Routledge Applied Corpus Linguistics

INTERDISCIPLINARY RESEARCH DISCOURSE

CORPUS INVESTIGATIONS INTO ENVIRONMENT JOURNALS

Paul Thompson and Susan Hunston

ROUTLEDGE

Interdisciplinary Research Discourse

Interdisciplinary Research Discourse: Corpus Investigations into Environment Journals provides cutting-edge insights into the nature of communication in interdisciplinary research domains. Using a corpus of nearly 12,000 articles taken from 11 journals, this book addresses the key questions that surround writing for an interdisciplinary audience. This books also explores:

- the ways in which writers write if they are writing for an interdisciplinary audience as well as for a specialist disciplinary audience;
- the different natures and instances of the term 'interdisciplinarity'; and
- whether an analysis of the rhetorical contexts in which research is relayed to interdisciplinary audiences is critical to understanding interdisciplinary research activities and communications.

Written by two leading figures in the field of Corpus Linguistics, this is an essential text for researchers and upper-level undergraduates working in the areas of Corpus Linguistics, Discourse Analysis and Linguistics in areas of interdisciplinary communication.

Paul Thompson is Senior Lecturer and Deputy Director of the Centre for Corpus Research at the Department of English Language and Linguistics at the University of Birmingham, UK.

Susan Hunston is Professor of English Language at the University of Birmingham, UK.

Routledge Applied Corpus Linguistics

Series Editor: Michael McCarthy

Michael McCarthy is Emeritus Professor of Applied Linguistics at the University of Nottingham, UK, Adjunct Professor of Applied Linguistics at the University of Limerick, Ireland and Visiting Professor in Applied Linguistics at Newcastle University, UK. He is co-editor of the Routledge Handbook of Corpus Linguistics, editor of the Routledge Domains of Discourse series and co-editor of the Routledge Corpus Linguistics Guides series.

Series Editor: Anne O'Keeffe

Anne O'Keeffe is Senior Lecturer in Applied Linguistics and Director of the Inter-Varietal Applied Corpus Studies (IVACS) Research Centre at Mary Immaculate College, University of Limerick, Ireland. She is co-editor of the Routledge Handbook of Corpus Linguistics and co-editor of the Routledge Corpus Linguistics Guides series.

Series Co-Founder: Ronald Carter

Ronald Carter (1947–2018) was Research Professor of Modern English Language in the School of English at the University of Nottingham, UK. He was also the co-editor of the Routledge Corpus Linguistics Guides series, Routledge Introductions to Applied Linguistics series and Routledge English Language Introductions series.

Editorial Panel: IVACS (Inter-Varietal Applied Corpus Studies Group), based at Mary Immaculate College, University of Limerick, is an international research network linking corpus linguistic researchers interested in exploring and comparing language in different contexts of use.

The Routledge Applied Corpus Linguistics Series is a series of monograph studies exhibiting cutting-edge research in the field of corpus linguistics and its applications to real-world language problems. Corpus linguistics is one of the most dynamic and rapidly developing areas in the field of language studies, and it is difficult to see a future for empirical language research where results are not replicable by reference to corpus data. This series showcases the latest research in the field of applied language studies where corpus findings are at the forefront, introducing new and unique methodologies and applications which open up new avenues for research.

Other titles in this series

Corpus Perspectives on the Spoken Models used by EFL Teachers
Angela Farrell

Corpus Linguistics and the Analysis of Sociolinguistic Change
Language Variety and Ideology in Advertising
Joan O'Sullivan

IVACS
Inter-Varietal Applied Corpus Studies

More information about this series can be found at www.routledge.com/series/RACL

Interdisciplinary Research Discourse

Corpus Investigations into Environment Journals

Paul Thompson and Susan Hunston

Routledge
Taylor & Francis Group
LONDON AND NEW YORK

First published 2020
by Routledge
2 Park Square, Milton Park, Abingdon, Oxon OX14 4RN

and by Routledge
52 Vanderbilt Avenue, New York, NY 10017

Routledge is an imprint of the Taylor & Francis Group, an informa business

British Library Cataloguing-in-Publication Data
A catalogue record for this book is available from the British Library

Library of Congress Cataloging-in-Publication Data
A catalog record for this book has been requested

ISBN: 978-1-138-06745-5 (hbk)
ISBN: 978-1-003-00934-4 (ebk)

Typeset in Times New Roman
by Apex CoVantage, LLC

Contents

Figures

Tables

Boxes

Acknowledgements

This initial research for this book was conducted on a project called 'Investigating interdisciplinary research discourse: the case of Global Environmental Change' that was funded by the Economic and Social Research Council (UK), August 2013 to November 2015 (ES/K007300/1). The research fellows were Akira Murakami and Dominik Vajn, and their input and insights throughout the project were invaluable. The project was a collaboration with Elsevier, UK, and we are grateful to Andrew Plume, Sarah Huggett and Matthew Richardson of the Elsevier Research team for their assistance and for their interest in the project throughout. We also acknowledge the support given to us by Adam Kilgarriff and the Sketch Engine company for giving us space on the Sketch Engine servers for our corpus.

We are grateful to Ron Carter for encouraging us to write this book for the Applied Corpus Linguistics series and it is a great sadness to us that he did not live to see the book reach publication. To the series editors, Mike McCarthy and Anne O'Keeffe, we offer our gratitude for their guidance and their feedback on the final draft of the book. Last, but by no means least, we would like to thank our colleagues at Birmingham and beyond who, over the last few years, have listened to our talks and have helped us to refine and clarify our thoughts.


Taylor & Francis
Taylor & Francis Group
http://taylorandfrancis.com

1 Introduction

1.1 What this book is about

In the most general terms, this book is about the discourse, in published research articles, of academic research that is interdisciplinary. It uses a variety of corpus investigation methods, some traditional, some innovative, with the twin aims of, firstly, characterising interdisciplinary discourse and, secondly, providing researchers with insights into what helps to make interdisciplinary research discourse successful. This is potentially a massive enterprise, and the book sets a number of limits on its endeavour.

As a case study in interdisciplinary research discourse, the book investigates research articles published in journals related to the study of the environment (see Chapter 3 for a detailed description of the corpus). Articles from four journals over a ten-year period comprise the main corpus used in this book: the Birmingham-Elsevier Environment, or BEE4 corpus. Mention is also made in the book of a larger corpus from 11 journals: the BEE11 corpus. Chapter 2 refers to another corpus compiled of internet texts related to the concept of 'discipline'.

The field of environmental studies was selected as the area of research because it draws on expertise in both the natural sciences that study the natural world (e.g. how plants grow, what happens to forests when rainfall changes, the effect of temperature changes on marine life, how insecticide use affects the wild bird population) and the social sciences that study what human beings do and why (e.g. how farmers respond to changes in the environment, how international agreements affect carbon emissions, what people think about recycling, what economic triggers will reduce energy consumption). Some of the journals in the larger BEE11 corpus are drawn from only one discipline, such as biology or economics, others focus on a fairly restricted topic, such as transport or pollution, others cover a broader spectrum, such as rural or urban systems, or environmental change. The BEE4 corpus consists of journals that represent natural sciences and social sciences, monodisciplinary and interdisciplinary research, and disciplines that are proximal or distal. Table 1.1 shows the four journals and how they relate to each of these distinctions. (See the following discussion and Chapter 3 for a more detailed description of the corpus.)

The book has a number of potential audiences. The primary audience is our fellow corpus linguists. For this community we introduce some innovative

Table 1.1 The BEE4 corpus.

Journal	*Natural/social science*	*Mono-/inter-disciplinary*	*Disciplines proximal/distal*
Agriculture, Ecosystems and Environment	Natural	Inter	Proximal
Global Environmental Change	Social and Natural	Inter	Distal
Plant Science	Natural	Mono	Proximal
Resource and Energy Economics	Social	Mono	Distal

methodologies and also contribute to the discussion of how corpus data is to be interpreted in a discourse context. Our accounts of how the research was carried out presuppose knowledge of corpus methods and thus address our primary audience. Our research has the potential to be of interest also to researchers contributing to interdisciplinary research, and also to teachers of students learning to produce professional academic texts. These issues are not addressed in each chapter, but we return to them in our concluding chapter.

In this opening chapter we give more context for our choice of topic, outline the main themes of the book and suggest what the book may have to offer the fields of academic discourse analysis and Corpus Linguistics.

1.2 Interdisciplinary discourse as a topic for research

There are a number of reasons for studying interdisciplinary research discourse. A major reason is that it is an unusual thing to do: there are a large number of corpus investigations into the discourse of academic research, but most of this focuses on discrete disciplines, often comparing, say, writing in a natural science with that in a social science (for a review, see Hyland 2006; see also Samraj 2008; Durrant 2017). The corpora that underpin studies of this kind tend to be composed of texts that fit securely within a given discipline (e.g. Omidian, Shahriari & Syanova-Chanturia 2018: 4); journals that blur the edges between disciplines, as it were, are avoided. In this study we take the opposite view and, as noted, focus on a field that attracts research from a variety of perspectives.

The previous paragraph makes the assumption that studying research discourse itself is of interest, so it worth expanding somewhat on why this might be so. Many of the early studies in this field had teaching as their main justification; from the 1970s onwards, English for Specific or Academic Purposes was a profession that attracted some of the most creative minds in Linguistics, well placed to carry out innovative investigations into the language they were required to teach. These studies were based on a variety of paradigms, from discourse (Dudley-Evans 1978) and rhetoric (Bazerman 1988) to genre (Swales 1981, 1990; Bhatia 1993) and Corpus Linguistics (Johns 1994). It quickly became apparent that the importance of such research went well beyond classroom needs and added valuable insight into how discourse interacts with community, identity and ideology. In this, academic

discourse acted as a microcosm for discourse as a whole. From the 1990s, it became increasingly clear that how academic writing is done and how research is reported construes disciplinary communities and individual identities (Halliday & Martin 1993; Hunston 1994; Swales 1990, 2004; Hyland 2012). Most significantly, perhaps, understanding the reciprocal relationship between the language and the ideology of scientific and other research paradigms has rivalled in importance the growing understanding of the similar relationship between business, the media and society (Fairclough 1995; Fairclough, Cortese & Ardizzone 2007). Research into academic discourse is therefore not a niche or peripheral activity. Rather it explains how we human beings gain understanding of our (natural, social and conceptual) world and how the language used influences or even constrains that understanding.

The avoidance of interdisciplinary discourse as a topic of linguistic research is difficult to defend in the current research climate because interdisciplinary research is often talked of as the solution to the world's problems. For example, at a local level, the University of Birmingham website publicises the Institute for Global Innovation (www.birmingham.ac.uk/research/global-goals/igi; accessed 20 August 2019) with this introductory sentence:

> At the Institute for Global Innovation (the IGI), our aim is to inspire, support and deliver world-leading, **multidisciplinary research** that addresses the world's most pressing challenges.
>
> (emphasis added)

At a national level, United Kingdom Research and Innovation (www.ukri.org/about-us/strategic-prospectus; accessed 20 August 2019) comments that:

> Realworld problems are inherently **multi- and interdisciplinary**, and we need to . . . [collaborate] across disciplinary, organisational and national boundaries.
>
> (emphasis added)

Beyond national boundaries, a European Union Policy Brief (Allmendinger 2015: 4) states that:

> More often than not, scientific breakthroughs happen to take place at the border of present academic disciplines, or to even go beyond these, thus resulting in an enormous proliferation of **interdisciplinary specialties** and many new ways in the production, dissemination, and use of scientific knowledge.
>
> (emphasis added)

In spite of the consensus that research that crosses discipline boundaries is important, especially if that research is to engage with the solutions to practical problems, it is informally acknowledged to present difficulties. Disciplines are cultures, and bringing people together from different disciplines presents challenges analogous to those involved in, say, international business meetings (Handford 2013). Issues

might include power differentials between the various disciplines and groups involved; different assumptions about what a research agenda consists of; use of different terminology, or use of the same terms with different meanings. As Choi and Richards (2017: 3) point out:

> One of the most basic challenges in interdisciplinary engagement arises from the fact that different disciplines have very different ways of understanding things, dealing with things and representing things.

Researchers who specialise in interdisciplinary projects sometimes use the metaphor of 'speaking a different language' to summarise the challenges involved in these enterprises. This metaphor implies differences that are both profound and superficial. They are profound in that interacting with a speaker of another language involves extensive study (for example, to learn new words for familiar objects) and a willingness to change one's mindset (for example, to make grammatical distinctions that are not made in one's own language). They are superficial in that translation between languages is always possible, even if sometimes not easy: it is assumed that *Harry Potter and the Philosopher's Stone* tells the same story about the same characters in each of its 75+ translations (en.wikipedia.org/wiki/Harry_Potter_in_translation; accessed 20 August 2019).

There is a growing tradition of research into the nature of interdisciplinary research, with conferences, papers and books devoted to the topic (e.g. Frost & Jean 2003; Barry & Born 2013; Choi & Richards 2017). Types of interdisciplinarity are distinguished: terms such as 'multidisciplinarity', 'interdisciplinarity' and 'transdisciplinarity' are used. In this book, we do not make these distinctions, and we use 'interdisciplinary' as a blanket term. What we look for in this research are accounts of interdisciplinarity that help us to make sense of our own data (see Chapter 2 for further discussion). Of greatest significance is that we adapt the proposal by Barry and Born (2013) to distinguish three relationships between disciplines (integrative-synthesis, subordination-service, agonistic-antagonistic) to explicate choices in writing. We note that where the writing represents the disciplinary relationship as integrative, the disciplines are expressed as on an equal footing, each contributing to a joint research effort. The stance expressed towards any potential conflict is conciliatory. Where the relationship is represented as supportive, one discipline is at the service of the other, providing information or methods that the other needs but not on an equal footing with it. Where it is expressed antagonistically, one discipline casts doubt on the workings of the other.

As discussed in Chapter 2, relatively little of the research into interdisciplinary research practices is carried out from a linguistic perspective. Notable exceptions include Teich and Holtz (2009) and Muguiro (2019), both of whom compare single disciplines (such as Education and Neuroscience) with the 'interdisciplines' (such as Educational Neuroscience) that bring the two together. Teich and Holtz (2009) examine the profiles of several lexical items in two disciplines (Linguistics and Computer Science) and their interdiscipline (Computational Linguistics), relating differences in use to the concept of Field as a component of register in Systemic-Functional

Linguistics. Muguiro compares three interdisciplines with each other, adopting Barry and Born's (2013) model. A rather different kind of study of interdisciplinary discourse is reported in Choi and Richards (2017). The authors focus on spoken interactions between researchers from different disciplines. They highlight the challenges of interdisciplinary collaboration and what helps the successful outcome of such interactions. They also offer a helpful reflection on the nature of disciplines and of interdisciplinarity. Following Kellert (2008: 36–39), Strathern (2004: 45) and Turner (2000: 51) they list a number of metaphors used to map out the disciplinary territory, including 'nations' (with hard or soft borders), 'tiles' (interlocking or overlapping), 'languages' (comprising grammars and ways of talking), 'families' (with lineages, inheritances and change over time), and 'cartels' (with a vested interest in maintaining the status quo) (Choi & Richard 2017: 18). The conclusion from these metaphors is that disciplines are both social entities and epistemic ones. To some extent, the differences between them are organisational and contingent; this means that productive conversations between disciplines are hampered by issues such as power differentials and social unfamiliarity. On the other hand, the differences include genuine presuppositions about what research questions are asked, how research is carried out, what counts as evidence and so on; this means that power and unfamiliarity are not the only barriers to productive interaction (Choi & Richard 2017). Studying the language of disciplines, and of disciplines in interaction, allows both the social aspects and the epistemic aspects of disciplinary dialogue to be identified. In the terminology used in Systemic Functional Grammar (Halliday & Martin 1993; see also Teich & Holtz 2009), language both constructs social relations as Tenor and construes the physical and social world as Field.

In the next section we summarise how our research was carried out and the place of previous approaches in it.

1.3 Our approach to interdisciplinary research discourse

The chapters of this book will present a number of studies of the BEE4 corpus (see section 1.1), with some reference to the BEE11 corpus which forms part of the methodology for Chapter 9 especially. An important point to make is that these studies are heuristic, in that we do not set out with an expectation or hypothesis of what we will find. We use the research discussed in section 2 of this chapter as a lens through which to interpret our findings but not as a means of setting an agenda. What we have done, in essence, is to identify traces in the corpus output of the models that characterise interdisciplinary research.

As noted at the beginning of this chapter, our investigations include both traditional (corpus) methods and innovative ones. For the most part the traditional methods are biased towards the qualitative while the newer methods place a greater reliance on the quantitative. Figure 1.1 shows our methods along a qualitative-quantitative continuum. (See section 1.4 for further discussion of the implications of this research for Corpus Linguistics, and vice versa.)

It is relatively simple to compare the various journals in the BEE4 corpus, to demonstrate similarity and difference. It is less simple to tell a story around those

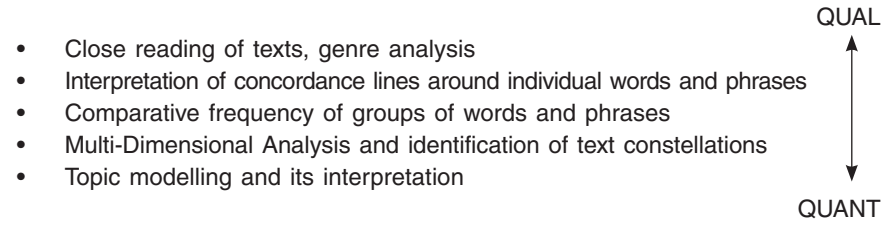

Figure 1.1 Methods used in the studies in this book.

journals and the nature of interdisciplinarity in discourse. We apply to the data we have collected (the journal articles) corpus investigation methods that rearrange and process that data. Our challenge is then to make sense of the rearranged data.

In this book, we draw three conclusions about the interdisciplinary discourse in our corpus. Firstly, the writers in the corpus write themselves an identity that is more or less interdisciplinary or not interdisciplinary at all. In other words, the identity of the writer as 'interdisciplinary' is not a factor of their professional circumstances or position within a research institution; rather it is fluid and a factor of how they write. This is not a surprising observation as it is in line with what is known about identity in academia and beyond: that it is discoursally constructed. In the most tightly focused journal in our corpus, *Plant Science*, identity choice is invisible, or at least unremarkable, because it is uniform among the articles. Writers do not present discipline as a contentious issue. In the most diverse journal, *Global Environmental Change*, many writers make that same choice, but there is a sub-set of writers who make discipline an issue, constructing their identity as part of an interdisciplinary exchange. It is perhaps worth adding here some complementary information about what we have not discovered in our study. We have not identified a set of language features that neatly divide interdisciplinary journals from monodisciplinary ones. We cannot say 'all interdisciplinary texts have a significant number of occurrences of features x, y and z, and therefore the presence of these features reliably identifies interdisciplinary texts'. It would probably be surprising, and a little suspect, if we could.

The second point is that all journals are different from each other, and all journals present a degree of internal variation. This variation is more apparent in the social sciences than in the natural sciences, but it is most apparent **between** the social sciences and the natural sciences, as all studies of academic discourse have demonstrated. In this book also, many of the differences we have identified between the four journals are attributable to the natural science–social science divide. In addition, in the journal *Global Environmental Change*, which combines natural science and social science research, there is evidence of similar internal variation, suggesting that writers bring into the journal their own disciplinary norms. One of the challenges addressed in this book is how to quantify variation of this kind. This theme is relevant to several chapters, but it is addressed most explicitly in Chapter 9 (Multidimensional Analysis) and Chapter 10 (Topic Modelling).

Thirdly, and rather more significantly, there are observable traces in the discourse of different stances taken in interdisciplinary work. In some cases we observe a writer accommodating to an interdisciplinary readership. In other examples we see writers treating alternative disciplinary approaches as significant or problematic. The fact that some interdisciplinary research involves an integration of approaches whilst other such research involves conflict and reflexive questioning leads to identifiable differences between integrative or supportive ('we can help you') and antagonistic ('you have got it wrong') stances. Of our four journals, only *Global Environmental Change* has substantial amounts of the antagonistic stance. *Agriculture, Ecosystems & Environment* has more expressions of an integrative or supportive stance.

The remaining chapters in this book will develop the arguments made here. To repeat, however, these arguments are not starting hypotheses for the study. Rather, the starting point is a series of exploratory investigations of the BEE4 corpus. The investigations aim to characterise the corpus and to produce findings that can be interpreted in terms of what is known about academic discourse and the nature of interdisciplinary research. The arguments presented in this section are the outcome of the investigations and are stated here to provide a road map for understanding the individual chapters of the book.

1.4 Our approach to Corpus Linguistics

As noted previously, our study uses a variety of methods from the range offered by Corpus Linguistics. These include targeted investigations of individual words, based on the manual scrutiny of concordance lines, right the way through to quantitative methods that produce graphs and box-plots. The point made in the previous section is that consistent messages emerge from this mixture of methods.

Planning and carrying out these studies has raised for us a number of issues relating to Corpus Linguistics as an approach to the study of language. These might be summarised under the headings of 'corpus design' and 'corpus methods'.

1.4.1 Corpus design

Good corpus design takes into account representativeness, size and balance (McEnery & Hardie 2011: 8). It is, however, also the art of the possible, with the availability of texts often a limiting factor. Availability can be affected by legal issues (such as copyright or ethical restrictions) and technical ones (such as how much resource is necessary to process the texts for inclusion). In this case, the publisher Elsevier kindly provided us with access to all the contents in 11 journals. For those journals, then, we have a near-saturation corpus (Oakley 2016), for a ten-year period, and we have preferred this to stratification (McEnery & Hardie 2011: 10) for identifying target texts. In other words, rather than selecting texts from many journals, we have focused on comprehensive coverage of selected journals. One of the reasons for this is that it is not clear what the categories for stratified selection would be. The apparently simple distinction between 'monodisciplinary' and

'interdisciplinary' is far from simple in practice (see Chapter 3), and in any case is confused by the known variation between disciplines and the unknown variation between types of interdisciplinarity. In short, there were no agreed and clear categories to select from. In any case, we were aware that research in linguistics that presupposes categories runs the risk of simply confirming those divisions, and we wanted to leave open the possibility of new categories emerging (Sealey 2010).

The corollary of this decision is that all the journals we use are intended for an international audience and are published by the same publisher. We therefore cannot comment on variability caused by 'house style' differences. We have similarly not distinguished the country of residence of the writers of each text so have no information about international variability. We adopt the position that the 'small cultures' of academic communities are more important in this respect than national cultures are (Holliday 1999).

Our work with the full BEE11 corpus (see Chapter 9) demonstrates the potential of working with a very large corpus. However, we have discovered that to make meaningful statements we needed to work with a more focused set of sub-corpora. This is because the processes are only semi-automatic, and involve considerable human judgement. For most of the studies in this book we use the BEE4 corpus, consisting of all the full-length research articles published in the four selected journals over a decade. However, because the journals differ in the number of articles published each year, each of the BEE4 sub-corpora is a very different size. This in turn means that the BEE4 corpus is not balanced, either in terms of the number of constituent texts or in terms of the number of tokens (Oakey 2009). To a large extent this disparity can be overcome through the usual normalisation techniques, but it does pose difficulties in the more automated studies. In our Multidimensional Analysis (Chapter 9), for example, the possibility exists that groupings of texts appear significant in one (smaller) sub-corpus but a similar grouping may be hidden within another (larger) sub-corpus.

In short, our corpus design is based on a decision to focus on the complete contents, over a decade, of a number of journals selected to contrast with each other in definable ways while addressing the same broad field of study. We would defend this decision as principled, even though it is based on concerns other than the need for representativeness and balance.

1.4.2 Corpus methods

The methods of corpus investigation constantly evolve, arguably in response to a diverse set of ambitions: to be maximally relevant in societal terms (Baker, Gabrielatos & McEnery 2013); to make statements that are quantitatively supported, and to do so by employing the most robust and relevant set of statistics (Gries 2019); to go beyond hypothesis-testing and to make discoveries that change our view of what languages and discourses are like (Sinclair 2004). Perhaps what sets Corpus Linguistics apart from other disciplines in the social sciences or the humanities, including other branches of linguistics, is that methodological invention and conceptual invention often go hand in hand. Examples would include Biber's (1988)

development of the concept of 'dimension', using factor analysis, and Sinclair's (1991) use of 'key word in context', or concordance lines, to develop the notion of 'unit of meaning'.

Methods in Corpus Linguistics are often presented as dichotomies. Studies that link text-based research and corpus-based research, for example, could be divided into those that start with text and use a corpus to test the generalisations proposed (Partington, Duguid & Taylor 2013) and those that use corpus methods to identify the 'best' texts to investigate in more detail (Baker & Levon 2015). A more general dichotomy is that between 'top down' and 'bottom up'. Top-down research begins with preconceptions about what it will be fruitful to investigate and restricts the corpus searches to those items. Bottom-up research tries to avoid such preconceptions and instead uses mostly quantitative methods to explore a corpus without knowing what will be found. The danger in top-down research is that only what is searched for will be found, and existing knowledge confirmed. The danger in bottom-up research is that it is so exploratory that it is difficult to interpret the results and to make sense of the data generated.

Some of the research reported in this book is top down. This is true particularly of Chapters 6, 7 and 8. Each of these chapters takes a set of words and phrases, selected in line with preconceptions about the topics we are researching, and engages in a series of quantitative (how many instances in each sub-corpus) and qualitative (in what collocational and functional contexts) studies in order to (1) describe differences between and within the sub-corpora and (2) identify the traces of what is known about types of interdisciplinarity, as noted in section 3 of this chapter. The more quantitative aspects of our study are bottom up and heuristic. They are described in Chapters 9 and 10. In keeping with bottom-up, methodologically driven research, these chapters introduce innovations. In Chapter 9 we adapt Multidimensional Analysis so that the compared groups of texts are *a posteriori* (an outcome of the research method) rather than *a priori* (input to the research method) (Thompson, Hunston, Murakami & Vajn 2017). In Chapter 10 we experiment with a version of Topic Modelling to see what this approach to content analysis can offer to supplement more traditional corpus methods (Murakami, Thompson, Hunston & Vajn 2017).

In all of the these studies, even those that are qualitative, we adopt an approach that might be described as corpus-based rather than text-based. That is, although we consider words and phrases in the context of the texts in which they are found, we do not move from a corpus to the investigation of specific individual texts. For the most part, for our qualitative work, we use extended concordance lines to examine the use of the target words in a random selection of texts. This mirrors the approach of, say, Baker et al. (2013) but not that of Baker and Levon (2015). In addition, however, we do adopt text-based methodologies to investigate the structure of research articles in the various journals (Chapter 4) and the structure and language of selected introductions (Chapter 5).

Our conclusions – that interdisciplinary identity is discoursally constructed, and that the various types of relations between disciplines leave identifiable traces in the corpus – arise from all the studies undertaken and reported in this book. Our

conclusion in relation to Corpus Linguistics is that we have needed the bottom-up studies in order to move it forward and open new avenues for research, but we have also needed the top-down studies to make more definitive and usable statements. In our own work we observe a productive tension between being open-ended and not limiting ourselves to predictable hypotheses, on the one hand, and finding meaningful things to say, on the other.

1.5 The organisation of this book and terminology

This book is organised in three main parts. This chapter and the next introduce the main concepts in book and review the notion of 'discipline' in university and internet discourse. Chapter 3 describes the BEE4 corpus in detail. The following five chapters present text and corpus evidence from a top-down perspective. Chapters 4 and 5 take a genre perspective and describe the journal article structures and aspects of the introduction sections in them. Chapters 6–8 explore the use of selected words and phrases in the BEE4 corpus: words related to environmental issues; status markers; and code glosses. The final part of the book expands the methodology and introduces the innovations noted previously: Chapter 9 deals with our revised version of Multidimensional Analysis; Chapter 10 reports our work on Topic Modelling. Chapter 11 concludes the book.

We end this chapter with some notes on terminology and conventions used throughout the book:

- For the most part we use abbreviations to refer to the four journals in the BEE4 corpus. The abbreviations and the journals to which they refer are:
    - AEE. *Agricultural Ecosystems and Environment*
    - GEC. *Global Environmental Change*
    - PS. *Plant Science*
    - REE. *Resource and Energy Economics*
- We use the terms 'natural science' and 'social science' for broad epistemological divisions. There is room for debate over terminology here. Some writers use 'physical science' or 'life science' instead of 'natural science', but in this book we use the latter term to refer to research that concerns systems other than human ones.
- We use the term 'journal article' or 'article' to refer to the texts that comprise the BEE corpus. The term 'paper' could be used, but we have selected 'article' as the preferred term. We avoid the term 'research article', as this indicates a specific genre and not all the articles in the corpus belong to that genre. (See Chapter 4 for further discussion of genre and text organisation.) The term 'text' is reserved for the short segments discussed in Chapter 10.
- Each journal article in the BEE corpus is identified by journal, date and name of first author. For example, 'GEC 2004 Jenkins' would refer to an article from the journal *Global Environmental Change*, published in the year 2004 and written by one or more authors, the first named of which is Jenkins.

In line with Corpus Linguistic practice, the source article is identified for examples, but not for concordance lines. A list of all the articles from which examples are taken can be found in the Appendix to this book.

- The articles in the corpus have been stripped of reference lists and equations, to avoid confounding frequency counts. All equations and symbols have been replaced by 'EQSYM'. For the most part, we have retained this in examples. In many cases we have replaced in-text non-integral citations with '(references)' and the year information in integral citations with '(ref)'.
- For the research reported in this book, the BEE4 and BEE11 corpora have been accessed via The Sketch Engine set of corpus analysis tools (www.sketchengine.eu/bibliography-of-sketch-engine/; Kilgarriff et al. 2014). We are grateful to the Sketch Engine team for permission to use this excellent resource.

References

Allmendinger, J. (2015). *Quests for interdisciplinarity: A challenge for the ERA and HORIZON 2020*. Directorate-General for Research and Innovation (European Commission).

Baker, P., Gabrielatos, C., & McEnery, T. (2013). *Discourse analysis and media attitudes: The representation of Islam in the British press*. Cambridge: Cambridge University Press.

Baker, P., & Levon, E. (2015). Picking the right cherries? A comparison of corpus-based and qualitative analyses of news articles about masculinity. *Discourse and Communication*, 9(2): 221–236.

Barry, A., & Born, G. (Eds.). (2013). *Interdisciplinarity: Reconfigurations of the social and natural sciences*. London: Routledge.

Bazerman, C. (1988). *Shaping written knowledge: The genre and activity of the experimental article in Science*. Madison: University of Wisconsin Press.

Bhatia, V. (1993). *Analysing genre: Language use in professional settings*. London: Longman.

Biber, D. (1988). *Variation across speech and writing*. Cambridge: Cambridge University Press.

Choi, S., & Richards, K. (2017). *Interdisciplinary discourse: Communicating across disciplines*. London: Palgrave.

Dudley-Evans, T. (1978). *Nucleus engineering*. London: Longman.

Durrant, P. (2017). Lexical bundles and disciplinary variation in university students' writing: Mapping the territories. *Applied Linguistics*, 35(3): 328–356.

Fairclough, N. (1995). *Critical discourse analysis: The critical study of language*. London: Longman.

Fairclough, N., Cortese, G., & Ardizzone, P. (2007). *Discourse in contemporary social change*. Bern: Peter Lang.

Frost, S., & Jean, P. (2003). Bridging the disciplines: Interdisciplinary discourse and faculty scholarship. *The Journal of Higher Education*, 74(2): 119–149.

Gries, S. T. (2019). 15 years of collostructions: Some long overdue additions/corrections (to/of actually all sorts of corpus-linguistics measures). *International Journal of Corpus Linguistics*, 24(3): 387–414.

Halliday, M. A. K., & Martin, J. (1993). *Writing science: Literacy and discursive power*. Pittsburgh: University of Pittsburgh Press.

Handford, M. (2013). *The language of business meetings*. Cambridge: Cambridge University Press.

Holliday, A. (1999). Small cultures. *Applied Linguistics*, 20(2): 237–264.

Hunston, S. (1994). Evaluation and organisation in a sample of written academic discourse. In M. Coulthard (Ed.), *Advances in written text analysis* (pp. 191–218). London: Routledge.

Hyland, K. (2006). Disciplinary differences: Language variation in academic discourses. In K. Hyland & M. Bondi (Eds.), *Academic discourse across disciplines* (pp. 17–42). Frankfurt-am-Main: Peter Lang.

Hyland, K. (2012). *Disciplinary identities: Individuality and community in academic discourse*. Cambridge: Cambridge University Press.

Johns, T. (1994). From printout to handout: Grammar and vocabulary teaching. In T. Odlin (Ed.), *Approaches to pedagogic grammar* (pp. 293–313). Cambridge: Cambridge University Press.

Kellert, S. (2008). *Borrowed knowledge: Chaos theory and the challenge of learning across disciplines*. Chicago: University of Chicago Press.

Kilgarriff, A., Baisa, V., Bušta, J., Jakubíček, M., Kovář, V., Michelfeit, J., Rychlý, P., & Suchomel, V. (2014). The sketch engine: Ten years on. *Lexicography*, 1: 7–36.

McEnery, T., & Hardie, A. (2011). *Corpus linguistics*. Cambridge: Cambridge University Press.

Muguiro, N. (2019). *Interdisciplinarity and academic writing: A corpus-based study of three interdisciplines*. Unpublished PhD thesis, University of Birmingham, Birmingham, UK.

Murakami, A., Thompson, P., Hunston, S., & Vajn, D. (2017). What is this corpus about? Using topic modelling to explore a specialised corpus. *Corpora*, 12(2): 243–277.

Oakey, D. (2009). Fixed collocational patterns in isolexical and isotextual versions of a corpus. In P. Baker (Ed.), *Contemporary corpus linguistics* (pp. 142–160). London: Bloomsbury.

Oakley, L. (2016). *An investigation into the representations of sexuality in sex education manuals for British teenagers, 1950–2014*. Unpublished PhD thesis, University of Birmingham, Birmingham, UK.

Omidian, T., Shahriari, H., & Syanova-Chanturia, A. (2018). A cross-disciplinary investigation of multi-word expressions in the moves of research article abstracts. *Journal of English for Academic Purposes*, 36: 1–14.

Partington, A., Duguid, A., & Taylor, C. (2013). *Patterns and meanings in discourse: Theory and practice in Corpus-Assisted Discourse Studies (CADS)*. Amsterdam: Benjamins.

Samraj, B. (2008). A discourse analysis of master's theses across disciplines with a focus on introductions. *Journal of English for Academic Purposes*, 7(1): 55–67.

Sealey, A. (2010). Probabilities and surprises: A realist approach to identifying linguistic and social patterns, with reference to an oral history corpus. *Applied Linguistics*, 31(2): 215–235.

Sinclair, J. (1991). *Corpus concordance collocation*. Oxford: Oxford University Press.

Sinclair, J. (2004). *Trust the text: Language, corpus and discourse*. London: Routledge.

Swales, J. (1981). Aspects of article introductions. In *ESP Monographs 1*. Aston University: Language Studies Unit.

Swales, J. (1990). *Genre analysis: English in academic and research settings*. Cambridge: Cambridge University Press.

Swales, J. (2004). *Research genres: Explorations and applications*. Cambridge: Cambridge University Press.

Teich, E., & Holtz, M. (2009). Scientific registers in contact: An exploration of the lexico-grammatical properties of interdisciplinary discourses. *International Journal of Corpus Linguistics*, 14(4): 524–548.

Thompson, P., Hunston, S., Murakami, A., & Vajn, D. (2017). Multi-dimensional analysis, text constellations, and interdisciplinary discourse. *International Journal of Corpus Linguistics*, 22(2): 153–186.

2 Disciplines and interdisciplinarity

2.1 Introduction

As noted in Chapter 1, this book is about interdisciplinary research discourse, but to understand what 'interdisciplinary' means and what the term 'interdisciplinary' covers, we first ask what disciplines are, and what it means to refer to something as 'disciplinary' (and what sorts of things we describe as 'disciplinary').

Following Giannoni's (2010: 21) distinction, disciplines might be defined both from the point of view of epistemology (what is studied and how it is studied) and from the point of view of social structures (when, where and why it is studied). Further distinctions can be drawn. Klein (2006), for example, argues that while the former is a functional differentiation, the latter construes disciplines as power systems, expressed through institutional structures and allocations of resources, prestige and privileges. Holley (2009) subdivides the 'what', distinguishing between a discipline as (1) a field of study and (2) the body of knowledge that is associated with a given field of study. Taking the discipline of economics as an example, one might therefore talk about doing economics research, about reading the economics literature or about the values and practices of an economics research community. For our purposes, however, the notion of a discipline as a discoursally constructed concept is particularly important, and for that reason we prioritise academic studies of disciplinary language (section 2.5) but we also examine how the notion of discipline is talked about unreflectively in public documents from the academy and beyond (section 2.2.2).

It becomes abundantly clear that although people may claim to know perfectly well what a discipline is, when examined in any detail, the notion can disappear into fuzziness. How do we know when a field of study constitutes a discipline? Core disciplines, such as physics, chemistry or philosophy, have long been acknowledged as such, but is a more recent field of study, astrophysics, a discipline, a sub-discipline or an interdiscipline? One university webpage sees it as a discipline: 'On the course, we will cover theoretical, observational and instrumental areas in astrophysics and other scientific disciplines' (www.cardiff.ac.uk/study/postgraduate/taught/courses/course/astrophysics-msc, accessed 21 August 2019), while a University of Copenhagen webpage regards astrophysics as a branch of physics which requires specialist knowledge of 'the links between astronomy and

other scientific disciplines' (https://studies.ku.dk/masters/physics/specialisations/astrophysics/, 21 August 2019). In the one view, astrophysics is given the status of a discipline, in the other it is seen as an interdiscipline, drawing on physics, astronomy and 'other scientific disciplines'.

As we shall see in this chapter, the term 'discipline' is polysemous. The indeterminacy surrounding disciplines has generated ways of talking about them that draw on metaphors that are both helpful and distorting. Applied linguistic research, however, has tended to avoid the debate about what disciplines are and has treated disciplines as given.

We approach these issues first by examining the ways that the words 'discipline', 'disciplines' and 'disciplinary' are used in public discourses. For this purpose, we first take a corpus linguistic approach to the exploration of these words (where the word 'discipline' is used in relation to academic activities) and look at evidence of the word and related terms in corpus data, to see what contexts the words are used in and for what purposes.

2.2 A corpus study of *discipline*, *disciplines* and *disciplinary*

Our aim is to identify the 'preferred ways' of talking about disciplines and disciplinarity, and the chief ways in which people conceive of disciplines. 'Discipline' is an abstract entity and it is discoursally constructed. Through investigation of the linguistic contexts in which the words *discipline*, *disciplines* and *disciplinary* are used, we aim to identify the metaphors that underlie the conceptualisation of 'discipline' and also to map the discourses in which different concepts are employed. In the following, we report the collocational and phraseological patterns around the three words. We also draw links between these patterns and the discussions and theories in the broad research literature on disciplines and disciplinarity.

2.2.1 The corpus

The data set that we use is a purpose-built corpus of texts downloaded from the internet. The purpose here is to gather evidence of how particular words are used and we have opportunistically chosen to collect our evidence from a particular medium, the World Wide Web, because of the ease of access, and also because of its widespread use. The sample collected does not contain transcripts of everyday spoken language use, nor (obviously) does it contain evidence from written communications that are not available on the internet, and in that sense it is a partial view on public discourses. It is, however, adequate for our purposes and, being focused on selected words, better than available alternatives. Large corpora of general language use are not fit for this purpose because they do not contain enough instances of the particular words that we are interested in. The 100 million word British National Corpus, for example, contains 6563 instances of the noun lemma DISCIPLINE (either *discipline* or *disciplines*) but the meanings are undifferentiated and many of them involve senses that our beyond the scope of our interest, such as in 'The aim is to provide a taste of the disciplines

and dilemmas of designing jewellery in this unique metal' and 'Only within the Parliamentary Labour Party was there an attempt at collective discipline'. Equally, although there are 1087 instances of the adjective *disciplinary* in the BNC, the majority of these refer to legal processes, associated with the words *tribunal*, *hearing*, *committee* and *action*. In our opportunistic corpus (which we call the 'Discipline Corpus'), we have captured 788 instances of the word *discipline*, 1339 of *disciplines*, and 1118 of *disciplinary*, nearly all of which relate to academic/research activity.

The Discipline Corpus was created using the freeware computer programme 'BootCat' (Baroni & Bernardini 2004). The internet was searched for relevant documents using a range of 'tuples' (sets of search terms). The default setting in this tool is that the user must specify at least five search items and the program then generates permutations of how these five items can be combined in groups of three. For example, if we take five items, A, B, C, D and E, then possible three item permutations are ABC, ABD, BCD and so on. These permutations are then used to form search queries in which the three items are linked by the Boolean operator 'AND' to specify that all three of the items must be found in a webpage to match the query. At this stage the user can choose to remove some of the permutations. Once the list of permutations is finalised, the search engine (in this case, Bing) then finds pages on the internet that feature all the terms in the given permutation, with each of the permutations in turn, and returns a list of page titles and URLs. This list has to be checked carefully to find any duplicates (the result of pages being retrieved by more than one search query) and other potentially redundant material, so that these can be removed before directing the program to download all of the pages that remain on the listing. The program then retrieves the pages, removes extraneous detail and creates a single corpus file, in which each text is preceded by the URL of the source text, as a heading (so that every text is retrievable).

As we are interested in uses of *discipline* in the broad sense of academic field or branch of knowledge and not in uses relating to general behaviour training (as in 'the discipline of writing') or punishment ('facing disciplinary action'), we used the following as 'seeds':

transdisciplinary	medicine	social sciences
environmental	discipline	culture
interdisciplinarity	humanities	science
university	disciplines	genre
disciplinarity	researcher	academic
higher education		

From a list of 40 tuples that the programme generated we then selected 30 tuples to search for, as shown in Table 2.1.

551 texts were downloaded and then thinned down to remove duplicated texts (where the same content had been published twice or more on different webpages) and any texts which were too short, which were on irrelevant topics, or which were simply bibliographies (lists of book titles) and which thus inflated the occurrence

Table 2.1 The 30 tuples used to search for texts for inclusion in the Discipline corpus.

interdisciplinarity/academic/science	'social sciences'/disciplinarity/ideas
interdisciplinarity/environmental/genre	researcher/transdisciplinary/science
genre/humanities/discipline	environmental/discipline/culture
humanities/disciplinarity/medicine	environmental/disciplinary/culture
transdisciplinary/'social sciences'/medicine	disciplines/interdisciplinarity/'social sciences'
environmental/science/transdisciplinary	environmental/disciplinarity/transdisciplinary
medicine/genre/discipline	university/disciplines/humanities
disciplinarity/culture/humanities	knowledge/disciplinary/research
environmental/transdisciplinary/genre	discipline/disciplinarity/environmental
disciplines/science/environmental	disciplinarity/medicine/genre
interdisciplinarity/university/environmental	medicine/humanities/discipline
'social sciences'/medicine/discipline	disciplines/'higher education'/university
academic/knowledge/discipline	university/writing/discipline
discipline/writing/value	academic/disciplinary/practices
innovation/disciplines/management	community/disciplinary/research

of particular phrases that appeared in oft-cited titles, such as Ken Hyland's monograph *Disciplinary Discourses* (Hyland 2000). The final corpus contains 479 texts, and 526,983 tokens.

It should be emphasized that the corpus contains a random sample of texts from the internet and is not structured to have balanced samples of texts from different categories. The majority of texts in the corpus can be described as:

- Announcements of talks, conferences
- Information about university courses
- Blogs
- Research articles and abstracts of articles
- English for Academic Purposes (EAP) and Writing in the Disciplines (WiD) pages
- Wikipedia articles

2.2.2 Corpus analysis

We start by making some observations about the immediate collocates of these words. To do this, we report on bigrams and trigrams that include one of the three key words and which have both high frequency and, importantly, range, in the corpus. By 'frequency' we mean the number of occurrences of particular sequences of alphabetical characters in all the corpus, and by 'range' we refer to the number of texts that a particular sequence is found in. If a sequence occurs many times but only in one or two texts, then it can be seen to have extremely limited range; the more texts that it occurs in, the greater the range. In our investigation we are interested at this stage to find sequences that are markedly frequent and which also have a range factor of at least eight (that is, they occur in at least eight different texts) for bigrams, and six for trigrams. In reporting the frequency and range each

Table 2.2 Bigrams in the Discipline corpus for *discipline*, *disciplines* and *disciplinary*.

	Bigrams
discipline (N=788)	the discipline (f=133; range=60) a discipline (93; 58) discipline and (58; 40) one discipline (42; 28) discipline of (34; 24) your discipline (31; 8) discipline is (24; 12)
disciplines (N=1339)	different disciplines (89; 50) disciplines and (117; 70) the disciplines (154; 67) of disciplines (130; 72) other disciplines (65; 43) academic disciplines (67; 51) across disciplines (57; 35) between disciplines (38; 27) disciplines to (48, 38) disciplines are (40; 34)
disciplinary (N=1118)	cross-disciplinary (133; 49) of disciplinary (115; 49) disciplinary research (72; 35) multi-disciplinary (66; 36) inter-disciplinary (57; 16) disciplinary boundaries (51; 30) disciplinary knowledge (36; 19) disciplinary perspectives (25; 15) disciplinary approaches (24; 21) disciplinary writing *along with* disciplinary discourse disciplinary conventions (28; 7)

sequence is shown with two numbers in brackets, where the first number indicates frequency and the second indicates the range.

For the three words, *discipline*, *disciplines* and *disciplinary*, the main bigram pairings are as shown in Table 2.2.

The most frequent *discipline* bigrams have either *the* or *a* before *discipline* and we can add that they are usually preceded by a preposition or by a verb; the most frequent prepositions will be shown in the trigram results for *discipline*. The pairing of *one* with *discipline* suggests that the writers are talking about a range of disciplines (*one* implies 'others'), and inspection of the concordance lines confirms this, with phrases such as 'vary from one discipline to another', 'the transfer of methods from one discipline to another', 'can be used in more than one discipline' and 'beyond the scope of any one discipline'. Disciplines are treated in these cases as discrete entities and occasionally perceived as being limited in scope.

One other *discipline* bigram that warrants discussion is *your discipline*, which stands out from the other bigrams by the use of a personal pronoun. The pairing has low range, which indicates that it occurs in only a limited number of types of

text. Inspection of concordance lines confirms this; these occurrences come from pages about 'Writing in the Disciplines', where the page is addressed to student writers, or to subject teachers who are being advised how to help their students with their writing, with reference (in the corpus) to 'writing conventions', 'rhetorical features' and 'genres'. Invariably, an assumption is made that students belong to disciplines, that they know what these are and that each discipline has conventions and genres that are known to the expert members of the discipline. In one case, though, the use of *your discipline* occurs in a forum in which the discussants present their ideas about the difference between 'discipline' and 'field', and make observations about *my discipline* in comparison to *your discipline*. Other than this, however, *your discipline* is restricted in use to a single discourse domain.

For *disciplines* bigrams, *the disciplines* is among the most frequent, with 45 occurrences following *in* (*in the disciplines*; see the following trigrams). Many of these come from 'Writing in the Disciplines' and are connected to the educational movement in academic writing support and teaching to prepare students for discipline-specific writing tasks and ways of thinking, as opposed to a broad notion of 'academic writing'. Again, there is an assumption in this phrasing that disciplines have distinct ways of thinking and of writing, and that they are identifiable. The framing of actions *in the disciplines* construes disciplines as social environments, in which mental process activities ('writing in the disciplines' or 'thinking in the disciplines') take place. Those activities are particular to those environments. In some cases, the discipline is seen as a community which is highly exclusive, with arcane practices and language that serves to restrict membership to a select few, almost like a secret society. One is *accepted into a discipline* as shown in the following concordance: 'being accepted within a discipline requires consistent displays of allegiance to a discipline's orthodoxy'.

