunt and Harvey are able to identify a set of distinctly different discourses around eating disorders, concluding with a discussion of how such discourses contribute towards the potential (dis)empowerment of those who have internalised or come into contact with them.

In Chapter 8, Jack Hardy studies members of a particular discourse community, university students who engage in the practice of writing academic discourse. Hardy is interested in the extent to which language use among community members is affected by both the amount of time spent among the community and also whether specific sub-communities (based on topic of study) are likely to differ from one another. Using a corpus approach popularised by Doug Biber called Multi-Dimensional Analysis, Hardy analyses the content of student writing to identify how they differ according to four dimensions. For example, for the first dimension considered (involved academic narrative vs. descriptive informational discourse), philosophy and education students tended to use more features associated with involved narratives. These include first and third person pronouns, *wh-* words, *to-*clauses and various verbs. However, biology and physics students tended to skew towards the polar opposite of this dimension, using more features associated with descriptive informational discourse: nouns, adjectives and longer words. A notable difference was also found between final year undergraduates and first year graduates, with the former being more 'involved' and the latter being more 'informational', although in their second and third years graduate students move back towards being 'involved' again. Similar patterns were found for the other three dimensions examined, suggesting that the move from undergraduate to graduate status leads to a somewhat extreme difference in writing style, which gradually edges back towards 'middle ground', perhaps due to an over-compensatory initial set of expectations about what 'graduate' writing ought to be like.

Dan McIntyre and Brian Walker take a different perspective on discourse in Chapter 9 which considers discourse presentation in a corpus of Early Modern English news reports. Using a hand-categorised dataset, the authors compare presentation of speech, writing and thought in Early Modern and Present Day English, finding higher frequencies of writing and thought presentation in the Early Modern corpus. By conducting qualitative studies of the parts of the corpus which contain such cases and relating them to the context of writing for news during the periods under study, McIntyre

Copyrighted material – 978–1–137–43172–1

Copyrighted material – 978–1–137–43172–1

and Walker are able to provide credible explanations for their findings. For example, particular patterns in news reporting may have been due to the higher penalties for publishing dissenting voices in a period before press freedom was established, or may be linked to the ideological stance of certain authors, who wished to use their articles as propaganda. Particularly interesting are cases where journalists report on the thoughts or emotional states of the people who they are writing about, people who were sometimes living in different countries to the journalist. McIntyre and Walker point to this phenomenon as contributing towards a fictionalising and thus more dramatic effect of news, indicating how much (or little) journalistic discourse practices have changed, compared to the present day.

The focus in Chapter 10 is on a multi-perspectival analysis of how creative practice occurs in a tertiary art and design study. Darryl Hocking uses a range of text types in order to explore this topic including student briefs, transcripts of studio tutorials and informal studio interactions. Accordingly, the corpus analysis was a component of a larger study which involved ethnographic analysis, metaphor analysis and discourse-historical analysis. Initial read-throughs of the dataset identified the word *ideas* as particularly salient (and a subsequent frequency and keyword analysis also showed the word to be important in some of the datasets examined). The tool NVivo was then used to categorise how participants characterised ideas in the dataset, leading to an analysis which combines qualitative examination of sections of transcribed talk with concordance analyses of relevant collocational pairs such as *your ideas*. The findings are also interrogated in relation to socio-historical context, with Hocking drawing on a discussion of how the concept of *ideas* has been characterised by other leading artists over time, helping to explain the discourses around the word that were found in the corpus under study.

The final part of the collection (categorising discourse more generally as representations) begins with Alan Partington's Corpus Assisted Discourse Analysis of representations of the Arab world in a number of English language newspapers from the UK, the United Arab Emirates and Egypt. A diachronic aspect of the analysis compares patterns found in 2010 (prior to the revolutionary wave of demonstrations and protests widely referred to as the Arab Spring) and 2013. In Chapter 11 Partington examines patterns around grammatical agency to show how a collectivised phrase like the *Arab world* is represented in 2010 as an active subject in material processes as well as a senser in mental processes. The phrase is frequently found to be cast in the role of an audience, reactive to outside stimuli. In order to test whether this feature is found of other, equivalent constructions, Partington also looks at the *western world*, concluding that of the two, the Arab world is represented as more passive and as an audience/recipient, somewhat hypersensitive and likely to take offence. However, this 'passive audience' representation is much less frequent in the 2013 data. Partington's chapter ends with a discussion on negativity and prejudice in news reporting, noting that negativity is a news

Copyrighted material – 978–1–137–43172–1

Copyrighted material – 978–1–137–43172–1

value and indicating how comparative corpus analyses (such as between different newspapers or across different time periods) can be useful in showing that negativity is markedly frequent and drastic in a particular outlet.