The community has established norms for behaviour and members of their community are expected to follow those conventions, and it is knowledge and exercise of those conventions that constitutes membership.

The second most frequent pairing is *of disciplines* which, from a reading of concordance lines with *of disciplines* as search term, is part of a noun + 'of disciplines' sequence, in which the most common nouns are *cooperation*, *integration*, *range* and *variety*. References to *range* and *variety* tend to occur in university promotional websites that provide information about courses that are available at the university, using the promise of variety to attract potential applicants. In these pages, a comparable synonym for *disciplines* is *subjects*. Disciplines are also sometimes simply institutional groupings. A university or other public body may speak of groupings within a university structure as being a discipline, interchangeable with *department*, a component of a school or college, as in: 'For any disciplines or degrees not listed below, please begin looking at the main list of disciplines and majors'.

For *disciplines*, there is an emphasis on differentiation between groupings as can be seen in the uses of *different*, and *other*, as was also observed for *one discipline*. This is evident also in the prepositions that occur before *disciplines – across* and *between*. Where *in* ('in the disciplines') implies a sense of membership, and the use of a personal pronoun potentially adds to that a sense of group identity

('your discipline'), the uses of *across* and *between* express spatial and dynamic metaphors. Collaborations can occur, integration can be achieved, research spaces created 'across disciplines'; interactions, and interplay take place 'between disciplines', and there are interfaces and intersections, as well as concepts that 'travel between disciplines' and knowledge that is transferred.

With *disciplinary* the emphasis is primarily on research activity that is not constrained (cf, *disciplinary boundaries*) by a single discipline, leading to 'cross', 'multi' or 'inter' disciplinary work. 'Boundaries' employs the metaphor of a limited area, a containing space, that one can extend beyond, work across or transcend. *Disciplinary approaches* often appear either as 'limitations' to be confronted or they are diversified as *multidisciplinary approaches*. Only *disciplinary knowledge* is used positively about the individual discipline, as in 'teaching disciplinary knowledge is important', although in a few instances the uses are negative (for example, 'cannot be adequately addressed by disciplinary knowledge').

That *disciplinary* is so often used in hyphenated forms (*inter-disciplinary*, *cross-disciplinary*) is notable as it indicates that *disciplinary* occurs more frequently in these hyphenated forms than *disciplinary* by itself. To check this, we looked in the 15-billion-word EnTenTen 15 corpus (accessed through Sketch Engine) and found that the frequency of both the unhyphenated and the hyphenated forms *interdisciplinary* and *inter-disciplinary* together is 322,222 (21.5 times per million words). By contrast, the adjective *disciplinary* occurs 127,194 times and in a large number of these instances it refers to general behaviour training or to punishment ('facing disciplinary action'). In general internet discourse, then, *interdisciplinary* (or *inter-disciplinary*) is a term used much more frequently than *disciplinary*.

In the same corpus (EnTenTen 15), we see that the nouns which are most frequently pre-modified by *interdisciplinary* are:

research (19,000)	*course* (3962)
approach (18,503)	*perspective* (3802)
team (14,517)	*field* (6099)
study (7801)	*artist* (2005)
collaboration (7539)	*curriculum* (1660)
nature (4006)	*center*/*centre* (1019)

The word *interdisciplinary* is frequently used both in connection with research (*approach*, *team*, *study*) and with teaching (*course*, *curriculum*). In addition, it can also refer to activity that takes place between researchers (*collaboration*); the groupings of people from different disciplinary backgrounds working on projects (*team*); individuals (*artist*) who take more than one disciplinary approach; the broad area of research (*field*); and to institutional groupings (*centre*).

By comparison, the nouns that are pre-modified by *disciplinary* are:

boundaries (4264)	*approaches* (872)
perspectives (1690)	*practices* (784)
backgrounds (1238)	*areas* (709)
knowledge (1007)	*fields* (579)

Readings of the concordance lines confirm what has been observed here, that these uses are often in the context of breaking down, or finding ways to transcend, boundaries, or bringing together approaches or perspectives, and they employ geospatial and behavioural metaphors. We now turn to examine the trigrams in the Discipline Corpus with highest frequency and/or range (Table 2.3).

As can be seen in Table 2.3, trigrams are less numerous than bigrams (which is not surprising because each trigram is made up of two bigrams) but the trigrams provide clearer evidence of particular phrasings, or preferred ways of saying things. Before *the discipline* the prepositions *of* and *in* are most frequent. Inspection of the concordance lines reveals that 46 out of the 133 occurrences of *the discipline* are in the pattern NOUN + *of the discipline*. The head of the NP often refers to characteristics of the discipline such as *activity*, *discourse*, *ethos* and *life*, and includes reference to the 'conventions'. The references to *ethos* and *life* construe disciplines as cultures. We also find references to the *state of the discipline*, the *future of the discipline* and the *history*, where disciplines are seen as temporal entities, with sometimes an intimation of a negative future, as in *fragmentation* though there are examples of a potentially positive development as in 'transformation of the discipline'. In both cases, *the discipline* is construed as an entity that is bound to change.

As a discipline is used with a name of a knowledge field, such as *ecology*, *management* or even *science* and *humanities*, to identify the entity as a discipline. The texts tend to focus on the historical development of a particular branch of knowledge ('anthropology also attained clear identity as a discipline';

Table 2.3 Trigrams in the Discipline corpus for *discipline*, *disciplines* and *disciplinary*.

	Trigrams
discipline (N=788)	of the discipline (46; 21) as a discipline (21; 16) in the discipline (21; 12) of a discipline (19; 15) a single discipline (14; 8) in a discipline (12; 10)
disciplines (N=1339)	in the disciplines (45; 21) from different disciplines (36; 24) range of disciplines (26; 24) disciplines such as (18; 15) from other disciplines (18; 12) of the disciplines (18; 12) across the disciplines (14; 9) variety of disciplines (13; 11)
disciplinary (N=1118)	of cross disciplinary (18; 12) of the disciplinary (13; 10) disciplinary approaches to (7; 5)

'sociology developed as a discipline in the second half of the twentieth century'). In these instances, disciplines 'emerge', are 'framed' and one discipline 'can rival' another.

The second most frequent trigram for all three words is *in the disciplines* and this has been discussed previously. The third most frequent is *from different disciplines* and the nouns that precede this trigram are mainly referring to people: *students*, *researchers*, *experts*, *groups*. Invariably the texts are discussing multidisciplinary activities, and draw on the concept of disciplines as groupings of researchers with a distinct set of methods and perspectives.

One final observation is that, in this corpus, disciplines hardly ever have agency attributed to them; they 'embody', they 'incorporate', 'emerge', they 'come into being' (all one instance each), but otherwise *discipline* predominantly indicates a grouping, a training or a way of building knowledge, that either has attributes or is developed.

In conclusion, this study of an opportunistic corpus made of texts taken from the internet has shown that:

- Spatial and temporal metaphors are used in relation to disciplines – disciplines have boundaries, these boundaries can be transcended, concepts travel between disciplines, and the disciplines change, develop, emerge and fragment.
- Disciplines are construed either as groupings which have norms for communication, for methodologies and for behaviour, or they are equated to social environments, with a tendency to be exclusive.
- A discipline is often construed as a discrete grouping, differentiated from other groupings.
- Disciplines typically do not have agency attributed to them.
- Institutions may use (in public-facing webpages) the term 'discipline' to refer to courses (training) that they offer, or to a body of knowledge.

2.3 Mapping the disciplines

As noted previously, there is a considerable body of research devoted to identifying parameters of difference between disciplines. The use of metaphors such as 'mapping the disciplines' to describe disciplines and disciplinarity permeates this research literature. As Krishnan (2009) observes, metaphors are a convenient means for talking about the highly abstract concepts of disciplines and knowledge, and it is difficult to avoid them. Spatial metaphors are already implicit in the ways that groupings of academics and academic subjects are described, in words such as *departments*, *fields* and *areas*. In one of the most influential books on disciplines and disciplinarity, Becher (1989) wrote of 'academic tribes' and 'territories', invoking both anthropological or geopolitical metaphors. In the former, apprentices are socialized into a disciplinary culture by learning the ways of the tribe (Becher & Trowler 2001; Ylijoki 2000); and in the latter disciplines have 'territories', 'fields' and 'frontiers', which members 'annex', 'map', and 'explore' (Becher 1989: 36).

Becher's (1989) book is the fullest elaboration of his part anthropological, part geopolitical mapping of disciplines. The book can be seen as a rejection of C.P. Snow's binary opposition of the arts and the sciences, and it drew substantially on the work of Biglan (1973), who had placed disciplines at different points in a three-dimensional model, with the three axes running from 'pure' to 'applied', from 'hard' to 'soft', and from 'life' to 'nonlife'. To those three, Becher added a further axis of 'rural' versus 'urban', which he described as a social component (as opposed to a cognitive component); the rural research communities are spread out with less overlap, as distinct from urban, closely situated and more competitive populations. These were not only the only frameworks that have been advanced: Pantin (1968), for example, divided disciplines into 'restricted' and 'unrestricted' sciences, while Toulmin (1972) distinguished between 'compact' and 'diffuse' disciplines.

While Toulmin's categories have informed analyses such as MacDonald's (1994) monograph on academic writing in the social sciences and the humanities, Becher's work has been the more influential and is frequently used in linguistic studies of disciplinary discourses. These studies tend to make reference to two of the four axes: those of 'pure-applied' and 'hard-soft'. They then categorise disciplines as being in one of the four quadrants. For example, Harwood (2005) looked at first-person pronoun use in research articles in Business & Management, Computing Science, Economics, and Physics, with these categorized respectively as soft-applied, hard-applied, soft-pure and hard-pure. While this is a convenient way to differentiate between disciplines, there is a danger, as Trowler (2014) has argued, that it imposes an essentialist view on disciplines, in which disciplines are seen as distinct from each other with each one having essential properties. Trowler proposed an alternative view based on Wittgenstein's 'family resemblances' in which distinguishing elements can be found in common in all members of the category but not necessarily at the same time.

The conceptualisation of knowledge as landscape, according to Becher (1990), can lead to notions of geopolitical entities with divisions between sciences and humanities, for example, being seen to be analogous to North–South divisions. He argued that disciplines should not be conceived solely on a single axis of interpretation. Taking disciplines as geo-political entities (that is, a discipline is like a nation), one can also observe how diverse that country is in its population, its geographical features, its culture, its history and so on. Some parts of that population may feel that they have more in common with peoples in other nations than in their own; Becher provided the example of a mechanical engineer who shared more with mathematicians studying fluid mechanics than with other engineers.

Metaphors are necessary but they also influence and constrain the arguments that are advanced (Becher 1990). Streeten (1976: 147–148), for example, noted that territorial metaphors can create exclusive ways of thinking, such as when academics talk about 'poaching' on other people's territory, or of intellectual imperialism. Brew (2008) questioned the dominance of anthropological metaphors in discussions of disciplinary and interdisciplinary identity, on the grounds that such metaphors imply that the groupings are distinct and that members of that

group share a sense of identity. Brew's interviews with researchers showed that they do not view their identities as fixed but rather that they move between fine-grained and coarse-grained levels of identification ('Old Norse Icelandic Studies', as one example goes, versus 'European Languages and Literature' [Brew 2008: 429]) depending on the context. Brew also argued that identities are rhetorically constructed – researchers construct their identities through their texts. In place of the anthropological metaphors, Brew advocated, following Lyon (1992), the adoption of metaphors of fluidity, to talk of 'rivers', 'flow' and 'confluence'; more recently, Muguiro (2019) has proposed talking of 'oceans' and 'seas'.

In our view, 'discipline' is a central notion when approaching academic discourse and groupings and at the same time it resists easy definition. While comparing disciplines, particularly those which are self-evidently different, is a valuable pursuit, we propose that terms of comparison need to be treated with caution and that we should avoid essentialist characterisations that overlook diversity and dynamism.

2.4 Interdisciplinarity

To date, there has been very little linguistic research conducted on interdisciplinarity. The study by Oakey and Russell (2014) referred to later in this section is one example, and Choi and Richards (2017) is a book-length treatment of spoken communication in interdisciplinary research contexts. Both of these focus on spoken interdisciplinary discourse.

In recent decades, interdisciplinarity has been promoted as an alternative to disciplinarity – researchers working exclusively within their own discipline – which is viewed as a restriction on productive research. An argument in favour of interdisciplinarity is that it encourages researchers to pool knowledge, and it has the potential to solve problems that individual disciplines cannot solve alone (Jacobs & Frickel 2009). At the same time, there are numerous barriers to be overcome: Franks et al. (2007) listed institutional and professional impediments (such as funding, hiring, career development, recognition), along with disciplinary insularity, among the barriers to interdisciplinarity. As Graff (2015: 5) observed, though, there is 'no single path to interdisciplinarity, no single model, no single standard for successful development'.

The adjective *interdisciplinary* frequently is used in connection with either research, in relation to activities and groups, or with teaching, in relation either to what is taught (e.g. 'an interdisciplinary subject') or the course (e.g. 'an interdisciplinary programme'). Much has been written on the subject of interdisciplinarity; Graff (2015) estimated that 300–400 articles about interdisciplinarity are published every year. In these articles, attention is paid to aspects such as the management of interdisciplinary teams, or of institutional groupings that aim to be interdisciplinary, to theorisation of what interdisciplinarity consists of, to discussions of how to encourage and sustain interdisciplinary research, to reports of the experiences of researchers in attempting to work with researchers from other disciplines.

Van Dusseldorp and Wigboldus (1994) distinguish between what they termed 'narrow' and 'broad' interdisciplinarity. An example of the former would be a team made up of different types of natural scientist: agronomists, soil scientists and climatologists. As the researchers would share the same broad research paradigms and approaches to knowledge development that apply in the natural sciences, communication between them will be simpler. Broad interdisciplinarity on the other hand brings together researchers working in a range of paradigms, methods and disciplines, such as agronomy, biotechnology, economics and social policy. In broad disciplinarity, they propose, communication is more difficult. Kelly (1996) made a comparable distinction between 'narrow' and 'wide' interdisciplinarity based on epistemological similarities and differences among participating disciplines. This is a distinction that we take up in Chapter 3, adopting the terms 'proximal' and 'distant' to describe the relationship between different disciplinary groupings within journals.

One form of interdiscipline is the hybrid subject which fuses two disciplines (or more) and which may (or may not) over course of time become a distinct discipline. Muguiro (2019) looked at three interdisciplines: Educational Neuroscience, Economic History, and Science & Technology Studies. Analysing texts in a corpus of 450 research articles, she examined features of citation practices and adjectives of importance in order to determine whether the writing of articles in an interdiscipline more closely aligns with one of the two disciplines represented. She developed a rich thesis that the three interdisciplines represent in turn each of the interdisciplinary relationships that Barry and Born (2013) posit: a subordination-service mode in the Educational Neuroscience articles, an integrative-synthesis one in Economic History, and an agonistic-antagonistic one in Science & Technology Studies.

Barry and Born (2013) argued that interdisciplinarity is more than simply a combination of disciplines or a synthesis of existing disciplinary approaches. Their view focused on the relationships between the disciplines. In the second of these, there is at least a putative parity between the disciplines, an attempt to integrate, whereas in the first, the subordination-service mode, one or more of the disciplines involved is included to compensate for, or substitute for, a lack in the master discipline(s), and Barry and Born posited that social sciences are often tasked with this role in collaborations with the natural sciences. The third mode of interdisciplinarity is the agonistic-antagonistic in which there is neither synthesis nor an agreed division of labour, but rather an opposition from one set of researchers towards the practices and theories of another.

Language is often said to be a major obstacle in interdisciplinary research work. Researchers in different disciplines are seen to speak different languages (cf, Roy et al. 2013). Specialist terminology can be opaque, as are also some ways of formulating concepts. In some cases, a particular word can be used with different senses. Oakey and Russell (2014: 388) gave the example of an astrophysicist and a high energy physicist struggling to understand each other until they realised that they had diametrically opposed ideas of what 'efficiency' means in relation to a particular statistical method. For the former, 'efficiency' related to how many items

could be excluded by the statistical method ('rejection efficiency') while for the latter it was related to how many could be included ('selection efficiency').

Language can be used as an analogy to describe different levels of interdisciplinary collaboration. Stevenson et al. (2013) described disciplinary research as characterised as monolingual discussions, in which members of a particular community share a 'specific set of approaches, standards and language', whereas multidisciplinary research can be understood as multilingual discussions in which little translation is provided. They followed this by suggesting that interdisciplinary research has the potential to lead to integration and the use of pidgins which allow the researchers to communicate while maintaining their own disciplinary languages, and, finally, transdisciplinary teams establish their own communities that have their own creole languages.

2.5 Corpus studies of disciplinary variation

In this section of the chapter, we briefly survey the corpus-based studies of disciplinary variation and of written interdisciplinary research discourse.

Disciplinary variation has attracted a considerable amount of attention over the decades, with the majority of research taking place in the field of English for Academic Purposes (EAP). EAP is rooted in language teaching and support provision for international students either before or during a programme, and teaching is usually focused on the specific needs of students within programmes of study. Where the instruction addresses English for Specific Academic Purposes (ESAP) then descriptions of the grammatical, lexical and rhetorical features of communication in specific subject areas are needed. An example of the textbooks developed for EAP teaching in the 1970s is the OUP English in Focus series, which included English language textbooks in specific subject areas such as mechanical engineering, agriculture or basic medical science (Gillett 2019). Research in the early years of EAP tended to focus on register features but latterly moved towards studies at a discoursal or genre level, which focused more on patterning, rhetorical moves, stance and phraseology (Flowerdew 2002), from the 1980s onwards. With this turn has come an increased use of the concept of 'disciplines', where the subject area is not only distinguished by specific vocabulary and by preferred grammatical structures but also by different discourse and cultural practices.

As observed in the Discipline corpus investigation, the expression 'disciplinary variation' is chiefly used by EAP practitioners and researchers, and it is usually based in corpus research. A key book in this field is Hyland's (2000) *Disciplinary Discourses*, a collection of corpus-based studies of disciplinary variation, using a corpus of research articles in eight disciplines. Hyland analysed a range of text genres (research articles, book reviews, scientific letters, abstracts and textbook chapters) in Molecular Biology, Magnetic Physics, Mechanical Engineering, Electronic Engineering, Philosophy, Sociology, Marketing and Applied Linguistics. What we can notice here, as in many other studies, is an assumption that these are disciplines and they are of comparable status. Within the studies that are presented as chapters in that book, Hyland makes comparisons between biology and

physics – for example, removing the distinctions that the texts actually come from Molecular Biology and Magnetic Physics – and he provides no discussion of whether Molecular Biology is characteristic of the majority of biology research writing or not. This does not invalidate Hyland's research, but it does raise two questions: How representative of a discipline are the corpus samplings that corpus variation studies create? And, secondly, within any given disciplinary grouping (biology, for example) how much variation is there?

Hyland's work has inspired many studies in which two or more 'disciplines' are studied: Biber, Conrad, Reppen, Byrd and Helt (2002), for example, look at spoken and written student language in three disciplines (Business, Education, Engineering), while Peacock (2014) examines the use of modal verbs in 600 research articles in 12 disciplines that are distinguished as science or non-science. These are but two disciplinary variation studies selected at random from a large number. The point that we wish to make is that the studies tend to choose disciplines to compare and then describe each discipline as belonging to a larger grouping, such as Peacock's 'science' versus 'non-science'. The groupings are typically based on broad classifications of knowledge systems, or they can reflect local institutional structures, as in the case of some large-scale academic corpora. For example, the MICASE corpus (Simpson & Swales 2001) uses 'academic divisions' to group disciplines: Biological and Health Sciences; Humanities and Arts; Social Sciences and Education; Physical Sciences and Engineering. These divisions reflect the institutional structures within the university where the corpus was constructed: the University of Michigan at Ann Arbor. The British Academic Spoken English (BASE; Thompson & Nesi 2001) and British Academic Written English (BAWE; Nesi & Gardner 2012) corpora then employed the same four-way categorisation in order to complement MICASE, labelling them as 'disciplinary domains', which were called 'Arts and Humanities', 'Life Sciences', 'Physical Sciences' and 'Social Sciences'. As these are based on institutional structures, they are not always compatible with the frameworks taken from other institutions. The corpus that Coxhead (2000) created in order to build the Academic Word List, by contrast, was divided into 'Arts', 'Commerce', 'Law' and 'Science', which reflects the institutional organisation of faculties at the New Zealand university at which she was based. Ackermann and Chen (2013), in turn, apply another set of categories to the academic corpus that they have built, comprised of 'Applied sciences and professions (AS); Humanities (HM); Social sciences (SS); and Natural/Formal sciences (NS)'. While each system of categories has its rationale and its explanatory potential, the variation makes comparisons difficult.

The disciplinary variation literature has focused on a broad range of rhetorical and linguistic features in a variety of genres, such as the use of first-person pronouns (Harwood 2005) and self-mentions (McGrath 2016); at stance in research articles (Jiang & Hyland 2015; Silver 2003; Stotesbury 2003), in theses (Charles 2003) and in undergraduate writing (Lancaster 2016); at how citation is used in research articles (Hu & Wang 2014) and theses (Thompson 2005); at the incidence of metadiscourse (Hyland 1997); at the uses of lexical bundles in research articles (Esfandiari & Barbary (2017), in research article abstracts (Omidian, Shahriari &

Siyanova-Chanturia 2018) and in student writing (Durrant 2015); and the uses of phraseological patterns in research articles and book reviews in two disciplines (Groom 2005). Key to many of these studies are questions about how writers position themselves in relation to, firstly, the status of the propositions that they present, and, secondly, to the readers that they are communicating with. Academic writing is seen not simply as transactional, as a transfer of information, but also as interaction, or as rhetorical action, in which writers make decisions about how they can persuade their readers and also about how they want to present themselves, what identity they choose to construct within their writing (e.g. Hyland 2015).

While these studies assume a commonality of genre across disciplines, Nesi and Gardner (2012) focused on the notion of genre itself. They examined the genres of assessed writing across the disciplines in higher education, using the broad disciplinary domain classification mentioned previously (Arts and Humanities, Life Sciences, Physical Sciences and Social Sciences). They identify 13 genre families (Nesi & Gardner 2012: 34), each of which contains variants, and which occur with differing degrees of frequency across disciplinary domains (and subsets thereof) and across the years of undergraduate study. They then divide the genres by broad social purpose: demonstrating knowledge and understanding (explanations and exercises); developing powers of informed and independent reasoning (critiques and essays); developing research skills (research reports, literature surveys, methodology recounts); preparing for professional practice (e.g. case study); writing for oneself (e.g. reflective writing). Nesi and Gardner's main purpose was to describe the genres of assessed writing across the disciplines, but one of the points that emerge strongly from their work is that starkly different emphases are placed on student writing by the various disciplines, with argumentation privileged in some areas while case studies or reports predominate in others.

The corpus-based research on disciplinary variation has tended to regard disciplines as unproblematic entities which can be characterized as having distinctive ways of writing and of thinking, and this is partly an effect of an approach to comparisons that looks for recurrence of features across large bodies of texts. Barry, Born and Weszkalny (2008: 26–27) caution against this, however: 'disciplines are' they write, 'routinely characterized by internal differences; the existence of a discipline does not always imply that there is acceptance of an agreed set of problems, objects, practices, theories or methods'. A small number of corpus-based studies of intradisciplinary variation have emerged recently. McGrath (2016) looked at self-mentions in 18 history and 18 anthropology research articles. She found that 'I' was used more frequently in the anthropology articles and ascribed this to knowledge-making practices of the discipline but then pointed out that there was considerable variation within the discipline in how frequently first-person pronouns were used and the types of roles that the writers constructed for themselves through self-mentions.

Another exploration of variation within as well as between disciplines is reported in Gray (2015). Gray conducted a Multidimensional Analysis of research articles in six disciplines (Philosophy, History, Political Science, Applied Linguistics, Biology, Physics) and she added further level of differentiation in each sub-corpus

by sampling, where possible, articles that were predominantly 'qualitative', 'qualitative' or 'theory' research articles. Gray's study shows that discipline is not the sole factor responsible for variation in research articles but that article type (in this case, research orientation) is also a powerful contributory factor.

2.6 Studies of interdisciplinary research discourse

In section 2.3 we noted that there are a large number of studies of the concept of interdisciplinarity. In contrast, there have been very few corpus studies of interdisciplinary research discourses to date. In the period 2013–2015, we conducted, in collaboration with the scientific publisher Elsevier (UK) who gave us access to large amounts of research articles, a corpus study of interdisciplinary research discourse, funded by the Economic and Social Research Council of England (ESRC). The corpus that was created for that project will be described in some detail in Chapter 3 and it is a sub-section of that corpus that is used in the studies that will follow, in Chapters 4 to 10. That project addressed the problem of how to carry out corpus research when it was not possible to identify comparable corpora. The major articles arising from that project focused on issues of methodology and made two substantial proposals relating to Multidimensional Analysis (Murakami, Thompson, Hunston & Vajn 2017) and Topic Modelling (Thompson, Hunston, Murakami & Vajn 2017). These approaches will be applied again to the four journals, in studies that are presented as Chapters 9 and 10.

Two other corpus-based studies of interdisciplinary research discourse are Teich and Holtz (2009) and Muguiro (2019), both of which have looked at the emergence of new interdisciplines, through the interaction of two established disciplines. The Muguiro study has been discussed previously. Teich and Holtz (2009) posed the question 'What are the linguistic effects of two disciplines coming into contact with each other?', taking computer science (CS) as the base and looking at articles in computational linguistics, bio-informatics, computer-aided design and micro-electronics which emanate from the combination of CS with, respectively, linguistics, biology, computer-aided design and micro-electronics. The article reported corpus-assisted systemic-functional analyses of two nouns, algorithm and model, in terms of the verbs that they appear with, the roles they play ('senser', 'goal', and so on) and the grammatical constructions they occur in. The results indicated that the interdisciplines have a greater versatility than the four 'pure' disciplines, but less versatility than computer science, in the range of verbs used, and that of the four interdisciplines computational linguistics evidences the greatest versatility. As the study looks at only two nouns, however, it is limited in its generalisability. What is notable is that both Muguiro (2019) and Teich and Holtz (2009) approached interdisciplinarity by looking at interdisciplines; they identified research activity that combined two (or more) distinct disciplines in a new, usually emergent, area of research. Our work differs from theirs in that we do not look at interdisciplines (identifiable new groupings) but at journals in which researchers from different disciplinary backgrounds publish work which is aimed at an interdisciplinary audience.

2.7 What 'interdisciplinarity' means in this book

This book presents a set of linguistic investigations of published research articles written in journals written for either monodisciplinary or for interdisciplinary audiences, in order to find out whether writing for an interdisciplinary audience is distinct from writing for specialist audiences. The approach that we take is to use corpus analysis and close reading techniques, working with a sizeable corpus of articles. Given that the previous sections have indicated that disciplines and disciplinarity are fuzzy concepts, and that categorisations are open to contention, it is necessary to clarify our approach. We recognise that researchers may have a range of affiliations, and more than a single disciplinary identity, but we aim to distinguish between writing for an audience that is primarily composed of researchers from the same broad disciplinary area (in one case, resource economics and in the other plant sciences) and writing which we characterise here as 'interdisciplinary' in that, even though the research reported in the articles may not be strictly 'interdisciplinary' (in the sense we detail later) it is written for an audience of researchers from different disciplinary backgrounds, and draws on a broader range of research literatures.

As we observed in the Discipline corpus study, a high proportion of the occurrences of the adjective 'disciplinary' occur in discussions of 'multidisciplinary', 'interdisciplinary' or 'transdisciplinary' research. While there is considerable debate over the meanings of these three terms, the commonly agreed definitions are that multidisciplinary work refers to the collaboration of disciplines in which each discipline remains within its standard ways of acting, interdisciplinary work involves a degree of integration of the disciplines involved, but with each researcher retaining their disciplinary perspective, and transdisciplinary work involves researchers working together within a shared conceptual framework (Grey & Connolly 2008; Huutoniemi 2010). While the distinctions are useful for discussing research and teaching groupings and activities, they seem less relevant for the categorisation of journals, however. Many of the articles that appear in the interdisciplinary journals studied in the following chapters, for example, are single-authored, and do not necessarily report on work that has been conducted in an interdisciplinary team. Whether the authors draw on approaches or theories from different disciplines or whether they are individuals who work in interdisciplinary research centres, the common ground is that they are writing for a journal that has a mixed disciplinary audience. As such, we follow Elsevier (2015) in using the term 'interdisciplinary' (in 'interdisciplinary research discourse') as an umbrella term to cover 'multidisciplinary', 'interdisciplinary' and 'transdisciplinary'.

References

Ackermann, K., & Chen, Y. (2013). Developing the Academic Collocation List (ACL): A corpus-driven and expert-judged approach. *Journal of English for Academic Purposes*, 12(4): 235–247.

Baroni, M., & Bernardini, S. (2004). *BootCaT: Bootstrapping corpora and terms from the web*. Proceedings of LREC 2004 (pp. 1313–1316). Paris: ELRA.

Barry, A., & Born, G. (Eds.). (2013). *Interdisciplinarity: Reconfigurations of the social and natural sciences*. London: Routledge.

Barry, A., Born, G., & Weszkalnys, G. (2008). Logics of interdisciplinarity. *Economy and Society*, 37(1): 20–49.

Becher, T. (1989). *Academic tribes and territories: Intellectual enquiry and the culture of disciplines*. Buckingham: Open University Press.

Becher, T. (1990). The counter-culture of specialisation. *European Journal of Education*, 25(3): 333–346.

Becher, T., & Trowler, P. (2001). *Academic tribes and territories: Intellectual enquiry and the cultures of disciplines* (2nd ed.). Buckingham: Open University Press/SRHE.

Biber, D., Conrad, S., Reppen, R., Byrd, P., & Helt, M. (2002). Speaking and writing in the university: A multidimensional comparison. *TESOL Quarterly*, 36(1): 9–48.

Biglan, A. (1973). The characteristics of subject matters in different academic areas. *Journal of Applied Psychology*, 57: 195–203.

Brew, A. (2008). Disciplinary and interdisciplinary affiliations of experienced researchers. *Higher Education*, 56(4): 423–438.

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(4): 313–326.

Choi, S., & Richards, K. (2017). *Interdisciplinary discourse: Communicating across disciplines*. London: Palgrave.

Coxhead, A. (2000). A new academic word list. *TESOL Quarterly*, 34(2): 213–238.

Durrant, P. (2015). Lexical bundles and disciplinary variation in university students' writing: Mapping the territories. *Applied Linguistics*, 38(2): 165–193.

Elsevier. (2015). *A review of the UK's interdisciplinary research using a citation-based approach*. Retrieved August 26, 2019, from https://webarchive.nationalarchives.gov.uk/20170712122715/www.hefce.ac.uk/pubs/rereports/Year/2015/interdisc/

Esfandiari, R., & Barbary, F. (2017). A contrastive corpus-driven study of lexical bundles between English writers and Persian writers in psychology research articles. *Journal of English for Academic Purposes*, 29: 21–42.

Flowerdew, L. (2002). Corpus-based analyses in EAP. In J. Flowerdew (Ed.), *Academic discourse* (pp. 95–114). London: Longman.

Franks, D., Dale, P., Hindmarsh, R., Fellows, C., Buckridge, M., & Cybinski, P. (2007). Interdisciplinary foundations: Reflecting on interdisciplinarity and three decades of teaching and research at Griffith University, Australia. *Studies in Higher Education*, 32(2): 167–185.

Giannoni, D. S. (2010). *Mapping academic values in the disciplines: A corpus-based approach*. Bern: Peter Lang.

Gillett, A. (2019). *Published EAP materials: A history*. Retrieved August 29, 2019, from www.uefap.net/materials/materials-history

Graff, H. (2015). *Undisciplining knowledge: Interdisciplinarity in the twentieth century*. Baltimore: Johns Hopkins University Press.

Gray, B. (2015). *Linguistic variation in research articles: When discipline tells only part of the story*. Amsterdam: John Benjamins.

Grey, M., & Connolly, C. (2008). "Coming together, keeping together, working together": Interdisciplinary to transdisciplinary research and nursing. *Nursing Outlook*, 56(3): 102–107.

Groom, N. (2005). Pattern and meaning across genres and disciplines: An exploratory study. *Journal of English for Academic Purposes*, 4(3): 257–277.

Harwood, N. (2005). "Nowhere has anyone attempted . . . In this article I aim to do just that": A corpus-based study of self-promotional I and we in academic writing across four disciplines. *Journal of Pragmatics*, 37: 1207–1231.

Holley, K. (2009). Interdisciplinary strategies as transformative change in higher education. *Innovative Higher Education*, 34(5): 331–344.

Hu, G., & Wang, G. (2014). Disciplinary and ethnolinguistic influences on citation in research articles. *Journal of English for Academic Purposes*, 14: 14–28.

Huutoniemi, K. (2010). *Interdisciplinary environmental research: Frameworks for Interdisciplinary Environmental Analysis (FIDEA), Helsinki*. Retrieved from www.fidea.fi

Hyland, K. (1997). Persuasion and context. *Journal of Pragmatics*, 30: 437–455.

Hyland, K. (2000). *Disciplinary discourses: Social interactions in academic writin*g. London: Longman.

Hyland, K. (2015). Genre, discipline and identity. *Journal of English for Academic Purposes*, 19: 32–43.

Jacobs, J., & Frickel, S. (2009). Interdisciplinarity: A critical assessment. *Annual Review of Sociology*, 35: 43–65.

Jiang, F., & Hyland, K. (2015). "The fact that": Stance nouns in disciplinary writing. *Discourse Studies*, 17(5): 529–550.

Kelly, J. (1996). Wide and narrow interdisciplinarity. *The Journal of Education*, 45(2): 95–113.

Klein, J. T. (2006). A platform for a shared discourse of interdisciplinary education. *Journal of Social Science Education*, 5(2): 10–18.

Krishnan, A. (2009). What are academic disciplines? Some observations on the disciplinarity vs. interdisciplinarity debate. ESRC National Centre for Research Methods, Working paper 03/09. Retrieved August 21, 2019, from http://eprints.ncrm.ac.uk/783/

Lancaster, Z. (2016). Expressing stance in undergraduate writing: Discipline-specific and general qualities. *Journal of English for Academic Purposes*, 23: 16–30.

Lyon, A. (1992). Interdisciplinarity: Giving up territory. *College English*, 54(6): 681–693.

MacDonald, S. (1994). *Professional academic writing in the humanities and social sciences*. Carbondale, IL: Southern Illinois University Press.

McGrath, L. (2016). Self-mentions in anthropology and history research articles: Variation between and within disciplines. *Journal of English for Academic Purposes*, 21: 86–98.

Muguiro, N. (2019). *Interdisciplinarity and academic writing: A corpus-based study of three interdisciplines*. Unpublished PhD thesis, University of Birmingham, UK.

Murakami, A., Thompson, P., Hunston, S., & Vajn, D. (2017). "What is this corpus about?" Using topic modeling to explore a specialized corpus. *Corpora*, 12(2): 243–277.

Nesi, H., & Gardner, S. (2012). *Genres across the disciplines: Student writing in higher education*. Cambridge: Cambridge University Press.

Oakey, D., & Russell, D. (2014). Beyond single domains: Writing in boundary crossing. *Handbook of Writing and Text Production*, 10: 385–411.

Omidian, T., Shahriari, H., & Siyanova-Chanturia, A. (2018). A cross-disciplinary investigation of multi-word expressions in the moves of research article abstracts. *Journal of English for Academic Purposes*, 36: 1–14.

Pantin, C. (1968). *The relations between the sciences*. Cambridge: Cambridge University Press.

Peacock, M. (2014). Modals in the construction of research articles: A cross-disciplinary perspective. *Ibérica*, 27: 143–164.

Roy, E., Morzillo, A., Seijo, F., Reddy, S., Rhemtulla, J., Milder, J., Kuemmerle, T., & Martin, S. (2013). The elusive pursuit of interdisciplinarity at the Human: Environment interface. *BioScience*, 63: 745–753.

Silver, M. (2003). The stance of stance: A critical look at ways stance is expressed and modeled in academic discourse. *Journal of English for Academic Purposes*, 2(4): 359–374.

Simpson, R., & Swales, J. (Eds.). (2001). *Corpus linguistics in North America*. Ann Arbor, MI: University of Michigan Press.

Snow, C. P. (1959). *The two cultures and the scientific revolution*. Cambridge: Cambridge University Press.

Stevenson, D., Shaw, G., Wise, P., Norton, M., Druzin, M., Valantine, H., & McFarland, D. (2013). Transdisciplinary translational science and the case of preterm birth. *Journal of Perinatology*, 33: 251–258.

Stotesbury, H. (2003). Evaluation in research article abstracts in the narrative and hard sciences. *Journal of English for Academic Purposes*, 2(4): 327–341.

Streeten, P. (1976). The meaning and purpose of interdisciplinary studies. *Interdisciplinary Science Reviews*, 1(2): 144–148.

Teich, E., & Holtz, M. (2009). Scientific registers in contact: An exploration of the lexicogrammatical properties of interdisciplinary discourses. *International Journal of Corpus Linguistics*, 14(4): 524–548.

Thompson, P. (2005). Points of focus and position: Intertextual reference in PhD theses. *Journal of English for Academic Purposes*, 4(4): 307–323.

Thompson, P., Hunston, S., Murakami, A., & Vajn, D. (2017). Multi-dimensional analysis, text constellations and interdisciplinary discourse. *International Journal of Corpus Linguistics*, 22(2): 153–186.

Thompson, P., & Nesi, H. (2001). The British Academic Spoken English (BASE) corpus project. *Language Teaching Research*, 5(3): 263–264.

Toulmin, S. (1972). *Human understanding: The collective use and evolution of concepts*. Princeton: Princeton University Press.

Trowler, P. (2014). Depicting and researching disciplines: Strong and moderate essentialist approaches. *Studies in Higher Education*, 39(10): 1720–1731.

Van Dusseldorp, D., & Wigboldus, S. (1994). Interdisciplinary research for integrated rural development in developing countries: The role of social sciences. *Issues in Integrative Studies*, 12: 93–138.

Ylijoki, O.-H. (2000). Disciplinary cultures and the moral order of studying: A case-study of four Finnish university departments. *Higher Education*, 39(3): 339–362.

3 The BEE4 corpus

3.1 Introduction

This book presents a set of studies of four journals in which we identify and then explore differences and similarities in the structuring and writing of texts in academic journals. As our interest is in the features of research writing that addresses audiences beyond disciplinary fences, we compare journals that are written for an interdisciplinary audience with those written for a monodisciplinary audience. The first task, therefore, is to identify journals that fit into these two categories, or which can be placed towards either end of the monodisciplinary–interdisciplinary cline. In addition to the placement of the journals along a disciplinarity axis, we select journals that are made up of articles which appear to be broadly similar in their approaches to research (what we will call 'proximal disciplines') and those that consist of articles that represent a diversity of approaches (which we term 'distal disciplines'). These distinctions are explained later in the chapter.

Identifying a journal as either monodisciplinary or interdisciplinary is not as easy a task as it might be thought. One solution may be to look at the name of a journal but this is not always a reliable guide to the contents. Similarly, one can ask researchers in one or more fields to identify journals that they regard as interdisciplinary, but this runs the risk of partiality. Another approach is to look at the statements that journal editors publish about the aims and scope of a journal to find out what identity they want to project, what role interdisciplinarity has for the journal, and what audience they seek to address and what audience they seem to construe through the choices about content and style of the statement. It is worth noting the importance of the role of the editors in journals – while in many international journals articles are reviewed by two or more reviewers and they will make recommendations over whether or not articles are publishable, it is the editors who decide on journal policies and who decide the directions that the journal will take.

The approach we have taken is based on citation statistics. We look at the citation data for a journal, and find out what range of research fields are both referred to and also cited by articles in that journal, as this shows the range of disciplines that are drawn on and which in turn then also draw on the journal in question. A journal that contains articles that cite research in a wide array of disciplinary areas and which is also cited by researchers in other areas would be judged to be more

'interdisciplinary', as it contains articles that are written by researchers that engage with research in several disciplines, and furthermore, the audience is shown to be a broad one in that researchers from several disciplinary domains are able to refer to these articles. On the other hand, a journal that cites a very broad range of disciplinary areas but is not cited in return appears to aspire to interdisciplinarity but fails to achieve it, on the basis that there is little evidence that the readership of that journal comes from a broad range of disciplines.

This chapter explains the procedures that were used for choosing journals that are monodisciplinary or interdisciplinary. We have little precedent for this within corpus work. As we observed in Chapter 2, corpus research has focused on studies of disciplinary variation rather than on variation within and between interdisciplinary journals. Consequently, researchers such as Hyland (2000), Hardy and Friginal (2016) and Yakhontova (2006) (to name but a few) have focused on building corpora that represent different disciplines. Where corpus studies of journal articles in different disciplines has typically sampled articles, taking, for example, ten articles from ten journals in a discipline (Giannoni 2010) in order to represent research articles in that discipline, we instead take all of the full-length articles written in our chosen journals over a ten-year period, in order to avoid imposing too many assumptions upon the data. In previous studies, selection of articles has been based on criteria such as the number of citations an article has received (highly cited articles have been selected, presumably in order to exemplify highly rated articles); or the prototypicality of the article in relation to a particular model of organisation, such as the conventional 'Introduction – Methods – Results – Discussion' (IMRD) model (discussed in more depth in Chapter 4). In the one case, the choice of criterion leads to the exclusion of the articles that are less cited (and we usually do not know the reasons why an article was not cited) and in the other to a possible over-emphasis on one approach to knowledge development.

We are concerned in this book primarily with the exploration of writing about research that is targeted at interdisciplinary audiences. The procedure we have used for selection of the journals is based on the citation profiles of the articles in the journals, as we explain further later, and therefore is based upon evidence of intertextuality rather than on concepts of intended audiences. To compensate for this, in the second half of the chapter, we consider the declarations that each of the journals make about their intended readership, in their statements of aims and scope. These statements are formalised announcements that journal editors construct for publication either within volumes of the journal or, in more recent years, on the journal website. The statements are of interest in that, on one level, the editors may explicitly identify the type of readership that they envisage for the journal, while, on another level, the terms in which the aims and scope are articulated and the ways in which they are framed are indicative of the audience that the editors see themselves as addressing. If a statement contains, for example, high levels of specialised terminology to describe theories, methods, and research areas, then the expectation underlying such statements is that the readership understands such terminology and is a specialist audience. In addition to demonstrating the types of audience that the journal sees itself as addressing, the statements of aims and scope also provide useful insights into the research orientations of each journal during the ten-year period that is represented in the corpus.

3.2 Selection of the journals: bibliometric data

All four of the journals that are discussed in this chapter are published by the international scientific publisher, Elsevier. For the studies reported in this book, we have created (with the assistance of Elsevier UK) a corpus of all the research articles published by the four journals in the ten-year period 2001 to 2010. This results in a collection of over 5000 research articles from this period, with two of the journals accounting for a higher proportion of the total than the other two, as will be seen here. This corpus is named the BEE4 corpus, with BEE standing for 'Birmingham-Elsevier Environmental' and the digit 4 added to indicate that it contains four journals and to distinguish it from the 11-journal corpus that we analysed in previous studies (Murakami, Thompson, Hunston & Vajn 2017; Thompson, Hunston, Murakami & Vajn 2017), which we will refer to as the BEE11 corpus. The difference between BEE4 and BEE11 will be explained in section 3.3.

These four journals were originally chosen by looking at the citation behaviour statistics that Elsevier compiles and publishes in Scopus. Put simply, Elsevier collects information on all the references that are made within each research article in any given journal and they also collect information on how many times, and by whom, a research article is cited by other authors, both from within that journal and from other journals worldwide. By classifying the journals and the other publications into broad subject groups, Elsevier is then able to see whether a research article makes references to other texts from a broad or narrow range of subject domains, and, conversely, whether texts written in a small or large range of subject domains make reference to that article. An article that cites texts from a broad range of subject areas is considered to be relatively multidisciplinary while one that makes references almost exclusively to articles and books in the same subject area as itself can be assumed to be strongly monodisciplinary.

The four journals that we have selected here are all attached to the broad category of 'Environmental Sciences'; that is, they deal with issues relating to the environment (or environments). The term 'environmental science' refers to the study of the effects of natural and unnatural processes, and of interactions of the physical components of the planet on the environment but it also can include, more loosely, research that looks at the political, moral and social aspects of human–environment interactions. The four journals that we have chosen are:

- *Agriculture, Ecosystems and Environment* (AEE)
- *Global Environmental Change* (GEC)
- *Plant Science* (PS)
- *Resource and Energy Economics* (REE)

While they all fall under the heading of Environmental Sciences, their citation profiles are distinct and help to place the journals at different points on a cline between monodisciplinarity and interdisciplinarity.

Figure 3.1 shows the citation data for our four journals in two periods, 1996–2000 and 2009–2013. We use these ranges because it gives an indication of what

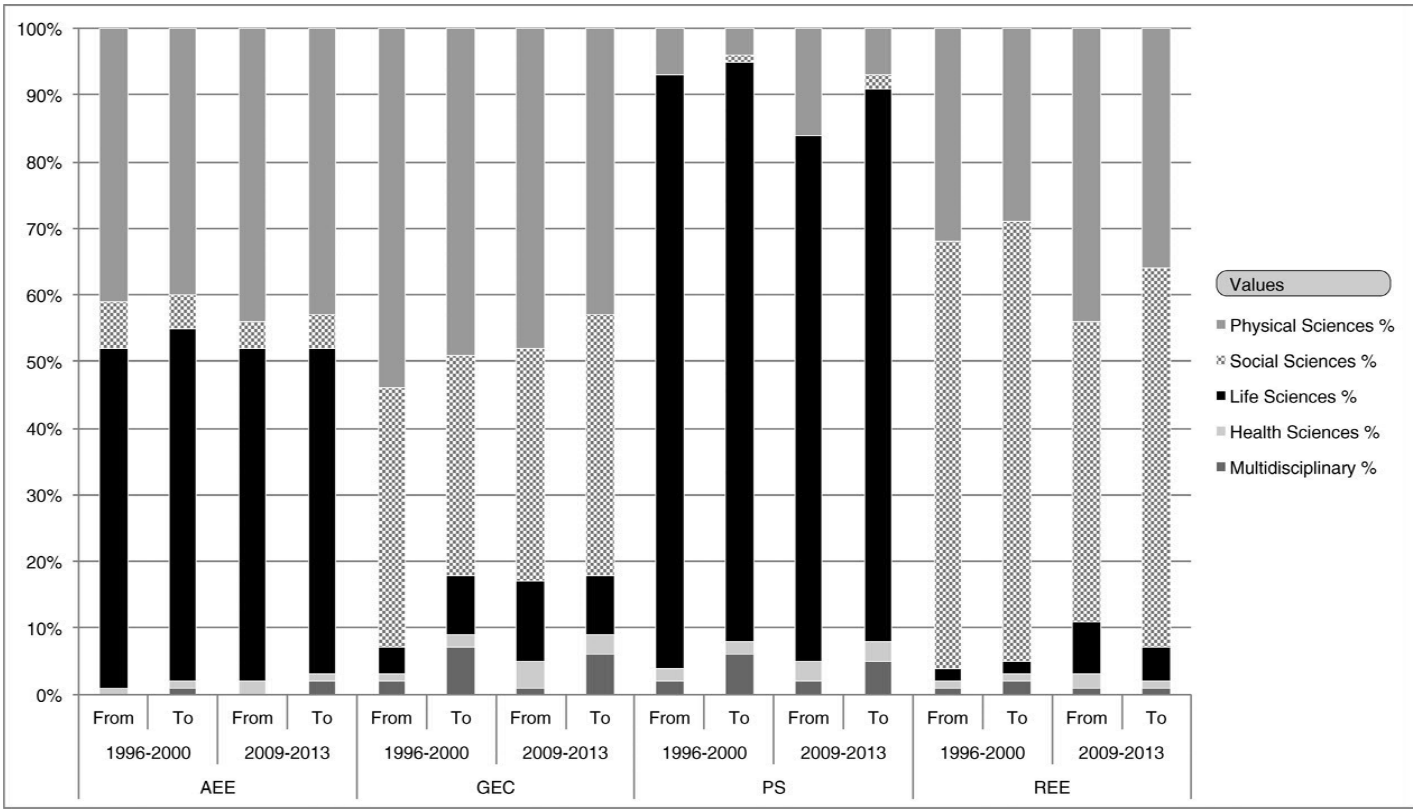

Figure 3.1 The disciplinary domains most cited by, and most citing, the four journals in two time periods, 1996–2000 and 2009–2013, expressed as percentages of all citations in a period. 'From' indicated that the citations were made **by** the journal, to other journals; 'To' indicates that the citations were **to** the journal, from other papers/books.

the journals were like in the periods leading up to the beginning and then at the end of the 2001–2010 period. In the chart, we can see the proportions of citations in each of the four journals to journals in broad disciplinary domain categories. The data are taken from Scopus, Elsevier's abstract and citation database, and the domain categories shown here are those employed within Scopus: Physical Sciences, Social Sciences, Life Sciences, Health Sciences and a final category called 'Multidisciplinary', which is a somewhat loose category for research that is difficult to categorise. The citations are estimated in two ways: firstly, the citations made by articles in the journal to other journals (addressing the question 'Which domains do the specific journal authors choose to cite?'), and secondly citations of articles in the journal by authors of articles in other journals ('Which domains choose to cite the articles in the specific journal?').

We can see that one journal, *Plant Science* (hereafter, PS), refers to and cites predominantly one main group, that of the Life Sciences, and, if we take adherence to a single disciplinary domain to be a marker of monodisciplinarity, then it is most clearly, of all the four, monodisciplinary. *Global Environmental Change* (hereafter, GEC) appears to have the most varied profile – although it tends to cite mainly two domains in the first period (1996–2000) this tendency has changed in the second period with increased citation of the 'multidisciplinary' domain and also the Life Sciences. *Agriculture, Ecosystems and Environment* (hereafter, AEE) is predominantly Life Sciences and Physical Sciences, while *Resource and Energy Economics* (hereafter, REE) is predominantly based in the Social Sciences with a secondary relation to Physical Sciences. It is interesting to note, in passing, that REE is more cited by Life Sciences journals than vice versa.

So far, we have looked at the broad disciplinary domains referred to by articles in each journal but a more fine-grained perspective can be gained by looking at data for the subject areas that are cited. In Table 3.1 we can see the citation profiles for the four journals, here showing the subject areas that are referred to (we have excluded the citation **from** information) in the two periods.

Looking at the second set of data we can see that in PS the majority of references and citations are in the areas of Biochemistry, Genetics and Molecular Biology on the one hand, and on the other, Agricultural and Biological Sciences. Both of these subject areas come under the domain category of 'Life Sciences', and these two subject areas dominate the profile. The other journal that is at the monodisciplinary end of the spectrum, REE, looks more blended when we view the disciplinary domain data (it is a blend of Social and Physical Sciences) but when we look at the subject area information we can see that it is predominantly 'Economics, econometrics and finance', with over 50% of the citations in that category, although it is interesting to note that the percentage has dropped markedly between the two periods. 'Environmental Science' is the second most common category.

Of the four journals, PS has only three subject areas with three or more percent of citations in one or the other period, and then REE and AEE have six each. The crucial difference between them is that REE is dominated by one subject area (Economics), and is thus treated as a monodisciplinary journal, while we place AEE in the interdisciplinary category. The final journal, GEC, is the most diverse of the four, with ten subject areas referred to by articles in the journal relatively frequently.

Table 3.1 The subject areas most cited by the four journals in two time periods, 1996–2000 and 2009–2013, expressed as percentages of all citations in a period. The criterion for inclusion is that the percentage must be a value of three or more in either period.

Agriculture, Environment and Energy

Subject area	*1996–2000*	*2009–2013*	*Percentile change*
Agricultural and Biological Sciences	45	43	–2.15
Environmental Science	32	33	0.85
Earth and Planetary Science	6	8	2.05
Social Sciences	3	3	–0.01
Immunology and Microbiology	4	2	–1.63
Biochemistry, Genetics and Molecular Biology	3	2	–1.03

Global Environmental Change

Subject area	*1996–2000*	*2009–2013*	*Percentile change*
Environmental Science	33	31	–1.31
Social Sciences	21	26	4.29
Earth and Planetary Science	16	8	–7.37
Agricultural and Biological Sciences	7	7	0.66
Economics, Econometrics and Finance	5	6	1.09
Multidisciplinary	5	5	–0.34
Arts and Humanities	3	5	1.99
Medicine	2	3	0.98
Energy	3	2	–0.8
Business, Management and Accounting	3	2	–0.28

Plant Science

Subject area	*1996–2000*	*2009–2013*	*Percentile change*
Biochemistry, Genetics and Molecular Biology	46	42	–3.48
Agricultural and Biological Sciences	42	41	–0.33
Multidisciplinary	4	4	0.26

Resource and Energy Economics

Subject area	*1996–2000*	*2009–2013*	*Percentile change*
Economics, Econometrics and Finance	54	42	–11.9
Environmental Science	20	23	3.11
Engineering	6	8	1.85
Social Sciences	6	7	1.64
Agricultural and Biological Sciences	2	4	1.87
Mathematics	3	2	–0.59

The picture that emerges from the citation profiles, therefore, places PS at one end of the spectrum, as the most clearly monodisciplinary of the four journals, with REE appearing as a monodisciplinary journal that engages with other disciplinary domains but is predominantly economics-based. GEC is the more varied of the two interdisciplinary journals, and therefore the most dissimilar to PS, on the evidence so far considered, and AEE is in the area between REE and GEC.

3.3 Corpus details

As explained in Chapter 1, the initial Birmingham-Elsevier Environment corpus consists of articles from 11 journals, and we call this the BEE11 corpus. This was the corpus created for a two-year ESRC-funded project into interdisciplinary research discourse, and analysis of that corpus has been presented in two articles, Thompson et al. (2017) and Murakami et al. (2017). Details of the number of texts, and the token count for each journal sub-section, in the BEE11 and the BEE4 corpora are shown in Tables 3.2 and 3.3. As can be seen, the number of files in a journal varies widely, with REE having the lowest number (219) while PS in BEE4

Table 3.2 Corpus information, BEE4 for the period 2001–2010.

Journal	*Number of files*	*Tokens*
AEE	1784	7,848,705
GEC	475	2,643,273
PS	2046	7,289,014
REE	219	1,421,647
BEE4	*4524*	*19,202,639*

Table 3.3 Corpus information, BEE11, for the period 2001–2010, for all journals except GEC, which is for the period 1991–2010.

Journal	*Number of files*	*Tokens*
Agriculture, Ecosystems & Environment (AEE)	1784	7,848,705
Advances in Water Resources (AWR)	1151	6,602,544
BioSystems (B)	1034	4,436,812
Computers, Environment and Urban Systems (CEUS)	365	2,070,056
Environmental Pollution (EP)	3570	14,409,138
Global Environmental Change (GEC)	675	3,971,520
Journal of Rural Studies (JRS)	331	2,667,967
Journal of Strategic Information Systems (JSIS)	168	1,204,412
Plant Science (PS)	2046	7,289,014
Resource and Energy Economics (REE)	219	1,242,819
Transportation Research Part D: Transport and Environment (TRTE)	358	1,421,647
All journals	*11,703*	*52,963,583*

has 2046, and in BEE11 the journal *Environmental Pollution* published 3570 full-length research articles in that ten-year period.

An important point to note is that the BEE4 corpus contains all the research articles for the ten-year period, 2001–2010, but the BEE11 corpus contains all the articles in a 20-year period for one of the journals: *Global Environmental Change*. The ESRC-funded project focused on the *Global Environmental Change* journal and therefore a wider time span was selected for that project. For most of the work in this book, the BEE4 corpus was used unless otherwise stated. Some of the investigation, however, was conducted on our corpus data through the Sketch Engine interface (Kilgarriff et al. 2014), where the *Global Environmental Change* data is the original 20-year collection. In that case, we are dealing with four journals but one of them has a larger time span (of 20 years) and we refer to this as the BEE4+ corpus.

3.4 Aims and scope statements

As is the case with most international journals, Elsevier journals publish a statement of the aims and scope of the journal both on the journal website and also within the printed copy of at least one journal issue each year. The aims and scope statement in a journal is an indication of editorial policy and the type of readership that the journal is aimed at. Elsevier's website in the period 2004 to 2008 had a section for each journal that was labelled 'audience' where the editors could describe the target audience, and we make reference to these descriptions in the following sections. This feature, however, disappeared when Elsevier redesigned its website in 2008.

As explained previously, the purpose of this survey of the statements of aims and scope for each journal is to identify both the stated audience of the journal and also the audience implied by the tone and content of each statement piece. These statements are an indication of whether or not the journal is aimed at a broad interdisciplinary audience and also an indication of what 'interdisciplinary' is conceived to be. We begin with the two interdisciplinary journals: AEE and GEC.

3.4.1 Aims and scope: AEE

At the beginning of the 2000s, AEE provided a statement of aims and scope (see Box 3.1).

This statement positions the journal as interdisciplinary (it publishes papers that 'develop and apply interdisciplinarity') and it is worth noting that the goal is to span **scientific** disciplines. This is restated in the third sentence which identifies the intended readership as 'scientists'. Although the final sentence of the second paragraph refers to economic and sociocultural perspectives as well as biological and physical, the emphasis throughout is more on the latter two than on the economic and sociocultural. This is evident also in the statement of audience given on the Elsevier journal webpage (see Box 3.2).

In 2006 the journal modified its statement of aims and scope (see Box 3.3).

Box 3.1 ***Agriculture, Ecosystems & Environment*** **aims and scope statement, 2004.**

Agriculture, Ecosystems & Environment deals with the interface between agriculture and the environment. Preference is given to papers that develop and apply interdisciplinarity, bridge scientific disciplines, integrate scientific analyses derived from different perspectives of agroecosystem sustainability, and are put in as wide an international or comparative context as possible. It is addressed to scientists in agriculture, food production, agroforestry, ecology, environment, earth and resource management, and administrators and policy-makers in these fields.

The journal regularly covers topics such as: ecology of agricultural production methods; influence of agricultural production methods on the environment, including soil, water and air quality, and use of energy and non-renewable resources; agroecosystem management, functioning, health, and complexity, including agro-biodiversity and response of multi-species ecosystems to environmental stress; the effect of pollutants on agriculture; agrolandscape values and changes, landscape indicators and sustainable land use; farming system changes and dynamics; intergrated [sic] pest management and crop protection; and problems of agroecosystems from a biological, physical, economic, and socio-cultural standpoint.

Box 3.2 ***Agriculture, Ecosystems & Environment*** **audience statement, 2004.**

Audience: Scientists in Agriculture, Forestry, Ecology and the Environment, Administrators and Policy-Makers in these fields

Box 3.3 ***Agriculture, Ecosystems & Environment*** **aims and scope statement, 2006.**

It seeks to provide a science forum to discuss how agricultural practices influence the broader environment and how changes outside agriculture affect agricultural systems. Emphasis is placed on innovative and original research that elucidates the link between agroecosystems (association of crops, pastures, livestock) and the environment (including energy, air, water, land). This two-way interface or link cuts both ways as it elucidates the influence of agriculture on the environment but also characterizes how changes in the latter impact agroecosystems.

The phenomena that are dealt with are primarily at the level of complex systems and 'the environment'. Although agricultural practices evidently involve human beings, there is no direct mention of human activity. In other words, the emphasis is on systems and on science.

3.4.2 Aims and scope: GEC

GEC has a shorter history than AEE. The statement of aims and scope for the *Global Environment Change* journal in its first issue (1990) can be found in Box 3.4.

The statement emphasises the focus on human impacts both in the sub-title (Human and Policy and Dimensions) and also in the third of the three sentences making up the statement (where both 'human contributions' and 'human responses' are mentioned). The statement also indicates the processes and phenomena that are the concern of the journal rather than the names of academic fields of study, and attention is focused on which real-world phenomena constitute the subject of enquiry, rather than on branches of the discipline.

In 2004 the statement was changed to make the interdisciplinary nature of the journal much more evident (see Box 3.5).

While AEE aims to bridge scientific disciplines within the natural sciences, GEC has the goal of spanning the social **and** natural sciences. The definition of global environmental change that the journal works by is also clearly stated, as 'the outcome of processes that are manifest in localities, but with consequences at multiple spatial, temporal and socio-political scales'. It is also interesting to note that the scope statement mentions 'issues' that the journal deals with – this indicates that the journal is strongly problem-oriented.

Finally, the explicit statement of the audience that is envisaged for the journal is shown in Box 3.6.

Box 3.4 *Global Environmental Change* aims and scope statement, 2001.

Global Environmental Change: Human and Policy Dimensions is an international journal that addresses the human ecological and public policy dimensions of environmental processes that threaten the sustainability of life on Earth. These processes include but are not limited to: deforestation, desertification, soil degradation, species extinction, sea level rise, acid precipitation, groundwater contamination, destruction of the atmospheric ozone layer, atmospheric warming/cooling, marine pollution, nuclear hazards, the emergence of new technological hazards, and the worsening effects of natural disasters. The journal emphasizes human contributions to worldwide environmental changes and explores the diversity of human responses to impacts of global change.

Box 3.5 *Global Environmental Change* aims and scope statement, 2004.

Global Environmental Change: Human and Policy Dimensions is an international, interdisciplinary journal spanning the social and natural sciences. It publishes high-quality original theoretical and applied research and review articles across the entire field of global environmental change. Areas include biodiversity and ecosystem services, water resources, climate change, international agreements, North-South relations, land use and cover change, institutions and governance. The journal interprets global environmental change to mean the outcome of processes that are manifest in localities, but with consequences at multiple spatial, temporal and socio-political scales. The journal addresses issues of public policy, economics, equity, risk, and resilience, science policy, international development, and health and well-being.

Box 3.6 *Global Environmental Change* audience statement, 2004.

Academics and researchers working in the policy sciences, environmental sciences and social sciences; policy makers and managers in government agencies, intergovernmental organizations that are affected by or concerned with global change; and organisations affected by or concerned with global environmental change strategies.

The audience is made up of not only those involved in the research world, broadly speaking, but also of policy makers and managers in a range of organisations (the types of which are more clearly specified than in the audience statements for AEE) and also organisations. This is clearly a more diverse audience than that mentioned by AEE.

3.4.3 Aims and scope: PS

We turn now to the two monodisciplinary journals. The first of these is PS, which describes itself as an international journal of experimental plant biology. The audience that is identified for this journal, by contrast with that for GEC, is exclusively researchers: plant biochemists, physiologists, molecular biologists and plant virologists.

The statement of aims and scope uses technical language (for example, words such as ‘protists’, ‘cyanobacteria’, ‘enzymology’, ‘phytopathology’) and assumes that readers belong to the plant biology community (see Box 3.7).

This statement is then followed by a list of specific areas that are typically covered by the journal, such as Enzymology and Intermediary Metabolism, Nucleocytoplasmic Interactions or Ultrastructure and Cellular Fractionation.

In the middle of the decade, the aims and scope statement changed and the journal became more defined in its policies for exclusion rather than stating simply what it includes (see Box 3.8).

Box 3.7 *Plant Science* aims and scope statement, April 2005.

An international journal of experimental plant biology which publishes, in the minimum time, papers in plant physiology and biochemistry, genetics and molecular biology, cell biology, and related areas.

Papers describing experimental work with multicellular plants and plant-like microorganisms (protists) including algae and fungi, and blue-green algae (cyanobacteria) are appropriate.

Submission of manuscripts in the broad areas of biochemistry (including enzymology), bioenergetics, cell biology, development, genetics, molecular biology, physiology and structure in relation to function, as well as fundamental work in these areas related to agriculture and phytopathology are invited.

Audience

Plant biochemists, physiologists, molecular biologists and plant virologists

Box 3.8 *Plant Science* aims and scope statement, December 2006.

Plant Science is not a publication vehicle for preliminary observations, for studies that are merely confirmatory of results accomplished in other plant systems, or that are addressed only to a specialized readership. Submitted manuscripts describing studies of preliminary nature, that are merely confirmatory in scope or of limited scope and interest to the general readership of the journal will be returned without formal review. As a general policy, the journal will no longer consider manuscripts just describing the cloning, sequencing and expression patterns of gene sequences that have been identified and characterized in other plant species, expression of a protein without proper characterization of the recombinant product and native form, the isolation, purification and characterization of enzymes extensively analyzed in other plants, description of molecular markers and plant genetic diversity without any relevant biological information, cell culture and/or transformation of plant species (or closely related species) already established in the literature and basic transcriptomic approaches with no further relevant functional characterization of genes of interest.

The statement indicates that articles must be written for a general readership (they must not be 'of limited scope and interest to the general readership of the journal) but it is clear that this is a general readership of biologists working in plant science. These scientists are expected to understand the types of submissions that are not invited – they will know about expression of proteins, isolation, purification and characterization of enzymes and so on, and regard these as basic aspects of scientific research in plant science.

3.4.4 Aims and scope: REE

The fourth and final journal is REE, which according to its subtitle is 'A Journal of Resource, Energy and Environmental Economics' and its audience is 'Energy Economists, Environmental Economists and Resource Economists'. The journal defines itself very clearly as an economics journal even though, as seen previously, many of the references and citations made in the journal are to texts in the physical sciences (environmental science) as well as social sciences (economics), and the proportion of the former relative to the latter increased during the period 2001–2010.

The editorial of a 2002 special issue of the journal explained that the issue was dedicated to George Tolley, 'one of three editors who, in the late 1970s, founded Resources and Energy, the predecessor to REE. . . . With the move from a multidisciplinary orientation to an economics disciplinary journal in 1993 (along with a change in title), George was instrumental in assuring the intellectual integrity of the newly revitalized journal' (Kolstad 2002: 1). This reinforces the identity of the journal as a monodisciplinary economics journal.

The aims and scope statement that appeared at the front of the printed collection in the early 2000s can be found in Box 3.9.

The statement explains what the journal aims to be (a forum), what the subject matter is and the questions that are addressed. These are clearly economic

Box 3.9 *Resource and Energy Economics* aims and scope statement, 2002.

Resource and Energy Economics provides a forum for high level economic analysis of utilization and development of the earth's natural resources. The subject matter encompasses questions of optimal production and consumption affecting energy, minerals, land, air and water, and includes analysis of firm and industry behavior, environmental issues and public policies. Implications for both developed and developing countries are of concern.

The journal publishes high quality papers for an international audience. Innovative energy, resource and environmental analyses, including theoretical models and empirical studies are appropriate for publication in *Resource and Energy Economics*.

Box 3.10 *Resource and Energy Economics* aims and scope statement, 2002.

Resource and Energy Economics publishes papers that advance economic theory and empirical methods to gain novel insights into environmental problems. We welcome innovative and high-quality papers that enhance our understanding of the economic causes and consequences of environmental and natural resource problems, and that contribute to discussions on improving environmental policy and resource management.

[. . .]

The ideal *Resource and Energy Economics* article starts from a well-defined economic problem, develops a transparent analytical model, and uses the appropriate theory, econometrics, or numerical techniques to give a novel perspective on the problem. Papers without strong links to environmental and resource theory are beyond the scope of *Resource and Energy Economics* and will be returned to authors without review.

questions, but stated in simple terms. It also describes in positive terms the types of articles that are encouraged (see Box 3.10).

The editors portray the journal as problem-oriented, similar to GEC with its identification of 'issues'. The ideal article is also described as being one that develops an analytical model for examining the problem and applying techniques or theory to the analysis of the problem. The focus is on the economic 'causes and consequences' but it is likely, given our knowledge of the citation profiles for REE that article authors must also have an understanding of environmental problems and thus refer to, and are cited by, the Physical and Life Science literature.

3.5 Conclusion

From the survey of the aims and scope statements for the four journals, we can see that the categorisation of the journals as 'interdisciplinary' or 'monodisciplinary' is supported. In the monodisciplinary group, PS defines its audience exclusively in terms of scientific researchers in the area of plant science, while REE states clearly that it is an economics journal, aimed at an audience of 'Energy Economists, Environmental Economists and Resource Economists'. In the interdisciplinary set, the statements for GEC indicate that the journal is aimed at a varied audience, comprising not only researchers (in policy, social or environmental sciences) but also policy makers, managers and organisations of differing sizes. AEE also addresses a non-researcher audience but its researcher community is composed mainly of natural scientists.

In the case of AEE, we speculate that the cultural variation between those who make up the readership for the journal is less varied than it is for the readership of GEC. By this, we mean that the readers of AEE are more likely to have shared

epistemological bases about how research is conducted, about the approaches used by researchers to test, form, challenge and construct new knowledge, and about the ways in which knowledge is communicated. Conversely, in GEC, it is likely that a consequence of the diversity of readership is that authors will have to 'unpack' their concepts, techniques and arguments as they cannot assume a shared epistemological base. Both journals are interdisciplinary journals, but in the one journal the epistemological distance between authors and audience is less than it is in the other journal. We propose that the former is a 'proximal' journal, in terms of its sharing of broad research orientation, while the latter is 'distal'. That is, 'interdisciplinary journals' is not a homogeneous category, and the journals may vary on a number of parameters. Here we suggest one parameter, that of closeness of epistemology, in the concept of 'proximal' and 'distal' journals. This distinction is similar to the narrow and broad interdisciplinarity that Van Dusseldorp and Wigboldus (1994) discussed, where 'narrow' describes contexts in which the disciplines share more or less the same paradigms and methods, while 'broad' refers to teams which face difficulties in communication because the members have different paradigms and knowledge cultures. We choose to use different terms ('proximal' and 'distal') as our focus in this book is on writing for interdisciplinary audiences rather than on interdisciplinary research team activities.

A further question, however, is whether the proximal/distal parameter exists only in interdisciplinary research contexts, or whether it is a factor within disciplines as well. As discussed in Chapter 2, Becher (1989) distinguished between convergent and divergent disciplines, where he postulated that the more convergent a discipline is, the greater the degree to which members of a discipline share common models and practices, and vice versa in the case of divergent disciplines. The concept of 'proximal' and 'distal' journals maps well onto Becher's terms, except that we apply our terms not only to disciplinary groupings but to journals too. If we consider the parameters of proximity and of mono-/interdisciplinarity as two axes then we can postulate that the four journals in our corpus can be placed in different points along the two axes as shown in Figure 3.2.

This conceptualisation is taken up in the following five chapters where we test and elaborate on the distinctions that have been postulated here. In Chapter 4, we investigate the types of articles that appear in each journal, looking at the section headings for all articles, in order to explore the degrees of conventionalisation that exist in each journal and also the diversity of roles that headings can play in each journal. Chapter 5 contains analysis of introduction sections taken from a sample of articles for each of the four journals, and Chapter 6 presents a discussion of how a number of significant terms are used by researchers in each of the four journals. In Chapter 7, we examine the use of status markers (Hunston 2011) within the journals in order to see whether and in what ways the journals differ in their uses of these markers and also the roles that these markers play in the research discourses of each journal. Chapter 8 investigates the uses of code glosses.

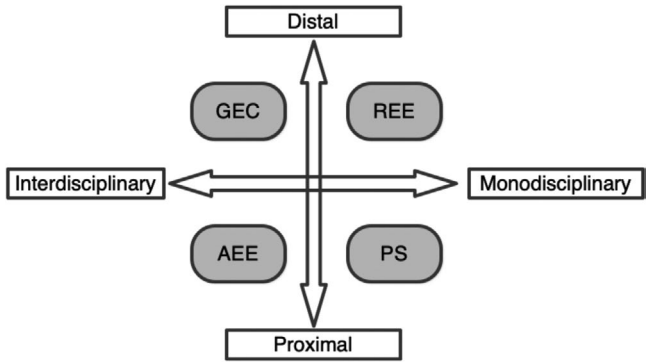

Figure 3.2 The four journals placed at different points along the axes of disciplinary status and proximity.

References

Becher, T. (1989). *Academic tribes and territories intellectual inquiry and the cultures of disciplines*. Milton Keynes: SRHE and Open University Press.

Giannoni, D. S. (2010). *Mapping academic values in the disciplines: A corpus-based approach*. Bern: Peter Lang.

Hardy, J. A., & Friginal, E. (2016). Genre variation in student writing: A multi-dimensional analysis. *Journal of English for Academic Purposes*, 22: 119–131.

Hunston, S. (2011). *Corpus approaches to evaluation: Phraseology and evaluative language*. London: Routledge.

Hyland, K. (2000). *Disciplinary discourses: Social interactions in academic writing*. London: Longman.

Kilgarriff, A., Baisa, V., Bušta, J., Jakubíček, M., Kovář, V., Michelfeit, J., Rychly, P., Suchomel, V. (2014). The sketch engine: Ten years on. *Lexicography*, 1: 7–36.

Kolstad, C. (2002). Preface. *Resource and Energy Economics*, 24: 1–2.

Murakami, A., Thompson, P., Hunston, S., & Vajn, D. (2017). What is this corpus about? Using topic modelling to explore a specialised corpus. *Corpora*, 12(2): 243–277.

Thompson, P., Hunston, S., Murakami, A., & Vajn, D. (2017). Multi-dimensional analysis, text constellations, and interdisciplinary discourse. *International Journal of Corpus Linguistics*, 22(2): 153–186.

Van Dusseldorp, D., & Wigboldus, S. (1994). Interdisciplinary research for integrated rural development in developing countries: The role of social sciences. *Issues in Integrative Studies*, 12: 93–138.

Yakhontova, T. (2006). Cultural and disciplinary variation in academic discourse: The issue of influencing factors. *Journal of English for Academic Purposes*, 5: 153–167.

4 Headings in the four journals

4.1 Introduction

Chapter 3 has shown how the four journals in our corpus differ in their aims and scope statements. The four journals appear, from the evidence, to address different audiences, and we proposed that they can be placed in different windows of a quadrant that is split along clines of monodisciplinarity–interdisciplinarity and proximity–distance. We now move on to explore the differences between these four journals, using both corpus and discourse analytic methods.

First, however, it is necessary to discuss exactly how similar the texts in each journal are to texts in the other three. The texts in the corpora are all 'articles' but what exactly is a journal article, and are all journal articles alike in their main purpose, their organisation and their sense of audience? Are the articles published in the four journals all 'research articles' or is there a nomenclature of article types that includes 'research article' as one type?

4.2 Organisation of research articles

A seminal text in the applied linguistics research literature on journal articles is Swales' (1990) book *Genre Analysis*. The main focus of that book was on the genre of the 'research article', or 'RA' for short, and the aim was to describe the prototypical organisation of such articles. The type of article that Swales analysed was the scientific report which typically consists of sections entitled 'Introduction', 'Methods', 'Results', 'Discussion' and 'Conclusions', although in some cases the results and discussion sections are joined together. As Gross, Harmon and Reidy (2002: 161–186) have shown, this is the format of scientific report that became conventionalized during the 20th century, as the natural sciences became increasingly institutionalized and as the need grew for faster production, review and reception of new scientific knowledge. This type of article structure is known as the 'IMRD' format, or sometimes as 'IMRaD'. It is the conventional form for empirical research reports in many fields of study. For example, Sollaci and Pereira (2004) report that original research articles in four leading medical journals that they studied (the *British Medical Journal*, *JAMA*, *The Lancet* and *The New England Journal of Medicine*) have used the IMRD structure exclusively since

1985. One of these four, the *British Medical Journal*, states explicitly on its website that submissions to the journal must follow the IMRaD style, and they should also include inter alia 'a succinct introduction that focuses – in no more than three paragraphs – on the background to the research question' (www.bmj.com/sites/default/files/attachments/resources/2018/05/BMJ-InstructionsForAuthors-2018.pdf, accessed 21 August 2019). This emphasis on succinctness is common in the physical and life sciences. The *Astronomy and Astrophysics* journal, for example, in its 'Aims and Scope' statement proposes that 'most scientific papers have the same structure' and that 'authors should have good reasons for deviating from it' as 'the goal of a scientific paper is not to impress the readers by poetic language but to transfer facts and new insights as lucidly as possible' (www.aanda.org/for-authors/author-information/paper-organization, accessed 21 August 2019). Efficient communication of facts and insights, then, is perceived to be the aim of scientific report writing. Interestingly, the format that the journal itself recommends is a slight deviation from the standard IMRD model of other scientific journals – in *Astronomy and Astrophysics*, a section called 'Observations', 'Calculations' or 'Mathematical derivations' replaces the 'Methods' section. However, this is a format that suits the discipline and which has become conventionalised. Fundamentally, the components remain the same, albeit in a different order.

Lin and Evans (2012) argue, however, that researchers in applied linguistics have tended to make assumptions about RA types and to assume that all empirical research articles follow the same IMRD pattern. When selecting articles for inclusion in a corpus, Lin and Evans observe, genre analysts are not clear about their criteria for choosing articles, although they tend to be explicit in their criteria for choosing journals, using information such as the journal impact factor rankings. In some cases, researchers have excluded articles that do not conform strictly to the IMRD headings (the example they give is Nwogu 1997), and this runs the risk of overlooking variations on the IMRD form. In their study of articles in a range of subject areas, Lin and Evans demonstrate that there are a number of variations around the IMRD pattern such as 'Introduction – Literature Review – Method – Results and Discussion – Conclusion (ILM[RD]C)', IM[RD]C, IMRDC and ILMRDC.

Lin and Evans were primarily concerned with criteria for data selection when considering which articles to be included in a corpus, and their article casts light on a problem involved in generalising from a large set of data. Van Enk and Power (2017) raise further questions about what is assumed to be 'research' in genre approaches to the research article, and they recount the problems that they encountered in trying to use the IMRD model in relation to research articles in the field of Education. The six sub-fields of Education that van Enk and Power investigated were found to differ widely in their epistemological orientations, and they found that trying to shoehorn all six into an IMRD structure was an impossible task. The most extreme disjuncture occurred in the case of Philosophy of Education, but they note also instances in the medical education journal where researchers adhered to an IMRD pattern even when they had no methods or results to report. In other words, then, there are journal articles that do not follow an IMRD model

and, in addition, there are articles that use the IMRD labels but in which one or more sections do not contain material that fits the label.

The corpus that has been established for this book contains all the articles published in four journals in a ten-year period, and we cannot assume that all the articles are of the same type (that is, that they follow the IMRD research article structure) nor that they perform broadly speaking the same functions. From a methodological point of view, one option would be to choose only those articles in each journal that belong to the given pre-determined category of IMRD RA, but the decision that we have made is to take all the full-length articles that have been published in the journals and then describe what variety we can find in each of these journals.

The IMRD format is not the only one used for articles published in academic journals, as a brief review of the internet and of journal contents pages can show. Let us take the example of economics. A guide to writing economics research articles on the University of Kansas website, for example, describes the six components of an economics research article as being, in the same order: introduction, data, model, estimation technique or methodology, findings, and conclusion (http://writing.ku.edu/sites/writing.drupal.ku.edu/files/docs/Guide_Writing_Economics.pdf, accessed 21 August 2019). A check on the table of contents for issues of three economics journals, however, shows that this model is not strictly adhered to, and the section headings are not conventionalised to the degree that empirical research articles in many science journals appear to be, based on an informal survey of three journals. The three journals that we surveyed are *Econometrics*, *Financial Economics* and the *Journal of International Economics*. For each journal, a regular issue was examined (with a deliberate decision to avoid special issues), from the year 2017. The one regularity was that all 23 articles began with an Introduction, and they all ended with a section called 'Conclusion' or 'Concluding remarks'. The average number of sections was not fixed at six, though the majority (N=19) did have five or six. Only six articles had a second section with the word 'data' in the heading, which is not to say that the sections are not about data but that, even if they are, the section does not have a standardised label.

Powers (2012), in a set of slides for a research methods in economics course at Harvard, draws a distinction between empirical and theoretical articles in economics, and further distinguishes between empirical articles that are observational and those that are experimental. For empirical articles, she suggests a pattern of Introduction, Data, Results, an optional section on Estimation, and Conclusion, while a theoretical article will contain an introduction, a literature review, a model that is developed through argumentation, and a conclusion. The division of articles into theoretical and empirical accords with Gray's (2015) observation that, in relation to variation in article types, discipline does not tell the whole story – in her study, Gray collected representative articles for her corpus in three categories: quantitative, qualitative and theoretical.

What, then, are the categories of article that can be found in academic journals, other than the IMRD type of research article? The publishing house, Springer, on its website (www.springer.com/gp/authors-editors/authorandreviewertutorials/

writing-a-journal-manuscript/types-of-journal-articles/10285504, accessed 21 August 2019), places journal contributions across its range of journals into five broad categories: 'Original Research', 'Short reports or Letters', 'Review Articles', 'Methods' and 'Case Studies'. It elaborates on the category of 'Original Research', noting that it includes 'full Introduction, Methods, Results, and Discussion sections' and that it is variously known as 'Original Article', 'Research Article'or 'Research', depending on the journal. From this description, the category of 'Original Research' article is equated to the IMRD model, and the alternatives are the apparently less substantial 'Short reports', the review articles, articles that are focused on methods, and the case studies.

A publisher that deals explicitly with a much wider range of article types is the open access scientific publisher Frontiers (www.frontiersin.org/, accessed 21 August 2019), which lists over 20 types of 'article' that can be published in their journals, with categories including 'Methods', 'Protocols', 'Review', 'Perspective', 'Data Report', 'Evaluation', 'Systematic Review' and 'Policy and Practice Reviews'. Frontiers currently publishes just under 60 journals and the list of article types is applied to all the journals, and consequently is quite loosely defined so that it can fit all journals. The description of 'review' articles, for example, does not distinguish between different types of review, other than to stipulate that the reviews should present the 'state of the art'. The taxonomy is useful for showing the diversity of article types but it does not explain much about how the different types of article are structured, unlike the IMRD model for research reports.

The IMRD article, as remarked previously, is a conventionalised text type that performs the function of allowing scientists to communicate their research concisely. The fact that an article is expected to follow this convention allows the reader to hold expectations of how the article will be framed and how it will develop. In some scientific disciplines, the order of the elements may vary – in some biology journals, the 'Methods' section can appear at the end of the article rather than after the Introduction (see, for example, the open access online journal *Plos Biology*); provided all articles in the journal follow this sequence, readers know where to find detailed information about methods.

Another aspect of the IMRD article is that typically the headings for each section say little about the content of each section. 'Methods', for example, tells the reader that the section will provide information about the methods that were used in the research, but it does not indicate which methods were used, whether they were innovative or standard methods in the field, or whether the researchers used quantitative, qualitative or mixed methods. Such additional information in a heading may be regarded as redundant in a research context in which the emphasis is on economy of words, and in which clear conventions exist, within a convergent community. Our claim, therefore, is that in a less convergent community there will be greater diversity in the organisation of articles, a broader range of research article types and different practices in the writing of headings for sections.

Van Enk and Power (2017) observed that headings are not always a reliable index of what is contained in a section, as, for example, when authors have used a conventional heading from the IMRD format, in order to follow the requirements

of the journal, even though the content of the section does not fit with the heading. It is also possible that what is implied by a term used in a section heading in one disciplinary domain may not be the same that is implied in a different field of research. The word 'observations', for example, can refer to what is seen, in one context, or can indicate that the section contains comments in another context. Even so, the degrees to which different journals adhere to the IMRD pattern can indicate the extent to which articles congregate around conventional approaches to empirical research that are perceived to be 'scientific'. Journals that do not contain a predominance of IMRD (or close equivalent) are likely to be less conventionalised and to be rather more distal in relationship between articles and research approaches.

To date, there has been little linguistic analysis of article section headings. Typically, in corpus work, headings are retained in the files produced for inclusion in a corpus, but they are annotated to identify them as 'headings', and not as part of the running text. The status of the headings in a broad corpus analysis is not clear – for example, it is not clear whether to include the words and phrases from headings in a description of the 'language' of academic articles or not. Do headings play a discourse role in the text, interacting with the paragraphs that follow them (and even those that precede them) or is their role primarily to label a text-part?

Rebeyrolle, Jacques and Pery-Woodley (2009) identified Rhetorical Structure Theory (Mann & Thompson 1988) as a linguistic theory of discourse that can accord headings a discourse role, and they illustrated this with Taboada and Mann's (2006: 426) use of the 'Preparation' relation to indicate the relation between a heading and the following sentence. The example given is that of a title ('The perception of apparent motion') and the first sentence of an abstract ('When the motion of an intermittently seen object is ambiguous, the visual system resolves confusion'). The title is what Mann and Thompson label the 'nucleus' and the sentence components act as 'satellite', where the relation between nucleus and satellite is that of 'Preparation'. The title/heading is a signal that helps a reader to build an interpretation of the text as a developing whole. An example from an article in the REE journal is as follows:

> The consequences of asymmetric information.
> We now consider the case where information is asymmetric in the sense that the value of the preference for the environment of each country is private information.
>
> [REE 2008 Hénin]

The 'asymmetric information' of the heading is picked up in the first line of the section in the 'case where information is asymmetric', with a clear lexical link between the heading and the first line. While we can say that the IMRD headings provide labels for the contents of the section, this type of heading acts both as a description of the content of a section and additionally interacts with the language of the ensuing text, sharing, in this case, exact lexical items, while in other cases the links may be between lexical items from the same semantic fields. For example, in the next extract, the 'threats to biodiversity' of the heading are picked

up in the following sentence by the words 'species extinction', which belong to a semantic field of 'threats to biodiversity':

> Threats to biodiversity.
> Human activities have dramatically accelerated the global rate of species extinction (reference).
>
> [REE 2004 Armsworth]

These two examples represent a type of heading in which a discourse relation between the section heading and the following text is linguistically marked in the heading and the opening sentence of the section.

A second type of heading that Rebeyrolle et al. (2009) describe is that of heading as 'theme' where the heading specifies the central topic of the ensuing section. Once again, the role of the heading is to help the reader to build a model for interpretation of the text as a developing whole, seeing how the parts relate to the whole.

In the following sections, we examine the first-level headings for all the articles in each of the four journals in our corpus in the period 2001–2010, with the aims of assessing the degree of variation in article organisation within each journal and also to examine the linguistic features of the headings.

4.3 Case study: headings in the four journals

We postulate that the amount of conformity in first-level section headings for research articles is indicative of the levels of conventionalisation within a journal and the readiness of researchers who write for that journal to adhere to those conventions. The use of an IMRD pattern of organisation is associated with a 'scientific' approach to research, and also linked to empirical research.

4.3.1 Headings in the two interdisciplinary journals

Looking at the two interdisciplinary journals first, we note that in the headings for the *Agriculture, Ecosystems & Environment* journal, there is a relatively high level of conventionalisation around the IMRD model of organization, with 83% of the articles in the period 2001–2010 using the headings 'Introduction', 'Methods', Results', 'Discussion' and (optionally) 'Conclusions' (Table 4.1).

The remaining 17% are composed of a variety of patterns with 62 articles (3.5%) using the word 'site' in one of the headings, as in the following example:

- Introduction
- Study site and GPP estimation from eddy flux measurements
- Description of the VPM model
- Estimation of VPM model parameters
- Results
- Discussion and conclusion

[AEE 2009 Yan]

Table 4.1 Percentage of articles in *Agriculture, Ecosystems & Environment* that have an IMRD structure.

Agriculture, Ecosystems & Environment	*Hits*	*%*
Number of files	1784	100
With 'Introduction'	1775	99
'Introduction' then 'Methods'	1531	85
IMRD(C)	1481	83

This is a model-based study, as described previously, with a section on the site of the study, which is relatively common for location-specific work in environmental studies. In several articles, the second section is dedicated to a contextualisation of the study, either in research terms (a review of the theoretical literature) or of the specific location of the study (e.g. 'Background: the tsunami disaster and mangrove replanting in Thailand'). These sections which contextualise the study could be treated as 'Literature Reviews'.

In stark contrast, only 7% of the articles in GEC use the IMRD pattern of organisation, and only 13% of the articles feature a 'Methods' chapter after the 'Introduction'. 18% of the articles do not even label the opening section as 'Introduction'. In the 77% of articles that start with an introduction but do not have a 'Methods' section following immediately, there is no clear pattern of what heading comes in second position (that is, it is not simply a question of another conventionalised set of section headings being used in place of IMRD) (Table 4.2).

There are eight articles that have 'Data' as the second heading, 14 that have a 'Background' section, but the majority have a heading that announces the theme of the section, as in the following:

- A brief history of phenological recording in Britain
  [GEC 2009 Lawrence]

- Climate change in the media
  [GEC 2010 Uusi-Rauva]

- Poverty in a challenging hydro-climate
  [GEC 2008 Enfors]

In most cases, these are headings for literature reviews or discussions of theory. Several of these headings are expressed as questions. Overall, in GEC, there are 68 question marks used in the first-level headings, in 475 articles, as compared to

Table 4.2 Percentage of articles in *Global Environmental Change* that have an IMRD structure.

Global Environmental Change	*Hits*	*%*
Number of files	475	100
With 'Introduction'	391	82
'Introduction' then 'Methods'	62	13
IMRD(C)	33	7

none in PS, five in 219 REE articles, and 32 in 1784 AEE articles. An example of the use of question marks in GEC is the following:

> Introduction
>
> - What is a Foucauldian analytics of government?
> - The making of the Earth System as a natural kind
> - The Anthropocene imagery
> - Implications: where does Earth System Science take us?
> - Conclusions
>
> [GEC 2009 Lövbrand]

This is clearly a type of article that is different from the conventional IMRD research article and it constructs the discourse as interactional (questions are posed, answers are proffered) rather than simply a recount of research actions. The headings construct the readers as active members of the discourse, and the headings are presented as more than labelling acts that follow a convention – they are a part of the dialogue that is established in the article.

4.3.2 Headings in the two monodisciplinary journals

Of the two monodisciplinary journals, PS is the most conventionalised, as would be expected on the basis of the citation profiles, and the aims and scope statements, discussed in Chapter 3. 97% of the articles follow the IMRD pattern and are labelled exactly as such. The heading 'Introduction' is given to all but seven of the first sections in the 2046 articles in the 2001–2010 period, and only 22 of those articles do not have a 'Methods' section immediately after the 'Introduction' (see Table 4.3).

The other monodisciplinary journal, REE, in contrast, has only one article out of 219 that follows the IMRD(C) pattern, and only four articles that have a 'Methods' section immediately after the 'Introduction'. A far greater number of articles (97) have a section called 'Model' (or similar) immediately after the 'Introduction' which, as discussed previously, is common in Economics research articles, and is consistent with the emphasis on economics in the journal (see Table 4.4).

Table 4.3 Percentage of articles in *Plant Science* that have an IMRD structure.

Plant Science	*Hits*	*%*
Number of files	2046	100
With 'Introduction'	2039	99.7
'Introduction' then 'Methods'	2007	98
IMRD(C)	1988	97

Table 4.4 Percentage of articles in *Resource and Energy Economics* that have an IMRD structure.

Resource and Energy Economics	*Hits*	*%*
Number of files	219	100
With 'Introduction'	213	97
'Introduction' then 'Methods'	4	2.8
'Introduction' then 'Model'	97	44
IMRD(C)	1	0.5

However, the number of articles following an 'Introduction – Model' structure is surprisingly low, at 44%. Of these 97 instances, too, many are in the form of 'The/A [defining characteristic] model' where the characteristic is a name or an adjective, such as 'A spatial decision model'. In some cases, the model is described not only in relation to the type of model but also its application to real-world activities, as in 'A bioeconomic model of salmon stocking', but these are rare.

Alternatives to the word 'model' are found in headings that mention 'scenarios', 'games', 'simulations' and 'framework', which all refer to theoretical models, but the choice of term is presumably defined by one's theoretical approach – one's adherence to game theory, or to scenario analysis, for example. Between scientific empirical research articles there can be differences in theoretical approach but this does not manifest itself in the choice of headings, where researchers opt to use the conventional headings, IMRD, but in REE the headings indicate that authors prefer to distinguish their models and their alternative approaches. In this, the articles can be interpreted as being typical of social science activity in the desire to differentiate research approaches that are based in distinct epistemologies (see Chapter 2) whereas in the natural sciences researchers are content to assert a uniformity of a 'scientific approach' that transcends differences.

The words 'discussion' and 'analysis' (25 occurrences) are relatively infrequent (31 occurrences in 219 articles), but the word 'results' (78) is more frequent. In the comparative absence of conventionalised headings (relative to PS and AEE), we

find articles with headings that refer to the specific phenomena that are the subject of investigation as in the following example:

- Introduction
- Dependence of hailstorm damage on temperature and precipitation
- Hailstorm damage insurance and data on hailstorm damage used in the analysis
- Estimation results
- Extrapolations of hailstorm damage using climate change scenarios
- Conclusions

[REE 2010 Botzen]

In this example, three of the six headings refer directly to the subject of the article (hailstorm damage) and make clear what treatment of the topic is presented in each section. One or more headings in articles in REE are often related to specific real-world phenomena or locations, as in 'Municipal water pricing and inefficiencies' or 'The Arctic National Wildlife Refuge', and economic models or theories are applied to these phenomena.

4.3.3 Analysis of a sub-corpus of headings

In this section, we explore the language of the first-level headings of all the articles in the BEE4 corpus, treating this as a sub-corpus. As has been observed, the most conventionalised of the four journals in terms of article structure and heading labelling is PS, and the least conventionalised is GEC. With this lower degree of conventionalisation in GEC comes a higher degree of linguistic diversity in the headings. Firstly, we can make an observation about the average number of headings per article in each journal, and the average number of tokens per heading, as shown in Table 4.5.

GEC and REE share quite similar profiles, with a greater number of headings per article and more tokens per article than the other two journals. As observed in

Table 4.5 Average numbers of first-level section headings per article, and tokens per heading for the four journals. (A substantial number of papers in PS have only three sections – Introduction, Methods, Results & Discussion – and thus the average is below four.)

Journal	*Average number of first-level section headings*	*Average number of tokens per heading*
Agriculture, Ecosystems & Environment	4.7	1.8
Global Environmental Change	5.7	3.9
Plant Science	3.9	1.7
Resource and Energy Economics	5.7	3.2

section 4.3.2, many of the articles in REE have the one of the words 'framework', 'model', 'scenario', 'estimation' in the headings as economics research features high levels of modelling. One distinction between REE and GEC is that, while these words account for 7.3% of the words in the REE journal headings, they account for only 2.1% of the words in the GEC headings. This suggests that GEC may contain articles written by economists but also that GEC contains a greater variety of articles, by researchers other than economists.

As we have observed in the previous section, headings in REE often provide information about what distinguishes the type of models used, and what the model is applied to. The majority of headings are noun phrases in all the journals and the noun phrases used in GEC and in REE tend to be more complex (in comparison to the single word IMRD headings typical in the other two journals) with premodification and/or postmodification evident in headings such as the following:

> Theoretical measures of compensating variation
> [REE 2009 Chattopadhyay]

> The trade-off between social efficiency and income
> [REE 2009 Antelo]

> Evolution of approaches to vulnerability
> [GEC 2006 Adger]

> Factors driving the development of a National Adaptation Strategy
> [GEC 2010 Biesbroek]

In the two examples for GEC, the noun phrase contains a labelling noun (*approaches*, *factors*) that describes a theoretical research construct and such nouns are common in the GEC headings, and also to an extent in the REE headings. Placing focus on theoretical research constructs, we have suggested, may be characteristic of social sciences research in general, where researchers tend to differentiate rather than converge. The use of such labelling nouns is a mark, therefore, that both journals contain social science research articles.