In Chapter 12 Paul Baker and Tony McEnery analyse a corpus of tweets that were made about a controversial television documentary series called *Benefits Street*, which contributed towards a wider debate about social class, inequality and unfairness in the UK. Unlike newspaper data, Twitter data is subject to less overt regulation and can also act as a useful indicator of public feeling. Our analysis began with the categorisation of the top 100 keywords, which were then subjected to collocational and concordance analyses, leading to the identification of three main discourses in the corpus (some historically ancient), along with associated discourse communities who participated in specific linguistic and social practices. While some keywords tended to index a particular discourse (e.g. a collocational network of the keyword *fags* showed that it was generally used to criticise people who received government benefits as wasting their money on cigarettes and other items deemed as non-essentials), other words and phrases were found to have contested meanings, such as *depression* or *people on benefits*. Our chapter ends with a discussion of how the medium of interaction (Twitter) impacts on the ways that discourses are articulated and circulated.

Also taking a critical discourse analysis perspective, in Chapter 13 Sally Hunt examines representations around agency and gender in a corpus of the popular Harry Potter novels written by the author JK Rowling. Rather than focussing on perhaps more traditional lexis (such as frequencies and collocations of words like *boy* and *girl*), Hunt instead looks at the actions that are afforded to male and female body parts, initially noting that many male body parts are more frequently referred to than female body parts (and are also more frequent compared to general British English). However, it is in conducting concordance analyses of words like *hand*, *feet*, *fingers*, *legs*, *face* and *arms* that Hunt is able to present a compelling picture of male bodies that are consistently constructed as more active and more likely to be exposed to danger (but less vulnerable). She concludes that female characters are excluded from having a meaningful impact in the world, an implicit and perhaps subconscious message given off in the book is therefore that agency is not for girls.

Finally, in Chapter 14 Amanda Potts uses a semantically tagged corpus in order to examine how identity is constructed in a corpus of American newspaper articles which referred to the natural disaster Hurricane Katrina. The semantic tagging acts as a helpful technique of data down-sampling, enabling Potts to focus on a smaller number of highly frequent predicational strategies, following the discourse historical approach to critical discourse analysis. Looking initially at collocates of the frequent word *people*, she shows how discourses around race and class intersect, leading to a phenomenon she calls deviancy doubling, whereby othered qualities (such

Copyrighted material – 978–1–137–43172–1

Copyrighted material – 978–1–137–43172–1

as *poor* and *black*) of social actors are compounded together with the effect of distancing such actors from the in-group. Potts notes that in favouring emphasis on certain identity characteristics, other characteristics (such as emotional or physical states) are backgrounded, while there is also less emphasis on more unifying and empowering constructions in the articles. Potts concludes that people affected by Hurricane Katrina are ultimately constructed as threats to social order and overall national welfare.

We hope that readers find this collection of chapters to be illuminating and thought-provoking, offering a range of contemporary perspectives on the combination of corpus linguistics and discourse analysis, a field which we feel has come of age with the publication of this volume.

Notes

1. An examination of the four British members of the Brown family (written corpora of 1 million words each from 1931, 1961, 1991 and 2006) finds the combined frequencies of *you*, *your*, *yours*, *yourself* and *yourselves* to be 4,654, 5,090, 5,069 and 6,237 respectively.
2. For example, a Google search of the words *islamophobic British press* retrieves images of British newspapers with front page headlines like 'Muslims tell us how to run our schools', 'Ramadan a ding-dong', 'Muslims tell British: Go to hell' and 'BBC put Muslims before you'.

References

Adolphs, S. and Carter, R. A. (2013) *Spoken Corpus Linguistics: From Monomodal to Multimodal* (London: Routledge).

Baker, P. (2005) *Public Discourses of Gay Men* (London: Routledge).

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

Baker, P., Gabrielatos, C. and McEnery. T. (2013) *Discourse Analysis and Media Attitudes: The Representation of Islam in the British Press* (Cambridge: Cambridge University Press).

Baker, P., Gabrielatos, C., Khosravinik, M., Krzyzanowski, 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.

Brown, G. and Yule, G. (1983) *Discourse Analysis* (Cambridge: Cambridge University Press).