Another noticeable feature of headings in GEC is the use of words ending '-ing', such as *facilitating*, *exploiting* and *coping*. The uses of such words suggests a concern with processes. There are 357 instances of words ending '-ing' in GEC, of which 'concluding' accounts for 24. Omitting *concluding*, on the grounds that it does not refer to a research process or a process in the real world, and also omitting the word *during*, and then converting the figures to percentages so that we can see how frequent '-ing' words are, proportionally in headings in each journal, we find that '-ing' words account for 3.62% in GEC, 3.06% in REE, but only 1.08% in AEE and 0.31% in PS. This emphasis on processes further acts to distinguish the two social sciences journals from the other two. However, if we then look at the range of different words used, we find that there are 134 types (different word

forms) in GEC headings as opposed to 47 in REE. Among the words occurring in GEC that do not occur in REE are 'mapping' and 'assessing', which on the surface derive from some forms of spatial analysis and possibly from approaches to evaluation that are not exclusively quantitative. The prevalence and diversity of words ending '-ing' suggest that GEC is a journal that contains a wider range of article types.

4.4 Conclusion

In this chapter, we have observed that not all research articles follow the IMRD model of organisation. We proposed that adherence to the IMRD pattern indicates the degree to which empirical research articles in a journal converge, and that journals which differ from IMRD (or a close equivalent) tend to be less conventionalised and rather more distal in relationship between articles and research approaches.

We have taken a look at a part of research articles that has received little attention in the literature, that of the first-level headings in an article, and we have investigated conformity and variation. The two natural sciences journals tended to conform towards the IMRD model, while the two social science journals exhibited considerable variation. The IMRD headings are so heavily conventionalised that they say nothing about the specific content of the section, and they could easily be replaced by numbers. The headings in GEC and REE by contrast are often indicative of the topic of the article, and, while they are predominantly noun phrases, there are also instance of interrogative forms that can be seen as overt interaction with the reader and which also interact with the following text in a way that the IMRD headings do not. Finally, we also observed that the range of headings provides evidence of greater variation of article type in GEC than in REE.

References

Gray, B. (2015). *Linguistic variation in research articles: When discipline tells only part of the story*. Amsterdam: John Benjamins.

Gross, A., Harmon, J., & Reidy, M. (2002). *Communicating science: The scientific article from the 17th century to the present*. West Lafayette: Parlor Press.

Lin, L., & Evans, S. (2012). Structural patterns in empirical research articles: A cross-disciplinary study. *English for Specific Purposes*, 31: 150–160.

Mann, W., & Thompson, S. (1988). Rhetorical structure theory: Toward a functional theory of text organisation. *Text*, 8(3): 243–281.

Nwogu, K. (1997). The medical research paper: Structure and functions. *English for Specific Purposes*, 16(2): 119–138.

Powers, D. (2012). How to write a research paper in Economics. *Lecture Slides for Econ 191 Module*, University of California at Berkeley. Retrieved August 21, 2019, from https://eml.berkeley.edu/~webfac/eichengreen/e191sp12/powersecon191_1-17-12.pdf

Rebeyrolle, J., Jacques, M.-P., & Pery-Woodley, M.-P. (2009). Titres et intertitres dans l'organisation du discours. *Journal of French Language Studies*, 19(2): 269–290.

Sollaci, L., & Pereira, M. (2004). The introduction, methods, results, and discussion (IMRAD) structure: A fifty-year survey. *Journal of the Medical Library Association*, 92(3): 364–371.

Swales, J. (1990). *Genre analysis: English in academic and research settings*. Cambridge: Cambridge University Press.

Taboada, M., & Mann, W. (2006). Rhetorical structure theory: Looking back and moving ahead. *Discourse Studies*, 8(3): 423–459.

Van Enk, A., & Power, K. (2017). What is a research article?: Genre variability and data selection in genre research. *Journal of English for Academic Purposes*, 29: 1–11.

5 Introductions (and beyond) in four journals

5.1 Introduction

Chapter 4 explored generic variation in the journals in the BEE4 corpus. In the first half of this chapter we maintain the focus on text and genre in the same corpus, this time investigating the introductions from a small number of texts. This is followed by two word studies using the BEE4 corpus as a whole. Our aim is to discover how writers use the initial section of their texts to position themselves and their research in relation to their readership. We find that writers exercise choice, firstly about whether to identify themselves as insiders or outsiders in relation to the readership community, and secondly about what their research is relevant to, and the scale of its significance. By exercising these choices, writers are able to position their articles as interdisciplinary or not. They thereby construct their identities as more or less interdisciplinary.

We define 'introduction' as the first section in an article. Mostly this is headed 'Introduction', but in some cases there is an alternative or no heading (see Chapter 4). For this reason, the presence or absence of an 'Introduction' heading is not criterial, and it is simply assumed that the paragraphs occurring at the beginning of an article up to the first or next heading constitute the introduction.

In carrying out this study we bear in mind the distinction noted by Barry and Born (2013) between interdisciplinary work that consists of one discipline assisting another and that where one discipline challenges the assumptions or working methods of another. However, our argument is that these differences do not lie only in the type of project undertaken but in the writer positioning and identity, in line with our view that interdisciplinarity is discoursally constructed. In this chapter we exploit the fact that the purpose of an article introduction is to establish the interest and relevance of the article to the readership, and it is in this section, therefore, that the writer identity as 'colleague', 'supporter' or 'challenger' is first apparent. In the introductions we investigate we see overt statements of support and of challenge, as well as introductions that make no such overt statement and simply assume shared practices. In other words, the introduction may construe the article as outside the mainstream of the journal readership's interests and practices or as inside it. More generally, introductions may treat shared interest as taken for granted, or may devote considerable resource to establishing such common

ground. In short, the introduction construes the research as more-or-less aligned to other research in disciplinary terms.

To illustrate these ideas, we begin with the introduction to an article from *Plant Science*. The title of the article is 'Identification and characterization of differentially expressed ESTs in date palm leaves affected by brittle leaf disease' (PS 2010 Saidi). The analytical framework we use is the CARS (Creating a Research Space) model proposed by Swales (1981, 1990). This model analyses research article introductions using three moves: establishing the field, identifying the niche and occupying the niche. It suggests that the work an introduction does is to identify the research topic, and possibly also to assert its importance, to identify a 'gap' in existing research, and to promise to fill that gap in the rest of the research article. More detail is given in section 5.2.

This article has been selected because it has an unremarkable, if not necessarily typical, introduction. The introduction consists of 23 sentences (identified by the presence of full stops) spread over five paragraphs. Table 5.1 gives an idea of the content of the introduction.

In this introduction, the research is made relevant to non-academic concerns through the identification of date palms as an important crop (sentences 1–3). This aspect is not stressed, however: for example, we are not told how many people are totally reliant on the crop, or what level of suffering might result if the crop fails, and there is no proposal to study what date palm farmers think, say or

Table 5.1 The introduction to PS 2010 Saidi.

Sentences	*Summary*	*Sample quotation*	*CARS move*
1–3	Date palms are an important food crop.	*Dates are a major food and income source for major populations . . .*	1 Establishing the field
4	Problem of brittle leaf disease (BLD).	*. . . date orchards are being decimated by a new disease . . .*	
5–13	Details of BLD.	*The disease evolved in three stages.*	
14–17	Research on possible causes of BLD.	*. . . suggesting that BLD is a nutritional disorder . . .*	
18–20	Account of genetic research into disease.	*. . . plant sciences have advanced . . . allowing molecular level studies. . . . SSH . . . has been successfully used in higher eukaryotes to study the molecular response to different diseases.*	
21–22	Absence and importance of molecular pathway research.	*. . . no transcriptomic studies have been reported. . . . It is therefore of importance to elucidate the molecular pathways . . .*	2 Identifying the niche
23	Account of research procedure.	*. . . we used SSH to construct subtractive cDNA libraries between healthy and affected date palm leaves.*	3 Occupying the niche

do. Instead the study is mainly expressed as relevant to researchers interested in applying molecular level studies to the study of disease in plants. An inconsistency is reported in research results (sentences 14–17); this is resolved (sentences 18–20) by the use of a novel method: transcriptomic studies, a version of which (SSH) is the method that is used in this research. This method is located within 'plant sciences' (sentence 18). 'Success' in this introduction relates to studying the responses of plants to disease on a molecular level (sentences 18–20), 'importance' relates to conducting research of a similar kind; this is success and importance from the point of view of researchers who share aims and approaches with the article authors. In short, the authors present themselves as fully aligned with their readership. Common ground is assumed, as are shared goals. The introduction might be characterised as uncontroversially monodisciplinary, in that it contains nothing that is expressed as outwith disciplinary expectations. It is against the background of this conventional introduction that we look at introductions that potentially diverge from it.

The first part of this study is qualitative in methodology and for the most part involves close reading of specific texts. This text-based approach is then supplemented by an account of how individual words and phrases may be used as a starting point for further corpus-based investigations.

5.2 Introduction structures

As noted previously, the most influential work on research article introductions has been carried out by Swales (1981, 1990, 2004), and has been much cited and utilised since (e.g. Samraj 2002; Ozturk 2007; Loi 2010). Swales' account of introductions is located within a tradition of studying genres as staged responses to a communicative need. Hyon (1996) has argued that genre studies derive from three schools of thought: studies of New Rhetoric (e.g. Miller 1984); the Systemic Functional Linguistics tradition (e.g. Hasan 1996); and studies related to English for Specific Purposes (e.g. Bhatia 1993). Although these three traditions have different areas of emphasis, they agree that the texts belonging to a genre respond to a shared context and purpose and follow a shared pattern of elements or moves. It is this concern with contextualised strategic structure that characterises genre studies.

Swales' study of the structure of introductions culminated in the three-move CARS model (Swales 1990). This model has obvious appeal for people teaching novice academics how to write – it is a simple model that works. However, the model is much more than just a 'recipe' for how to write an effective introduction. It serves as an example of a convention that has emerged in response to a need: to make research appear valuable to a community. Swales' description of the moves (1990: 141) makes it clear how this is achieved:

- Move 1 establishes the topic of the research, defined more broadly or more narrowly. There may be an overt signal of centrality (e.g. 'this area is important'), or a less overt one (e.g. 'a lot of people are researching this these days'). The move establishes the relevance of the research, in general

terms, to the journal readership, and asserts common ground with that readership. It may also, as in the 'date palm' example in Table 5.1, establish the 'real world' relevance of the research or its relevance to a much broader group of researchers.

- Move 2 establishes a 'niche', that is, a need for research, because of a gap or an unanswered question. A broader or narrower group of researchers is assumed as experiencing that need.
- Move 3 explains how the present article will occupy the niche established by the gap or question in move 2. Move 3 asserts both the importance and the originality of the research and outlines the aims and/or methods, questions answered and so on. It does this in more or less detail: in some introductions the argument or findings of the research are summarised; in others they are not.

The three moves can be interpreted as a mechanism for the writer to position their work in relation to a community: in Hyland's terms (2012) to establish their proximity to a discipline and their position in relation to it. That interpretation risks essentialising research communities and disciplines, however. An alternative interpretation is that the moves serve to construe a community and the researcher's contribution to it. In other words, the introduction defines and positions the community for which it is written as well as the author's stance in relation to it.

A preliminary question to be answered in this chapter is: what introduction structures are observed in the BEE4 corpus? To answer the question, 12 articles from each journal were selected and their structures identified using the CARS model. The articles were selected using a pseudo-random method, avoiding special issues.[1] Two articles from each of 2000, 2002, 2004, 2006, 2008, 2010 were taken. In each introduction, transition places between moves were identified, marked by signals such as *however* and *in this paper*, then the content of the sentences between those transitions was scrutinised to establish the move. Table 5.2 illustrates this process with three examples per journal, showing the transitions between moves and the move label given. Identification of moves is fairly straightforward, with clear transitions occurring in most cases.

Table 5.3 shows the move structure of each of the 48 articles. As can be seen from that table, most articles follow a 1, 2, 3 structure. In some cases, one of the possible moves did not occur. The most commonly 'missing' move is 2. As can be seen in Table 5.3, some introductions in GEC and PS have moves 1 and 3, but no move 2. In many articles more than one instance of one or more of the moves was found. The most extreme instances of this are in REE. The article REE 2010 Aronsson, for example, has no fewer than eight moves in a recursive pattern: 1, 2 + 1, 2, 3 + 1, 2, 3.

It would appear, then, that the writers of all 48 articles adhere to a common approach towards structuring their introductions. Perhaps surprisingly, there are no major differences between the journals, with the exception that REE introductions frequently outline the structure of the article in the introduction and also summarise the argument in some detail (in line with Lakic's 2010 study of introductions in

Table 5.2 Examples of move transitions.

	Move 1	*Move 2*	*Move 3*
AEE			
2000 Maul	The Mississippi Alluvial Valley is an expansive intensively farmed area . . .	However, little information is available on the quality of water retained on fields.	Thus, the objectives of this study were . . .
2006 Arriaga	Assessing risk related to genetically modified organisms has been extensively discussed . . .	However, very few authors have approached assessing risk to biodiversity . . .	In this paper, a methodology to address risk assessment for biodiversity . . . was developed
2010 Acosta-Martinez	Water availability is a significant factor limiting agricultural production . . .	Thus, it is important to evaluate beyond 5 years the responses of soil microbial communities . . .	This study intends to evaluate the long-term sustainability . . .
GEC			
2002 Harrington	. . . concentrated animal agriculture became the dominant model for beef production.	Numerous studies have investigated . . . but relatively few researchers have addressed agriculturists' perceptions . . .	The purposes of this paper are to (1) provide an overview . . . and (2) analyze perceptions . . .
2006 Bakkenes	Global and European climate have [sic] changed notably in recent decades.	Several studies have tried to determine . . . Outcomes of these studies show considerable uncertainty.	Here we avoid this debate by evaluating the consequences . . .
2010 Zeitoun	Global food trade is receiving growing attention from environmental researchers . . .	Until recently, however, the magnitude of food trade . . . has only been guessed at.	This paper presents and discusses the research that quantifies that trade . . .

(*Continued*)

Table 5.2 (Continued)

	Move 1	*Move 2*	*Move 3*
PS			
2000 Jelenska	Plants, like other higher organisms, develop owing to the processes of cellular division, cell expansion and differentiation.	Although may stages of this process were described in detail, the precise chain of events leading to the appearance of a new organ still remains to be determined.	We have isolated four clones. . . . In this study we describe . . .
2006 Wu	Silicon in rice plants can increase photosynthesis . . .	Most research concerning the comparison . . . is conjectural. . . . There is no study directly using . . . Because roots are the uptake organ . . . a systematic analysis . . . is necessary.	The objectives of this study were (i) to compare . . . and (ii) to identify . . .
2010 Zhu	Phosphorus is an essential element required for various cellular processes . . .	Although miR399 has been identified . . . , our knowledge on miRNA regulation of P nutrition . . . is limited.	Using microarray as a tool, it is possible to perform high-throughput profiling. . . . Taking this advantage, we have determined the spatial expression patterns . . .
REE			
2002 Dupont	The world's fisheries are currently threatened on many fronts . . .	While the benefits and difficulties . . . have been assessed by many researchers, virtually no attention has been given to evaluating changes in fishing capacity brought about by the introduction of individual harvesting rights . . .	In this study, we examine harvesting capacity in Canada's . . . groundfish fishery. . . . Our findings suggest that . . .
2006 Feng	In recent years we have witnessed increasing interest in . . .	In spite of the potential . . . there is limited research on the efficiency of bankable permits.	Our work differs from previous studies . . . in the following three aspects.
2010 Aronsson	The importance of international cooperation . . . is widely recognised.	A relevant question is how the climate policy can be implemented . . .	In the light of these observations, the purpose of this paper . . .

Table 5.3 Introduction structures.

AEE		*GEC*		*REE*		*PS*	
2000 Downie	1, 2, 3	2000 Klooster	1, 3	2000 Damania	1, 2, 3, 3	2000 Jelenska	1, 2, 3
2000 Maul	1, 2, 3	2000 Haites	1, 3	2000 Considine	1, 2 , 3	2000 Migge	1, 3
2002 Choudhary	1, 2, 3	2002 Yohe	1, 3, 3	2002 Dupont	1, 2, 3	2002 Gairi	1, 2, 3
2002 Smart	1, 2, 3	2002 Harrington	1, 2, 3	2002 Chermak	1, 2, 3	2002 Shelton	1, 2, 3
2004 de Paz	1, 2, 3	2004 Boykoff	1, 2, 3	2004 Anderson	1, 2, 3, 1, 2	2004 Moraes	1, 2, 3
2005 Fiener	1, 3	2004 Kim	1, 2, 3	2004 Dijkgraaf	1, 3, 1, 3	2004, Trainotti	1, 3
2006 Raisei	1, 2, 3	2006 Bakkenes	1, 2, 3, 3	2006 Feng	1, 2, 3	2006 Dong	1, 2, 3
2006 Arriaga	1, 2, 2, 3	2006 Piao	1, 2, 3	2006 Maddison	1, 2, 3	2006 Wu	1, 2, 3
2008 Yadav	1, 2, 3	2008 Risbey	1, 2, 3	2008 Aprahamian	1, 2, 3	2008 Zhan	1, 2, 3
2008 Farahbakhshazad	1, 2, 3, 2, 3	2008 Bitzer	1, 2, 3	2008 Albers	1, 2, 1, 3, 1, 3, 2, 3	2008 Arencibia	1, 2, 3
2010 Menendez	1, 2, 3	2010 Zeitoun	1, 2, 3	2010 Schaffner	2, 3, 1, 2, 3, 2, 3	2010 Saidi	1, 2, 3
2010 Acosta-Martinez	1, 2, 3	2010 Pereira	1, 2, 3	2010 Aronsson	1, 2, 1, 2, 3, 1, 2, 3	2010 Zhu	1, 2, 3

economics articles). In other words, whatever differences in strategy are to be observed, they are not apparent from the introduction structure itself. The next section therefore takes a rather different approach.

5.3 Constructing difference: cases from a special issue

It seemed necessary at this point in the research process to look for unusual introductions rather than standard ones. The aim was to identify extreme cases of introductions that positioned themselves as 'other', as outside the mainstream of the journal's readership. If extreme cases could be found, we would then interpret what writers were doing in those cases, and to look again at the standard introductions for less overt traces of similar practices.

To identify the cases of 'otherness' we looked at the special issues between 2000 and 2010 in GEC. There are a total of 13 special issues in those ten years. These were precisely the issues that had been avoided in making the initial selection of 12 articles (see note 1). One issue, a 2009 issue (volume 19, issue 2) on 'Traditional Peoples and Climate Change', offered the most promising material. This issue consists of an introductory editors' article, six articles specific to the special issue, and 11 articles designated 'Regular Papers'. Of the six special issue articles, three stand out. There are: GEC 2009 Byg; GEC 2009 Lawrence; GEC 20009 Turner. The articles are noticeable because they very obviously position themselves as outside the mainstream of articles in the journal; in particular they present themselves as using a disciplinary approach to the questions of the journal that is different from the norm. Moreover, the three articles rather nicely illustrate the types of interdisciplinarity identified by Barry and Born (2013). As noted in Chapter 1, we take the useful ideas proposed by Barry and Born (2013) and adapt their terminology to describe the written outcome of research rather than to working methods. Using this adapted terminology, GEC 2009 Byg is an 'antagonist' article, while GEC 2009 Lawrence and GEC 2009 Turner are 'integrative' ones, with GEC 2009 Turner presenting itself as 'supportive' also. It is important that this difference is not inevitable or a consequence of the research itself – indeed, the research in each article is rather similar in nature – but is the result of choices by the writers as to how to position themselves and their work. Each of these article introductions will now be described in some detail.

GEC 2009 Byg: 'Local perspectives on a global phenomenon: climate change in Eastern Tibetan villages'

The introduction to this article consists of seven paragraphs. Move 3 begins at paragraph 4, with a statement about the current research: *Our study was concerned with . . . in China's Yunnan province*. Paragraphs 4–6 describe the situation in China, and paragraph 7 outlines the methodology used in the study: *In order to better understand . . . we conducted semi-structured interviews*. Preceding this extended move 3 there are three paragraphs that argue for the importance of investigating local perceptions of climate change (which the article itself does), in the context

of the failure of science adequately to model the social and economic impact of climate change. This constitutes move 2: establishing a gap in research, though the 'gap' is the inadequacy of the whole scientific research paradigm. There is in addition a short move 1 (the first sentence of the article): *Global climate change models have increased greatly in number and quality . . . thereby improving the scientific understanding*. In short the structure is:

Move 1 (sentence 1): 'Climate change has been increasingly and successfully modelled'. This move establishes the field and the extent to which research is successful.
Move 2 (paragraphs 1–3): 'Research based on physical science is inadequate. Local studies of traditional peoples are needed'. This move identifies a gap or problem with current research.
Move 3 (paragraphs 4–7): 'This article studies one such group of people that is badly affected by climate change'. This move offers the current research as a solution to the problem.

As explained previously, the introduction positions the article as antagonistic to physical science. This antagonism is expressed in a number of phrases, such as:

- *. . . cannot be adequately understood relying on science alone . . .*
- *. . . adequate modelling is impossible . . .*
- *. . . positivist science, which pretends to maintain the position of a neutral observer . . .*
- *. . . mostly absent from scientific studies and models . . .*
- *. . . issues that have been overlooked by science . . .*
- *. . . cannot be estimated through models . . .*

The evaluation of 'science' is overt and consists of negative statements of ability/possibility (*cannot*, *impossible* and the denial of 'adequacy'), statements of absence (*absent*, *overlooked*) and even implications of deceit (*pretends to*). This is contrasted with the value of the current approach:

- *. . . can promote the understanding . . .*
- *. . . can make valuable contributions in gaining a better understanding . . .*

These phrases are saturated (Martin & White 2005; Hunston 2011) with evaluative meaning: most of the words in them contribute to the positive evaluation (*can* + *promote* + *understanding*; *can* + *valuable* + *contributions* + *gaining* + *better* + *understanding*).

The whole is made relevant to research by filling a gap in knowledge, but is also made relevant to policy concerns and to the needs of ordinary people:

- *. . . there remains much uncertainty about . . .* [relevance to research]
- *. . . has largely been ignored . . .* [relevance to research]

- . . . *is important from a policy point of view* . . . [relevance to policy]
- . . . *the actual impacts of climate change on people's lives* . . . [relevance to people]

In short, the article uses the standard introduction format to construe an opposition between science disciplines and research relating to people and to position the second as more useful than the first. There is overt negative evaluation of a research paradigm that is uniquely scientific in orientation and saturated positive evaluation of the alternative perspective.

GEC 2009 Lawrence: 'The first cuckoo in winter: phenology, recording, credibility and meaning in Britain'

Like GEC 2009 Byg, this article overtly construes a novel approach to the topic of climate change (phenology in this case) in the introduction, comparing it with conventional 'science' approaches. Unlike that article, however, it offers an approach that is construed as complementary, adding to rather than supplanting other methods. The tone is conciliatory rather than combative.

As with GEC 2009 Byg, the introduction is relatively long: eight paragraphs. The first four paragraphs explain phenology, stating that this is what the article is about and locating phenology as complementary to scientific knowledge; paragraphs 5–6 exploit the complementarity of the two approaches in the context of the opportunities offered by the special issue; paragraphs 7–8 outline the focus (Britain) and method (unspecified analysis of interviews and bulletin board posts) used in the article. Although this article does not include unambiguous signals of transition between moves, moves can nonetheless be proposed. Our interpretation is:

Move 1 (paragraphs 1–4): Establishes the field of phenology as the topic of the article.
Move 2 (paragraphs 5–6): Identifies the novelty of the article in bringing together phenology and conventional science.
Move 3 (paragraphs 7–8): Identifies the scope and method of the article.

Unlike GEC 2009 Byg, this article takes its own discipline as its starting point and asserts the value of multiple viewpoints, without offering its own discipline as a remedy for shortcomings of conventional science. Instead, the article offers evidence for collaboration between two disciplines – *the rate of increase in papers about <u>both</u> climate change <u>and</u> phenology is even faster* (original emphasis) – and suggests that its own contribution is to increase understanding – *provides an opportunity to . . . understand better how context, experience and fact interact in a mutual process*.

In place of the critique of science, then, are statements about the consistency and complementarity of phenology and science. This is first presented as an anticipated but rejected opposition between two kinds of knowledge (example 1).

1 Indigenous knowledge and scientific knowledge are often presented as a dichotomy. . . . Agarwal invites us to reject this dichotomy. . . . Instead, knowledge creation is an evolving process that interacts with the evolving context.

[GEC 2009 Lawrence]

Also rejected is the proposal to move away completely from the paradigms of conventional science (example 2).

2 [W]e should be cautious in moving too far from the philosophy and structured experimental approaches that characterise science in its formal sense.

[GEC 2009 Lawrence]

In both examples, there is an implied dialogue (Martin & White 2005) between the writers and a particular readership: one that would share the authors' view of the insufficiency of traditional scientific research alone but who would hold a more antagonistic view of natural science disciplines.

There is a succession of phrases emphasising the compatibility of the two approaches:

- *. . . allowing for an active testing-out of knowledge that resonates with scientific practice*
- *. . . explore the characteristics of citizen science as a co-produced hybrid of local and scientific knowledge*
- *. . . at the interface between scientifically regulated knowledge formation processes and subjective personal experience*
- *. . . both parts of knowledge production*

One intriguing aspect of the GEC 2009 Lawrence introduction that is in contrast to GEC 2009 Byg is the use of first-person plural pronouns (*I* and *we*). In the GEC 2009 Byg introduction, there is one instance of *our* (***our*** *study*) and one of *we* (***we*** *conducted . . . interviews*). In these instances the pronoun is exclusive of the reader and refers only to the research team. There are no instances of *I*. The GEC 2009 Lawrence introduction contains five instances of *we/us/our* and seven instances of *I/me/my* (excluding instances in quotations). These pronouns construct three groups of people: (1) people living in southern England who observe the arrival and departure of swifts (*The air becomes full of their shrill calls and* ***we*** *know that the height of summer is here . . .*), and humankind more generally (***our*** *interactions with nature*); (2) researchers in the field of environmental change (*Agarwal invites* ***us*** *to reject . . .*), especially those with a social science perspective (***we*** *should be cautious in moving too far from . . . science in its formal sense*); and (3) the researcher herself (***my*** *conceptual framework . . .* ; ***I*** *explore the roles . . .*). This unifying of 'we ordinary people' and 'we researchers' enacts the theme of the article: the importance to science of lay-persons' observations. The alignment of the writer with a research community counteracts the 'otherness' of her message.

In particular, her message to social scientists that they should not reject scientific methods is softened by the inclusion of herself in the social science community (***we*** *should be cautious* . . .). We shall return to this point.

GEC 2009 Turner: 'It's so different today: climate change and indigenous lifeways in British Columbia, Canada'

One of the notable features of this article, authored by Turner and Clifton, is that the second author is not an academic researcher but is a leader in the community that is being studied. This joint authorship asserts an equal partnership between both participants in the research process and a rejection of the more 'scientific' researcher/subject model. The introduction consists of six paragraphs. The final paragraph constitutes move 3, outlining the article: *This paper discusses adaptations to environmental change*. . . . The need that the article answers (the gap in research) is articulated in the previous 2 paragraphs: *there is a need to understand* . . . , together with a statement of the relevance of studies of indigenous people that positions those people as sources of information rather than as the object of study: *Turning for help and insight to Indigenous Peoples makes great sense*. The move 1 that provides the background to this stated need explains not the present state of research but the situation of the people of British Columbia. In short we have:

Move 1: 'The indigenous people of British Columbia observe climate change that increases feelings of vulnerability'.
Move 2: 'Global research into climate change will benefit from using the observations of the British Columbia people'.
Move 3: 'This article explains what that local knowledge is and how it can be used'.

As with GEC 2009 Lawrence, the starting point from which the article is launched is not a general research question but an 'outsider' perspective: indigenous people in Canada observe the weather in the case of Turner; local inhabitants in England observe wildlife in the case of Lawrence. GEC 2009 Turner, however, makes fewer attempts to negotiate the intersection of its discipline with the mainstream, offering only one rationale for the importance of their study: *Turning for help and insight to these Indigenous Peoples makes great sense*.

The three articles described here might be said to be strikingly 'interdisciplinary', not because of the disciplines they espouse but because of the way they position themselves as outside a perceived mainstream. GEC 2009 Byg does this by problematizing science and asserting the superiority of the researchers' own approach. GEC 2009 Lawrence does it by asserting the complementarity of the two approaches. GEC 2009 Turner does it, almost as an aside, by offering the authors' research as assistance to others.

In terms of generic structure, it has been stated that the three-move CARS model satisfactorily accounts for the structure of these introductions. However, within

that model there are optional 'steps', meaning that each move might be realised in a number of different ways. For example, move 1, Establishing a territory, is said by Swales (1990: 141) to be expressed by one or more of three steps: Claiming centrality; Making topic generalisations; Reviewing items of previous research. For interdisciplinary articles of this kind it may be that another step is necessary for the analysis: Specifying discipline. Making this small adjustment would illustrate how the three introductions both conform to the standard introduction structure and yet adapt it to their own needs.

We now look more briefly at the other three articles in the targeted special issue. They are less remarkable in their approach to the ones discussed previously, but they do also position themselves as outside the expectations of the readership of the journal.

GEC 2009 Bridges: 'Living on the margin: ethnoecological insights from Marshall Islander at Rongelap atoll'

The three moves in this article are:

Move 1: An account of the particular features of atolls.
Move 2: Atolls are stated to be simpler (than continents) in terms of available resources. It is implied that they are therefore a useful site for research.
Move 3: This research explores two hypotheses.

The starting point for the article is the researchers' own interest in atolls, but these are then made relevant to the question of climate change in continents. In other words, continental climate change is positioned as the mainstream and climate change related to atolls as the relevant 'outside' interest. This is achieved through one sentence (example 3).

3 Life on an atoll generally operates at a different spatial scale than on continents.

[GEC 2009 Bridges]

The importance of the 'outsider' view is asserted in a cleft sentence (example 4).

4 It is in this local context that the potential impact of climate change must be viewed.

[GEC 2009 Bridges]

In other words, the message that is overt in GEC 2009 Turner ('We are outsiders but our research is relevant to you') is present but less overt in this article.

GEC 2009 Salick: 'Eastern Himalayan alpine plant ecology, Tibetan ethnobotany, and climate change'

The three moves in this article are:

Move 1: The importance of studies of alpine botany in the context of climate change.
Move 2: An account of the GLORIA network.
Move 3: The research question for the article.

This introduction positions itself firstly as being about alpine botany, making the case for the importance of this study in the context of climate change as a whole. The novel research focus (ethnobotany as a discipline and 'what are the human impacts' as a research question) appears in move 3 (example 5).

5 To GLORIA's ecological mandate, we have added Ethnobotany to ask: What are the human impacts of climate change in the Himalayan Alpine?
[GEC 20009 Salick]

There is a very implicit message here that ethnobotany is a discipline, or research method, that complements standard scientific ones.

GEC 2009 Grabherr: 'Biodiversity in the high ranges of the Alps: ethnobotanical and climate change perspectives'

The three moves in this article are:

Move 1: The relevance of Alpine plants to human use.
Move 2: The attested effect of climate change on Alpine plants.
Move 3: The applicability of ethnobotany to the Alps.

This article orientates itself towards GEC 2009 Salick, comparing the Alps with research carried out in the Himalayas. In this sense the Salick work on Tibet is taken as the 'given', to which the Alps work is the 'new' (example 6).

6 [O]ne might expect that a considerable portion of this [Alps] flora is ethnobotanically relevant as documented for the Himalayas . . . ; may be [as] relevant . . . for the Alps as . . . for the Himalayas.
[GEC 2009 Grabherr]

The discipline of ethnobotany, which GEC 2009 Salick positions as 'outsider' is here simply mentioned: *ethnobotanically mentioned*; *from an ethnobotanical perspective*. In other words, this article positions ethnobotany as the mainstream and the study of the Alps as the novel 'outsider'.

It is now possible to summarise what has been gained from this text-based study of articles from the special issue of GEC. Writers use the three moves from the CARS model to construe a research community, assert the importance of their research to it, and represent their research as standing in various relationships to

that community. It is that relationship that is important from the point of view of interdisciplinary discourse. In most cases the research is presented as coming from an 'outsider' discipline; thus the relationship is construed as interdisciplinary. In one case (GEC 2009 Byg), the relationship is presented as antagonistic. In another (GEC 2009 Lawrence), it is presented as complementary. In another (GEC 2009 Turner) the current research is presented as assisting mainstream research. The other three articles make less effort to negotiate a relationship between their research and the mainstream of the journal. The starting point for those journal articles is their own concerns, with the special issue interpreted as an opportunity to present this research to another audience.

The 'take-away' messages from this study are (1) the readership for each article, and the relationship between the disciplines represented are constructed discoursally; (2) discoursal construction involves establishing both convergence and divergence between disciplines, for example convergence of purpose and divergence of approach; (3) a special issue that is designed to present research from a particular academic community to a wider community is a good entry point for this investigation, but the messages taken from it can be applied more widely. This wider application is taken up in the next section.

5.4 Strategies in the introductions of AEE

The previous discussion has indicated that writers of introductions have the option of establishing, more or less explicitly, the common ground that exists between themselves and the imagined readership of the journal. They also have the option of construing their research approach – the questions and the methods used – as diverging from an accepted norm (or not). In the preceding discussion some extreme examples were considered as raising awareness of how these options might be exercised.

To take this discussion further we re-examined the selected 12 introductions from AEE to identify how the options are exercised and how this leads to strategies of introduction. AEE was selected as the target journal because, as argued in Chapter 3, it is a journal that is interdisciplinary, but the disciplines that contribute to it are proximal rather than distal in approach. From the 12 introductions, three strategic approaches to introduction writing can be identified, and these are exemplified in the following sections.

5.4.1 The 'narrow relevance' strategy

Sample article: AEE 2000 Maul.

This introduction takes a low-key approach to establishing relevance, does not relate the research to interests from other disciplines, and only marginally relates the research to the non-research community. It might be said to be inward-looking and to construct itself as monodisciplinary.

The first sentence of the introduction is a description of the place being studied (example 7).

7 The Mississippi Alluvial Valley is an expansive intensively farmed area within the southeastern United States characterized by highly erosive alluvial soil.

[AEE 2000 Maul]

This description includes size and location, but also indicates the relevance of the place to the journal (*farmed area*) and suggests a possible problem (*highly erosive*). There is, however, no overt statement of centrality. Importance is implied in a statement at the end of the first paragraph where a problem (*uncontrolled runoff*) is related to aquatic life (*damaging effects on aquatic invertebrate and fish communities*). A further statement at the end of move 2 relates the current research to research and non-research interests (example 8).

8 Knowledge pertaining to dynamics of sediment retention . . . could benefit both landowners and watershed researchers.

[AEE 2000 Maul]

In contrast to other articles, the statements of centrality are implicit rather than strongly stated, and are not placed at the beginning of the introduction. No person-related problem is indicated at the beginning of the introduction. Overall, the range of people influenced by the issue dealt with in the article is limited to landowners and watershed researchers. There is no statement of research method in the introduction. This is a journal article that stays within its own disciplinary boundaries, making little attempt to include a broader readership.

5.4.2 The 'broad relevance' strategy

Sample article: AEE 2006 Raiesi

This article is at the opposite end of the scale in terms of establishing convergence of interest. The introduction contains seven words (tokens) indicating importance (e.g. *crucial*, *important*, *necessary*, *significant*), as opposed to only one token in AEE 2000 Maul. Move 1 of the introduction, comprising four paragraphs, establishes the centrality of the topic of the article. Although the topic is 'carbon and N mineralization', the location of the study is a 'wetland ecosystem', and the article begins by establishing the importance of such places (example 9).

9 Wetland ecosystems are an important natural resource in the biosphere. . . . These ecosystems are particularly important in conservation biology . . . the global C cycle . . . and global climate changes.

[AEE 2006 Raiesi]

The relevance of the particular topic of the article is then established (example 10).

10 Carbon and N mineralization processes are of great important in maintaining soil quality and fertility, and hence agricultural sustainability.
[AEE 2006 Raiesi]

After an account of how the process is affected by agricultural practices such as ploughing, there is a summary statement, shown as example 11.

11 Plant residue management is, therefore, crucial in wetland soils.
[AEE 2006 Raiesi]

It might be said, then, that considerable work is carried out in the introduction to identify convergent interests, establishing the relevance of this particular research to other researchers and to the non-research world of agriculture. There is, however, no statement of research method or approach in the introduction. Instead the novelty of the article is formulated in terms of the research question it addresses, shown as example 12.

12 [I]t is necessary to understand the residue decomposability . . . and more importantly the interaction between soil cultivation and plant residue.
[AEE 2006 Raiesi]

Although the work being done here suggests that the article is expected to lie outside the readers' normal experience, it is not made clear whether or not the difference is a disciplinary one.

5.4.3 The 'divergent method' strategy

Sample article: AEE 2008 Farahbakhshazad

Three introductions from the AEE sample present their methodology as diverging from a norm, and each of them is about modelling as a method. In each of the three articles, modelling is presented as an alternative to, or supplement to, field research. In AEE 2008 Farahbakhshazad, field experiments are first described positively (example 13) and then negatively (example 14).

13 Field experiments play a key role in obtaining first-hand information about the effects of alternative management practices.
[AEE 2008 Farahbakhshazad]

14 However, most field experiments require extensive time and resources.
[AEE 2008 Farahbakhshazad]

The proposed method is then offered as a practical supplement to field experiments (example 15).

15 [P]rocess-based models have been developed and adopted to assist the policy making process in agricultural studies.
[AEE 2008 Farahbakhshazad]

In the other 'modelling' articles, the necessity of adopting this method is asserted even more strongly (examples 16 and 17):

16 Insight into these dynamics can only be accomplished through modelling.
[AEE 2008 Yadav]

17 Simulation models and GIS are powerful tools by themselves, but coupled together allow N modelling at a regional scale.
[AEE 2004 de Paz]

To sum up: in each introduction from the AEE sample, an amount of 'work' can be identified that the writer does in order to establish either the relevance of the question to be answered, or topic to be addressed, and to indicate the relationship of the proposed method to more established methods. The evidence offered to support this assertion – that introductions can be distinguished in terms of the amount of work done in this respect – is entirely qualitative, and judgements about it is mostly subjective. The amount of work cannot be quantified. What is being offered here is a suggestion of strategy types rather than a measurement. Bearing that in mind, however, it appears that the AEE introductions are relatively parsimonious in terms of the arguments they produce in favour of the relevance of their research and even more so in terms of the extent to which they stress the novelty and divergence of their research methods.

As noted previously, what is at stake here is not the 'reality' of the research environment but the way that environment is construed by the writer. When a writer indicates that the norm of research is to undertake field experiments but that modelling offers a useful short-cut alternative, this construes an environment in which experiments and modelling are discrete methods of enquiry but compatible and complementary methods. Other researchers, undertaking experiments or constructing models, may perceive themselves as belonging to the same or different disciplines.

5.5 Strategies in the introductions of GEC

The previous exercise was repeated with the 12 GEC introductions. (This is the original 12, not the special issue ones.) None of these introductions uses a 'narrow relevance' strategy. All of them include work to establish the relevance of their study to broader concepts. Indeed it might be said that the 'broad relevance'

strategy is the default one in GEC and the way in which the writers for this interdisciplinary journal position their work in relation to their readers. Two examples are given here. The first (GEC 2004 Boykoff) is an article that indisputably introduces a novel area of study (newspaper discourse) to the readers of the journal. The second (GEC 2006 Piao) has been selected as a contrast because it reports a conventional scientific project but nonetheless makes this relevant to the GEC readers.

GEC 2004 Boykoff: 'Balance as bias: global warming and the US prestige press'

As noted previously, this article sits outside the mainstream of research in environmental change because it studies newspaper discourse rather than either physical entities (soil, rivers, plants) or social entities (agriculture, policy, consumers). Its structure may be summarised as follows:

Move 1: 'The US prestige press misrepresents the science on global warming'.
Move 2: 'Investigating the media's portrayal of global warming is crucial'.
Move 3: 'The article will argue that the press fails to support environmental science'.

What is interesting in this introduction is what it does not say. In its four paragraphs (relatively short for this journal), there is no account of research method. Neither is there any statement that criticises other researchers for ignoring the issue of journalism, or arguments for a necessary alignment of interests. Rather, it is the press that is criticised for failing to transmit to the public accurate information about environmental change (examples 18 and 19).

18 This paper explores the notion that the US prestige press . . . has contributed in significant ways to failed discursive translations regarding global warming.

[GEC 2004 Boykoff]

19 The continuous juggling act journalists engage in often mitigates against meaningful, accurate and urgent coverage of the issue of global warming.

[GEC 2004 Boykoff]

The introduction includes statements of the importance of the research, and therefore of the discipline (example 20), but does not explicitly contrast this with other relevant disciplines.

20 [I]nvestigating the mass media's portrayal of global warming is crucial.

[GEC 2004 Boykoff]

In short, this paper introduces an unusual discipline (linguistics) to the journal readership, but the introduction does not explicitly locate this discipline among the others that the readers may encounter.

GEC 2006 Piao: 'NDVI-based increase in growth of temperate grasslands and its responses to climate changes in China'

By contrast with GEC 2004 Boykoff, this article is a more conventional science-based study of grasslands in China and the ways they have changed as a result of climate change, using the Normalized Density Vegetation Index (NDVI) as the methodology. This methodology has been used extensively before; the novelty of the study is the quantification of local changes in vegetation and correlating this with climate change data. The structure of the introduction is:

Move 1: 'China's temperate grassland is important. It can be studied using NDVI'.
Move 2: 'Previous research in this area has been inadequate'.
Move 3: 'The current study will use NDVI, correcting for previous inadequacies'.

Moves 2 and 3 of the introduction can be said to address a monodisciplinary or insider readership. Move 2 has typical indicators of 'gap', shown in examples 21 and 22:

21 [T]he linkage between climate change and vegetation growth . . . has not been adequately quantified.
[GEC 2006 Piao]

22 [N]o effort has been made to investigate the effects of climate change on the relationships between vegetation growth and climate variables.
[GEC 2006 Piao]

Move 3 outlines the purpose of the study, the research questions and the scope of the study. In move 1, however, steps are taken to make the research relevant to a broader audience. The grasslands are described as important economically (*supports the world's largest population of sheep and goats*) and ecologically (*plays an important role in the regional climate and reducing soil loss*). Move 1 finishes with a statement of relevance of the research, which is both academic (*understanding of the grassland ecosystem carbon cycle*) and practical (*the sustainable use of grassland resources*). Moreover, particular words and phrases are used to add emphasis to the evaluations of relevance. The word *dramatic* modifies *climate change* in example 23; the phrase *not only . . . but also* links the two important roles of the grasslands in example 24; the phrase *both . . . and* links the two aspects of importance of the research in example 25.

23 [T]he climate of this region has experienced **dramatic** change in the past several decades.

[GEC 2006 Piao]

24 [China's temperate grassland] **not only** supports the world's largest population of sheep and goats, **but also** play an important role in the regional climate.

[GEC 2006 Piao]

25 [S]tudies on the impact of climate change . . . are of significance to **both** the understanding of the grassland ecosystem carbon cycle **and** the sustainable use of grassland resources in China.

[GEC 2006 Piao]

The work done in this introduction, therefore, focuses on the applications of the topic rather than the novelty of the method. A considerable amount of work is done to stress the importance of grasslands to human well-being and the importance of this research to the interests of the GEC readership.

5.6 Summary: studies of introductions

So far in this chapter, 48 introductions from the BEE4 corpus have been investigated to establish their structure, using the CARS model. Five of these introductions (three from AEE and two from GEC) have been examined in greater detail, along with six additional introductions from a special issue of GEC. The upshot of the study is that writers can choose whether to orientate their study to a narrow or broad audience. If a broad audience is chosen, the introduction can place the article as inside or outside the mainstream. If outside the mainstream, the article can be placed as 'antagonistic', as 'integrative', or as 'supportive'.

Various features of language and discourse structure are used to achieve these various positionings. The introduction can begin from the point of view of the writer's own discipline or particular concern, subsequently linking that to more general concerns, or it can begin with the more general statement about concern for a given environmental issue, subsequently linking that to the topic of the article. The introduction can be more or less self-reflexive about discipline, either naming disciplines and considering how they stand in relation to one another, or omitting to do so. The introduction can be more or less overt in the evaluative language it uses, both in articulating a gap in research, and therefore the shortcomings of previous research, and in describing its own contribution.

5.7 Beyond the introduction: construing community through pronouns

We now pick up two of the themes noted from the investigation of individual journal articles discussed previously, exploring those language phenomena in the BEE4 corpus as a whole, that is, in the complete articles rather than only in the introductions.

The first item to be studied is the first-person pronoun *we*, prompted by its use in GEC 2009 Lawrence. In that article, *we* is used to refer to both 'people in general' and 'the research community'. This is in the context of an article that construes itself as interdisciplinary and conciliatory or integrative. The specific role of *we* in this construal inspires the corpus study. From each journal, 100 sample lines of *we* are scrutinised to identify the groups of people construed. The 100 lines are obtained using the Sketch Engine function that randomly selects the specified number of lines from the total available. Table 5.4 shows the classification of referents used and the numbers of instances in each.

In all the four journals, most instances of *we* refer to the research team – that is, the writers of the article. The verbs following *we* (i.e. of which *we* is the subject) are physical research actions (e.g. *calculated*, *collected*, *found*, *measured*), mental research procedures (e.g. *assume*, *conclude*, *consider*, *suspect*) or discoursal actions (e.g. *clarify*, *discuss*, *report*, *suggest*). In other uses of *we*, the pronoun refers to the research community more generally, as in *we need to know* or *we might expect*. The third use is to refer to human society, not limited to researchers, as in *We have eliminated most human predators*. In some cases, *we* is indeterminate and may refer to the research team alone or to the community. This is particularly the case when a modal verb is used (e.g. *We can infer from this . . .*), where it may be the writers of the article or the writers and their readers together who form the inference. Table 5.4 gives the number of instances from the sample of 100 in each journal referring to each group. The 'missing' instances in GEC and REE are where *we* occurs in a quotation from another person rather than in a statement averred by the article writers themselves.

Table 5.4 shows that although the journals are largely consistent with each other, GEC is the most different from the others in having a smaller proportion of instances that do not refer to the research team alone. Examples of instances that refer to human society include philosophical questions about the relationship between human society and the ecosystem it inhabits, examples of how human society impacts the environment, and examples of human perceptions of the environment (see examples 26 to 28).

26 [W]ho are we and what special qualities do we possess?
[GEC 1996 Van Asselt]

27 Mobile pollution impacts are affected by how much we travel.
[GEC 2004 Carmichael]

28 Is it worse than we thought?
[GEC 2008 Risbey]

Table 5.4 Number of referents of *we* in 100 sample lines from each of four journals.

	AEE	*GEC*	*PS*	*REE*
Research team	98	76	100	96
Research community	2	10	0	3
Human society	0	6	0	0
Total	100	92	100	99

What is notable here is not so much that GEC talks about human society, which is to be expected, but that it does so from the position of within rather than outside that society, using *we* to do so. This affirms a continuity between the research community and the human community it researches. The usage is at odds with the other journals.

Writers in GEC also use *we* to refer to the research community more frequently than those writing for the other journals. Again, what is interesting is not that the research community is talked about, but that writers choose to align themselves overtly with that community. From the ten examples selected from the 100-line sample, shown in Figure 5.1 as sentence-mode concordance lines, two observations can be made. One is that seven of the instances perform a directive, recommending a course of action or thought (Figure 5.1a, directives in bold).

1. However, if we are going to ask people to make substantial sacrifices for reducing greenhouse gas emissions, then **let us do so for the right reason:** the threat of irreversible and non-substitutable loss of natural capital if we fail to reduce emissions.
3. When we can adequately document impacts, then **we can address likely individual and societal adjustment strategies** for these sub-populations, strategies that will no longer marginalize women and children, placing them in the terra incognita of forgotten casualties.
4. The global problematique in which we find ourselves thus offers an opportunity **to go beyond past conceptual limitations.**
6. In order to meet the challenges of the Anthropocene, **we need to develop models and methods** that are capable of illuminating the interactions between biophysical forces and anthropogenic drivers.
7. **The scientific community has to create a push** so that we see more studies that provide concrete example of how to operationalize the use of remote sensing to monitor IEAs—a number that is too small today.
9. However, **we should be cautious about combining emissions** from a wide range of sources about which there are different degrees of uncertainty.
10. Given the abundance of this fuel within the continental USA, **we should expect that** considerations of economics and national energy security will weigh heavily in favour of coal as the fuel of choice for meeting rising energy demands.

Figure 5.1a Directives in a sample of ten instances of *we* from GEC.

The second is that five of the instances make explicit reference to disciplines or to knowledge-making (Figure 5.1b, relevant references in bold).

In other words, the writers are performing face-threatening acts (Brown & Levinson 1987), in telling the community to do something, and are doing so in the context of sets of disciplinary expectations. The alignment mechanism of *we* is a way of mitigating that face-threatening act. Line 4 of Figure 5.1b, repeated here as example 29, illustrates the performance of a directive in the context of disciplinary thinking and is thus a central example of the phenomenon.

29 The global problematique in which we find ourselves thus offers an opportunity to go beyond past conceptual limitations.

[GEC 1992 Dahlberg]

This corpus study confirms that the use of *we* constitutes an expression of overt community solidarity. In the ten instances shown in full in Figure 5.1, common ground is asserted at the same time as a challenge is expressed to accepted research agendas. As illustrated in GEC 2009 Lawrence, this appeal to common ground would seem to be a marker of integrative interdisciplinary writing.

2. If we turn to the reflexive views of the world characteristic of **some branches of social science and the humanities,** we are often looking for the feedback loops and cybernetic connections which link our cognitions of the natural world with those of society.
4. The global problematique in which we find ourselves thus offers an opportunity to go **beyond past conceptual limitations.**
5. Re-thinking the **kinds of knowledge** we make.
6. In order to meet the challenges of the Anthropocene, we need to **develop models and methods** that are capable of illuminating **the interactions between biophysical forces and anthropogenic drivers.**
7. **The scientific community** has to create a push so that we see **more studies that provide concrete examples** of how to operationalize the use of remote sensing to monitor IEAs—a number that is too small today.

Figure 5.1b Explicit reference to disciplines or knowledge in a sample of ten instances of *we* from GEC.

5.8 Beyond the introduction: construing conflict with *inadequate*

At the beginning of this chapter, it was argued that article introductions can be used to assert common ground ('we are interested in the same things') and methodological difference ('we do things differently') or even opposition ('we do things better'). In this second corpus study, we wish to explore in more detail that elusive phenomenon of statements of opposition, recognising that these statements may not occur in the introductions to articles. The difficulty is finding phraseology that may lead to identification of these statements – that is, reducing the observation of a phenomenon to words that can be the starting point for corpus searches. As a preliminary attempt, we use the word *inadequate*, as this was used in GEC 2009 Byg, the 'antagonistic' article discussed previously.

Table 5.5 shows the raw and normalised (per million words) frequency of *inadequate* in the four journals. It is substantially most frequent in GEC, less frequent in AEE, and infrequent in PS and REE. The speculation that it is associated with interdisciplinary discourse in general and with agonistic disciplines in particular would seem to be supported. In looking at the instances in more detail we draw a distinction between 'research entities', such as *data* or *technique*, and 'non-research entities', such as *policy* or *enzymes*. The non-research entities may be social (e.g. *inadequate regulation*) or physical (e.g. *inadequate irrigation*). However, it is the research entities that are particularly interesting, because describing these as *inadequate* has the potential to construe antagonism between disciplines or approaches.

In REE, the entities judged 'inadequate' are roughly equally split between research and non-research:

- Research entities: *data* (2), *quality* (2), *scope sensitivity*, *spatial sampling*. (total: 6)
- Non-research entities: *pollution reduction*, *policy*, *environmental regulation*, *sites (where species live)*, *level of support*. (total: 5)

In PS, there are more instances overall. The entities judged 'inadequate' are mostly non-research entities, with one-third of the total being research entities.

- Research entities: *culture medium composition*, *RFLP technique*, *conditions*, *evaluation*, *amount of protein recovered*, *use of guaiacol*, *yield of antigen*, *strategy*, *approach*, *evidence*. (total: 10)

Table 5.5 Frequency of *inadequate* in four journals.

AEE		*GEC*		*PS*		*REE*	
frequency	pmw	frequency	pmw	frequency	pmw	frequency	pmw
136	13.86	194	41.15	30	3.29	11	7.08

- Non-research entities: *carpel development*, *chemical controls*, *seedling emergence*, *amount of NA*, *amounts of necessary precursors*, *irrigation* (2), *expression of the relevant transcription factors*, *activities of enzymes*, *mechanism*, *methionine levels*, *proline*, *Pi supply* (2), *rates of N acquisition*, *sink capacity*, *supply of reductants*, *antioxidant system*, *promoter*, *uptake*. (total: 20)

In both journals, inadequacy in research is related to specific aspects of methodology, such as *use of guaiacol*, as well as broader methodological issues, such as *evaluation*.

Because of the greater frequency of the word *inadequate* in GEC, a different procedure was followed in order to establish the relative frequency of 'research entities' and 'non-research entities' modified by the adjective. As a matter of convenience, the concordance lines were first sorted so that the words occurring immediately to the right of *inadequate* appeared in alphabetical order (R1 sorting); the same lines were then sorted so that the words occurring immediately to the left of *inadequate* appeared in alphabetical order (L1 sorting). These two sorting procedures made patterns in the data easier to see, but they had the additional benefit of separating nouns that are modified attributively (e.g. *an inadequate guide to policy*) and those that are modified predicatively (e.g. *the definition was inadequate*). Attributive adjective use is associated with information that is given, or shared ground. Predicative adjective use is associated with information that is new. Thus, using an evaluative adjective predicatively makes a stronger statement of evaluation. The lists of research and non-research entities, where *inadequate* is attributive and predicative, are shown in Table 5.6a. Table 5.6b compares the number of tokens.

In GEC, *inadequate* is more often used to describe non-research entities than research entities. Taking all the noun phrases together, there are more instances of attributive *inadequate* than predicative. Another noticeable difference is that the attributive use is more likely to occur in recurring phrases, such as *inadequate water supply*/*services* (four instances), whereas each of the predictive uses is unique. However, when non-research entities are evaluated, the adjective is by far most frequently attributive and thus 'taken for granted' (e.g. *inadequate water supply*), whereas when research entities are evaluated, the adjective is equally frequently attributive (e.g. *inadequate metrics*) and predicative (*the available data is inadequate*). The greatest strength of criticism, therefore, is reserved for research entities, whereas the dire state of the non-research world is assumed.

Consistent with the interdisciplinary nature of the journal, the non-research entities assessed as inadequate in GEC include both physical entities (*land*; *calories per day*; *livestock feed* etc.) and human ones (*modern institutions*; *interventions*; *decisions about pesticides* etc.). Turning to the research entities, although some relate to details (e.g. *data from tropical regions*), most are to do with more general concepts (e.g. *ocean models; concepts in the literature*). Most importantly, and in line with the observation that GEC contains some discourse that is antagonistic to physical science, there are some instances of judgements of whole research

Table 5.6a Nouns modified by *inadequate* in GEC.

	inadequate *is an attributive adjective*	inadequate *is a predicative adjective*
Research entities	attempt at putting numbers; guide to policy; data; characterisation; collection of data; communication (between disciplines); coverage; data; evidence; focus on equity; guidance (provided by data); job (done by climate assessments); knowledge; measurements; metrics; models; perceptions; sampling; scientific knowledge; understanding (4).	definition; explanations; the state of scientific knowledge; this assumption; approaches which deal with physical or biological systems; single-issue approaches; data from tropical regions; spatial representation in models; available data; ignoring adaptation; conventional explanation; models of these processes; view of human nature; knowledge base; specification of the model; system; ocean models; concepts in the literature; data; impact assessment which only consider biophysical relationships; standard forms of multivariate classification; model interfaces; social science tools; amount of data; a scientific mode of inquiry.
Non-research entities	access; allocation of funds; support; attention (3); bookkeeping; budget(ary support) (2); capacity/ies (3); cash resources; certification (3); control of risks; diffusion of knowledge; electricity supply systems; farming implements; financial support; fisheries management; form of wages; funding (2); government (policy and practice) (3); management strategies; healthcare (2); housing (2); information (2); infrastructure (3); irrigation; labour (2); land (management and processes) (3); law enforcement; legal frameworks; livelihoods; local supplies; management of waste; incineration; treaty; occupation of territory; oil production; participation of stakeholders; pest control; planting seeds; policies (4); political will; property rights; recharge of ground water; resources; responses; services; settlement patterns; sewage (disposal) (3); shelter; soil occupation; staffing; support (2); technologies; training; transport (2); infrastructure; warmth; water (4).	guard staff; water treatment facilities; sums (of money); a mere statement of policies; harvest; recent legislation; funding levels; guard bases; resources; people's efforts (at recycling); incentives for renewables; livestock feed; non-fossil alternatives; approaches to infrastructure; a goal of 450 parts per million; waste handling facilities; arrangements for grazing control; social systems; information; interventions; infrastructure; modern institutions; calories per day; normal adaptation measures; mitigation of acute hazards; sulphur controls; range of fallow; (crop) yields; adaptation measures; natural regeneration; training programs; decisions about pesticides; compensation for damage; voluntary agreement; safety margins; Indian IPCC interface; age group feels; original plan; services; building codes; information supplied; measures; clustering agreements; investment; education; bargaining; mitigation and adaptation (separately); outcome of choices; fishery regimes; resources; land; measures; system of governance; status quo of governance; single issue approaches.

Table 5.6b Number and proportion of tokens of nouns modified by *inadequate* in GEC.

Type of entity	*Number of nouns (tokens) modified by attributive* inadequate		*Number of nouns (tokens) modified by predicative* inadequate		*Total*	
	Number	*Percentage*	*Number*	*Percentage*	*Number*	*Percentage*
Research	23	12.2	25	13.3	48	25.5
Non-research	85	45.2	55	29.3	140	74.5
Total	108	57.4	80	42.5	188	100

domains as inadequate: *the state of scientific knowledge*, *approaches which deal with physical or biological systems (only)*, *single-issue approaches*, *view of human nature* (that ignores cultural context), *impact assessment which only considers biophysical relationships*, *a scientific mode of inquiry*. There is also one instance of equivalent critique of social science: *social science tools*.

The same procedure of separating attributive and predicative uses is followed in the study of AEE (Tables 5.7a and 5.7b).

Note that in Table 5.7a *knowledge* and *representation* appear in both 'research' and 'non-research' lists because the knowledge, for example, may be held by researchers (e.g. *Species data . . . may suffer from . . . inadequate knowledge of the taxa under study*) or by farmers (e.g. *The question is why farmers . . . had not taken part. . . . Inadequate knowledge and hostility . . . seem to be the main reasons*).

The proportions of use are roughly similar to those in GEC, with non-research entities evaluated as inadequate more frequently than research entities, and the adjective used attributively more frequently than predicatively. However, the differences are not as extreme, and in particular the proportion of research entities evaluated predicatively is noticeably higher. Overall, the predicative or 'new information' criticism is a more likely choice in AEE than in GEC and this is particularly marked where research entities are being critiqued.

The non-research entities that are evaluated as inadequate include both physical entities (e.g. *amounts of N in the synthetic fertilizer*; *rainfall; nutrient availability*) and human or social ones (e.g. *centralized water supply system*; *fertility conservation methods*; *technical support*). Although research entities are more likely than in GEC to be evaluated predicatively, in all cases it is specific details of the current or previous research that is evaluated as inadequate, rather than whole research domains. In particular, and in contrast with GEC, there is no challenge to science.

To summarise: evaluations using the adjective *inadequate* are more frequently found in interdisciplinary than monodisciplinary journals. In both the interdisciplinary journals, non-research entities evaluated in this way include both physical and social entities. Research entities are also evaluated using this adjective. For the most part these constitute aspects of previous research, but in GEC this evaluation is extended to include physical science approaches as a whole.

Table 5.7a Nouns modified by *inadequate* in AEE.

	inadequate *is an attributive adjective*	inadequate *is a predicative adjective*
Research entities	data; attention paid to off-site interactions; characterization of the soil profile; coupling of atmospheric and oceanic processes; criterion for assessment; experimental data; representation of regional vegetation types; information (2); knowledge (2); match; monitoring of technology adoption; national statistics; quality (of sensor); representation (2); sample replication; sensor frequency response; values.	replicates; data (2); model performance; database (2); information; datasets; monofactorial light-response function; defining model uncertainty; land-use intensity indicators; design of evaluation studies; studies; partial budgeting approach; two-dimensional model; measures; data; methods; calculation of the NAAR; evaluation of systems; time-span; majority of studies; grid resolution; abundance values; sampling frequency; results; definition of indicators.
Non-research entities	agricultural policy; amounts of N, P or K in the synthetic fertilizer; availability of FYM; available Cr; capacity; centralized water supply system; compensation; conservation levels; conservation practices; cover; coverage of fruit; crop and nutrient management; establishment of the legumes; external inputs; fertility; fertility conservation methods; field control; finance; investment; knowledge (2); land management; legume biomass; flows (of water); management (4); mechanization; mineral N; N inputs/ supply (5); number of extension workers; P fertilization; production; provision; rainfall (2); representation; return of nutrients; root systems; sanitation practice; soil aeration (2); soil management (2); stocking rate; supplies (3); technical support; technologies; use of fertilizers; water (4); weed control; Zn intake.	management measures; pollution control; availability of irrigation water; K inputs; amounts; P release; nutrient availability; 3m wide riparian habitat; monocultures; B bassiana; level of production; available sources of organic matter; P recycled; traditional fallows; N treatment; plant growth periods; soil fertility.

Table 5.7b Number and proportion of tokens of nouns modified by *inadequate* in AEE.

Type of entity	*Number of nouns (tokens) modified by attributive* inadequate		*Number of nouns (tokens) modified by predicative* inadequate		*Total*	
	number	*percentage*	*number*	*percentage*	*number*	*percentage*
Research	20	15.7%	27	21.3%	47	37.0%
Non-research	63	49.6%	17	13.4%	80	62.9%
Total	83	65.4%	44	34.6%	127	99.9%

5.9 Conclusion

This chapter has investigated a selection of introductions from journal articles in the BEE4 corpus. Close text analysis of those introductions has been supplemented by a small number of corpus investigations. Although this chapter has been more text-based than other chapters in this book, it is consistent with them in proposing that in the article introductions we observe disciplinary identities being construed, writers positioning themselves and their research as within or outside the mainstream, and writers adopting various attitudes towards other research paradigms.

The structure of the introductions studied was observed to be broadly the same, with the CARS model working well as an explanation of the rhetorical moves being employed. Differences lay in the amount of 'work' done in the introductions to establish the importance of the research to the readership. It was noted that writers could stress common ground and/or could focus on the unique contribution of their research. The readership could be construed as people working within a similar paradigm or as a very different group of researchers. Finally, writers could take a conciliatory stance towards other approaches or an antagonistic one. The antagonistic stance is relatively rare but noticeable when it does occur. It was noted that in the journals announcing themselves to be interdisciplinary – AEE and GEC – some of the writers position themselves as interdisciplinary and others do not.

Starting from an interesting use of *we* in one of the introductions, this pronoun then formed the basis of a corpus search of the four journals. It was noted that *we* can be used strategically either to elide or to emphasise the distance between the writer and other research communities. Both uses are most common in GEC.

In an attempt to investigate further the expression of antagonism in the journals, the word *inadequate* was the subject of the next corpus search. This study illustrated two points: the interdisciplinary journals AEE and GEC assess both natural and social entities as *inadequate*, reflecting the dual focus of both journals. AEE offers most criticism of research entities: methods, models and so on. However, only GEC includes instances where a whole research paradigm, as opposed to details of methodology, is described as *inadequate*. GEC has been shown to be uniquely self-reflexive, in that authors exercise the option of positioning themselves both outside and in opposition to other disciplines.

Note

1 The aim was to identify the first and fourth issue from selected years, and to pick the second original research article in each of those issues. An exception was made if the issue in question was a special issue – in that case the next issue was chosen. The procedure was not entirely straightforward, with the result that the selection of papers was not entirely consistent. However, as the aim is simply to randomise the selection of papers across the decade and to avoid 'cherry-picking', the result is adequate.

References

Barry, A., & Born, G. (Eds.). (2013). *Interdisciplinarity: Reconfigurations of the social and natural sciences*. London: Routledge.

Bhatia, V. (1993). *Analysing genre: Language use in professional settings*. London: Longman.

Brown, P., & Levinson, S. (1987). *Politeness: Some universals in language usage*. Cambridge: Cambridge University Press.

Hasan, R. (1996). *Ways of saying, ways of meaning: Collected papers of Ruqaiya Hasan* (C. Cloran, D. Butt & G. Williams, Eds.). London: Cassell.

Hunston, S. (2011). *Corpus approaches to evaluation: Phraseology and evaluative language*. London: Routledge.

Hyland, K. (2012). *Disciplinary identities: Individuality and community in academic discourse*. Cambridge: Cambridge University Press.

Hyon, S. (1996). Genre in three traditions: Implications for ESL. *TESOL Quarterly*, 30: 693–722.

Lakic, I. (2010). Analysing genre: Research article introductions in economics. *Journal of Linguistics and Intercultural Education*, 3: 83–100.

Loi, C. K. (2010). Research article introductions in Chinese and English: A comparative genre-based investigation. *Journal of English for Academic Purposes*, 9(4): 267–279.

Martin, J., & White, P. (2005). *The language of evaluation: Appraisal in English*. London: Palgrave.

Miller, C. (1984). Genre as social action. *Quarterly Journal of Speech*, 70: 151–167.

Ozturk, I. (2007). The textual organisation of research article introductions in applied linguistics: Variability within a single discipline. *English for Specific Purposes*, 26(1): 25–38.

Samraj, B. (2002). Introductions in various disciplines: Variations across disciplines. *English for Specific Purposes*, 21(1): 1–17.

Swales, J. (1981). Aspects of article introductions. In *ESP Monographs 1*. Aston University: Language Studies Unit.

Swales, J. (1990). *Genre analysis: English in academic and research settings*. Cambridge: Cambridge University Press.

Swales, J. (2004). *Research genres: Exploration and applications*. Cambridge: Cambridge University Press.

6 Words in context

Environment, *science*, *important*

6.1 Introduction

Chapter 5 ended with a study of two words, selected because they had played important roles in some of the research article introductions we investigated. In this chapter we carry out further studies of individual words. The words are selected because of their perceived likely importance in our corpus of environmental journal texts. They are: *environment*, *science/scientists/scientific* and *important*. We use these example words to demonstrate how some of the observations about interdisciplinary research surveyed in Chapter 2 and further discussed in Chapter 5 can be shown to be reflected in word choice and phraseology in the journals. We focus on two issues arising from the previous chapters: the range of meanings or usages a word may have in different research contexts, and how many of these occur in a specific journal; and the presence or absence of evidence of antagonism between disciplines.

As elsewhere in this book, we focus on the BEE4 corpus, comprising the two monodisciplinary journals PS and REE and the two interdisciplinary journals AEE and GEC. Overall our findings confirm the division into two pairs and also clarify the difference in the nature of the interdisciplinarity in AEE and GEC.

6.2 One word, several meanings: the case of *environment*

One of the most frequent comments made by interdisciplinary researchers is that their colleagues in other disciplines 'speak a different language' (e.g. Bauer 1990: 112). As a consequence, dialogue between such researchers requires a degree of 'multilingualism', or the presence of an interpreter. A light-hearted quiz on the website of the journal *Nature* (www.nature.com/news/how-interdisciplinary-are-you-1.18362, accessed 27 July 2018) draws attention to this with a number of questions about word meanings. People taking the quiz are asked to select which meaning(s) of given words they regularly use in their research. For example, *cell* is '(a) a place to hold a prisoner, (b) a source of electrical energy (battery), (c) the basic unit of a living organism, and/or (d) a set of experimental conditions, often representing one of several treatment groups'. *Division* is '(a) the process

by which cells multiply, (b) a mathematical operation, and/or (c) discord within a group'. The more meanings a participant can honestly say they use, the higher their Interdisciplinarity Score.

The examples used in the *Nature* website are very clearly polysemous words, and it is difficult to imagine a scenario in which their use would cause confusion or dispute within a research group. It is even difficult to imagine a research project in which more than one meaning of a cited word is used (for example, a project combining research into prison conditions and cellular biochemistry). Many words, however, exhibit more subtle distinctions of meaning. The noun *struggle*, for example, can refer to conflict or competition (*the struggle for market share*) or to a difficult task (*it was a struggle to put food on the table*) (Hunston 2018). The two meanings are not wholly different, but neither are they exactly the same. Informal interviews with interdisciplinary researchers suggest that terms used with slightly different meanings are more likely to cause problems than those that are clearly polysemous. An example given to us by a researcher in Sports and Exercise Science is the noun *well-being*, which is used by many people with a general or common-sense meaning of experiencing physical health and contentment but by others with much more specific meanings related to philosophy or psychology. It is reported that interdisciplinary research that aims to measure degrees of well-being encounters problems when some researchers involved regard the term as a general but simple idea and others see it as a specific, complex and contentious term. An example from Manor-Binyamini (2011: 2007) is *success experience*, which for teachers refers to 'coping with the syllabus' but for educational psychiatrists means achieving mental health.

An example from our own field is the word *grammar*, which has a meaning in linguistics that is both specific, in that it contrasts with other terms such as *discourse* or *lexis*, and contentious, in that different schools of thought propose competing models of grammar (or syntax). Outside linguistics it is sometimes used to mean something like 'the way things happen or are organised'. Some examples of this are: *the grammar of institutions*, *the very grammar of exploitation and oppression*, *a single conceptual grammar of civilizational order*, *the basic grammar of America*, *the historical grammar of secularism*. (These examples are from the enTenTen15 corpus, accessed via the Sketch Engine site, using the search phrase *grammar of*.) Speaking personally, we find this general use difficult to understand, because the more specific use gets in the way. If a research project were proposed to investigate 'the grammar of secularism', a linguistic grammarian might find it difficult to see what their involvement might be.

One of the markers of an interdisciplinary journal might be expected to be the use of different meanings of multi-meaning words, whereas a monodisciplinary journal might include only one meaning. This possibility will be explored in relation to the word *environment*. It occurs with the frequencies shown in Table 6.1.

Table 6.1 Frequency of the word *environment* in four journals.

Journal	*Raw frequency*	*Normalised frequency pmw*	*Frequency order*
AEE	2260	230.27	2
GEC	3044	645.62	1
PS	781	85.69	4
REE	285	183.38	3

The meanings of *environment* are illustrated by examples 1–6, taken from a random 100 lines sampled from the BEE4 corpus.

1 . . . with a particular focus on the capacity of existing legal and administrative structures to protect the **environment**.

[GEC 1993 Jasanoff]

2 Moreover, the **environment** of the developing embryos can be modified by adding specific signalling molecules.

[PS 2003 Baldan]

3 The aquatic **environment** in rice fields is largely influenced by the growth of the rice crop.

[AEE 2005 Vromant]

4 The practical difficulty of incorporating the **environment** into traditional measures of productivity is the task of assigning weights to the output of emissions.

[REE 2003 Aiken]

5 . . . supporting institutions including those that shape the regulatory **environment**.

[AEE 2007 Pascual]

6 . . . as a response to experienced or expected changes in the societal . . . **environment**.

[GEC 2009 Pahl-Wostl]

The meanings might be differentiated in two ways. There is an obvious difference between references to the natural world in examples 1–4 as opposed to the social-political world in 5–6. Within examples 1–4, however, there is a difference of scale and specificity. In example 1 and example 4, the *environment* is a large-scale entity, without obvious boundaries. The referent in example 1 is all the natural areas of India, in example 4 the referent is potentially worldwide but from the reference to *emissions* would appear to be restricted to the breathable air (as opposed to rivers,

wildlife etc). In example 2 it is a small-scale entity, the immediate surroundings of an embryo. It exists only in relation to that embryo and without the embryo it would not be *environment*. In example 3 the scale is mid-way between the two. The *aquatic environment* is an entity in its own right, in contrast to example 2, but it has defined boundaries, being restricted to rice fields.

Each of the examples suggests a different topic of research and possibly a unique epistemology. 'The environment' as an entity of interest to governments and researchers seeking to minimise (the impact of) human-caused environmental change occurs in examples 1 and 4. The same word in examples 2 and 3 does not refer to this entity, although they do refer to the natural world. The word in examples 5 and 6 is unrelated to the issue of changes in the physical environment except by chance: in example 5, the regulations concerned do relate to the natural environment and how human activity impacts on it, and the topic of the journal article quoted in example 6 is the interaction between societal and natural changes. It is reasonable to propose that the various meanings in examples 1–6 relate to research orientation. Therefore, in looking at journals we might ask which meanings predominate in each journal, but we might also ask to what extent the instances in a given journal are consistent as to meaning. Answering the first question will indicate differences between journals. Finding multiple meaning types in one journal may be an indication of the presence of several disciplines.

To investigate *environment* in the BEE4 corpus, three separate studies were carried out. In each case the aim was to determine the range of meanings of the word. First, the words occurring immediately before *environment* were identified; to avoid a long list of single occurrences, an arbitrary cut-off of four occurrences in each journal was applied. Some of these words are grammatical words (e.g. *the*, *and*, *of*, *their*), others are lexical (e.g. *agricultural*, *aquatic*, *changing*). The lexical modifiers were then examined with the aim of assessing the degree of variation – that is, the extent to which the modifiers belong to similar or very different semantic sets. Second, all instances of the bigrams *its environment* and *their environment* in each journal were examined to ascertain the referent for the possessive. Again, it is the range of referents that is of interest. Finally, those instances where *the environment* is preceded by a preposition such as *to* or *of* were identified. These prepositions occur because they collocate with a preceding noun or verb, as in *adaptation to the environment* or *depend on the environment*, so they can be used as a probe to investigate types of responses to the environment.

For reasons of space, only the main findings of these three studies will be presented here.

Table 6.2 lists the adjective, noun and verbal modifiers of *environment* (i.e. omitting grammatical items) occurring at least four times in each journal, in alphabetical order.

Collocating adjectives include those identified as:

- Geographic regions, such as *Australian*, *Arctic*, *Sahelian*;
- Large-scale classifications of the natural environment, such as *aquatic*, *desert*, *forest*, *humid*, *marine*, *tropical*;

Table 6.2 Lexical modifiers of *environment* in four journals.

AEE	*GEC*	*PS*	*REE*
abiotic	aquatic	adverse	better
agricultural	Arctic	aquatic	cleaner
anaerobic	better	aqueous	decision
aquatic	biophysical	arid	global
Australian	broader	cellular	improved
biotic	built	changing	local
biophysical	business	contaminated	regulatory
changing	changing	controlled	
competitive	decision-making	culture	
controlled	different	different	
crop	dynamic	drought	
desert	economic	external	
drier	enabling	greenhouse	
ecological	European	growth	
economic	external	light	
enabling	global	lipid	
European	healthy	natural	
external	human	osmotic	
favourable	institutional	plant	
field	international	reducing	
forest	living	root	
free	local	rooting	
future	marine	saline	
GIS	national	salinity	
Global	Nations	single	
growing	natural	soil	
heterogeneous	new	stress	
humid	physical;	stressful	
internal	policy	surrounding	
local	political	vitro	
lowland	regional	water	
marine	regulatory		
Mediterranean	-rich		
nations	Sahelian		
natural	semi-arid		
new	social		
particular	socio-		
pH	Stockholm		
physical	surrounding		
plant	terrestrial		
policy	UN		
political	urban		
production	water		
rich	wider		
rural	world		
socio-economic			
soil			
specific			
surrounding			
suitable			
stable			
terrestrial			
tropical			
understorey			
water			
wider			

- Classifications of the small-scale natural environment, such as *anaerobic*, *biotic*, *saline*, *-rich*;
- References to research contexts, such as *controlled*, *greenhouse*, *in vitro*, *natural*;
- References to parts of plants and plant-related processes, such as *cellular*, *growing*, *osmotic*, *root*;
- References to human contexts, both geographic and social, such as *socio-economic*, *urban*, *built*, *human*, *decision-making*, *regulatory*;
- References to the environment as changing or varied, such as *changing*, *different*, *dynamic*, *stable*;
- Evaluations of environments as either positive or negative, such as *favourable*, *rich*, *stable*, *changing*, *stressful*, *contaminated*.

The monodisciplinary journal REE is unusual in having relatively few modifiers of the word. Instead, it has an atypically high proportion of instances of *the environment*. This phrase accounts for 58.5% of instances of *environment* in REE, as opposed to 34.3% in AEE, 32% in GEC and 18% in PS. The phrase *the environment* predominantly refers to an undifferentiated, large-scale, physical entity. The other monodisciplinary journal, PS, has more modifiers occurring at least four times, possibly because it is a larger sub-corpus, but the range of modifiers is again relatively narrow. It includes those relating to *environment* as a small-scale, natural entity, to research contexts, and to plants and their processes. In other words, both the monodisciplinary journals refer predominantly to one meaning of *environment* only, as indicated by the modifiers, and it is a different meaning in each journal. Neither journal specifies when the environment is the physical one, leaving that as the default option.

The interdisciplinary journal GEC has a high proportion of modifiers relating to different aspects of the human experience, including geographic classifications (*urban*, *built*) and aspects of human society (*business*, *economic*, *institutional*, *political*, *regulatory*) as well as physical aspects (*living*, *healthy*). There are also modifiers relating to the natural environment (*marine*, *semi-arid*, *physical*, *biophysical*). This suggests, as expected, that the journal includes several meanings of the word, with a predominance of meanings related to human institutions. What is perhaps most striking, however, is that a sub-set of the modifiers specify whether the environment under discussion is physical or social (e.g. *biophysical*, *economic*, *political*, *social*). This implies a contrast with other kinds of environment, an awareness that what the environment in question consists of cannot be taken for granted. This in turn suggests a non-homogeneous readership.

The other interdisciplinary journal, AEE, has the longest list of modifiers and the widest variety. It includes instances of all the categories in the bullet-point list earlier in this section. The modifiers relating to the natural environment include those relating to large-scale classifications (*aquatic*, *forest*, *lowland*, *terrestrial*, *tropical*) and those relating to small-scale contexts (*anaerobic*, *biotic*, *growing*, *pH*, *soil*). There are references to the research process (*controlled* and its opposite *natural*), where that research is clearly carried out on non-human entities. The

adjectives relating to human activities include those indicating agricultural activity (*agricultural*, *crop*, *field*, *rural*) and the physical/social distinction found also in GEC (*biophysical*, *economic*, *political*, *socio-economic*). As with GEC, this suggests a diverse readership for which such specification is needed.

In carrying out this study, no consideration has been given to the relative frequency of the adjectives in the various journals, nor to the range of texts that each adjective is found in. This is deliberate. The aim is not to find what is more or less typical of each journal but to identify variety where it exists within journals. In AEE, for example, the phrases *understorey environment* (meaning the part of a forest that lies below the tree canopy) and *political environment* occur only four times each, in two texts each. The point is not that AEE often contains references to these entities (*understorey* occurs more frequently in AEE than in the other journals, but *political* occurs more frequently in GEC), but that texts relating to different research orientations co-exist in the journal, confirming it as an interdisciplinary journal.

Turning to the study of *its environment* and *their environment*, it is noted that both bigrams occur at least four times in all the journals except REE. In REE only *their environment* occurs (there are no instances at all of *its environment*.) Overall, the possessive refers to three kinds of entity. One consists of human individuals, communities and societies, as in example 7.

7 It gave the researchers valuable insights into the way **the local people perceive their environment**.

[AEE 2001 Zurayk]

The second comprises plants or animals, as in example 8.

8 Some **plant species** are able to modify **their environment** between establishment and maturity.

[AEE 2005 Fedoroff]

Also noticeable are references to organisms or systems; these are abstractions that may be species, communities or individuals used in modelling (see example 9).

9 They also afford the modeller the analytical tractability necessary to incorporate spatial and temporal interdependencies between **agents and their environment** through movement and decision making.

[AEE 2005 Manson]

As in the study of modifiers, the focus here is on the variety of referents within a single journal. We look at the monodisciplinary journals first. PS has only 13 examples of these bigrams, all of which refer to plants as species or to parts of plants. Examples of the referents include *cell*, *fruit*, *species*, *plant*, *organisms*. The most frequent phraseology is *adapt/response to its/their environment*, and most instances of the bigrams (8 out of 13) stress that plants modify themselves

in order to accommodate to their environment. The other frequent meaning made is the interaction between a plant and its immediate environment. REE has only seven examples. Of these, the referent in six cases is a characterisation of people: *first generation*, *household*, *consumers*. The referent in the remaining instance is *pond-breeding salamanders*. The verbs used with the phrase are *improve*, *value* and *change*, indicating that people both interact with and modify their environment. In line with the previous study of modifiers, then, both the monodisciplinary journals use these phrases with single meanings and a consistent phraseology.

The interdisciplinary journal AEE has 39 instances of the phrase, of which two-thirds have plants, animals or systems as referent, while the other one-third of instances refer to people and communities. Frequent phraseologies construe meanings of relationships/interactions between an organism or community and its environment, perceptions of environment, responses to and modification of the environment. Whereas in PS the phraseologies construe the entity (the plant) as being influenced by the environment, and in REE the phraseologies construe the entity (people) as influencing their environment, in AEE both phraseologies are present.

Most instances of the phrase occur in the other interdisciplinary journal, GEC, which has 74 instances. The vast majority, 57, have people as the referent, with 14 having systems as the referent and only three referring to plants or animal species. Most frequent meanings are around relationships and interactions between the entity and its environment and changes in the entity in response to the environment.

What is striking in GEC, however, is the variety of characterisations of human beings and their society in the referents of the possessive. These range from *forest-dwelling people*, to *economic units called households*, to *decision makers*. Some examples are given here:

- *the influence of* ***humans*** *on their environment*
- *the way some* ***decision makers*** *see their environment*
- *studies of* ***forest-dwelling people*** *and their environment*
- ***Sweden and most other western nations*** *have . . . [prioritized] their environment*
- *a lack of connection felt by* ***the public*** *to their environment*
- *whether* ***people*** *were actually perceptually aware of polluted air in their environment*
- *humans, in* ***economic units called households****, alternately modify their environment*
- *the stresses that* ***a society*** *exerts on its environment*

This combination of 'people in general', 'people in social units', 'people defined by the place they live', and 'people defined as non-expert' reflects the breadth of topics of the articles in the journal, and the range of their research orientation, even when they could all be defined broadly as social science.

The third study is about prepositions. In the BEE4 corpus, a preposition occurs immediately before *the environment* in 67.6% of all instances of the noun phrase

(1394 out of 2062). (The figures for the individual journals are: AEE 74.1%; GEC 61.3%; PS 76.5%; REE 66.4%, suggesting a broad uniformity.) The prepositions *of*, *on* and *to* together account for 62.1% of these preposition instances (867 out of 1394). They constitute the three most frequent prepositions preceding *the environment*: *of the environment* occurs 322 times in the BEE4 corpus, *on the environment* occurs 308 times and *to the environment* occurs 237 times. Table 6.3 gives more information about the relative frequency of the prepositional phrases in each journal. It shows a familiar pattern: GEC and PS at opposite ends of the frequency spectrum, AEE and REE lying between the two extremes. The interdisciplinary journals AEE and GEC have a higher frequency than their monodisciplinary counterparts.

As noted, particular prepositional phrases are known to complement adjectives, nouns and verbs that share aspects of meaning, though this is a one-to-several relationship rather than one-to-one (Hunston & Francis 2000). For example, a sample 100 instances of *on the environment* from the BEE4 corpus shows that the phrase complements the following sets of words:

- *consequences*, *effect*, *effects*, *impact*, *impacts*, *influence*, *load* (e.g. *effect on the environment*)
- *damage*, *pressure*, *risk*, *stress* (e.g. *pressure on the environment*)
- *dependence*, *dependent*, *depends* (e.g. *depends on the environment*)
- *focus*, *information*, *literature*, *results*, *studies* (e.g. *information on the environment*)

Selecting the prepositions, then, is a way of accessing these meanings. Grouping the words as shown in the bullet-point list allows low-frequency words, such as *consequences* or *influence*, to count alongside the more frequent *impact*. In addition, complementation is a structural concept rather than a strictly collocational one, so these grouped words include examples such as *the impact of intensive agriculture on the environment* as well as the more obvious *impact on the environment*. Focusing on the sequences of this type that precede *the environment* casts further light on the kind of entity that the 'environment' is.

Because meaning is a slippery concept, exact quantitative information cannot be given. However, an examination of the three phrases in the four journals suggests that the following meanings are particularly frequent:

- Meanings associated with 'damage' occur with all the prepositions. Examples of phrases include: *degradation of*, *impact on*, *damaging to*.
- Meanings associated with 'protection' occur with *of*. Examples of phrases include: *conservation of*, *management of*, *protection of*.
- Meanings associated with physical leakage of elements occur with *to*. Examples of phrases include *emissions to*, *loss (of sth) to*, *discharge (of sth) to*, *loss (of sth) to*.
- Meanings associated with human awareness, knowledge or research occur with all the prepositions. Examples of phrases include *perception of*, *research on*, *attention to*.

Table 6.3 Frequency of three prepositions preceding *the environment* in four journals.

Journal	*AEE*			*GEC*			*PS*			*REE*		
Raw frequency total	386			393			36			52		
Normalised total	39.33			83.35			3.95			33.46		
Preposition	*of*	*on*	*to*	*of*	*on*	*to*	*of*	*on*	*to*	*of*	*on*	*to*
Raw frequency individual	99	152	135	190	131	72	9	6	21	24	19	9
Percentage	25.6%	39.3%	34.9%	48.3%	33.3%	18.3%	25.0%	16.6%	58.3%	46.1%	36.5%	17.3%

- Meanings associated with the state, components, or characteristics of the environment occur with *of*. Examples of phrases include *aspects of*, *state of*, *sensitivity of*, *vagaries of*.
- Meanings associated with connectivity between human populations and the environment occur with *to*. Examples of phrases include *link to*, *related to*, *ties to*.

The four journals can be looked at independently. However, because frequencies in PS and REE are low, only AEE and GEC will be considered here.

In AEE, meanings associated with 'damage' or 'detrimental impact' are highly frequent. They account for 91.4% of instances of *on the environment* (the most frequent of the phrases in this journal), and phrases such as *damaging to*, *risk to*, *contamination of* and *degradation of* are also frequent. The 'physical leakage' meaning is found frequently with *to*, accounting for 41.4% of the instances of that preposition, and 'state' words occur with *of*, accounting for 46.4% of the total instances of that preposition.

GEC shows a more varied profile. There is, as in AEE, a preponderance of phrases associated with 'damage' (e.g. *degradation of*, *effect on* and *damage to*). The proportions are slightly lower, though, accounting for 73.2% and 29.1% of the instances of *on* and of *to*, respectively. The 'state' meaning is also frequent with *of*, accounting for 35.2% of instances of the preposition. GEC includes more instances of *the environment* as a topic of research, discussion, and thought, with phrases such as *conception of*, *agreement on* and *approach to*. These account for between 12% and 15% of the relevant preposition. GEC also includes more instances of phrases indicating 'improvement', such as *care of*, *protection of*, and of phrases indicating connections, such as *link to*, *ties to*.

In summary, this section has looked at the word *environment* from a number of perspectives. The word occurs with different levels of frequency in each of the BEE4 journals, suggesting that the concept is more central to GEC and AEE than to REE and PS. It was also noted that the word has different, though related, meanings and that researchers might relate to these meanings in different ways. Five different nuances of meaning were suggested at the beginning of this section, in examples that are repeated here:

A general, unbounded natural entity (example 10):

10 . . . with a particular focus on the capacity of existing legal and administrative structures to protect the **environment**.

[GEC 1993 Jasanoff]

A general, unbounded entity: air (example 11):

11 The practical difficulty of incorporating the **environment** into traditional measures of productivity is the task of assigning weights to the output of emissions.

[REE 2003 Aiken]

A specific, small-scale entity existing in relation to another (example 12):

12 Moreover, the **environment** of the developing embryos can be modified by adding specific signalling molecules.

[PS 2003 Baldan]

A mid-range natural entity (example 13):

13 The aquatic **environment** in rice fields is largely influenced by the growth of the rice crop.

[AEE 2005 Vromant]

A social entity (examples 14 and 15):

14 . . . supporting institutions including those that shape the regulatory **environment**.

[AEE 2007 Pascual]

15 . . . as a response to experienced or expected changes in the societal . . . **environment**.

[GEC 2009 Pahl-Wostl]

Overall, it is apparent that plants are represented as adapting to the environment, while human being are representing as exerting change on the environment. Relating this to the journals, it has been shown that the various uses of the word distinguish the four journals. The monodisciplinary journal PS is not 'about' the environment at all. It is 'about' plants and the systems it comprises, and the word *environment* occurs to describe properties of plants. The other monodisciplinary journal, REE, has the highest relative frequency of the unmodified phrase *the environment*, suggesting that the term refers to an unnuanced entity that is large-scale, general and operates as a factor in economic modelling. The two interdisciplinary journals use the word in a variety of meanings, in particular relating to both human and natural phenomena. AEE has the widest range of modifiers, suggesting an awareness on the part of writers (or editors) that the term has multiple meanings and must be defined when used. GEC is also notable for its breadth of use: in each of the three studies it showed more variation than the other journals.

6.3 The notion of science

Preliminary investigations of the BEE4 corpus suggested that 'science' might be a contested term in the journals. For most writers in the corpus, science is the way that knowledge is built and problems solved. Where those problems cannot be solved, it is because current knowledge is inadequate and more science must be done. For many other writers, however, scientific research is only part of the solution to environmental problems, and scientists must liaise with others, such

Table 6.4 Normalised frequencies (pmw) of *science*/*social science*/*scientific*/*scientist*/*s* in four journals.

	AEE	*GEC*	*PS*	*REE*
Science (excluding *social science*)	28.93	519.00	21.94	12.23
Social science	2.04	41.78	0.11	0.64
Scientific	56.65	610.20	19.20	19.95
Scientist	2.04	21.85	0.44	00.00
Scientists	32.20	314.54	3.40	8.36
Total	121.86	1507.37	45.09	41.18

as social scientists or policy makers, in order to achieve change. For a minority of writers, science itself is part of the problem, and a radical shift in research paradigms is necessary if environmental disaster is to be avoided.

The positions outlined previously can be corroborated by the study of a group of words comprising the terms *science*, *social science*, *scientific* and *scientist/s*. Table 6.4 gives the normalised (per million words) frequency of each of the search terms in the four journals.

Table 6.4 shows that this word family is much more frequent in GEC than in the other journals and the same is true of each word or phrase individually. In each case, AEE is the journal with the next highest frequency, and the frequencies in PS and REE are very low. This variation is clearly not because of different topics in the journals. They all report scientific research conducted by scientists (or social scientists). It would appear, then, that the difference in frequency occurs because of differences in disciplinary stance rather than because of content.

There is clearly a great deal of material that could be reported on here. To keep it within reasonable limits, just two detailed studies will be reported here: a study of the word *scientists* in AEE and a study of the word *scientific* in GEC.

6.3.1 Scientists *in AEE*

Before reporting this study in detail, two observations about the use of corpus investigation techniques in relation to how groups of people are represented in discourse are in order. The first observation warns against under-interpretation: as shall be seen here, it is not sufficient to look at immediate collocates of a word in order to establish the discoursal significance of its use. For example, the word *scientists* may collocate with a research action such as *collect data*. This appears to be a neutral use. Looking in closer detail may indicate, however, that in a given article scientists are being criticised for collecting only a particular kind of data and ignoring more important kinds. This means that concordance lines must be scrutinised in detail, and this places limits on the amount of data that can be dealt with.

The second observation is a warning against over-interpretation. It might be thought that searching for the word *scientists* will yield all relevant information about scientists. To put this the other way round, it might be thought that if we want

to know how scientists are written about in the corpus texts we should search for the word *scientists*. This is not the case. The journals REE and PS are as much about the work of scientists as AEE and GEC are, yet the word *scientists* occurs in those journals relatively rarely. What we discover when we search for the word *scientists* is how people conducting research are talked about when they are referred to as *scientists*. Using the word in itself performs an act of potential 'othering' (Said 1978) and invites reflection on the actions and assumptions of the people involved. In other words, the word *scientists* is unlikely to be used when the actions and thoughts of scientists are being presented uncritically or unreflectively. Neutral discourse about scientists tends not to refer to them as *scientists* but as named individuals, or defined groups. For our study, this has both positive and negative consequences. On the negative side, it means that we cannot answer the question 'How does this journal talk about the work of scientists?' by looking at the word *scientists*. On the positive side, it means that if we are interested in challenges to the supremacy of the scientific paradigm and scientific knowledge, we might expect to find this by looking at the word *scientists*. In looking at the word *scientists*, therefore, we are not asking 'how are scientists represented in this corpus?' but we are asking 'what are the consequences for the representation of scientists of using the word *scientists*?'

There are 316 instances of *scientists* in AEE. From the alphabetically sorted concordance lines, the following generalisations can be made:

- Looking at the kinds of verbs that occur to the right of *scientists*, it is apparent that scientists are mostly represented as the actors in various research processes: *develop models*, *design surveys*, *focus research*, *have proposed a variety of theoretical frameworks*, *judge a model's success*. Similar verbs occur to the left of the phrase *by scientists*: *project conducted by scientists*, *models developed by scientists*. Scientists are also categorised by their field of expertise: *agricultural scientists*, *agronomy scientists*, *ASB scientists*, *natural scientists*, *social scientists*, *soil scientists*.
- The word *scientists* co-occurs with *and*. Scientists are linked with other groups, and it is stressed that they should work together effectively: *scientists and farmers/local people/policy makers*; *there has been little dialogue between scientists working . . .* ; *improved communication between scientists, local authorities*.
- There are some indications that scientists need to be more effective. These take the form of recommendations and accounts of help given: *scientists must articulate ways forward*, *scientists should forecast*, *enabling scientists to play this role better*, *it is necessary for scientists to provide . . . ,force scientists to work with empirical models*, *help scientists optimize field surveys*.
- In some cases, lack of certainty and lack of homogeneity are the theme, reflecting the human and contingent aspect of scientific research, and a departure from mainstream scientific tradition: *creates space for scientists to maintain conflicting opinions*, *has prompted scientists to question . . .* , *some scientists even dared to formulate their narratives*, *some scientists do not share this faith*, *call from some scientists to think more deeply*, *they led some scientists to escape from mainstream approaches*.

Table 6.5 Number of papers in AEE containing instances of the word *scientists*.

Number of instances	*Number of papers*
0	1638
1	95
2	28
3	10
4	5
5	2
6	1
7	1
10	1
21	1
25	1
30	1

As noted, a more detailed scrutiny of each concordance line is needed to give a more complete picture. It is also the case that some articles contain a very large number of the search word, while others contain only one or two. Table 6.5 shows the number of individual articles containing each number of instances.

The vast majority of articles contain no mention of the word at all. 95 articles contain only one use, 28 contain two uses and so on, with four articles containing between ten and 30 uses each. When the use of the word in each article is observed, three uses predominate. The highest figure is for references to collaboration between scientists and other groups of people, mainly farmers on the one hand and policy makers on the other. This is the use found in 63 articles. Example 16 gives an example of this use, in this case a successful collaboration between groups. Other examples recommend improved collaboration or bemoan the lack of such collaboration.