Burr, V. (1995) *An Introduction to Social Constructionism* (London: Routledge).

Caldas-Coulthard, C. R. (1995) 'Man in the news: The misrepresentation of women speaking in news-as-narrative-discourse', in S. Mills (ed.) *Language and Gender: Interdisciplinary Perspectives* (Harlow: Longman), pp. 226–39.

Candlin, C. N. (1997) 'General editor's preface', in B. Gunnarsson, P. Linell and B. Nordberg (eds) *The Construction of Professional Discourse* (Harlow: Addison Wesley Longman), pp. viii–xiv.

Clark, C. (2010) 'Evidence of *evidentiality* in the quality press 1993 and 2005', *Corpora* 5(2): 139–60.

Fairclough, N. (1992) *Discourse and Social Change* (Cambridge: Polity Press).

Copyrighted material – 978–1–137–43172–1

Copyrighted material – 978–1–137–43172–1

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.

Galtung, J. and Ruge, M. H. (1965) 'The structure of foreign news', *Journal of Peace Research,* 2(1): 64–91.

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

Harvey, K. (2012) 'Disclosures of depression: Using corpus linguistics methods to interrogate young people's online health concerns', *International Journal of Corpus Linguistics,* 17(3): 349–79.

Hunston, S. (2004) 'Counting the uncountable: Problems of identifying evaluation in a text and in a corpus', in A. Partington, J. Morley and L. Haarman (eds) *Corpora and Discourse* (Bern: Peter Lang), pp. 157–88.

Hunston, S. (2011) *Corpus Approaches to Evaluation: Phraseology and Evaluative Language* (London: Routledge).

Hunston, S. and Francis, G. (1999) *Pattern Grammar: A Corpus-Driven Approach to the Lexical Grammar of English* (Amsterdam: Benjamins).

Intellectual Property Office (2014) *Exceptions to Copyright: Research*. https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/375954/Research.pdf.

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* (London and New York: Routledge), pp. 129–49.

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* (Amsterdam: Benjamins), pp. 157–76.

Marchi, A. (2010) '"The moral in the story": A diachronic investigation of lexicalised morality in the UK press', *Corpora* 5(2): 161–89.

McEnery, T. (2005) *Swearing in English: Bad language, purity and power fro 1586 to the present*. London: Routledge.

Partington, A., Duguid, A. and Taylor, C. (2013) *Patterns and Meanings in Discourse: Theory and Practice in Corpus-Assisted Discourse Studies (CADS)* (Amsterdam: John Benjamins).

Reid, E. (1996) 'Informed consent in the study of on-line communities: A reflection on the effects of computer-mediated social research', *The Information Society,* 12: 169–74.

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

Stubbs, M. (1983) *Discourse Analysis: The Sociolinguistic Analysis of Natural Language* (Chicago: University of Chicago Press).

Stubbs, M. (1996) *Texts and Corpus Analysis* (London: Blackwell).

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

Taylor, C. (2010) 'Science in the news: A diachronic perspective', *Corpora,* 5(2): 221–50.

Teubert, W. (2001) '"A province of a federal superstate, ruled by an unelected bureaucracy": Keywords of the Euro-sceptic discourse in Britain', in C. Good, A. Musolff, P. Points and R. Wittlinger (eds) *Attitudes Towards Europe* (Abingdon: Ashgate), pp. 45–88.

Toolan, M. (2002) 'What is critical discourse analysis and why are people saying such terrible things about it', In M. Toolan (ed.) *Critical Discourse Analysis: Critical Concepts in Linguistics Vol. III* (London: Routledge), pp. 218–41.

Widdowson, H. G. (2004) *Text, Context, Pretext. Critical Issues in Discourse Analysis* (Oxford: Blackwell).

Copyrighted material – 978–1–137–43172–1

Copyrighted material – 978–1–137–43172–1

Index

Copyrighted material – 978–1–137–43172–1

Copyrighted material – 978–1–137–43172–1

Copyrighted material – 978–1–137–43172–1

Copyrighted material – 978–1–137–43172–1

Copyrighted material – 978–1–137–43172–1

Copyrighted material – 978–1–137–43172–1

Copyrighted material – 978–1–137–43172–1

Copyrighted material – 978–1–137–43172–1

Copyrighted material – 978–1–137–43172–1

Copyrighted material – 978–1–137–43172–1

Copyrighted material – 978–1–137–43172–1