16 . . . and participatory monitoring by farmers, **scientists** and decision makers helps to develop strategies to keep these indicators within critical thresholds.
[AEE 2001 Goma]

The next most frequent use simply reports the research undertaken by *scientists*, illustrated in example 17. This is the use in 40 articles.

17 For this reason, **scientists** classify pyrethroid chemicals as super-toxic and warn that they have the potential to significantly bioaccumulate in aquatic organisms.
[AEE 2006 Brady]

The only other use of high frequency comprises reference to interdisciplinary research, either between natural scientists and social scientists, or between natural scientists from different disciplines. Such research may be reported, recommended or problematized. In some cases references to a novel research method are interpreted as reference to research from a different discipline. Example 18 shows interdisciplinary research being reported, while example 19 is a recommendation for such research.

18 Agronomists and social **scientists** studying the maintenance of maize populations . . . distinguish three types of seeds.

[AEE 2007 Brush]

19 [This paper] considers ways that **scientists** from different disciplines can collaborate to determine the functions and value of agrobiodiversity.

[AEE 2007 Jackson]

Other minority uses of *scientists* are: in reports of research carried out on scientists themselves, reporting attitudes etc. (four articles); in descriptions of institutions (three articles); and in articles challenging the scientific paradigm (two articles). Table 6.6 summarises these figures.

Of particular interest, of course, are the three articles that contain over 20 instances of the word *scientists* each. These are described in turn here.

- AEE 2004 Pontius. 21 instances. The article offers a novel model for assessing land-use change. Example 20 gives an extended example, with three instances of the word. It illustrates that the article offers advice from the 'outside'. In other words, the writer is construed as a member of one discipline offering assistance to another discipline.

20 The problem is that **scientists** usually already know that persistence dominates the landscape. **Scientists** want to identify the dominant signals of land change. The methods of this paper allow **scientists** to identify the signals of change separately from any given level of persistence.

[AEE 2004 Pontius]

Table 6.6 Uses of *scientists* in AEE.

Use of scientists	*Number of papers containing this use*
Referring to collaboration between scientists and other groups	63
Scientists as actors in the research process	40
References to interdisciplinary science	28
Scientists as the object of research	4
Accounts of institutions	3
Challenges to the scientific paradigm	2

- AEE 2007 Tomich. 25 instances. The article reports a consultation of scientists involved in Alternatives to Slash and Burn (ASB), itemising opportunities and challenges afforded by work that is highly interdisciplinary and involving multiple non-scientific partners. Example 21 illustrates this. Interdisciplinarity and the relationship between scientists and other groups are therefore the topic of this article.

21 From the 'technological optimism' days, in which soil science, agronomy and other biophysical disciplines predominated, the mix of ASB **scientists** has steadily grown to include more ecologists, economists, geographers, and other social scientists. In parallel, the set of stakeholder has grown from an initial focus on farmers and partners in participating national agricultural and forestry research systems and universities to include policymakers at various levels, environmental NGOs and civil society groups.

[AEE 2007 Tomich]

- AEE 2009 Pollini. 30 instances. As with AEE 2007 Tomich, scientists themselves are the object of research in this article, and as with AEE 2007 Tomich, the scientists are concerned with ASB. The focus here is on the challenge to science itself posed by the research agenda, with scientists themselves questioning the effectiveness of the scientific approach to problem solving (example 22), and social scientists encouraging alternative approaches (example 23).

22 Some **scientists**, however, do not share this faith in predictive models and perfect design.

[AEE 2009 Pollini]

23 They encourage **scientists** to look on the side of the constructivist paradigm and to adopt flexible approaches such as adaptive management, social learning and action-research.

[AEE 2009 Pollini]

To summarise this account of *scientists* in AEE: although the word *scientists* occurs in only a minority of articles in AEE, it is proportionally more frequent in this journal than in the monodisciplinary natural science journal PS. The word is sometimes used in the context of reporting what researchers do and think, but it more frequently indicates a site of meta-disciplinary discussion. Articles that use the word refer to interactions between groups, either scientists influencing farmers, policy makers and the like, or scientists from different disciplines influencing each other. The scientific research paradigm itself is rarely challenged, and where natural scientists and social scientists are depicted as working together this collaboration is presented as harmonious.

6.3.2 Scientific *in GEC*

The adjective *scientific* occurs 2805 times in GEC. Sketch Engine's 'Word sketch' function gives a list of 25 nouns modified by *scientific*. In order of frequency these are: *community* (193), *knowledge* (193), *uncertainty* (146), *research* (106), *information* (95), *assessment* (80), *consensus* (70), *understanding* (62), *evidence* (54), *finding* (52), *literature* (52), *work* (28), *basis* (26), *credibility* (26), *advice* (23), *debate* (21), *report* (20), *input* (19), *expertise* (18), *certainty* (17), *controversy* (17), *discipline* (16), *discourse* (16), *inquiry* (14), *opinion* (14).

These nouns might be divided into:

- Research entities: *assessment*, *evidence*, *finding*, *input*, *inquiry*, *literature*, *report*, *research*, *work*
- Social entities: *community*, *discipline*
- Group intellectual entities: *certainty*, *consensus*, *controversy*, *debate*, *discourse*, *uncertainty*, *understanding*
- Outward-facing intellectual entities: *advice*, *credibility*, *expertise*, *information*, *knowledge*, *opinion*

The research and social entities are likely to be relatively neutral; the group intellectual entities feature both unity (*certainty*, *consensus*, *understanding*) and disagreement (*controversy*, *debate*, *uncertainty*). Of particular interest are the nouns categorised here as 'outward-facing'. These are entities that the scientific community presents to the rest of the world (e.g. *scientific advice*, *scientific information*, *scientific opinion*) or ways in which the rest of the world views the scientific community (e.g. *scientific credibility*, *scientific expertise*).

Two phrases from this latter group will be investigated in greater detail. In each case, the phrase turns out to be ambivalent. First, *scientific advice*. This may be treated as unproblematic – scientific advice is the straightforward consequence of scientific research – or as problematic – scientific advice is either flawed or ignored. Table 6.7 shows a categorisation of the 23 instances.

According to Table 6.7, over half the instances of *scientific advice* problematize the concept, usually by contrasting it with political expediency.

Turning to the more frequent phrase *scientific knowledge*, it is worth a special study, not only because of its frequency, but because it is a site of potential conflict. Examples 24 and 25 illustrate this.

24 The plan, however, failed to draw adequately on existing **scientific knowledge** of marine pollution sources.

[GEC 1991 Boxer]

25 Thus ignorance is endemic to **scientific knowledge**, which has to reduce the framework of the known to that which is amenable to its own parochial methods and models.

[GEC 1992 Wynne]

Table 6.7 Categorising instances of *scientific advice* in GEC.

Unproblematic	. . . a federal research institute providing technical and scientific advice.
	. . . strengthen the link between scientific advice and the policy world.
	The essence of the political debate, into which the IPCC scientific advice was being fed, was precisely about . . .
	What are the implications for policy of obtaining scientific advice from research institutions.
	. . . the key role of scientific advice in this issue.
	. . . requested its SBSTA to prepare scientific advice in order to integrate biodiversity considerations.
	It would help provide scientific advice required by decision makers . . .
	. . . might play in providing scientific advice on the climate change issue.
	. . . and provide the scientific advice necessary for the integration of biodiversity considerations . . .
Problematic	Thus, scientific advice on climate change, having to contend between an uncertain science and call for action.
	. . . intended to secure legitimate scientific advice, but did the result of this process create a sound consensual basis . . .
	. . . the basis for the claimed ‘intrinsic ambivalence’ of scientific advice.
	. . . the intrinsic ambivalence of scientific advice derived from the research community invites other interests . . .
	. . . arguing that the intrinsic ambivalence of scientific advice derived from the research community invites other intere.
	. . . the impact of scientific advice on the present negotiations seems to have been marginal.
	. . . exposed the limits of scientific advice and the essentially political nature of the IPCC process.
	. . . the idea of a climate treaty based on scientific advice was soon hotly debated.
	Limits imposed on scientific advice by politics.
	. . . there have been countless efforts to provide scientific advice to decision makers.
	. . . our desire to find powerful analytic tools that will provide scientific advice on climate change and sustainable . . .
	In fact, the bulk of their scientific advice came from members of the epistemic community.
	. . . all used scientific advice less for the purpose of informing their own actions than for the promotion . . .
	. . . some of the scepticism which scientific advice has to face on this issue.

Table 6.8 Number of papers in GEC containing instances of the phrase *scientific knowledge*.

Number of instances	*Number of papers*
0	348
1	51
2	10
3	12
4	7
5	2
7	2
8	1
42	1

Example 24 treats scientific knowledge as a given, presupposed, whereas example 25 challenges its very basis. The phrase, then, is potentially reflective of an agonistic approach to science, though not necessarily so.

In GEC, *scientific knowledge* occurs 207 times. As with *scientists* in AEE, the phrase is concentrated in a relatively few articles. Table 6.8 shows the distribution.

One article has a massive 42 instances and a further 12 articles have four or more instances each. The article with 42 instances (GEC 1992 Wynne) is entitled 'Uncertainty and environmental learning: reconceiving science and policy in the preventive paradigm'. The abstract summarises its critical approach to the model of knowledge in natural science disciplines (see example 26).

26 [T]he implications for current assumptions about **scientific knowledge** . . . a critical examination of approaches to characterizing different kinds of uncertainty . . . the key dimension of unrecognized indeterminacy in **scientific knowledge** . . . involves the possible reshaping of the 'natural' knowledge itself.

[GEC 1992 Wynne]

In this article, the phrase *scientific knowledge* is associated with critique of underlying assumptions, and the same appears to be true of the use of this phrase throughout the journal. Figure 6.1a, for example, shows all instances of the phrase *scientific knowledge is* in the journal as concordance lines sorted so that the words immediately to the right of the node phrase are in alphabetical order.

Some of these lines, repeated here as Figure 6.1b, indicate the constructed, contingent and partial nature of scientific knowledge.

Some, repeated here as Figure 6.1c, talk about how such knowledge is (mis) used or (mis)understood.

Some, repeated here as Figure 6.1d, stress the uncertainties and slow progress inherent in scientific knowledge.

rn ideology of science: that scientific knowledge is a public good, equally available and
fact that the advancement of scientific knowledge is a slow process - and response measures
social environment where new scientific knowledge is but one factor (Berkhout and Scoones 19
d from the processes whereby scientific knowledge is communicated to policymakers towards t
are taken as relative terms, scientific knowledge is considered more objective than other sy
lassifications on which that scientific knowledge is constructed. Before discussing these ki
ant western society. Western scientific knowledge is critically important; no one would deny
owever, we can see that when scientific knowledge is deployed in the public domain, the soci
derstanding about the issue. Scientific knowledge is elevated above local knowledge, essenti
lain differences in the ways scientific knowledge is generated and used – that is, the socia
iting to be read by evolving scientific knowledge is giving way to the idea that there are m
nition elusive. The state of scientific knowledge is growing impressively, but remains inade
ards the processes [whereby] scientific knowledge is integrated with policy concerns in comp
t point out that even though scientific knowledge is invariably incomplete this does not pre
onships of power in the ways scientific knowledge is made, mobilised and valorised. It is th
a problem when (as is usual) scientific knowledge is misunderstood and is institutionalized
research specialty. Yet the scientific knowledge is not fully determined by 'the facts' – w
s framed in the first place. Scientific knowledge is not independent from cultural factors,
it fails to acknowledge that scientific knowledge is often widely represented in policy circ
g public knowledge, of which scientific knowledge is one sort ... The point is that public k
183) points out, saying that scientific knowledge is open to social evaluation and negoti
in the production and use of scientific knowledge is receiving more an more attention. In th
y address the question: What scientific knowledge is required to predict future change reli
ssue of whether, in general, scientific knowledge is socially constructed (and as Wynne (199
ose of some, especially when scientific knowledge is stripped of the uncertainty it acknowl
ange must be postponed until scientific knowledge is sufficient for the task, or it leads
absolute correctness of all scientific knowledge is to be shunned. This being the case, pe

Figure 6.1a Scientific knowledge is in GEC, right sorted.

rn ideology of science: that scientific knowledge is a public good, equally available and
social environment where new scientific knowledge is but one factor (Berkhout and Scoones 19
lassifications on which that scientific knowledge is constructed. Before discussing these ki
derstanding about the issue. Scientific knowledge is elevated above local knowledge, essenti
lain differences in the ways scientific knowledge is generated and used – that is, the socia
iting to be read by evolving scientific knowledge is giving way to the idea that there are m
onships of power in the ways scientific knowledge is made, mobilised and valorised. It is th
research specialty. Yet the scientific knowledge is not fully determined by 'the facts' – w
s framed in the first place. Scientific knowledge is not independent from cultural factors,
g public knowledge, of which scientific knowledge is one sort ... The point is that public k
183) points out, saying that scientific knowledge is open to social evaluation and negoti
in the production and use of scientific knowledge is receiving more an more attention. In th
ssue of whether, in general, scientific knowledge is socially constructed (and as Wynne (199

Figure 6.1b Scientific knowledge as contingent.

d from the processes whereby scientific knowledge is communicated to policymakers towards t
are taken as relative terms, scientific knowledge is considered more objective than other sy
owever, we can see that when scientific knowledge is deployed in the public domain, the soci
lain differences in the ways scientific knowledge is generated and used – that is, the socia
ards the processes [whereby] scientific knowledge is integrated with policy concerns in comp
a problem when (as is usual) scientific knowledge is misunderstood and is institutionalized
it fails to acknowledge that scientific knowledge is often widely represented in policy circ

Figure 6.1c Scientific knowledge (mis)used or (mis)understood.

```
fact that the advancement of scientific knowledge is a slow process - and response measures
nition elusive. The state of scientific knowledge is growing impressively, but remains inade
t point out that even though scientific knowledge is invariably incomplete this does not pre
ose of some, especially when scientific knowledge is stripped of the uncertainty it acknowl
ange must be postponed until scientific knowledge is sufficient for the task, or it leads
absolute correctness of all scientific knowledge is to be shunned. This being the case, pe
```

Figure 6.1d Scientific knowledge uncertain and slow.

A second illustration of how the bigram is used is an examination of the three-word phrase *of scientific knowledge*, which occurs 62 times in the journal. The nouns that appear immediately to the left can be divided into these groups:

- Nouns meaning 'scientific knowledge is an object of academic study': *sociology* (10), *social analysis* (1), *studies* (1)
- Nouns meaning 'scientific knowledge is a constructed and political entity': *constitution* (1), *construction* (1), *co-production* (1), *creation* (1), *dominant regulatory discourses* (1), *certain kinds* (1), *'expert'/constructed/strategic nature* (4), *notion* (1), *power* (1), *the very production* (1), *status* (1), *technical structure* (1)
- Nouns indicating that 'scientific knowledge is a commodity that is given, received, and used': *application* (1), *communication* (1), *consumption* (1), *diffusion* (1), *incorporation* (1), *recipients* (2), *role* (4), *use/s* (4)
- Nouns meaning 'scientific knowledge is like a physical object': *body* (4), *borders* (1), *boundaries* (1), *domain* (1), *an object* (1), *any piece* (1)
- Nouns meaning 'scientific knowledge needs to be progressed and protected': *advancement* (1), *distortion* (1), *pursuit* (1), *objective validation* (1)
- Other nouns and phrases: *on the basis* (1), *in terms* (1), *state* (1)

For the most part, then, the phrase *scientific knowledge* is used when the body of information produced by scientific disciplines is being treated as contingent and potentially suspect. Although it is used in relatively few articles, its use nonetheless represents a key theme in the journal GEC. For comparison, the phrase occurs 25 times in AEE, twice in REE and never in PS. Its use in AEE is less confrontational than in GEC: in AEE, *scientific knowledge* is expressed as necessary, in need of development, and requiring translation into practical action. There is only one instance of *scientific knowledge is* (interestingly, rather like the GEC instances in that it indicates insufficiency: *scientific knowledge is not enough* . . .) and only two of *of scientific knowledge* (*input of*, *lack of*).

6.3.3 Conclusions about the notion of science

The two indicative studies in this section have drawn attention to the role that the words *science/scientific/scientists* plays in the interdisciplinary journals. The first observation was that all the words are more frequent in the interdisciplinary journals than in the monodisciplinary ones. This suggests that the reflexive use of the words – scientists talking about science – might be indicative of interdisciplinary discourse. This has been linked to the concept of 'othering': labelling scientists as such potentially distances them from the reader. Then, the word *scientists* was examined in detail in AEE. In that journal, the word appears to act as a magnet for discourse that construes disciplines as mutually supportive and academic disciplines as contributing to real-world action. Selecting the word selects those texts that construe interdisciplinarity in this way. The final study, *scientific* in GEC, presented us with a more antagonistic view of interdisciplinarity, with the scientific paradigm itself coming in for criticism.

The different approaches of the AEE and the GEC writers to interdisciplinarity are therefore revealed.

6.4 What counts as *important*

The final word chosen for this concordance study is an overtly evaluative one: *important*. Some of the investigations in this section are inspired by Muguiro (2019). Evaluations of importance perform a variety of functions, from discourse-related functions such as emphasising a statement or conceding a point, to research-related functions such as justifying a research methodology. Examples 27–31 illustrate the various functions, using a variety of adjectives. It is noticeable that in all these examples, a phenomenon or activity is evaluated as a route to the achievement of a goal (Hunston 1993). In other words, whatever the immediate rhetorical function of the 'important' word is, the sentence as a whole gives information about how a desired end point is to be reached. Everything around that goal – what it is, who will achieve it and how – is an important contributor to differences between texts.

27 Given our interest in modelling the competition for land, it is **important** to recognise that land is a heterogeneous endowment.

[REE 2009 Golub]

Discourse function: emphasising the truth of a statement.
Goal indicated: Researchers create a viable model.

28 Since light is **essential** for plants to provide energy for metabolism . . .

[PS 2001 Sharma]

Discourse function: asserting the value of a physical phenomenon ('light').
Goal indicated: Plants metabolise.

29 However, experiments over many seasons under natural field conditions are necessary to assess differences in yield . . . because the effects of UV-B radiation on growth and yield are strongly influenced by seasonal microclimatic conditions.

[AEE 2001 Kumagai]

Discourse function: asserting the value of a research method.
Goal indicated: Researchers achieve viable results.

30 [T]he performance of political leadership was **crucial** in the development of integrative solutions.

[GEC 1992 Andresen]

Discourse function: asserting the value of political action.
Goal indicated: Politicians develop solutions.

31 At this stage, it is **important** to expand the domain of investigation and communication: to include a wider range of social groups and disciplines in a sustained dialogue.

[GEC 1991 Wescoat]

Discourse function: asserting the value of interdisciplinary research and citizen science.
Goal indicated (from preceding sentence): Researchers assess the impact of climate change.

Examples 27 and 29 are about researchers and how they gain knowledge. Example 28 is about the natural world – plant metabolism – rather than the research world. Examples 30 and 31 integrate a specific research agenda with political action (example 30) and with interdisciplinary research and social groups (example 31). In all cases, even example 28, the evaluation of importance is used to justify a particular line of research. It is, therefore, a potential resource for justifying the work from one discipline to researchers in another. Differences in frequency and usage between the journals are therefore worth investigating.

6.4.1 Frequency

As illustrated previously, several adjectives can express the concept of 'important'. Eight such adjectives have been identified from the nine adjectives listed by Francis, Hunston and Manning (1998) in the construction 'IT LINK-VERB [IMPORTANT] THAT'. (The ninth, *significant*, is not included here because of its extensive use in reporting statistical measurements.) Table 6.9 shows the normalised frequency of each of the eight adjectives in each journal in the BEE4 corpus. In each word column, the underlined number shows the journal with the highest relative frequency of the word in question. In most cases, the journal is GEC, but *essential* is proportionally more frequent in PS. In each row, the figure in bold shows the most frequent word of the eight in each journal. In every case, *important* is the most frequent word. As a consequence, *important* is the most frequent word overall. The adjectives *imperative*, *paramount* and *vital* are of relatively low frequency. The words *critical*, *crucial*, *essential* and *necessary* are fairly frequent.

The order of the journals assessed by relative frequency of all the adjectives is: GEC (1438), REE (990), AEE (948), PS (829). For each pair of journals (GEC and REE; AEE and PS), the interdisciplinary journal has a higher frequency of the adjectives than the monodisciplinary one does.

Because the word *important* is the most frequent of the eight words in all the journals, the following studies will relate to this word only, with the caveat that this makes the study only illustrative. The studies focus on two phrases: *it is important to* and *important for/in* (i.e. the three adjective patterns: IT + LINK-VERB + ADJECTIVE + TO-INFINITIVE; ADJECTIVE + *for* + NOUN PHRASE; and ADJECTIVE + *in* + NOUN PHRASE (Hunston & Francis 2000).

Table 6.9 Normalised (pmw) frequencies of words meaning 'important' in four journals.

	Critical	*Crucial*	*Essential*	*Imperative*	*Important*	*Necessary*	*Paramount*	*Vital*	*total*
AEE	96.28	30.87	62.46	4.99	**607.36**	131.84	2.65	12.12	948.57
GEC	264.27	95.87	112.62	22.48	**918.16**	221.85	7.42	35.63	1438.30
PS	68.79	39.94	128.58	1.43	**471.21**	106.31	1.54	11.85	829.65
REE	126.76	50.19	43.11	3.22	**570.74**	193.04	0.64	3.22	990.92

6.4.2 It is important to

The phrase *it is important to* is used to indicate the value of an action, either one which the researchers have undertaken or one which is recommended. The action may relate to the discourse, as in *it is important to note*, or to the research world, as in *it is important to analyse*, or to the social world, as in *it is important to consult*. Table 6.10 shows the raw and normalised (per million words) frequencies of the phrase in each of the BEE4 journals.

As expected, the phrase is most frequent in the social science and interdisciplinary journals and least frequent, relatively, in PS. The most striking difference is between the interdisciplinary AEE and the monodisciplinary PS. Further investigation is carried out with concordance lines: 100 sample lines are examined in those cases where there are more than 100 overall – that is, for all the journals except REE.

Because this is a phrase with a variety of functions, and it is difficult to know how fine-grained to make the analysis, the first analysis uses Halliday's process types (Halliday 1994): 'material process' (effectively, research actions such as *analyse* or *select*), 'mental process' (such as *consider*, *note* or *take into account*), and 'verbal process' (effectively, placing emphasis on a statement, as in *it is important to point out*). Table 6.11a gives the results. Because there are only 66 lines for REE, additional percentage figures are given in brackets. One instance in PS is not a verb.

The mental process verbs cover three broad areas: 'something is important for our argument', such as *it is important to acknowledge*; 'something is important for us as researchers to ask or find out', such as *it is important to understand*; and the formulaic emphasising *it is important to note/notice*. In Table 6.11b, the first of these is termed 'argument', the second 'research' and the third '*note*'. Then, because the material processes and the mental: research processes both relate to research, these are added together in the 'Material + research' row. The mental:

Table 6.10 Frequency of *it is important to* in four journals.

	AEE	*GEC*	*PS*	*REE*
Raw frequency	296	236	138	66
Frequency pmw	30.16	50.05	15.14	42.47

Table 6.11a Frequency of verbs following *it is important to* in four journals.

	AEE	*GEC*	*REE*	*PS*
Material	38	30	13 (19.69%)	39
Mental	57	67	48 (72.72%)	43
Verbal	05	03	05 (7.57%)	17
Total	100	100	66 (99.98%)	99

Table 6.11b Frequency of verbs following *it is important to* in four journals, detail.

	AEE	*GEC*	*REE*	*PS*
Material	38	29	13 (19.69%)	39
Mental: argument	10	17	11 (16.66%)	01
Mental: research	23	20	17 (25.75%)	20
Mental: *note*	23	28	20 (30.30%)	21
Verbal	05	05	05 (7.57%)	17
Material + research	61	49	44 (66.66%)	59
Argument + verbal	15	22	23 (34.84%)	18

argument and the verbal processes similarly fulfil equivalent functions, so these are also added together in the 'Argument + verbal' row.

The results for the journals are not very different: in each case there are more instances of attaching importance (with this phraseology) to research actions and fewer instances of attaching importance to arguments. However, the difference is larger between the two natural science journals (AEE and PS) than between the two journals that include social science (GEC and REE).

6.4.3 Important for/in

The sequence 'x is important for/in y' is interesting in that it reveals the priorities for the community in which the texts are set. The approach to be taken here is to identify the prevailing 'semantic sequences' (Hunston 2008) around these phraseologies in each of the four journals. First, though, Table 6.12 gives the overall frequencies.

Unsurprisingly, GEC shows the highest relative frequency of use, but in this case PS has a surprisingly high frequency.

The next stage is to identify the sequences of meaning used in each journal. For instance, example 32 is analysed as the sequence:

'research entity (*results*)' + 'important (*may be important in*)' + 'application (*developing plant resistance*)'

32 [O]ur results may be **important** in developing plant resistance to viruses. [PS 2001 Shin]

This example both asserts the value of the research ('the results are important/useful') and implies the priority of the 'real-world' application of that research ('our research is valuable because it has applications'). The value-system of the discourse community of the journal is thereby established.

Not all uses of *important for/in* so obviously imply a value-system. Example 33, also from PS, exemplifies the sequence:

'natural entity (*enzymes*)' + 'important' + 'natural process (*chlororespiration*)'

Table 6.12 Frequency of *important for/in* in four journals.

	AEE	*GEC*	*PS*	*REE*
Raw frequency	778	437	625	93
Frequency per million words	79.27	92.69	68.57	59.84

33 Two thylakoidal enzymes, which are **important** in cholorespiration, have been molecularly characterised.

[PS 2007 Tallon]

Here the importance is restricted to the natural world, but it is implied that studying an important enzyme is itself of significance.

The full list of identified sequences is as follows:

1 'Research process' + 'important' + 'Application'
2 'Research entity/process' + 'important' + 'Research entity/process'
3 'Natural entity/process' + 'important' + 'Natural entity/process'
4 'Natural entity/process' + 'important' + 'Social entity/process'
5 'Natural entity/process' + 'important' + 'Research process'
6 'Social entity/process' + 'important' + 'Social entity/process'
7 'Social process' + 'important' + 'Research entity/process'
8 'Social entity/process' + 'important' + 'Natural entity/process'
9 'Intellectual process' + 'important' + 'Social entity/process'

It is not possible to undertake a reliable quantitative study of these sequences, because identifying them is a manual process, but an indication of their relative frequency in the four journals is obtained by classifying 20 randomly selected instances from each journal. The results are shown in Tables 6.13 to 6.16.

Of particular interest is the sequence 'natural entity/process' + 'important' + 'natural entity/process'. This sequence does not involve human activity at all and therefore represents the writer reporting on the natural world without representing it as influencing or being influenced by social or research processes. All the other sequences do involve those social or research processes, either as first or last element in the sequence. In one sense, then, the 'natural' + 'natural' sequence demonstrates one kind of stance towards the natural world and all the others demonstrate variations on another stance. In the first, the natural world is autonomous; in the others, the natural world is in some sense subservient to the human one. The indicative frequency of the 'natural' + 'natural' sequence in each journal is therefore instructive.

Table 6.13 (AEE) shows the largest number of sequence types, and this can be said to reflect the interdisciplinary nature of the journal. Natural, social and research entities are evaluated as important, and natural, social and research entities are identified as the rationales for that evaluation. In other words, the journal reports research on both natural entities (such as *fungi* or *earthworm activity* or

Table 6.13 20 instances of *important for/in* from AEE showing semantic sequences.

Natural entity/process	*Important*	*Social entity/process*
Winter and autumn	are relatively less important for	agriculture in this region
The fungi involved	are very important for	agriculture
Water relations and resistance to degradation	are also very important for	achieving the goals of sustainable soil management
Areas	important for	recreation and biodiversity
Natural entity/process	**Important**	**Natural entity/process**
Linear features	are important for	botanical diversity
Fast mineralisation of catch crop S	is important for	early plant growth
Summer fog	is important in	ameliorating moisture stress
Pesticide emissions	were more important in	HB_2
Vegetation structure	is architecturally important for	web builders
Earthworm activity	is important for	improving and maintaining soil fertility
Landscape features in the regions surrounding the target sites	are important for	conservation of species-rich grasslands
Social entity/process	**Important**	**Social entity/process**
Rural roads	are important for	the development of the northern highlands
Land use drivers that	are important for	Europe
Cropping activities	become important in	this region
Farmers' ages, educational backgrounds and values	are important for	decision-making in landscape management
Social entity/process	**Important**	**Natural entity/process**
Seasonal adoption of patterns of habitat use	may be important for	the corn bunting
Social entity/process (research input)	**Important**	**Research process**
The local scale of farm specialisation as well as	are important for	the prediction of land-use intensity
Natural entity/process	**Important**	**Research process/outcome**
A good resistance against the most important diseases	were considered more important	in the choice [of varieties]
Erosion risk and contributing distance	are important for	the variation in the transport factor
Research entity	**Important**	**Research process**
These uncertainties	are less important for	this study

resistance against disease) and social entities (such as *land use drivers* or *seasonal adoption of patterns of habitat use* or *rural roads*). The identified desirable outcomes involve human activity (such as *agriculture*), natural beneficiaries (such as *botanical diversity*), and research procedures (such as *choice of varieties*). The 'natural' + 'natural' sequence discussed previously accounts for just under half (seven) of the 20 examples. In short, it is the variety of instances that is important in this journal.

Table 6.14 (GEC) also represents an interdisciplinary journal. Here, though, the range of sequences is a little less: five in all. What is striking is that in only one of the 20 instances has 'natural entity/process' as the final sequence element (though two others have it as first element) and there are no instances of the 'natural' + 'natural' sequence. The most frequent sequence is 'social entity/process' + 'important' + 'social entity/process'. The social entities are varied in nature and include manufactured objects such as *refrigerators*, agricultural processes such as *changes in cropping intensity*, social relations such as *community bonds* and abstract concepts such as *power*. This reflects a diversity of approach in the journal. Table 6.14 includes the only examples of a sequence that identifies 'intellectual process'; this could be subsumed under 'social entity', but it does refer to people's skills and memories rather than their actions or relationships, so it is reasonable to treat it separately. A final point about Table 6.14 is that one of the instances recorded as 'research entity' + 'important' + 'research entity' has a specifically interdisciplinary focus. Example 34 gives the full context.

34 The continuing integration and dialogue between natural and social science along these lines of inquiry will further inform policy in coming decades. A more explicit concern with equity and justice will be **important** in furthering the study of global environmental change.

[GEC 2005 Adger]

This sequence might be described as 'social science concern' + 'important' + 'field of research'. It asserts the relevance of social science and in doing so implies contradiction of an assumption that natural science is the priority concern. The presence of this example, reflecting the character of many articles in GEC that we have noted elsewhere in this chapter, is indicative. Also indicative, however, is the apparent comparative rarity of such examples.

Table 6.15 shows 20 instances from the monodisciplinary journal PS. What is striking here is the predominance of the 'natural' + 'natural' sequence (17 of the 20 instances), reflecting the journal's concern with the natural world as a self-standing entity. The remaining instances justify research processes in terms of their applications to agriculture or to other research.

Finally, Table 6.16 exemplifies the monodisciplinary journal REE. As expected, the sequence elements are social rather than natural, reflecting the journal's concern with the social rather than the natural world. There are no instances of the 'natural' + 'natural' sequence, and only one of the 'natural' + 'social' sequence. What is noticeable also is the relatively high frequency (10 out of 20) of instances

Table 6.14 20 instances of *important for/in* from GEC showing semantic sequences.

Research entity/process	*Important*	*Research entity/process*
A more explicit concern with equity and justice	will be important in	furthering the study of global environmental change
Choice of year reflects the concern of when the climate change impacts	become important for	identifying historical responsibility
Uncertainty in GDP growth	is relatively important for	the uncertainty in energy intensity
Natural entity/process	**Important**	**Social entity/process**
Many fishes	important for	subsistence and commerce
Wildlife species	important in	subsistence-based hunting
Social entity/process	**Important**	**Social entity/process**
Refrigerators which	are important for	improving living standards
Technological developments	are important for	the performance of agriculture
Population density and change in population	important for	their impacts on the consumption and production processes
A transition towards a low-carbon energy system	is important for	simultaneously covering energy demand and protecting climate systems
Formal credit and insurance	were considered most important for	enhancing household's capacity to adapt to market and climatic changes
Community bonds and high levels of trust among households	are important for	decreasing vulnerability to climate change impacts
Changes in cropping intensity	are becoming more and more important in	safeguarding national food security
Which system components	proved important for	successful responses
Personal contacts	are important for	protection of urban green space
Most of these services	were important in	their lifestyles
The Mangrove Action Project	is important in	raising the profile of these issues
Power	was very important in	forging agreement
Intellectual process	**Important**	**Social entity/process**
The ability to anticipate and respond to dangers, opportunities and changes	is important for	safe travel
This tacit aspect of social-ecological memory	may well be as important for	management of ecosystem services
Social entity/process	**Important**	**Natural entity/process**
The non-physical consequences of travel	might be as important for	global environmental change

Table 6.15 20 instances of *important for/in* from PS showing semantic sequences.

Research process	*Important*	*Application*
The creation of new banana varieties	is important for	its [banana] improvement
Identification of P-efficient genotypes	are important for	parent selection in a traditional breeding program
Research process	**Important**	**Research process**
Discussing gene expression in terms of the degree of development	is important for	understanding the mechanisms involved . . .
Natural entity/process	**Important**	**Natural entity/process**
GE interaction effects	were very important for	BRR
The Ce terminus of calmodulin	is especially important in	the interactions with other proteins
Many motifs	potentially important for	GT
The function of the xanthophyll cycle	is important for	the photo-protection of Arabidopsis leaves
Which cis-elements	are important for	the tissue-specific expression from these regions
Additional domains in the large ARF-GEFs	are likely to be important for	their physiological function
The acidic subunits	are important in	formation of the multimeric structures
Two thylakoidal enzymes	are important in	chlororespiration
These pathways	are important for	the synthesis of nucleic acids
Its (PME) action	was considered to be important in	pectin polymer digestion
Position 462	important for	heat stability
Hsp90	is important for	the folding of protein kinases
The former [process]	is important for	determining the total K.+ and Na.+ in the plant
Such lipid modifications	are known to be important for	the membrane anchorage of Rabs
This activity [dephosphorylation]	can be important in	the response to stress
Novel genes	important for	halotolerance mechanisms
The regions	important for	the activation of transcription by auxin

Table 6.16 20 instances of *important for/in* from REE showing semantic sequences.

Social entity/process	*Important*	*Social entity/process*
Commercial highways	became important in	industrial siting
Information that lessened site severity	was also important in	valuation
The French delegation system	is in particular more and more important for	developing countries
Efforts in reducing emissions	have not been equally important for	all economic sectors
These effects	are potentially important for	the overall profitability of coordinating gas and permit exports
EC has no impact on GDP temporarily . . . and this	is important for	policymakers
Other issues	are equally important for	the size of emission abatement costs
The decoupling of coal use and carbon emissions	is particularly important for	regions with a large endowment of coal reserves
This increase [in urbanisation]	is important for	health damage from air pollution
Natural entity/process	**Important**	**Social entity/process**
The wild fish that	are important in	developing countries
Social process	**Important**	**Research entity/process**
The differences in the firm's investment decisions	are important for	the evaluation of the quota system's ability . . .
Such variation [between individuals]	can be extremely important in	determining population and community dynamics
Proven reserves	are consistently important in	explaining participation in the cartel
The absolute volumes of permit exports and gas exports	are important for	the sign [+/-] of this expression [equation]
Firms' strategic behaviour	is not included in what Porter considers important for	his argument
Research entity/process	**Important**	**Research entity/process**
Both distributional and allocational concerns	are thus important for	motivating the analysis undertaken here
The form of the water demand functions	is important for	deriving such welfare effects
Nonlinear weighting	is empirically important in	explaining choice behaviour
Statistical measure are available and this	can be important in	valuation of risk changes
The assumption regarding the sign of [equation]	turns out to be important in	signing the direction of welfare impacts

of sequences involving research processes. This suggests that a key focus of the journal is research methodology and model building: how research should be done as much as what that research finds out.

6.4.4 Conclusion: what counts as important

This study of aspects of the use of the word *important* has demonstrated the usefulness of the word in reflecting the difference between journals and in particular in highlighting the character of interdisciplinary journals. The word, and others with similar meanings, is more frequent in interdisciplinary journals. The study of the phrase *it is important to* does not reveal much beyond the obvious. It shows a slight difference between those journals that are about the natural world, whether inter- or mono-disciplinary, and those that include concerns for the social world. Examples from the latter (GEC and REE) are more likely to attach importance to elements in an argument than are those from the former (AEE and PS).

The study of *important for/in* is more instructive in terms of interdisciplinary writing. By classifying a sample of concordance lines into sets of 'semantic sequences', the wider diversity of the two interdisciplinary journals than the two monodisiciplinary ones is demonstrated.

There is clearly more that could be said about this word and its synonyms, but enough has been said here to show the kind of traces of research context that can be found in a corpus by studying an evaluative word and its phraseologies. The study has highlighted the unique character of each of the BEE4 journals. The particular character of the interdisciplinary journals has been shown to be, not the use of any individual kind of language, but the diversity of language use.

6.5 Conclusion

This chapter has demonstrated the benefits of examining in detail the behaviour of individual words. The focus of interest, and the methodology used, has been slightly different in each of the studies undertaken.

The first word studied, *environment*, is a key term (obviously) in the field of environmental studies. The focus here has been on the collocates of the word and what these show about its meaning. We have seen how its meaning is subtly different in the different journals, reflecting the various research agendas apparent in those journals. The two interdisciplinary journals have been shown to use the word with more variety of meaning, indicating the multiplicity of agendas appearing therein.

The group of words around *science* has been studied in relation to relative frequency and phraseology. Perhaps surprisingly, the words are more frequent in the interdisciplinary than the monodisciplinary journals, even though the concept of 'science' is important in all the journals. This has been interpreted in terms of a distance construed (or not) between a body of researchers and the journal readership. The study of *scientists* in AEE and the investigation of *scientific* in GEC have highlighted the alternative approaches to interdisciplinarity in the two journals,

with writers in AEE tending to adopt an integrative position whereas those in GEC are more likely to express antagonism towards the scientific paradigm.

The word *important* represents a site of evaluation. In this chapter, the lexical-grammatical patterns known to be used with the word have formed the basis of investigation, and this has been extended to include the concept of 'semantic sequence'. This methodology has been used to isolate what is considered important and why in each journal. The number and frequency of the semantic sequences per journal has again revealed the diversity of the interdisciplinary journals in comparison with the monodisciplinary ones.

This chapter has shown the value of investigating individual words in detail, and of using what is known about language patterning as a route into that detail. Chapter 5 made similar points about the additional words *we* and *inadequate*. Unfortunately but inevitably, time and space limit the number of investigations of this kind that can be demonstrated. Chapter 7 will continue this type of study but with words selected according to a particular theory of evaluative meaning.

References

Bauer, H. H. (1990). Barriers against interdisciplinarity: Implications for studies of Science, Technology, and Society (STS). *Science, Technology, & Human Values*, 15(1): 105–119.

Francis, G., Hunston, S., & Manning, E. (1998). *Collins cobuild grammar patterns 2: Nouns and adjectives*. London: HarperCollins.

Halliday, M. A. K. (1994). *An introduction to functional grammar* (2nd ed.). London: Routledge.

Hunston, S. (1993). Evaluation and ideology in scientific writing. In M. Ghadessy (Ed.), *Register analysis: Theory and practice* (pp. 57–74). London: Pinter.

Hunston, S. (2008). Starting with the small words: Patterns, lexis and semantic sequences. *International Journal of Corpus Linguistics*, 13: 271–295.

Hunston, S. (2018). *Towards one thousand constructions: Rethinking the learner's understanding of lexis and grammar*. Paper given at the TaLC conference, University of Cambridge, July.

Hunston, S., & Francis, G. (2000). *Pattern grammar: A corpus-driven approach to the lexical grammar of English*. Amsterdam: John Benjamins.

Manor-Binyamini, I. (2011). A model of ethnographic discourse analysis for an interdisciplinary team. *Journal of Pragmatics*, 43: 1997–2011.

Muguiro, N. (2019). *Interdisciplinarity and academic writing: A corpus-based study of three interdisciplines*. Unpublished PhD thesis, University of Birmingham, Birmingham, UK.

Said, E. (1978). *Orientalism*. London: Penguin.

7 Status and disciplinarity

7.1 Introduction

The concept of 'epistemic status' has been extensively described before (Hunston 1993, 2000, 2011, 2013). In this introduction we give a brief outline sketch to set the context for this study.

When research is described, writers calibrate the degree of confidence ascribed to each proposition. To demonstrate this, example 1 is an extract from a popular book about evolution (*A Rough Guide to Evolution* by Mark Pallen). In this extract, the writer discusses the possibility that life on earth originated elsewhere in the universe.

1 A key question is where did life originate? Most investigators make the parsimonious assumption that life originated here on Earth (geogenesis). However, the idea that life originated elsewhere and was then seeded here (panspermia or exogenesis) has a long and distinguished pedigree. . . . The discovery in the mid-1990s of what appeared to be fossilized bacteria in ALH84001, a meteorite of Martian origin, stimulated interest in a limited form of panspermia, whereby life might have originated elsewhere in our solar system. This idea has gained support from the realization that ice occurs on the Moon and Mercury.

(Pallen 2009: 150–151)

This extract contains a number of propositions, such as: *life originated here on Earth*; *life originated elsewhere*; *the idea of panspermia has a long history*; *ALH84001 contained fossilized bacteria*; *fossilized bacteria were discovered in ALH84001in the mid-1990s*; *ice occurs on the Moon and Mercury*; *the idea of limited panspermia has gained support from information about the Moon and Mercury*. Note that a single clause may contain several propositions. For example, the final clause in the extract (*This idea has gained support from the realization that ice occurs on the Moon and Mercury*) contains two propositions – 'ice occurs on the Moon and Mercury', 'the idea of panspermia has gained support from the realization that ice occurs on the Moon and Mercury' – and refers to a third from the preceding sentence – 'life originated elsewhere in our solar system'. The clause

combines the three propositions so that a 'realization' provides support to an 'idea' within a narrative of the development of knowledge.

From the way the extract is written, we as readers understand each statement in a particular way. We know that there are competing ideas ('life originated on Earth'; 'life originated elsewhere') which may or may not be true. We know that we are expected to believe that two things have been discovered: 'something like fossilized bacteria on a Martian meteorite' and 'ice on the Moon and Mercury', though we also know that the 'ice on the Moon' is a certainty whereas 'bacteria in a Martian meteorite' is only a possibility. We are likely to accept the historical account of belief in panspermia: 'it has long been believed by some', 'in the mid-1990s there was renewed interest in a limited form', and 'recent discoveries about the Moon and Mercury support the idea of panspermia'. However, we are unlikely, from reading the extract, to understand that either geogenesis or panspermia are proven facts and we might agree with 'most investigators' that it is safest to assume that life originated on Earth, until there is rather more evidence to the contrary.

The key vocabulary that tells us how to interpret the various propositions in example 1 (*question*; *assumption*; *idea*; *discovery*; *appeared*; *might*; *realization*) is the topic of this chapter. The BEE4 journals are compared in terms of that vocabulary. The next sections (7.2 and 7.3) theorise the notion that readers understand statements in particular ways because of the language used by the writer, using the concept of epistemic status. One way that status is conveyed is through the use of status markers (*assumption*, *idea*, *discovery* etc.), and the relative frequency of these markers across the four journals is the subject of section 7.4. A quantitative account of status marker frequency is given in section 7.5, and a qualitative account of selected markers of status is given in section 7.6.

The chapter makes two main points. Firstly, the frequency, phraseology and use of various status markers is a key distinguisher between natural science and social science disciplines (see also Teich & Holtz 2009). The comparison between the four journals bears this out. Secondly, the co-text around status markers can be another site of dispute between disciplines. Where present, this dispute suggests that the relationship between the disciplines is antagonistic rather than integrative (Barry & Born 2013), as noted in other chapters. As a consequence, an investigation of status markers can identify those journals where the nature of the interdisciplinarity is antagonistic.

7.2 The concept of status

In the literature on epistemic status (Hunston 1993, 2000, 2011) it is argued that all statements or propositions in academic journal articles, and indeed in texts more generally, are accorded an epistemic status. The analyst might label the status of a given proposition might be labelled using nouns: 'discovery', 'hypothesis', 'conclusion' etc. It might also be labelled using paraphrases: 'what is claimed to be', 'what might be', 'what is unlikely to be' etc. The concept of status has been defined as 'the averred degree and type of alignment between a . . . proposition and the world' (Hunston 2011: 92). The term 'alignment' means the relationship between what is said (the proposition) and how the world is. Something that is marked as

a *discovery*, for example, is directly aligned with the world: as readers were are positioned to accept the proposition as fact. Modifying a statement with *might be* or *appears to be*, on the other hand, indicates an alignment with a possible world. It is important to recognise that these alignments are not inevitable consequences of how the world is but are the consequences of language choices made by the writer. They are 'averred' (or claimed) alignments, not necessarily actual ones. Status tells us about how language constructs a view of the world, not about how the world is.

Status may be marked or unmarked. This makes it somewhat different from related concepts such as stance. Every proposition in an article is evaluated for status, whether this is marked or not. If it is unmarked it is inferred from the context and from the form of the proposition. In example 1, for instance, *This idea has gained support* . . . is unmarked and written as fact. Where status is marked it uses lexis including verbs (*assume*, *believe*, *conclude*, *find* etc.), nouns (*assumption*, *belief*, *conclusion*, *result* etc.) and adjectives (*possible*, *true* etc.). These are language resources that are also associated with approaches to interpersonal meaning. For example, the verbs associated with marking status are also discussed under headings such as evidentiality (Chafe & Nichols 1987), stance (Biber 2006; Charles 2006), modality (Halliday 1994) and engagement (Martin & White 2005). The nouns associated with marking status are often discussed in the context of signalling nouns (Flowerdew & Forest 2017), stance (Charles 2007), or cohesion (Francis 1986). Again, the notion of status is somewhat different from these other concepts because it is considered that status categories, such as hypotheses, research actions, observations and deductions, form the building blocks of scientific knowledge, so that their significance goes beyond interpersonal meaning and text organisation and relate to Field as well as Tenor (Halliday 1994). As this chapter considers status markers only, however, and not status in general, the similarities between the concepts are more apparent than the differences.

Some examples from the BEE4 corpus will now be given to explain further the concept of status (examples 2–5).

2 Analysis shows that the purified rHLF contains xylose but lacks sialic acid.
[PS 2002 Nandi]

3 The rapid advancement in seismic and drilling technologies over the history of the petroleum industry is consistent with this hypothesis.
[REE 2001 Thompson]

4 Pethick (ref) has claimed that a stalemate might develop between the requirements of freshwater and intertidal habitats conservation.
[GEC 2007 Turner]

5 Although we interpret this to mean that all of the Enod8 genes known so far are close together on the M. truncatula genome, it also suggests that the Enod8 and Enod8-like genes may not all be contiguous.
[PS 2002 Dickstein]

Example 2 contains two propositions. One, 'the purified rHLF contains xylose but lacks sialic acid', has a status marked by *the analysis shows that*. It is an interpretation of the result of an experimental process, analysis, and is marked as closely aligned with the world by the word *show*. The other proposition is the whole sentence: 'Analysis shows that the purified rHLF contains xylose', whose status is unmarked but depends on the context as a research article. The default interpretation is that this is fact rather than fiction and may be taken as directly aligned with the world, or fact.

In example 3 there is a proposition that must have been articulated prior to this sentence and which is given the status of hypothesis, with the marker *hypothesis*. That proposition is only possibly aligned with the world. The rhetorical purpose of this sentence, however, is to support the hypothesis, making its alignment with the world more certain. A second proposition, 'seismic and drilling technologies have advanced rapidly over the history of the petroleum industry' is nominalised as *the rapid advancement in seismic and drilling technologies over the history of the petroleum industry*. This is unmarked and treated as an assumed fact that is directly aligned with the world. The phrase *is consistent with* indicates the relationship between the assumed fact and the hypothesis, with the former supporting the latter.

In example 4, the proposition 'a stalemate might develop between the requirements of freshwater and intertidal habitats conservation' is attributed to Pethick. The use of the verb *claim* implies that Pethick has presented the proposition as likely to be true but that the current writer does not share that confidence. There is at least a potential mismatch between Pethick's view and the writer's. The proposition is further modified by the modal auxiliary *might*, with future reference. The status of the proposition is therefore an attributed claim about the future. It is aligned with a possible future world.

In example 5, there is again a reference to previous propositions, using the pronouns *this* and *it*. These previous propositions are not marked for status in this sentence, but the default interpretation is that they are the results of experiment. The proposition 'all of the Enod8 genes known so far are close together' is evaluated as possibly true, and marked explicitly as the outcome of interpretation by the authors: *we interpret . . . to mean . . .* The proposition 'the Enod8 and Enod8-like genes may not all be contiguous' is similarly evaluated as possibly true, by the choice of the verb *suggest*, but the interpretative action of the researcher is effectively hidden and the proposition is presented as dependent on the evidence alone. The certainty of the interpretation is, however, reduced by the modal auxiliary *may*.

These explanations of the status markers in the four examples are summarised in Tables 7.1a–d.

These examples show the process of building knowledge from a hypothesis or claim (examples 3 and 4) to the interpretation of experimentation (examples 2 and 5). They also show the interaction between propositions with different status evaluations. In example 2, actions marked as 'analysis' give credence to a proposition derived from that research action. In example 3, a 'hypothesis' is supported by an 'assumed fact'. The examples also show how the attribution of status is

Table 7.1a Status in example 2.

Analysis shows that the purified rHLF contains xylose but lacks sialic acid.		
Proposition	*Marker*	*Status*
the purified rHLF contains xylose but lacks sialic acid	analysis shows that	interpretation of experiment result

Table 7.1b Status in example 3.

The rapid advancement in seismic and drilling technologies over the history of the petroleum industry is consistent with this hypothesis.		
Proposition	*Marker*	*Status*
[unknown]	hypothesis	hypothesis
seismic and drilling technologies have advanced rapidly	none	assumed fact

Table 7.1c Status in example 4.

Pethick (2002) has claimed that a stalemate might develop between the requirements of freshwater and intertidal habitats conservation.		
Proposition	*Marker*	*Status*
a stalemate might develop between the requirements of freshwater and intertidal habitats conservation	Pethick has claimed might	attributed claim about future world

Table 7.1d Status in example 5.

Although we interpret this to mean that all of the Enod8 genes known so far are close together on the M. truncatula genome, it also suggests that the Enod8 and Enod8-like genes may not all be contiguous.		
Proposition	*Marker*	*Status*
[unknown: *this*, *it*]	none	experiment result
all of the Enod8 genes known so far are close together on the M. truncatula genome	we interpret . . . to mean	tentative interpretation of result
the Enod8 and Enod8-like genes may not all be contiguous	suggests that may	tentative interpretation of result

modified by word choice, such as *show* versus *suggest* (examples 2 and 5), or by the level of explicitness of researcher intervention: *we interpret* versus *[the result] suggests that* (example 5).

A true analysis of status in text is qualitative and, like Martin and White's (2005) study of Engagement, depends on contextual as well as formal information. This makes it difficult to carry out anything other than small-scale studies

of individual articles, using the sentence-level methodology demonstrated here. However, given the importance of status markers such as the verbs *show*, *suggest*, *claim* and *interpret* and nouns such as *hypothesis* to the expression of status, quantitative studies of the frequencies of these markers and qualitative studies of their usage patterns can act as proxies for more detailed studies of status when comparing corpora. Studies of this kind will form the basis of the rest of this chapter.

7.3 Why status matters

The kinds of status that are allocated to propositions are indicative of the ideology surrounding the research process. Evaluation of status is, therefore, connected with disciplines and their assumptions and, as will be shown in the sections that follow, with relations between disciplines.

Natural sciences and social sciences are said to differ in their construal of the research process, with natural sciences emphasising 'finding out' and social sciences prioritising 'arguing' (Becher 1989). Propositions in natural sciences disciplines are expected therefore to be relatively more frequently marked as certain or true, using verbs such as *SHOW* or *INDICATE* and nouns such as *observation* or *result*. Those in social sciences disciplines are expected to be relatively more frequently marked as open to debate, such as *ARGUE* or *argument* and *CLAIM* (noun and verb). This disciplinary difference is confirmed by, for example, studies of verbs by Charles (2006) and studies of signalling nouns by Flowerdew and Forest (2017). It is important, of course, that what is at stake here is relative frequency: all disciplines can be expected to have instances of each of *show*, *indicate*, *argue* and *claim*.

As noted, a proposition may be attributed to a person (*Pethick has claimed that . . .* ; *we interpret . . . to mean that . . .*) or to a non-human phenomenon (*the results suggest that . . .* ; *the presence of . . . shows that*). In the latter case, knowledge is construed as arising naturally from the research process or from observation of the natural world. As noted by sociologists of science from the 1980s onwards (e.g. Gilbert & Mulkay 1984), this form of discourse masks or at least mediates the human involvement inevitable in the interpretation of research results. There is some evidence that natural science disciplines are more likely than social science disciplines to mediate human involvement in the research process in this way (Charles 2006). A previous study of popular writing about evolution (Hunston 2013), based on the book from which example 1 is taken, has shown a pattern emerging in which the writer attributes ideas to people (*Darwin proposed that . . .*) but tends to attribute propositions which are accepted by the scientific community to non-human entities (*The results show that . . .*). In short, paying attention to the sources of propositions, as indicated by the subjects of status marker verbs, is ideologically revealing. It has been noted also that in social science disciplines, human attributees are not restricted to researchers (Charles 2006); in environmental science, groups such as politicians and policy advisors, farmers and members

of the general public all have statements attributed to them. An additional source of interest, as shall be shown later, is that researchers can be identified as having particular epistemological loyalties, indexing both inter- and intra-disciplinary divisions.

Finally, markers of status can be used to layer evaluations and to indicate degrees of difference between various groups participating in debate. Example 6 illustrates this.

6 Just as many in the more 'numerical' sciences insist that risk is an objective function, and that it is the perception of risk . . . that is the problem, so have many social scientists ignored the interplay between humans and natural and technological systems. Thus some of these scholars see risk as merely a reflection of the distribution of power in society.

[GEC 1992 Dovers]

The status markers *insist*, *ignore* and *see . . . as* construe a difference between the writer and other researchers and a critique of those researchers by the writer that is implied rather than stated outright. Table 7.2 summarises the differences in attributed belief. It will be noted also that the researchers to whom belief is attributed are distinguished by disciplinary labels: 'many in the more "numerical' sciences"' and 'many social scientists', and also that they constitute a somewhat vague group, with no specific authors cited.

Thus, status markers have a particular role to play in construing complex disagreement in academic discourse. Given what is known about disciplinary differences and about the role of status markers in construing disagreement, it is expected that the relative frequency of various status markers will vary between the journals in the BEE4 corpus.

Table 7.2 Attributed beliefs in example 6.

Current writer believes	*Attributed authors believe*	*Status marker*
Risk is not an objective function.	Risk is an objective function. (Attributed to: many in the numerical sciences)	insist
The problem of risk is not only a problem of perception.	The perception of risk is the problem (Attributed to: many in the numerical sciences)	insist
The interplay between humans and natural systems is important.	Any interplay between humans and natural systems is unimportant. (Attributed to: many social scientists)	ignore
Risk is not only a reflection of the distribution of power in society.	Risk is a reflection of the distribution of power in society. (Attributed to: many social scientists)	see . . . as

7.4 Identifying and counting status markers in the BEE4 corpus

The enterprise of investigating status markers quantitatively cannot be undertaken directly because no finite list of status markers exists. The issue is complicated by the fact that some potential markers are sometimes markers of status and sometimes not. For example, the verb *perceive* marks status when the agent is human (example 7), but not when it is non-human (example 8).

7 [T]he number of varieties . . . is certainly higher than the diversity actually **perceived** by the agriculturists.

[AEE 2002 Peroni]

8 . . . a specific interaction between the red signal, **perceived** by the phyB1 photoreceptor at least, and the DGT-dependent auxin signalling pathway . . .

[PS 2001 Kraepiel]

The only way accurately to identify status markers is to undertake a discourse or text-based study. In that case, only a relatively small corpus could be used. As the aim here is to examine status markers in a relatively large corpus, some margin of error has to be accepted, and conclusions drawn appropriately.

As a shortcut to obtaining quantitative information in this study, use is made of the fact that most status marking adjectives, nouns and verbs sometimes occur with appositive that-clause complementation (e.g. *certain that*,*conclusion that*, *suggested that*). A sufficiently complete list does exist of the adjectives, nouns and verbs occurring in English with these clauses (Francis, Hunston & Manning 1996, 1998), and this can be exploited to form a list of potential, though not actual, status markers. Using this resource, a list of 621 target words (lemmas restricted by part of speech) was compiled and their raw frequency and normalised frequency (per million words) in each of the four journals obtained. Of these, over 530 occur at least once in at least one journal, but only about 250 occur in all the four journals. A total of 121 occur over 100 times per million taking the four journals as a single corpus.

As noted, this method of identifying status markers is somewhat rough-hewn. Firstly, although most status markers are sometimes complemented by that-clauses, the converse is not true: not all words that are sometimes complemented by that-clauses indicate status. Consequently, before working further on the list of words, non-status words are manually removed from it. Secondly, as noted, when a word does sometimes mark status, it does not always do so. For example, the noun *conclusion* sometimes marks the status of a proposition but sometimes is a section header. The noun *position* sometimes marks attributed status (‘The person’s position on this is that . . . ’) but sometimes indicates physical location. The adjective *right* can indicate status (‘They are right to state that . . . ’) but often indicates the right side of a graph or a body. Without checking each concordance line, the proportion of instances that fall in or out of the category of status marker cannot be calculated. In

the discussion to follow, where it is known (from the observation of concordance lines) that a given word in a given journal is never or rarely a status marker, that word is omitted from the relevant lists (as with *right* in PS and REE). In this chapter conclusions are not drawn from the number of occurrences of individual status markers, therefore. Rather, conclusions are drawn from the relative frequency of groups of status markers, so that the importance of error in individual calculations is minimised, given that error in the form of over-reporting will certainly have occurred.

It may be asked why an alternative method, based on the occurrence of the appositive that-clause construction, is not used. It is almost certainly true that any 'word' + 'that-clause' construction in the corpus would mark status. This would certainly avoid over-reporting. On the other hand, this method would under-report the frequency of status markers. To illustrate this, consider the noun *justification/s*. This noun expands the category of status, as it indicates not only a degree of acceptance of the truth of a proposition, which is its status proper, but also the role of that proposition as a rationale for an argument or course of action. The lemma occurs in all four journals. It occurs most frequently in GEC (20.15 per million words), then in REE (8.36 pmw), AEE (3.46 pmw) and finally PS (0.55 pmw). This is to be expected, as the word is used most frequently to report how ordinary citizens or policy makers, rather than researchers, respond to environmental issues, and this is a concern for GEC and REE rather more than for AEE or PS. The lemma occurs 147 times in the four-journal corpus. It is followed by *for* 76 times, by *of* 12 times, by a to-infinitive six times and by a clause beginning with *that* only four times, of which only one is an appositive that-clause, the rest being relative clauses. Thus, the pattern that has been used to identify status markers is used with this particular lemma only once in all the four journals (example 9; evaluated proposition in italics).

9 Generally, reports that invite such unjustifiable inferences do not explicitly endorse these themselves: they simply relegate the caveats to the fine print, usually on the dubious **justification that** *policymakers and the general public are not sufficiently sophisticated with risk or uncertainty to appropriately contextualize the caveats*.

[GEC 2010 Boykoff]

On the other hand, the lemma is used in other ways that indicate the status of propositions in the article. In example 10 it is a prospective signalling noun.

10 [T]he ethical imperative to put the most vulnerable subject first has various **justifications**. [Subsequent sentences enumerate these justifications.]

[GEC 2010 Grasso]

In example 11 the evaluated proposition is not contiguous with *justification* and is given status by both *justification* and *belief*. In example 12, *with some justification* modifies the status of the proposition primarily given status by the verb *worried*. (Relevant propositions are in italics.)

11 A related **justification** for adaptation is a belief by some of its advocates that *our grandchildren's society will be more wealthy than today's . . .*
[GEC 1991 Kempton]

12 Social scientists worried, with some **justification**, *that human dimensions research would forever remain the handmaiden of research agendas set by the natural science world*.
[GEC 1997 Sanderson]

In many other cases, such as example 13, the propositions evaluated as justifications lie outside the article.

13 The diverse goals, **justifications** and institutional context present at the inception of assessment processes 'frame' assessment conduct and outcomes.
[GEC 2001 Farrell]

In short, this method of identifying status markers allows the recognition of lemmas which occur only rarely in the canonical 'word' + 'that-clause' pattern. On the other hand, the frequency count includes instances where no proposition in the article is evaluated (in Flowerdew's terms, the word is not a signalling noun) and status is given intertextually.

With all its shortcomings, then, the method of obtaining the frequencies of candidate status markers, based on their appearance in the lists produced by Francis et al. (1996, 1998), seems to be the best compromise for obtaining information from the BEE4 corpus. Two kinds of study are reported in the following sections: a quantitative one and a qualitative one based on selected markers.

7.5 Quantitative study of status marker frequency

As noted previously, from what is known about differences between disciplines, and in particular about differences relating to citation forms and stance forms, it might be predicted that disciplines will differ in terms of the frequency of various status markers, with natural science disciplines preferring those expressions that construe knowledge as arising uncontroversially from experimentation, as in *analysis shows that . . .*, and social science preferring expressions that foreground human intervention and dispute in the construction of knowledge, as in *Pethick has claimed that . . .* Investigating the comparative frequency of status markers in the four journals is therefore likely to lead to insights into their disciplinary character.

From the list of 250 lemmas that occur in all four journals, the most frequent 135 in the BEE4 corpus are identified, excluding those lemmas that are not potential status markers, but retaining those lemmas that are sometimes markers and sometimes not. The normalised frequency of each lemma in each of the four journals is recorded, and a total for the BEE4 corpus obtained. Table 7.3 shows the 30 most frequent markers in each journal, based on normalised frequency.

Table 7.3 The 30 most frequent status markers in four journals (nouns shown in italics).

AEE	*GEC*	*PS*	*REE*
show	*scenario*	show	assume
analysis	*risk*	*analysis*	show
find	*problem*	observe	consider
determine	*analysis*	find	*analysis*
calculate	*approach*	suggest	*information*
indicate	see	indicate	see
consider	*assessment*	report	*problem*
observe	*information*	determine	find
indicator	show	know	expect
estimate	consider	demonstrate	*risk*
suggest	*decision*	reveal	estimate
report	suggest	appear	imply
scenario	*knowledge*	confirm	*decision*
information	find	*perception*	*assumption*
approach	*concern*	maintain	determine
see	expect	see	*approach*
explain	*indicator*	consider	*probability*
argue	assume	calculate	*scenario*
assume	indicate	*observation*	suggest
risk	understand	expect	*equation*
expect	argue	*evidence*	indicate
predict	*concept*	*fact*	*fact*
establish	*consequence*	estimate	*knowledge*
assessment	estimate	predict	know
maintain	determine	explain	note
observation	*assumption*	*information*	mean
problem	*fact*	establish	*rule*
appear	*understanding*	*approach*	calculate
mean	mean	propose	*hypothesis*
conclusion	know	*finding*	observe

The lists in Table 7.3 show that in broad terms the same status markers occur in all the journals. Many relate to the process of carrying out research or interpreting results. Words such as *show*, *analysis*, *find* and *estimate* are found in all the lists. There are also differences. The words *risk* and *problem* are found in the AEE, GEC and REE lists but not the PS one. Some words implying human thought processes, such as *concern* and *understand* are found in the GEC list only, but others, such as *knowledge* and *decide* are also found in the REE list. Words connected with economic modelling, such as *assumption*, *probability* and *equation*, are restricted to the REE list.

It is noticeable that the ratio of nouns (shown in italics) to verbs in Table 7.3 is different for the four journals. The AEE list has ten nouns from the 30 total words, the GEC list has 16, PS has eight, and REE has 14. If the presence of noun status markers is an indicator of complexity of argument, this would suggest that GEC has the highest proportion of complex arguments and PS the lowest. Another contributing factor is also possible. For example, the noun *assessment* appears in both

the AEE and the GEC lists in Table 7.3. It is sometimes used as a straightforward nominalisation of the verb *assess*, as in example 14.

14 *P. ovale* and *P. malariae* are not included in this **assessment** because of the small number of cases and limited distribution.
[GEC 1999 Martens]

A congruent version of example 14 might be:

> 'We did not assess *P. ovale* and *P. malariae* because there are only a small number of cases in only a few areas'.

Frequently, though, use of the noun permits an evaluation of the method of carrying out assessment. In AEE these instances support an evaluation of the researchers' method of investigation, as in example 15. In GEC there are instances which cast doubt on a whole field of research, as in example 16.

15 This does not allow an **assessment** for the whole landscape, with about 100ha of abandoned land being currently out of farmer's [sic] control. However, we can assume . . . that the abandonment . . . differs more in timing than in process.
[AEE 2006 Mottet]

16 They assume a classical approach to risk analysis, i.e. that science can adequately assess risk according to the knowledge base of the day, and political decisions on risk management can then be based comfortably upon that **assessment**. But there is a growing body of criticism of that approach.
[GEC 1995 Munson]

Table 7.3 gives no information as to frequency, simply listing 30 markers from each journal. This is problematic, particularly given the different sizes of the journal sub-corpora. To get a more appropriate comparison, for each of 135 markers, the percentage of instances occurring in each journal is calculated. A lemma which is evenly spread across the journals will have a percentage of approximately 25% in each journal. One word that comes close to this even spread is the verb *find*, whose proportions across the journals are: AEE 28.6%; GEC 15.9%; PS 28.6% and REE 26.8%. An example of a much more skewed distribution is the word *assessment*, where the proportions are: AEE 18.6%; GEC 70.8%; PS 2.1% and REE 8.3%. The vast majority of instances, proportionally, are in GEC, and as we have seen, the word could be said to contribute to an agonistic approach to other disciplines. This makes it a more distinctive feature of GEC than Table 7.3 would indicate.

Table 7.4a shows percentage uses that are 'very high' (at least 50% of the total) and 'high' (between 30% and 49% of the total). Words that are unique to one journal list are shown in bold. Table 7.4b lists the words that are 'high' or 'very high' in two journals (these words are those appearing not in bold in Table 7.4a).

Table 7.4a Percentage use of status markers in four journals (unique words in bold).

AEE	*GEC*	*PS*	*REE*
'VERY HIGH' USE i.e. ≥ 50%	**scenario** 61.6% **assessment** 70.8% **knowledge** 55.1% **concern** 63.2% **understand** 51.6% **concept** 67.8% **understanding** 63.2% **perception** 66.1% say 51.4% **principle** 56.4% **perceive** 59.8% projection 56.2% **recognition** 57.6% **believe** 50.2% learn 55.7% **idea** 60.1% ask 52.5% **threat** 63.6% **agree** 58.6% **accept** 53.0% view 52.5% **law** 69.5% uncertain 51.1% **emphasize** 54.7% **consensus** 58.4% **notion** 66.2% **proposal** 64.8% **awareness** 62.7% **feel** 67.6% **statement** 67.4% **convention** 85.4% **opinion** 77.9% **attitude** 75.3% **intend** 56.9% **threaten** 63.2% prospect 59.4% aware 56.2% **acknowledge** 73.3% **danger** 69.6% question 52.4% rationale 55.5% tell 56.8% **surprise** 88.4%	**observe** 51.2% **confirm** 59.2% stress 50.5% **deduce** 82.0% **appearance** 61.3%	**assume** 63.9% **assumption** 60.4% **probability** 65.4% **equation** 70.2% **imply** 69.8% rule 52.5% **specify** 56.8% expectation 58.4% **proposition** 86.1% **suppose** 70.6% **discovery** 74.3% certainty 57.8% **recall** 70.4% forecast 55.4% **proof** 70.9%

	AEE	*GEC*	*PS*	*REE*
'HIGH' USE i.e. 30% – 49%	estimate 30.1% **calculate** 48.4% **indicator** 48.8% **explain** 37.3% argue 38.4% **predict** 32.6% establish 32.2% calculation 37.8% prediction 36.3% resolution 37.9%* recommendation 34.5% indication 34.4% sense 47.5%	see 36.0% information 34.1% problem 47.3% risk 47.9% **approach** 45.4% decision 43.7% indicator 37.7% fact 31.1% argue 39.9% mean 35.0% conclusion 30.1% evidence 30.1% establish 34.1% note 35.7% **consequence** 46.9% agreement 41.5% **finding** 33.4% **propose** 33.5% implication 46.5% illustrate 37.0% rule 31.2% theory 45.7% consideration 37.4% **respond** 46.4% **think** 49.7% calculation 34.8% **mention** 35.5% prediction 31.2% interpretation 40.0% state 37.5% stress 34.7% expectation 31.3% resolution 48.0%* decide 36.5% likelihood 38.6% belief 48.4% recommendation 43.0% recommend 30.5% confidence 35.8% **recognize** 49.9% indication 34.4% chance 43.4% anticipate 46.9% certainty 37.1% forecast 35.1% sense 33.2%	**show** 45.6% **analysis** 36.0% **suggest** 36.1% **determine** 31.3% **indicate** 36.4% **report** 44.4% **know** 33.6% **appear** 31.2% **demonstrate** 43.5% **maintain** 30.6% **reveal** 47.3% **apparent** 31.6% **verify** 45.2% **hypothesize** 34.2%	**consider** 39.8% see 39.0% information 42.6% problem 40.5% risk 38.7% estimate 41.9% **expect** 49.3% decision 46.6% fact 34.7% mean 34.8% **observation** 30.2% conclusion 31.5% evidence 32.3% note 38.8% agreement 32.6% implication 40.3% note 38.8% agreement 32.6% implication 40.3% **hypothesis** 48.9% **possibility** 41.6% illustrate 41.5% theory 45.0% consideration 46.3% **conclude** 39.3% say 40.8% projection 34.4% learn 38.8% **prove** 33.7% interpretation 37.5% ask 37.7% state 44.6% **explanation** 40.8% decide 49.8% likelihood 46.2% view 33.5% uncertain 37.5% belief 46.3% confidence 31.8% chance 35.0% **correct** 41.7% anticipate 35.9% prospect 34.0% aware 30.1% **discover** 37.8% **notice** 35.8% question 30.3% rationale 32.0% tell 36.5%

Table 7.4b Shared high-frequency status markers.

AEE & GEC	*AEE & REE*	*GEC & REE*
indicator	estimate	see
argue		information
establish		problem
calculation		risk
prediction		decision
recommendation		fact
recommend		mean
indication		conclusion
sense		evidence
		note
		agreement
		implication
		illustrate
		rule
		theory
		consideration
		concept
		say
		projection
		learn
		interpretation
		ask
		state
		expectation
		decide
		likelihood
		view
		uncertain
		belief
		confidence
		chance
		anticipate
		prospect
		certainty
		aware
		question
		rationale
		tell
		forecast

Table 7.4c Number of distinctive status markers in four journals.

AEE	*GEC*	*PS*	*REE*	*GEC&REE*	*AEE&REE*	*AEE&GEC*
4	40	18	22	38	1	9

Table 7.4c quantifies the information in Tables 7.4a and 7.4b, giving the number of unique high-frequency status markers in each journal and the number shared across two journals.

Table 7.4c shows most clearly that GEC is, in these terms, the most distinctive journal of the four, with AEE the least distinctive. GEC and REE share the most distinctive markers, while PS shares no high-frequency markers with other journals. A scrutiny of the lists in Table 7.4a suggests the following:

- AEE is distinctive in terms of talking about judgements about the future: *calculation*, *estimate*, *predict(ion)*, *recommend(ation)*, but most status markers occur with low proportional frequency.
- GEC is distinctive in terms of:
    - Markers connected with research theories, e.g. *approach*, *idea*, *notion*, *principle*, *theory*
    - Markers connected with human thought and communication, e.g. *agree*, *attitude*, *awareness*, *believe*, *concern*, *decision*, *feel*, *learn*, *opinion*, *perceive*, *recognition*, *think*, *understanding*, *view*
    - Markers related to risk, e.g. *danger*, *problem*, *risk*, *threat*, *threaten*
    - Markers connected with the research process, e.g. *calculation*, *conclusion*, *evidence*, *finding*, *implication*, *proposal*, *scenario*
    - Markers that occur in conjunction with a reflexive disciplinary stance, e.g. *assessment*, *knowledge*
- PS is distinctive in having high-frequency status markers connected with the empirical research process, e.g. *analysis*, *confirm*, *determine*, *hypothesize*, *indicate*, *observe*, *reveal*, *show*, *verify*. These status markers are those that presuppose the veracity of the scientific method.
- REE is high in markers that relate to researcher involvement in the (mathematical) research process, e.g. *assume/assumption*, *belief*, *conclude/conclusion*, *consider*, *decide*, *discover*, *expect*, *explanation*, *observation*, *question*, *specify*, *suppose*, *view*.

This part of the study characterises each journal. GEC and REE, the two journals that incorporate social science research, both separately and together have by far the greatest number of highly frequent types. This confirms one of the observations from Table 7.4c: that GEC and REE are similar to each other in this respect. In addition, GEC and REE are shown to be distinctive in markers that relate to human actions, perceptions, thought and communication. PS is distinctive in its emphasis on an objective research process. AEE is perhaps distinctive in its lack of distinction: markers in AEE tend to occur with around average frequency.

There is a limit to the information that can be gleaned from a quantitative study of this kind. Greater insight into the use of status markers in the four journals can be gained from a qualitative study of selected examples. This will be shown in the next section.

7.6 Qualitative studies of selected status markers

The quantitative information given in the previous section suggests that the journal GEC is the most distinctive of the four journals, as might be predicted by the fact that it is both interdisciplinary and represents distal disciplines. It also suggests that GEC and PS are the most different from each other, as might be predicted from the fact that the first incorporates at least some social science whereas the latter is focused on natural science, and a particular topic within natural science at that. To investigate these findings in more detail, some of the status markers will now be examined in greater detail. These are:

- The verbs *ARGUE*, *CLAIM*, *INDICATE* and *SUGGEST*
- The nouns *argument/s* and *observation/s*

The verbs form a cline between 'most certain' (*INDICATE*) to 'least certain' (*CLAIM*), with *ARGUE* and *SUGGEST* lying in between. Evidence for this cline is largely intuitive, but there are a number of supporting arguments. Previous studies suggest that a proposition marked as a 'claim' is often subsequently contradicted (Hunston 2004). An additional argument can be based on evidence from philosophy (Willett 1988), which points out that visual evidence (e.g. 'I know John is in the library because I have seen him there') is more reliable than hearsay (e.g. 'I know John is in the library because Mary told me he is there'), or indeed deduction (e.g. 'I know John is in the library because he is not here'). The verb *INDICATE* (like *SHOW*) most congruently (in Halliday's 2004 terminology) construes a visual process, and is therefore the most certain, whereas *ARGUE*, *CLAIM* and *SUGGEST* are most congruently verbal processes. All the verbs can occur with human subjects, which, as argued here, indicates that the proposition is contingent upon human subjectivity, but *INDICATE* and *SUGGEST* (only) are predicted to occur with non-human subjects also, indicating less contingency and therefore greater certainty. Finally, in Martin and White's (2005) terms, *INDICATE* and *SUGGEST*, relatively speaking, contract the discoursal space, discouraging disagreement, while *ARGUE* and *CLAIM* expand it, inviting disagreement or debate. Given the possibilities of choice of verb and choice of subject, there is a rhetorical scale here with '*The results indicate that . . .* ' at one end and '*The researchers claim that . . .* ' at the other.

In addition, as noted, the verbs *SUGGEST*, *ARGUE* and *CLAIM*, with human agents, open up a potential divergence between the writer and the cited person. In the case of *SUGGEST* the cited person is represented as being non-assertive about the truth of the proposition, whereas in the case of *ARGUE* and *CLAIM* the cited person is represented as asserting that the proposition is true, leading to potential disagreement between them and the writer.

Similarly, the selected nouns indicate different rhetorical choices. *Argument* construes a verbal process that expands the discoursal space and indicates a subjective form of knowledge construction. *Observation* construes a visual process, even when used metaphorically, that contracts the discoursal space and interprets knowledge construction as an objective process.

The contrast between the status markers studied is designed to highlight the differences between the four journals studied and also to bring into focus issues of (inter-)disciplinarity in them.

7.6.1 Four verbs: ARGUE, CLAIM, INDICATE *and* SUGGEST

As noted, these verbs are chosen because they reflect differences in the knowledge-construction process. It is predicted that they will occur with different frequency in the four journals and that contextual or phraseological uses of the words will also indicate differences between the journals.

Table 7.5 shows the normalised (per million words) frequency of each verb lemma in the journals.

Table 7.5 shows that *INDICATE* and *SUGGEST* are more frequent than the other two verbs in all the journals; these verbs are particularly frequent in PS and are least frequent in GEC. *ARGUE* shows a differential between GEC and REE on the one hand, where it is relatively frequent, and AEE and PS on the other hand, where it is less frequent. *CLAIM* is relatively rare in all the journals, but is most frequent in GEC.

Each verb in turn will now be examined in context, to identify the range of uses each one has. These uses are identified qualitatively, limiting the amount of data that can be processed. To get an approximate idea of comparative quantities, a 100-line random sample from each journal is taken (using the Sketch Engine sampling function) from the thousands of total instances. This means that the percentage figures obtained are indicative only, but the broad trends can be identified.

7.6.1.1 ARGUE

The vast majority of instances of *ARGUE* in the four journals (approximately 80%) are followed by a that-clause, as in example 17. A smaller proportion of instances are followed by *for* and a nominalised proposition, as in example 18, or as an adjunct-clause as in example 19. Other minor phraseologies include *argue about*, *argue against*, *argue in favour of*.

17 Many people **argue** that we have a moral imperative.
[REE 2004 Armsworth]

Table 7.5 Four verb lemmas in four journals: normalised (pmw) frequency.

	AEE	*GEC*	*PS*	*REE*
ARGUE	31.79	323.66	11.96	163.44
CLAIM	9.68	71.05	2.85	19.95
INDICATE	586.06	373.92	796.51	430.47
SUGGEST	470.80	590.90	873.31	480.66

18 Many people . . . **argue** for an approach rooted in longitudinal studies.
[AEE 2007 Zander]

19 The analysis, we have **argued** above, shows how the rich array . . .
[GEC 2001 Lambin]

The subject of the verb – that is, the source of the proposition – is typically animate, either an individual or a group. It may be another researcher (example 20), the writer themselves (examples 21 and 22), or groups such as policy makers or farmers (example 23). Propositions attributed to researchers including the writer are part of academic knowledge construction. Those attributed to groups such as farmers are evidence for knowledge but not part of the academic debate itself. People can be distinguished by their disciplinary, epistemic or other affiliation, as in example 24. Attributing an argument to a human speaker construes knowledge as construction by people and their words. Identifying the speaker by their affiliation highlights partisan interests and disputes, including disciplinary ones.

20 Adelman (ref) **argued** that . . .
[REE 2001 Thompson]

21 We **argue** that . . .
[PS 2004 Balconi]

22 In this study it is **argued** that . . .
[AEE 2002 Gerbens-Leenes]

23 Most farmers **argued** that . . .
[AEE 2005 Parikesit]

24 Environmental groups critical of government secrecy **argued** that . . .
[GEC 1992 Wynne]

Another frequent phraseology with *ARGUE* is a passive structure with a modal verb of possibility and an anticipatory *it*: *it can/could/might be argued that*. This is used to propose a tentative statement or hypothesis, as in example 25. The same modal auxiliaries are also used with general human subjects (example 26).

25 It can be **argued** that the structure of agricultural landscapes . . .
[AEE 2003 Nyffeler]

26 [O]ne could **argue** that each sector . . .
[AEE 2004 Dendoncker]

A less frequent, and unexpected, use of *ARGUE* is with a non-human subject such as a model, observation of the natural world, or empirical results, as in example 27.

Attributing an argument to a non-human source makes the verb process metaphorical and construes knowledge construction as a product of research processes rather than human thought.

27 Our results **argue** for a functional link between . . .
[PS 2005 Courdavault]

Table 7.6 summarises the frequency of these uses.

In three of the journals, expectations are met in that researchers are the most frequent source of argument (AEE 69%; GEC 69%; REE 86%; total of rows 1, 2 and 3). The debate between other researchers and the writers in constructing knowledge is enacted in these lines, and the personal nature of the research process is particularly apparent in REE, where the writer as subject of the verb accounts for 28% of the total. Non-researchers as source are more common in GEC (18%) than in the other journals. GEC also has the highest proportion (9%) of sources identified by epistemological or geographic affiliation. This highlights the contingency of knowledge construction upon intellectual or social background, something which is an important component of the interdisciplinary character of GEC.

The journal PS is different from the others in having a low proportion of human subjects of the verb. Instead, and contrary to expectation, research processes, theories and observations make up 41% of the total. There is also a relatively large number of 'hypothetical' uses with *can* etc. (28%). These are often used to set up a 'straw man' argument that is then contradicted, as in example 28.

28 It could be **argued** that the thermal aggregation of chloroplast stromal proteins . . . is being affected by heat shock proteins. . . . However, in vivo labelling showed that [the proteins] did not differ in the pattern of HSP synthesis.
[PS 2004 Ristic]

Table 7.6 Uses of *ARGUE* in four journals.

		AEE	*GEC*	*PS*	*REE*
	Total instances	312	1526	109	254
	Instances classified	100	100	100	100
1	Researcher/s as source	49	42	20	53
2	Researcher/s by discipline/geography/ group as source	3	9	2	1
3	Writer as source	17	18	9	28
4	Non-researcher as source	8	18	0	4
5	Hypothetical with *can/could/might*	17	11	28	12
6	Non-human source	6	1	41	1
7	Other	0	1	0	0

Note: The 'other' use in the GEC sample is a citation form taken from an article about language use.

In this journal, then, human intervention in the research process is mediated to the extent that human arguments are presented as arising directly from the methods and results that provoke them. Where human argument is described, it is likely to be confined to the speculative end of the research process.

7.6.1.2 CLAIM

The study of *CLAIM* was carried out using the same classification system as used for *ARGUE* previously. Table 7.7 shows the results and is a little more complex than Table 7.6. In three of the journals (AEE, PS and REE) there are fewer than 100 instances in total, so all instances have been scrutinised. In GEC, a random sample of 100 was taken, as for *ARGUE*. The scrutinised lines include instances where *CLAIM* is not a marker of status (e.g. *This provision will enable people to claim land . . .* [GEC 1994 Yarnal]) and some instances where *claim/s* has been mis-tagged and the word is a noun rather than a verb. These uses are excluded from analysis. In Table 7.7, the total instances, number of exclusions, and number of instances classified are shown. Then the number and percentage (of instances classified) of each source type is recorded. As with *ARGUE*, there are some instances of hypothetical uses such as *it could be claimed that . . .* , and phrases such as *there is no reason to claim . . .* are included in this category (Table 7.7, row 5).

As expected, human beings, mainly researchers, comprise most of the subjects of the verb in all the journals, although there are a few non-human sources such as *Models that claim . . .* (AEE 2002 van den Berg) or *a previous hypothesis, which claimed . . .* (PS 2010 Ye). There are no instances in any of the journals (including all 335 instances of *CLAIM* in GEC) of words such as *results*, *data*, *findings* as subject.

Table 7.7 Uses of *CLAIM* in four journals.

	AEE		*GEC*		*PS*		*REE*	
Total instances	95		335		26		31	
Excluded	11		20 (from 100)		0		3	
Instances classified	84		80		26		28	
Source	Number	%	Number	%	Number	%	Number	%
1 Researcher	41	48.8	22	27.5	17	65.3	10	35.7
2 Researcher/s by discipline/ geography/group	1	1.1	11	13.7	0	0	0	0
3 Writer	2	2.3	3	3.7	7	26.9	8	28.5
4 Non-researcher	28	33.3	36	45	2	7.6	6	21.4
5 Hypothetical	6	7.1	3	3.7	0	0	1	3.5
6 Non-human	6	7.1	5	6.2	0	0	3	10.7

What is noticeable from Table 7.7 is the relatively large proportion of the 'non-researcher' category (row 4), especially when this is compared with the equivalent number in Table 7.6. This suggests that non-researchers are more likely to be construed as 'claiming' than as 'arguing'. 'Claiming' implies disagreement on the part of the writer whereas 'arguing' is more neutral. Attributing a statement to a non-researcher using *ARGUE* implies a greater acceptance of the viewpoint expressed than does attributing the statement using *CLAIM*. Examples 29 and 30 illustrate this. In example 29 the implication is that the farmers are incorrect, while in example 30 there is a greater possibility that the farmers are correct. Note also that the statement attributed to the farmers in example 29 is expressed in non-technical terms, while the one in example 30 is expressed in technical terms (*soil nutrient content*, *micro-topography*). It might be said that a greater degree of respect is accorded the farmers, as co-workers, in example 30 than in example 29, where it is implied that the farmers impede efforts to improve their practice.

29 Sometimes farmers do not adopt the new technologies at all and at other times they may adopt but then abandon technology that does improve production. Many times this is because the farmers **claim** that there are no markets for the increased production . . .

[AEE 2001 Laloe]

30 This micro-variability was acceptable to the farmers who **argued** that a variation in soil nutrient content is well adapted to the micro-topography and to the variability of rainfall, and reduces the risk of crop losses.

[AEE 2001 Hoffmann]

A search in the four journals for *farmer/s* followed by either *CLAIM* or *ARGUE* returned 14 instances with *CLAIM* and five with *ARGUE*. Although there are few examples of either, the difference in frequency does suggest that the tendency across the journals as a whole is to present non-researchers in somewhat negative terms. This makes the articles in GEC that take an alternative view (as in the introductions discussed in Chapter 5, for example) more noticeably different from the mainstream.

This difference between *CLAIM* and *ARGUE* throws light on another interesting comparison between Tables 7.6 and 7.7. This relates to the category of 'researchers identified by their disciplinary, geographic, or other group affiliation' (row 4 in each table). Only GEC uses this category with notable frequency, and it is used with a slightly higher proportional frequency with *CLAIM* rather than *ARGUE*. This suggests that 'other' disciplines or stances are likely to be disparaged by according the propositions attributed to them the status of 'claim' rather than 'argument'. Examples 31 and 32 illustrate this use of *CLAIM*.

31 Though often **claiming** to have universal applicability, single-disciplinary . . . assessments still require exterior inputs from other disciplines in order

to address properly the dynamic and multi-layered nature of vulnerability and resilience.

[GEC 2010 Weichselgartner]

32 A little reflexivity indicates not only the often less than secure grounds of many of our (i.e. social sciences') interpretations; it also suggests that exactly the same argument could be directed at the social science community engaged in research on global environmental change as SBC directs at the natural scientists (i.e. that social scientists are mere opportunists **claiming** unrealistic policy relevance for their research).

[GEC 1995 Shackley]

7.6.1.3 INDICATE

INDICATE is a frequent verb with many thousands of occurrences. As can be seen from Table 7.5, it is relatively more frequent in AEE and PS and relatively less frequent in GEC and REE. The method of classification was as the same as for *CLAIM*, but in addition to the human/non-human classification a class of 'meta' uses was identified, in instances such as *the Y-axis indicates pixel radiance* (AEE), where no evaluation of status is given. Table 7.8 shows the results for each journal.

The results here are fairly uniform, with the non-human sources predominating, accounting for between 58% (GEC) and 92% (PS) of total occurrences (row 4). The meta uses (row 5), relating to diagrams and charts, also occur relatively frequently. Human sources are comparatively rare and non-researcher human sources very rare outside GEC. In GEC, the relatively high proportion of 'writer' instances is largely accounted for by instances of *our research indicates* rather than *I/we indicate*.

A further observation that emerges from the concordance lines is that although non-human subjects predominate, precisely what those sources are depends on the journal concerned. Table 7.9 gives a few examples from each journal. The most striking difference is between PS and REE. In PS, the sources are mostly results of

Table 7.8 Uses of *INDICATE* in four journals.

		AEE	*GEC*	*PS*	*REE*
	Total instances	5752	1763	7260	699
	Instances classified	100	100	100	100
1	Researcher/s as source	11	18	4	0
2	Writer as source	1	5	0	1
3	Non-researcher as source	1	8	0	2
4	Non-human source	73	58	92	85
5	Meta use	14	11	4	12

Table 7.9 Examples of non-human subjects of *INDICATE*.

Journal	*Examples*
AEE	confidence interval; our analyses; average yield of crop; evidence of redistribution; these findings; simulation modelling; identification; this pattern; our results; Fulmaria
GEC	the model; recent research; endowments of these inputs; these findings; integrated measurements; emissions compliance assessment; archival records; first results; a high score; similar symptoms; some ice core studies
PS	low correlation; the algorithm; our sequence analysis; the decrease in the ratio; these data; sequence comparison; these findings; these observations; recent reports; these results; electron microscopic studies; increase in RNase; loading plot and t-test
REE	estimation results; marginal effects of temperature; change in significance; the data in fig 3a; this result; phase dynamics; marginal effects; equations; tests; simple models; this result; our modelling structure; positive sign of proximity; the sign on this variable

experimental procedures, such as *the decrease in the ratio*, and what is indicated is an interpretation of these results. In REE, they are results, such as *change in significance*, or aspects of models, such as *our modelling structure*.

While undertaking this study of the subject of *INDICATE* it became apparent that a noticeably large number of instances were non-finite *indicating*, where the subject or agent of the verb is understood. Example 33 is typical.

33 The N-terminal portion of the both [sic] proteins resembles a signal peptide consensus sequence, which enables the proteins to penetrate the membrane, **indicating** that these GRPs may be located and functionally related to the cell wall or plasma membrane.

[PS 2001 Magioli]

Given that the non-finite *indicating* has no grammatical subject, it must be deduced that the source of the proposition 'these GRPs may be located and functionally related to the cell wall or plasma membrane' is the previous proposition 'the N-terminal portion of the proteins . . . enables the proteins to penetrate the membrane'. In other words, a piece of information that itself is the outcome of observation or experimentation leads to the deduction of another piece of information, with the human activity of deduction heavily masked. This structure, where the understood subject of the verb is somewhat vague, is particularly frequent, proportionally, in PS. As a proportion of the total instances of the lemma it accounts for 19.9% in AEE, 11.0% in GEC, 25.5% in PS and 13.7% in REE. Thus AEE and PS are similar to each other in this respect, and GEC and REE are similar to each other. (Of the 3286 instances of *indicating* in the four journals, only 10% are part of a finite verb phrase [e.g. *could be indicating*].)

A concern to be somewhat more precise in distinguishing non-human sources led to a somewhat different approach to the word *SUGGEST*, as shown in the next section.

7.6.1.4 SUGGEST

The categorisation of proposition sources used in the study of *SUGGEST* is as follows:

- Researcher. The source is a human researcher. Where the verb is non-finite or passive the source is deduced from context. Examples such as *It has been suggested that* are labelled as 'human source'. This category conflates the 'other researcher' and 'current writer' categories used in the previous studies.
- Non-researcher. This may be a policy maker or an actor such as a farmer or research participant.
- Text. These sources are the texts involved in research, such as *paper*, *article*, *study*, and also discoursal elements such as *the discussion in section 2* or *section 3 of this paper*.
- Research process. These sources are research processes, such as *analysis*, *experiment*, *observation*. We include here the words *data*, *evidence*, *findings* and *result*, even though these are in a sense outcomes, because the empirical process itself is lexicalised.
- Research outcome. These sources are the outcomes of research processes and are often observations of the natural world that are the consequence of experimentation (e.g. *a maize grain yield of . . . was realized, suggesting that . . .*). This category is close to the 'physical world' one, the difference being that the research result is the consequence of manipulation by the experiment whereas in the physical world examples the experiment may reveal but not cause the entity.
- Physical world. The proposition source is an object or process in the natural world. These objects are often observable through experimentation (e.g. *Analysis of correlations . . . revealed . . . strong relationship, suggesting that . . .*). In this example, if the agent of *suggesting* were *analysis* it would be classified as 'research agent: process', but as is *relationship*, that is a physical world abstraction revealed by analysis, it is classed as 'physical world'.
- Social world. The source is an entity in the social world, such as a farming practice or a distribution of income.

The results are shown in Table 7.10. Rows 1–7 show the figures out of 100 for each source category in each journal. Rows 8–10 show some of the source categories amalgamated to form three different focuses: argument, observation and experiment. 'Argument' (row 8) comprises human researchers and the texts that they produce. This therefore includes instances such as 'we suggest', 'many researchers

suggest', and 'previous studies have suggested'. Propositions are attributed to what people say or write and are represented as being the outcome of debate. 'Observation' (row 9) encompasses both research outcomes and either the physical or social worlds, as appropriate. It includes instances such as 'these results suggest', 'the higher rainfall suggests' and 'the increase in unemployment suggests'. Propositions are attributed to the observed social and natural worlds, without apparent human intervention. 'Experiment' (row 10) includes all stages in the research process including the outcome of research. It includes instances such as 'the model suggests' and 'the results suggest'. Propositions are attributed to the research world. Research outcomes (row 5) are included in both the 'Observation' and the 'Experiment' categories. It may be noticed that the total numbers for REE in Table 7.10 add up to 101. This is because, although 100 lines were considered, in one instance the agent is both *increasing fish prices* [social entity] *and depleted wild fish stocks* [physical entity], so the example is assigned to both categories.

As with the previous studies, the presence of a few instances of 'non-researcher' in the GEC and REE data (row 2) reflects the interest in some articles in those journals in non-academic actors. The numbers here are lower than in the equivalent row in Table 7.7 (*CLAIM*), in line with the observation that non-researchers tend not to be accorded the same status as researchers, with a few exceptions as noted previously. The difference between 'physical world' and 'social world' as agent (rows 6 and 7) also reflects differences in the topics of the journal. AEE and PS are both concerned with the natural world; GEC is concerned with both the natural world and the social; REE is concerned mainly with economic factors in the social world.

The high figure for REE in row 4 (Research process) is largely explained by the frequency of the phrases *the results suggest*/*the data suggests*/*the findings suggest*. As a test of this, another search was carried out for all instances of

Table 7.10 Uses of *SUGGEST* in four journals.

		AEE	*GEC*	*PS*	*REE*
	Total instances	4621	2786	7960	747
	Instances classified	100	100	100	100
1	Researcher	29	31	26	23
2	Non-researcher	0	2	0	1
3	Text	5	10	3	8
4	Research process	33	34	27	40
5	Research outcomes	14	4	11	8
6	Physical world	19	7	33	1
7	Social world	0	12	0	20
8	Argument (1 + 3)	34	41	29	31
9	Observation (5 + 6/7)	33	23	44	29
10	Experiment (4 + 5)	47	38	38	48

result/*results*/*data*/*findings* immediately preceding *SUGGEST* in each of the four journals. As a proportion of total instances in each, these phrases account for 7.3% of instances of *SUGGEST* in AEE, 2.6% in GEC, 11.2% in PS and 9.5% in REE.

Row 8 in Table 7.10 suggests that GEC is the journal with the highest proportion of human argumentation as the source of ideas, knowledge and interpretation. PS has the highest proportion of observation as the source (row 9), while AEE and REE appear to be most heavily focused on experiment as the source (row 10).

7.6.2 Two nouns: observation *and* argument

A similar process to that outlined in the previous section was followed in investigating two noun markers of status. As before, the aim is not simply to apply a predetermined categorisation system but to use observations of the sample concordance lines to draw more nuanced conclusions about the operation of status markers in the journals. The two selected nouns, *observation/s* and *argument/s*, represent the two approaches to knowledge construction that inform the natural and social sciences, respectively. Observation of the physical world and of the results of empirical work underpins natural science research, where it is assumed that the researcher is discovering how the physical world works. Building an argument underpins social science research, where the researcher is assumed to be construing knowledge.

7.6.2.1 Observation/s

Table 7.11 shows the raw and normalised frequencies of the lemma *observation/s* in each of the four journals. It is markedly frequent in PS and relatively infrequent in GEC and REE, with AEE falling in the middle. Table 7.11 also shows the proportional frequency of three phraseologies. The phrases 'the observation that' and 'this/these observation/s' are chosen because they are used in conjunction with the overt expression of a proposition, as in example 34, where 'organic fertilizers diminish herbivorous insect populations' is evaluated as an observation, and therefore aligned to the world (i.e. factual). The phrase 'observation/s of' nominalises the process of observation and can be used to relate that proposition to another. In example 35, two propositions – 'there are bacteria within cells' and 'P.s.pv.coriandricola is able to penetrate cell walls' – are evaluated as an observation and a deduction, respectively, with the second being dependent on the first.

34 Some evidence supports the **observation** that organic fertilizers diminish herbivorous insect populations (references).

[AEE 2001 Morales]

35 Further, the **observation** of bacteria within cells is evidence of the ability of P.s.pv. coriandricola to penetrate cell walls.

[PS 2001 Refshauge]

Table 7.11 Frequency and phraseology of *observation/s* in four journals.

	AEE		*GEC*		*PS*		*REE*	
Total *observation/s* raw frequency	2142		682		1718		371	
Total *observation/s* normalised frequency (pmw)	218.25		144.65		188.49		238.72	
	Number	%	Number	%	Number	%	Number	%
'*the observation that*'	29	1.35	14	2.05	114	6.63	9	2.42
'*this/these observation/s*'	149	6.95	65	9.53	497	28.92	35	9.43
'*observation/s of*'	238	11.11	98	14.36	212	12.33	22	5.92
Total percentage		19.41		25.94		47.88		17.77

Table 7.11 shows that the journal PS has a relatively larger proportion of the total instances of *observation/s* taken up by these phrases. We might conclude that PS is more overt than the other journals in specifying the propositions that are evaluated as observations. REE has the lowest proportion. The interdisciplinary journals AEE and GEC lie between the two extremes.

The second part of this study is based on a random sample of 50 instances of *observation/s* from each of the four journals. Each (expanded) concordance line has been scrutinised and allocated to a meaning type. Examples 36–39 are examples of the four meaning types identified in the journal AEE.

36 All **observations** were done by a single observer.

[AEE 2002 Kujawa]

Meaning type allocated: description of method.

37 [T]he data were unbalanced both in the number of transects . . . and the number of **observations** within each transect . . .

[AEE 2010 Oom]

Meaning type allocated: critique of method.

38 These **observations** substantiate reports that above-ground biomass and grain yield . . .

[AEE 2009 Muthuri]

Meaning type allocated: role of the observation in the research process.

39 This **observation** is of great importance with respect to climate change.
[AEE 2005 Paranychianakis]

Meaning type allocated: evaluation of observation.

Allocation of the 50 random concordance lines to these four categories shows that in AEE most instances of *observation/s* are concerned with how empirical research is carried out, and how well or badly it has been done (33/50 lines), with a smaller number of instances concerned with role of the observation in the research process (14/50 lines). A far smaller number (2/50) evaluate the observation subjectively, as important or worrying.

In carrying out the same study for GEC, initially the same categories were adopted. A total of 5/50 lines were part of the description of method, and a further 11/50 were a critique of method (which we shall return to in a moment). A total of 12/50 lines fitted the 'role of observation in research' classification, and 3/50 expressed subjective evaluation. Together these categories account for 31 of the 50 lines. Two further categories were required to account for all the concordance lines. The larger category (13/50) comprised instances where the observer is an individual other than a researcher. Examples 40 and 41 show this, with the agents of observation in 40 being people writing to national newspapers and those in 41 being animal herders. Note the respect accorded to the herders' experience in example 41 (comparing examples 29 and 30 in this chapter). There are no instances of this type in AEE.

40 Now, however, other national newspapers publish letters competing for stranger **observations**.
[GEC 2009 Lawrence]

41 The recent increased frequency of such years . . . suggests consistency with herders' **observations** of patchier and heavier rains.
[GEC 2010 Marin]

The second category that is unique to GEC, though with only 3/50 instances, is where the proposition that is evaluated as an observation is itself evaluative, as in example 42. In this instance, the proposition 'social scientific research on participation is fragmented' is an assessment of the quality of social science research. This proposition could have been assigned the status of a 'claim' or an 'argument', as it is clearly subjective, but instead it is described as an 'observation'.

42 An important **observation** is the fragmented nature of social scientific research on participation.
[GEC 2002 van Asselt Marjolein]

The presence of these examples raises awareness that in the other journals, the label *observation* is exclusively attached to propositions that are expressed as non-subjective.

The exercise of classifying each of 50 concordance lines gives rise to the identification of another phenomenon that is peculiar to GEC: calling into question the value of observation itself as a methodology. There are only two instances of this in the 50 lines (examples 43 and 44), and they have not been placed into a separate category because they fit into the 'critique of method' one, but they are worthy of mention. Example 43 reports an argument against observation, while 44, although it asserts the non-arguable nature of a particular observation, by that very assertion raises the point that some observations are regularly disputed (which is the point of the paragraph from which this example comes).

43 [I]t has been argued that the troposphere has not warmed over the satellite era . . . thus casting doubt on . . . the reliability of thermometer-based **observations** of surface warming.

[GEC 2005 Antilla]

44 The direct **observation** of a tree being cut down is highly unlikely to be disputed.

[GEC 2002 Jones]

In short, the word *observation/s* in GEC has a quite different profile from that in AEE, and this difference highlights the restricted meaning of *observation/s* in AEE.

Turning to the journal PS, where the word *observation/s* is particularly frequent (see Table 7.11), we see it nonetheless fitting a very restricted profile. So many instances of 'role of the observation in the research process' were identified (39/50) that they were initially subdivided in more specific types: the observation shows something (example 45), observations are (in)consistent (example 46), and an observation (dis)confirms a hypothesis (example 47). In addition, there are 9/50 instances of *observation* used in the description, though not the critique, of a method, and just 1/50 instance of *observation* being evaluated as *of particular interest*.

45 These **observations** indicate that ethylene acts to stabilize EIN3/EIL proteins.

[PS 2010 Huang]

46 However, our **observations** are also in agreement with those reported by Kuwabara et al (references).

[PS 2005 Pandey]

47 These **observations** agree with the proposal by Bartels (ref) that both the prevention of oxidate stress and the elimination of reactive oxygen species are the most effective approaches.

[PS 2005 Caravaca]

The final journal, REE, is noticeable in that the observations reported are of people rather than of plants or other natural phenomena. Perhaps because people behave in a less consistent way than plants do, observation of them tends to raise problems, and there is a relatively large proportion of instances (18/50) that critique methodology, either positively or negatively, as well as a large proportion of instances describing method (25/50). The remaining instances (6/50) are classified as 'the role of the observation in the research process'. In other words, although the focus of research in REE is very different from that in PS, the classes of meaning around the word *observation/s* are of the same types; differences in proportional frequency can be attributed to differences in the nature of the entities that are observed. Examples 48–50 illustrate the categories of method, critique of method and role in research, respectively.

48 The panel consists of six **observations** per individual as described above.
[REE 2003 Boxall]

49 Given the paucity of **observations**, it would be tempting to adopt the linear expenditure system.
[REE 2003 Maddison]

50 The **observation** that efficiency gains are largest for low budgets is in line with results from the conceptual model.
[REE 2010 Drechsler]

We may draw a number of conclusions from the study of *observation/s*. In PS, the word is used in accordance with expectation, construing the research process as a straight line between observation and knowledge. The concept of observations as the reliable bedrock of scientific research is not problematized. In REE, although the object of observation is quite different, and the practice of observing people rather than things is seen to be problematic, the role of observation as a basic step in the creation of knowledge is similarly assumed. The same is true for AEE, with the additional point that *observation/s* is sometimes subjectively assessed as 'important'. This is consistent with the point made in Chapter 5 that articles written for an interdisciplinary audience tend to be relatively explicit about the importance of the research undertaken.

GEC is the most distinctive in its use of *observation/s*, and here again we observe the discipline reflexivity and sometime antagonism that has been remarked upon in Chapters 5 and 6. The word is relatively low in frequency, but the range of uses is greater than in the other journals. A distinction is made between 'researcher' observation and 'lay person' observation, but when *observation/s* is used the former is not necessarily given a privileged position in the construction of knowledge. Indeed the concept of researcher observation itself, as the basis of knowledge construction, is called into question.

Given that the previous discussion is based on 50 lines only, as a test of how widespread this use of the word might be, all 250 instances of the singular *observation* have been scanned, and instances where the word has a discipline-reflexive

use are identified. All instances are shown in examples 51–58. In all cases except 52, *observation* is a non-count noun and occurs without a determiner. These examples are about observation as a phenomenon rather than as the label given to an in-text proposition. In 52, *observation* is preceded by *an*, but it is a hypothetical observation rather than a specific instance. The examples explain and critique the act of observing, an act which in the other journals is taken for granted.

51 For example, **observation** and theory are typically in reasonable balance in the physical science, theory dominates in many areas of economics, while the reverse is true in much of ecology.

[GEC 2005 Newell]

52 For instance, an **observation** might contradict a hypothesis, or strengthen the belief in it.

[GEC 2006 Welp]

53 Some of the approaches highlighted are reliant upon building understanding through experiential knowledge and **observation**, while other approaches follow a conventional hypothesis-deductive model.

[GEC 2008 Armitage]

54 Social-ecological memory extends beyond the merely extractive collection of ecological information to a deeply integrated connection between **observation** and meaning among groups of people (references) and their institutions (references).

[GEC 2010 Barthel]

55 We consider that those who seek to blend knowledge from the natural science, the social sciences and the humanities need to develop approaches that: utilise rich interactions between **observation** and theory . . .

[GEC 2005 Newell]

56 Indigenous or traditional knowledge is also the product of both **observation** and interpretation.

[GEC 2009 Lawrence]

57 Learning about the limits of the choice constraint requires not theory but **observation**.

[GEC 2002 Patt]

58 Seasonal changes in species behaviour and distribution are not amenable to rigorous scientific experimentation or controlled **observation**; and the participation of a range of possibly less scientifically trained observers also removes usual sources of standardisation.

[GEC 2009 Lawrence]

7.6.2.2 Argument/s

The final status-indicating word to be investigated in the journals is the noun *argument/s*. This noun has been selected because, like *observation*, it construes a stage in the construction of knowledge, but unlike *observation* it represents a stage that is overtly subjective. The study supports both these characterisations. Table 7.12 shows the overall frequency of the word in each of the journals and the normed frequencies. There is a large difference in the frequencies, with GEC and REE having a high frequency and AEE and PS a low frequency, making a very clear distinction between these pairs of journals. Table 7.12 also shows the proportion of instances ascribable to four key phraseologies: '*this/these argument/s*'; '*argument/s for/in favour of/against*'; '*argument* + that-clause' and 'adjective + *argument/s*'. Table 7.12 also shows the approximate percentage of the total instances in each journal accounted for by these phraseologies. (This figure is approximate and slightly inflated because it is arrived at by simply adding together the other percentages. Individual instances may occur in more than one category; for example *a compelling argument in favour of* . . . would be counted both as '*argument in favour of*' and as 'adjective + *argument*'.) This latter figure shows the importance of the identified phraseologies, which account for between, roughly, 60% and 85% of total occurrence. It also suggests that the range of contexts in which *argument/s* is used in GEC and REE is greater than in the other two journals, because the 'total accounted for' percentage is lower (though see the previous point about adjectives). PS appears to be the most uniform journal, with over 85% of the very few instances accounted for by the phraseologies.

Two further studies on this lemma have been carried out. All the modifiers, including adjectives, were extracted from 100 instances of *argument/s* in each journal, with the exception of PS where all 37 instances were used. The modifiers were then categorised according to meaning. The following range of meanings was identified:

- Deictic adjectives: *above*, *following*, *latter*, *second* etc. There are 11 types altogether and no obvious differences between the journals.

Table 7.12 Phraseology of *argument/s* in four journals.

	AEE	*GEC*	*PS*	*REE*
Total instances	110	715	37	230
Normed frequency (pmw)	11.21	151.65	4.06	147.99
This/these argument/s	13 (11.8%)	56 (7.8%)	9 (24.3%)	16 (6.9%)
N for/in favour of/against	26 (23.6%)	85 (11.8%)	9 (24.3%)	45 (19.5%)
N that/be that	11 (10%)	91 (12.7%)	5 (13.5%)	21 (9.1%)
Adj + N	33 (30%)	240 (33.5%)	9 (24.3%)	64 (27.8%)
Total accounted for	75.4%	65.8%	86.4%	63.3%

- Evaluative adjectives: *important*, *main*, *major*; *compelling*, *powerful*, *strong*; *challenging*, *legitimate*, *reasonable* etc. There are 29 types altogether and no obvious differences between the journals except that PS uses the fewest and GEC the most, as might be expected from the overall frequency of the noun.
- Domain-related adjectives: *economic*, *political economy*, *ethical*, *scientific* etc. There are seven types altogether, almost exclusively in AEE and GEC. This suggests the presence of a reflexive approach to disciplinarity, as discussed previously in relation to *observation*. We will return to this point.
- Specifying modifiers: *compensation*, *latency*, *replication*, *separate-taxon*, *terms of trade*, *government wastage* etc. There are 31 types altogether, with no type occurring in more than one journal. Phrases such as *the acid-growth argument* (PS) or *the guilt/indignation argument* (REE) clearly construe field-specific arguments that assume considerable shared familiarity with previous research. This is a minority category in all the journals except REE, which has 17 of the 31 types, accounting for 20 of the 100 sample lines. This suggests that of all the journals, REE is the one that, in this phraseology at least, construes the most homogenous audience.
- Terms used in modelling. In a very few cases in AEE and REE only, exemplified by *terminal leaves of function argument* or *chain-rule argument*, the word *argument* seems to have a specific meaning in the context of theoretical modelling.

As noted, domain-related adjectives are used in AEE and GEC, suggesting that a difference in disciplinary approaches is a concern. Of the seven adjective types, two are used in both journals: *economic* and *scientific*. The concordance lines for these bigrams (*economic argument/s* and *scientific argument/s*) are shown in Figures 7.1a and 7.1b.

What is noticeable is that, in AEE, *economic argument/s* and *scientific argument/s* are taken at face value and are represented as contributing to a debate. In GEC, whereas this is true of *economic argument/s*, it is not true for *scientific argument/s*. This phrase is more likely to be used to contrast the propositions referred to as *scientific arguments* with political or commercial considerations (see example 59) or to cast doubt on the presumed neutrality of science (example 60). All the occurrences challenge the concept of science having a privileged position in the construction of knowledge.

59 These [political disputes] are concerned with the validity of the **scientific arguments**, and also the questions of cost, equity, and responsibility.
[GEC 2008 van Aalst]

60 Therefore, to examine the ethical dimensions . . . one must be unable to untangle the ethical issues from the **scientific arguments** that often pretend to be ‘value-free’ but that on further examination are heavily value-laden.
[GEC 2003 Brown]

The economic argument for taking MFA seriously is a market-fa
, to find if there were scientific arguments in favour of domestic support to agric
ojects vying to provide scientific arguments favouring the implementation of TBI sy

Figure 7.1a Economic argument and *scientific arguments* in AEE.

abandoned. On the contrary, economic arguments and analyses are much needed to back
y use both conservation and economic arguments to justify the coercive exclusion of
per, would be to assess the economic arguments for and against the near-term targe
mportant in light of recent economic arguments made in favour of using a lower disc
ed response—on the basis of economic arguments alone. For the 550 ppm target, this
arly or default pathway; if economic arguments only are taken into account, rates i
t be determined by national economic arguments. As several key contributors have po
ut to achieve: a convincing economic argument for strong (global) mitigation action

nt in effect crippled the scientific argument against CFCs and left the ecological
ly butressed by pieces of scientific arguments. A twofold problem of expectations.
forced into neutrality by scientific arguments about the radiative forcing potentia
thin. Irrespective of the scientific arguments, the usefulness of reductionism as a
uch as solar variations), scientific arguments relating to human caused versus natu
 with the validity of the scientific arguments, and also the questions of cost, equ
 be important substantive scientific arguments which are not being properly aired i
e ethical issues from the scientific arguments that often pretend to be "value-free
gas concentrations). Both scientific argument and value judgements are seen to def

Figure 7.1b Economic argument and *scientific arguments* in GEC.

In the second further study, all instances of *this/these argument/s* were identified in the four journals and the meanings around them identified. The aim is simply to see what these meanings are and in particular to identify whether they are similar or different from the meanings associated with *observation/s*. Four meanings are identified:

- An argument exists or is made – e.g. *this argument holds that . . .* ; *this argument may be reciprocally presented*; *industry lobbyists have consistently put this argument forward*.
- An argument supports or is supported by evidence – e.g. *further weight is given to this argument by . . .* ; *this argument resonates with the idea that . . .* ; *these arguments support the hypothesis that . . .*
- An argument is the result of other research processes or leads on to other research processes – e.g. *based on these arguments*, *we assume . . .* ; *these arguments indicate that . . .* ; *this argument is not a correct basis for assuming . . .*
- An argument is assessed – e.g. *I would challenge this argument; this argument seems logical because . . .* ; *this argument has been challenged on at least three accounts*.

The relative frequencies of these meanings (set out in Table 7.13) are not very instructive, because any differences are dwarfed by the large difference in overall frequency. One point, however, is worth making. The most usual use of *this/these argument/s* in PS is in the context of 'support'. The examples show that arguments are set in contrast with facts, evidence and observations, with the latter forming the basis of accepting or discounting the argument (e.g. *the observation . . . refutes this argument* or *Evidence for this argument came from the fact that . . .*). In GEC, arguments are less likely to be supported or overthrown by fact and more likely to occur as part of a larger theoretical construct (e.g. *This argument assumes that . . .* or *. . . lays out the theoretical foundations for this argument*) without necessarily being shown to be true or false.

This study of *argument/s* has shown that, as expected, it is in many ways the corollary of *observation/s* and construes propositions as contributing to a

Table 7.13 Meanings around *this/these argument/s* in four journals.

	AEE	*GEC*	*PS*	*REE*
Total occurrences	13	56	9	16
Argument exists	2	14	1	3
Argument supports	4	8	6	3
Argument causes	4	18	0	3
Argument is assessed	1	12	1	5
Other	2	3	1	2

chain of knowledge-construction. Whereas *observation/s* is most frequent in PS, *argument/s* is most frequent in GEC. Further study of this word shows that, in PS, arguments are construed as dependent on fact/observation, whereas in GEC they are more likely to be construed as dependent on each other. It also reveals other contrasts between the journals, notably the community-building role of 'modifier + *argument/s*' in REE, and the consensus-challenging role of *scientific argument/s* in GEC.

7.7 Conclusion

This chapter has used quantitative and qualitative information about selected lemmas as a resource for investigating the characteristics of the four journals. The lemmas were selected from a list derived on grounds of linguistic form: all the adjectives, nouns and verbs on the list are attested to occur with appositive that-clauses. In addition, all the selected lemmas play a role in marking the epistemic status of propositions that combine to form disciplinary knowledge. It has been argued, therefore, that these lemmas play a significant role in the research process.

The quantitative research reported in the chapter shows which status markers are most distinctive to each of the journals studied. The qualitative research focuses on six individual lemmas, four verbs and two nouns, and carries out a number of studies, including categorising a sample of each lemma in each journal, and also examining concordance lines to identify distinctive features of each journal.

The normalised frequency list, of which Appendix B gives a flavour, shows considerable divergence between the journals. As might be anticipated from the discussion in Chapter 3, the journals PS and GEC are the most distinct from one another. For the most part, the journal REE is most like GEC, and the journal AEE is most like PS. This confirms a natural science–social science divide that appears to over-ride the monodisciplinary–interdisciplinary distinction, at least when studies are quantitative. Evidence for the pairing of GEC with REE (social science) and of AEE with PS (natural science) is given by:

- The presence of words connected with human thought, perception, action and affect in the lists of distinctively high frequency words in GEC and REE only (see Tables 7.4a and 7.4b)
- The relatively high frequency of *ARGUE*, *CLAIM* and *argument/s* in GEC and REE, and the relatively high frequency of *INDICATE*, *SUGGEST* and *observation/s* in AEE and PS (see Tables 7.5, 7.11 and 7.12)
- The relatively high frequency of human agents of the four selected verbs in GEC and REE, and the correspondingly high frequency of non-human agents of the four selected verbs in AEE and PS
- In GEC and REE only, the attribution of *observation/s* to non-researchers
- The relatively high degree of diversity in the phraseology around *argument/s* in GEC and REE compared to AEE and PS

It has been noted throughout this chapter, however, that the GEC is the most distinctive of the four journals, a finding again anticipated by previous chapters. As in those chapters, it has been noted that some aspects of the use of the selected lemmas can be described as discipline-reflexive. Some uses go further and offer a critical or antagonistic stance towards one discipline by another. Evidence for this is given by the following:

- Of the potential status markers that occur infrequently and/or in only one journal, most occur in GEC only. Many of these words refer to human feelings or make a judgement about the accuracy of information, such as *misunderstanding* or *fallacy*.
- GEC has the highest number of status marker types (see Table 7.4a). It has the highest number of distinctive status marker types (see Table 7.4c).
- With the verbs *ARGUE* and *CLAIM*, GEC has the highest proportion of subjects that delimit human agents in terms of discipline (see Tables 7.6 and 7.7).
- Uses of the noun *observation/s* in GEC include, uniquely, a problematization of the process of observation itself.
- The phrase *scientific argument* is used relatively frequently in GEC and has a discipline-reflexive role.

It is these characteristics – discipline reflexivity and disciplinary antagonism – that seem to mark the specific interdisciplinary nature of GEC. It might be contrasted with AEE, where the evidence from status markers point to a more integrative stance.

References

Barry, A., & Born, G. (Eds.). (2013). *Interdisciplinarity: Reconfigurations of the social and natural sciences*. London: Routledge.

Becher, T. (1989). *Academic tribes and territories: Intellectual enquiry and the cultures of disciplines*. Milton Keynes: Open University Press.

Biber, D. (2006). Stance in spoken and written university registers. *Journal of English for Academic Purposes*, 5(2): 97–116.

Chafe, W., & Nichols, J. (1987). *Evidentiality: The linguistic coding of epistemology*. Santa Barbara: Praeger.

Charles, M. (2006). The construction of stance in reporting clauses: A cross-disciplinary study of theses. *Applied Linguistics*, 27(3): 492–518.

Charles, M. (2007). Argument or evidence? Disciplinary variation in the use of the *Noun that* pattern in stance construction. *English for Specific Purposes*, 26(2): 203–218.

Flowerdew, J., & Forest, R. (2017). *Signalling nouns in English: A corpus-based discourse approach*. Cambridge: Cambridge University Press.

Francis, G. (1986). *Anaphoric nouns*. University of Birmingham and English Language Research, Birmingham, UK.

Francis, G., Hunston, S., & Manning, E. (1996). *Collins cobuild grammar patterns 1: Verbs*. London: HarperCollins.

Francis, G., Hunston, S., & Manning, E. (1998). *Collins cobuild grammar patterns 2: Nouns and adjectives*. London: HarperCollins.

Gilbert, G. N., & Mulkay, M. (1984). *Opening pandora's box: A sociological analysis of scientists' discourse*. Cambridge: Cambridge University Press.

Halliday, M. A. K. (1994). *An introduction to Functional Grammar* (2nd ed.). London: Routledge.

Halliday, M. A. K. (2004). *The language of science*. London: Continuum.

Hunston, S. (1993). Evaluation and ideology in scientific writing. In M. Ghadessy (Ed.), *Register analysis: Theory and practice* (pp. 57–74). London: Pinter.

Hunston, S. (2000). Evaluation and the planes of discourse: Status and value in persuasive texts. In S. Hunston & G. Thompson (Eds.), *Evaluation in text: Authorial stance and the construction of discourse* (pp. 177–206). Oxford: Oxford University Press.

Hunston, S. (2004). "It has rightly been pointed out . . .": Attribution, consensus and conflict in academic English. In M. Bondi, L. Gavioli & M. Silver (Eds.), *Academic discourse, genre and small corpora* (pp. 15–33). Rome: Officina Edizioni.

Hunston, S. (2011). *Corpus approaches to evaluation: Phraseology and evaluative language*. London: Routledge.

Hunston, S. (2013). Systemic functional linguistics, corpus linguistics, and the ideology of science. *Text and Talk*, 33: 617–640.

Martin, J., & White, P. (2005). *The language of evaluation: Appraisal in English*. London: Palgrave.

Pallen, M. (2009). *The rough guide to evolution*. London: Rough Guides.

Teich, E., & Holtz, M. (2009). Scientific registers in contact: An exploration of the lexicogrammatical properties of interdisciplinary discourses. *International Journal of Corpus Linguistics*, 14(4): 524–548.

Willett, T. (1988). A crosslinguistic survey of the grammaticalization of evidentiality. *Studies in Language*, 12: 51–97.

8 An explanatory code gloss

In other words

8.1 Introduction

It was noted at the beginning of Chapter 7 that the theme of interpersonal meaning in discourse has been addressed from a number of viewpoints and using varied terminology. The concept of 'metadiscourse' is an influential approach to the study of interaction between writer and reader in academic discourse. Hyland (2005: 3) ascribes the original use of the term to Zellig Harris (1959), and it came to prominence through the work of Vande Kopple (1985), Crismore, Karkkanen and Steffenson (1993), Hyland (e.g. 2000, 2005, 2010) and Hyland and Tse (2004). Although there are differences in detail among these approaches, the definition from Hyland and Tse (2004: 157) is fairly representative:

> Metadiscourse is defined here as the linguistic resources used to organize a discourse or the writer's stance towards either its content or the reader.
>
> (Hyland 2000: 109)

Hyland and Tse emphasise that the notion of community is essential to an understanding of metadiscourse. When using metadiscourse resources, the individual writer negotiates with the community, demonstrating competence and alliance with community norms 'to gain a more sympathetic hearing for their own views' (Hyland & Tse 2004: 163).

Hyland (2005) lists a number of metadiscourse resources, divided between 'interactive' and 'interactional' resources (following Thompson 2001). Among the interactive resources, whose function is to 'help to guide the reader through the text' (Hyland 2010: 128) are code glosses, which 'help readers grasp the meaning of ideational material' (Hyland 2010: 128). Code glosses include items such as *i.e.*, *in other words*, *for example* and *namely*. The explanatory function of code glosses suggests that they may be particularly useful resources for writers in interdisciplinary journals, and it is for this reason that we undertake a detailed study of one of these code glosses in the BEE4 corpus. As noted, the term 'code glosses' usually includes words and phrases expressing exemplification (*e.g.*, *for example*), but we restrict the category here to comprise only phrases that express paraphrase or explanation. Such phrases might include the very frequent expression *i.e.* (4130

instances in AEE) and the very infrequent expressions *that is to say* and *put (it) another way*, each of which occurs fewer than 20 times in the whole BEE4 corpus. We restrict ourselves here to four expressions: *i.e.*, *in other words*, *in short* and *this means (that)*. Table 8.1 shows the relative frequency of these phrases in the BEE4 corpus.

The abbreviation *i.e.* is used for quick glosses, often in parentheses. Each of the other phrases has a distinct canonical meaning. For *in other words* this might be expressed as 'to paraphrase'; for *in short* it is 'to summarise'; for *this means* it is 'the logical consequence is'. These uses are exemplified in examples 1–3.

1 These results indicate that the induced resistance in tomato is ineffective against the ladybird beetle. **In other words**, the ladybird beetle completely overcomes the induced resistance of the host plant, tomato.

[PS 2005 Shinogi]

2 Assessment capacity building efforts can improve the quality and scope of assessment outputs by regularizing and expanding important functions such as monitoring and data collection, calibration and analysis. **In short**, all of the large-scale assessment processes studied here influenced the conduct of scientific and technical research over time.

[GEC 2001 Farrell]

3 Projects may be supported by up to 30% of the investment costs. It is the intention only to support projects that would not have been carried out without a subsidy. In practice **this means** that only projects with a pay-back period exceeding 2 years . . . are supported.

[REE 2002 Bjorner]

What meaning might be attached to the presence or absence of these code glosses in a corpus? As noted by Hyland (2005), their presence enhances the interaction

Table 8.1 Explanatory code glosses in the BEE4 corpus (normalised pmw).

		AEE	*GEC*	*PS*	*REE*
i.e.	Raw frequency	4130	1138	1288	1049
	Normalised	420.80	241.36	141.31	674.98
in other words	Raw frequency	160	363	62	138
	Normalised	16.30	76.99	6.80	88.80
in short	Raw frequency	15	73	20	9
	Normalised	1.52	15.48	2.19	5.79
this means	Raw frequency	235	138	57	106
	Normalised	23.94	29.27	6.25	68.21
Total	Normalised	480.38	363.10	156.55	837.78

between writer and reader and is an indication of the attention paid to the reader by the writer. Specifically, the connection between two clauses or sentences is explicitly signalled (Hoey 2001) and is not left to the reader to deduce. Another important point is that the presence of the phrases is itself a signal that explanatory material is present in the text, material that for some readers may be redundant. For instance, there are two alternatives to example 1, given here as examples 1a and 1b.

1a These results indicate that the induced resistance in tomato is ineffective against the ladybird beetle. The ladybird beetle completely overcomes the induced resistance of the host plant, tomato.
1b These results indicate that the induced resistance in tomato is ineffective against the ladybird beetle.

In example 1a, the phrase *in other words* is omitted and the reader is left to deduce the connections between the two sentences by themselves. In example 1b, the whole second sentence is omitted and the reader deduces the consequences of the failure of induced resistance in the tomato plant without help. One reason for the presence of *in other words*, therefore, is 'to make the information more explicit'; the other reason is 'to give more explanation'. For both these reasons, a higher incidence of the code gloss phrases indicates a greater degree of interactivity and attention to reader needs. It might be hypothesised that articles in interdisciplinary journals would have a higher frequency of code glosses, reflecting a greater perceived need for explicitness and explanation than would be found in monodisciplinary journals. A counter hypothesis is that the code glosses would be more frequent in the social science journals than the natural science ones.

Hyland (2010: 136) finds that 'the more "soft knowledge" social science disciplines (in texts written by postgraduate students) employed more metadiscourse overall (56% of the normed count', but that interactive resources, of which code glosses are one, are more evenly spread. The normed frequencies (per 10,000 words) for code glosses (Hyland 2010: 139) show the social sciences Applied Linguistics and Public Administration sharing the highest frequencies (41.1. and 36.6 respectively) with Biology (36.0). Business Studies, Computer Science and Electronic Engineering have slightly lower frequencies (30.0, 32.3 and 30.7 respectively).

Table 8.1 shows less consistency in the BEE4 corpus than in Hyland's student corpus. Overall (see the 'Total' row), REE has the highest frequency of explanatory code glosses, and has the highest frequency of each expression. AEE comes second, but the large total is largely accounted for by *i.e*. The phrasal code glosses are less frequent. GEC has frequencies fairly close to REE, except that *i.e*. is much less frequent. PS has the lowest frequency of code glosses. Even excluding *i.e*., AEE has more instances of each code gloss than PS does. The rest of this chapter will look specifically at *in other words*, with particular attention to AEE and GEC to see whether the expected interdisciplinary readership appears to have influenced

this use. *In other words* has been chosen because its canonical meaning of 'to paraphrase or restate' would appear to make it of particular use to writers crafting their text for an interdisciplinary, and therefore partly non-expert, audience. It is also a full-phrase code gloss, unlike the more frequent *i.e.*, so its use may be considered less routine. Its use raises interesting questions about the relationship between the individual and the community that we shall return to in the conclusion to this chapter.

8.2 A quantitative study of *in other words*

We begin with a quantitative survey of *in other words* in the BEE4 corpus. As well as asking how frequent the phrase is in the journals, it is important to see how evenly distributed the phrase is within each journal. Table 8.2 shows the frequency of the phrase, the number of articles in each sub-corpus and the number and percentage of articles that include the phrase. In most cases an article will include one or two instances, but the bottom row of the table gives the number of articles in each journal that include four or more instances of the phrase.

Table 8.2 shows that *in other words* is more frequent, and present in a larger proportion of articles, in GEC and REE than in the other two journals. It is somewhat more frequent in AEE than in PS. Each journal except PS has a small number of articles that contain four or more instances of the phrase.

The abstract for each of the 'high use' articles has been examined. There is evidence that these articles introduce methods or approaches that might be novel for the reader. Comments on the four AEE articles are shown in Table 8.3.

In each case in Table 8.3 there is explicit mention of discipline, approach or a mixture of methods. This and the frequency information in Table 8.2 suggests that *in other words* may perform a specifically interpersonal function in these journals and is more likely to be used in contexts where an effort of explanation or justification has to be made, and/or where a diverse readership is anticipated. This discussion will be continued using qualitative evidence in the next section.

Table 8.2 Distribution of *in other words* in four journals, 2001–2010.

	AEE	*GEC*	*PS*	*REE*
Raw frequency of *in other words*	160	243	62	138
Normalised frequency of *in other words*	16.30	76.99	6.80	88.80
Number of articles in sub-corpus	1785	434	2047	219
Number of articles containing *in other words*	120	133	50	81
Percentage of articles containing *in other words*	6.7%	54.7%	2.4%	36.9%
Number of articles containing ≥ 4 instances of *in other words*	4	13	0	7

Table 8.3 Articles with ≥ 4 instances of *in other words* in AEE.

Code	*Title*	*Comment*
AEE 2004 Pontius	Detecting important categorical land changes while accounting for persistence	This article comes from a special issue that '[incorporates] landscape ecological and fragmentation analyses within remote sensing science'. The article abstract identifies a problem – 'many scientists fail to analyze the matrix according to its various components and thus fail to gain as much insight as possible' – and a solution – 'This paper examines the cross-tabulation matrix to assess the total change of land categories according to two pairs of components'. The article is positioned as introducing a novel type of modelling: 'An example of change . . . illustrates the methods. These methods enable scientists to focus on the strongest signals'.
AEE 2004 Bruijnzeel	Hydrological functions of tropical forests: not seeing the soil for the trees?	This article also comes from a special issue whose editorial promises that its articles 'bridge the gap' between research and practice. Having said that, there is no further evidence from the abstract that the article is positioned as applied research.
AEE 2007 Andrieu	Relationships between diversity of grassland vegetation, field characteristics and land use management practices assessed at the farm level	The abstract for this article is notable for its apparent mixture of methodologies. On the one hand there is statistical modelling: 'Regression analyses were performed to analyze the relations between the grassland vegetation types . . . , management practices . . . and the topographic . . . and topologic . . . characteristics of the fields'. On the other hand farmers have been interviewed: 'farmers' statements showed that the grazing and cutting management rules were mostly determined by the slope of the fields and the distance from the cowshed and, to a lesser extent, by the altitude and aspect'.
AEE 2009 Pollini	Agroforestry and the search for alternatives to slash-and-burn cultivation: from technological optimism to a political economy of deforestation	This is an article whose abstract declares a reflexive attitude to research approaches. The aim of the article is to study a completed 'multidisciplinary' research effort aimed changing farmers' attitudes to slash-and-burn. It is critical of natural science researchers: 'an excessive faith in the positivist paradigm of Western science maintained the illusion that perfect biophysical solutions could be designed', and concludes that there is a need to '[pay] more attention to disciplines that employ the narrative mode to depict realities . . . '.

8.3 A qualitative study of *in other words*

The qualitative study of *in other words* begins with 20 random lines from each of the BEE4 journals, studied in expanded context. The aim is to identify the range of uses of the phrase and to see what is involved in making that identification. One thing that becomes apparent is that in order to see what rhetorical work *in other words* is doing, an expansion of the concordance line is necessary. Often, looking back several sentences is necessary to specify the function. This is because *in other words* sometimes encapsulates (Sinclair 1993) quite a long stretch of text.

In the broadest terms, two uses of *in other words* can be identified. One is the canonical paraphrase or restatement, in which something is expressed first in one way, then another. Example 4 illustrates this. The phrase *the most direct transportation distances* paraphrases or explains the phrase *as the crow flies*.

4 In this Figure, distances . . . have been measured 'as the crow flies'; **in other words**, the most direct transportation distances.

[AEE 2003 Cowell]

The use of this paraphrase suggests that the reader may experience some difficulty in understanding the first formulation. In example 4, it is a vernacular idiom (*as the crow flies*) that is restated, quite possibly for an international audience. (A pedant might object that *as the crow flies* does not in fact mean 'the most direct transportation distances' but rather 'in a straight line, and therefore possibly shorter than the most direct transportation distances'!)

Terminology may be restated in this way, but restatement is more frequently used to paraphrase questions, accounts of method, descriptions of models, arguments and so on. Examples 5–7 illustrate some of these uses: a research question in example 5, a comment on method in example 6 and part of a model in example 7.

5 An important issue of the transfer of durable resistances to other species is, whether the durability of the resistance is also transferred. **In other words** is the durability an intrinsic characteristic of the resistance or is it also affected by other host factor . . .

[PS 2004 Lanfermeijer]

6 This is not quite as big a problem as it seems at first; since the observations are 5 or 10 years apart each data point embodies a lot of unique information. (**In other words**, the data are not as correlated as annual data would be.)

[GEC 2008 Weinhold]

7 Fourth, [in our model] the representative firm . . . gains expertise . . . by using the technologies in its production process. **In other words**, we include learning-by-using.

[REE 2003 Mulder]

It might be expected that a restatement using *in other words* would paraphrase a complex idea or a technical statement into a simpler one. That certainly does occur, as illustrated in example 8. However, perhaps surprisingly, the opposite also happens, and a simple statement is reformulated in longer phrases or more technical terms, as in examples 9 and 10.

8 Further, ISO 14001 provides an external benefit. . . . Therefore, firms adopt ISO 14001 when their expected long-term profit with certification is greater than without. **In other words**, the benefits of ISO 14001 adoption outweigh the cost.

[REE 2010 Nishitani]

9 Consequently, it is worth studying the impact of a better technology, or, **in other words**, an increase in the technology parameter EQSYM.

[REE 2010 Jiang]

10 These results suggested that grain yield increase might be genetically explained by either biomass enhancement of harvest index improvement. **In other words**, GY QTLS could be resolved in BM-determined QTLS, such as aGY-1–1 and qGY-3, and HI-determined QTLS like qGY-3.

[PS 2004 Zhang]

All of this suggests that the restatement use of *in other words* manages an accommodation between writer and reader, in which the writer takes responsibility for the reader's understanding of what is being said. The rephrasing of the statement, or question, offers the reader two opportunities to follow the argument. The rationale for this may simply be that the idea to be explained is perceived as complex, but there are also signs of audience design, as in example 4 where the rephrasing of *as the crow flies* suggests an awareness of an international readership. There is also the opportunity to address two distinct audiences. This point is of relevance to writing for an interdisciplinary journal, and we shall return to it.

We now turn to the second use of *in other words*. In the first use, discussed previously, the restatement does not move the argument forward, but circles the same ground as it were. In the second use, what follows *in other words* is an extension of a previous statement. Frequently found instances include an interpretation of results, as shown in example 11, and a next stage or upshot of an argument, as shown in example 12. In these examples, *in other words* might be paraphrased as 'what we conclude from this is . . . '.

11 An analysis of four suspension cultures did not reveal any relationship between the cytogenetic state of cell strains and their morphogenetic potential. **In other words**, diploidy cannot guarantee high morphogenetic potential.

[PS 2003 Zoriniants]

12 If the disturbances are small, nature will return to equilibrium. However, as soon as a threshold is passed, disturbances pose a threat to the functioning of nature. **In other words**, nature can to a certain extent be controlled.
[GEC 1996 van Asselt]

Why a phrase that might be thought to mean 'I am going to say the same thing in different words' is used to offer an interpretation or conclusion to what has gone before is a matter of speculation. It also raises the question of whether audience design is an issue here as it was with the restatement uses. What would motivate using *in other words* rather than *therefore* or *in short* in such context? There is no definitive answer to this, but it may be that using *in other words*, in preference to alternative such *therefore* or *in short*, frames what follows as an obvious consequence of what has gone before. This in turn construes the information as shared ground.

An important point to be made here is that in many instances the precise meaning of *in other words* is difficult to establish. What follows *in other words* may both carry the argument forward and also reframe that argument in different terminology. In example 13, for instance, there are two key propositions: 'gophers build new burrows based on where existing burrows are' and 'gopher distribution is influenced by intrinsic as well as extrinsic factors'. These propositions are linked by *in other words*. Arguably, both example 13a – a restatement – and example 13b – an extension – are possible paraphrases. Either way, 'intrinsic as well as extrinsic' is a neat summary of the argument that packages the finding in a way that would be accessible to someone who had not followed the preceding argument in detail.

13 After the first year of invasion, the numerical distribution became dependent on the previous distribution during the early phase of the invasion. The locations of new burrows (home ranges) becomes [sic] influenced more by existing burrows than by the spatial pattern of preferred resources. **In other words**, as the invasion progressed beyond the field edge and first few interior clusters, the gopher distribution in that field came to be influenced by factors that were intrinsic as well as extrinsic to the population.
[AEE 2001 Smallwood]

13a Gophers build new burrows based on where existing burrows are. **To say this another way**, gopher distribution is influenced by intrinsic as well as extrinsic factors.
13b Gophers build new burrows based on where existing burrows are. **We conclude from this** that gopher distribution is influenced by intrinsic as well as extrinsic factors.

This essential indeterminacy of meaning makes *in other words* a particularly interesting phrase to study. It also means, however, that its various functions are difficult to quantify, and the phrase is best examined qualitatively.

Finally, it might be argued that adding information with *in other words* performs an additional function of emphasising as well as explaining. Saying something twice makes it less likely to be overlooked. This is illustrated in example 14. In this example there are three paraphrases: *mediating effect* is rephrased as *influenced by*; *social welfare policy* is paraphrased as *distributive policies*; and *environmental outcome of economic growth* is rephrased as *effects of economic expansion on the environment*. It could be said that little is added or even explained by these rephrasings, which seem to be of an equivalent level of technicality, and therefore that the purpose of stating the question twice is simply to emphasise its importance.

14 An important question is whether social welfare policy may have an important mediating effect on environmental outcome of economic growth. **In other words**, to what degree, if any, are the overall effects of economic expansion on the environment influenced by a country's distributive policies?
[GEC 2001 Koop]

To summarise: *in other words* serves a variety of functions that are not reliably distinguishable. Each use of the phrase may be multi-functional. In all cases, however, the phrase affords the writer an opportunity to accommodate to reader expectation and needs. For this reason it might be expected to be part of the writer's repertoire when writing for an interdisciplinary journal. We explore this idea further by looking at case studies in the next section.

8.4 *In other words*: case studies

In a series of case studies (one for each journal), we explore how writers might use *in other words* to position the reader and so construe their own and the reader's identity. Although this work is not quantitative, examples have been chosen that scrutiny of concordance lines has identified as typical of each journal. We begin with the monodisciplinary journals (PS and REE) and continue with the interdisciplinary ones (AEE and GEC).

8.4.1 Drawing a conclusion from a result: PS

Example 15 exemplifies a frequent use of *in other words* in PS: drawing a conclusion from a set of experimental results.

15 These results suggest that the leaves absorb waves with high frequencies during waves passage, even more than in case of polymers (references). **In other words**, the leaves can transmit waves with relatively low frequency only, because of their filtering effect.
[PS 2002 Fukuhara]

The title of the article is 'Acoustic characteristics of botanical leaves using ultrasonic transmission waves'. The article concludes that acoustic measurements vary

according to the internal structure of the leaf, thereby revealing 'the fundamental characteristics of living leaves'. The research gap identified in the introduction is that applying ultrasonic techniques in agricultural and plant science has rarely been done. The article does not give an indication as to why the research is important or what application it might have. This should not be taken as a comment that the research is unimportant, rather it might be concluded that to a subject expert would not need to be told these things. Example 15 is the only instance of *in other words* in the article. This occurs in the section headed Results and Discussions. The section has a recursive structure of 'presentation of result' + 'conclusion from result'. Example 15 comes from the first of a number of sub-sections within Results and Discussions. This sub-section gives overall profiles of the various leaves tested (gum tree, vine, violet etc.) before more detailed results are given.

The phrase *in other words* in example 15 links two statements – *leaves absorb waves with high frequencies* and *leaves transmit waves with relatively low frequency only* – that are the obverse of each other, based on the antonomy of *absorb-transmit* and *high-low*. It becomes clear that what matters in the study is the rate at which acoustic waves pass through leaves. The result in example 15 is expressed first as a condition of absorption and then in terms of what matters more: the condition of transmission. The baseline result of the experiment is emphasised by being stated twice. The movement is not from more to less complex but from less to more significant. This 'upshot of result' use of *in other words* is typical of PS usage and is found also in the other journals, especially AEE.

8.4.2 Summarising a model: REE

Example 16 exemplifies a use of *in other words* that is typical of REE: summarising the outcome of a model. Often a great deal of text is encapsulated by *in other words*. Example 16 gives an instance where ten sentences are encapsulated; extracts only are given in the example.

16 Fig 1 shows the decision of the Northern firm on location. . . . The figure shows a case, where $\Gamma = 0.05$. A higher fixed cost of relocation simply expands the export region. . . . A larger e_0 shifts the cone . . . to the left, while a smaller e_0 lowers it to the right. . . . Lower pollution intensity also tilts the *t* curve downwards. . . . **In other words**, the more pollution intensive is the industry, the higher is the green tariff required to block delocation.
[REE 2007 Naghavi]

The title of the article is 'Can R&D-inducing green tariffs replace international environmental regulations?' The article discusses the question of which mechanism is best for ensuring that businesses minimise harm to the environment and that businesses from the global North do not take the decision to relocate their activities to the global South to avoid environmental regulation. The introduction explicitly refers to 'two important branches of literature in environmental economics' that study location decisions on the one hand and research and development

related to pollution control on the other. The article promises innovation in developing a framework or model that will connect the two.

Example 16 comes from the 'results' section of article, which shows what happens under various conditions specified by the proposed model. The figure referred to in example 16 is part of this framework and demonstrates how businesses are predicted to take decisions relating to location, based on the intersection of tariff and industry type. The description of the figure is extensive (ten sentences) and highly technical. The sentence introduced by *in other words* summarises the result in terms that are simple and are related to the significance of the result for the 'real world' rather than the model world. As with the PS example, the sentence is explanatory of results, but it greatly simplifies the result. It would be intelligible to a policy maker, for example. *In other words* might therefore be said to introduce an encapsulation and simplification of the research, as well as a statement of its applicability.

8.4.3 Reference to interdisciplinarity: AEE

The journal AEE has several 'typical' uses of *in other words*. The first, shown in example 17, shows a summary or conclusion of an extended account of results. This is similar to example 16, and might be described as introducing the 'takeaway message' of the research article.

17 In this study we can conclude that there is a systematic transition from Pasture/scrubland to Vineyard as Vineyard was systematically gaining from Pasture/scrubland and at the same time Pasture/scrubland was also systematically losing to Vineyard. The same was the case with the transition of Vineyard to Pasture/scrubland. **In other words**, there are tendencies for systematic exchanges between vineyard and pasture.

[AEE 2010 Manandhar]

The second is the somewhat curious case of a simple statement being reformulated in lengthier and more technical terms, as in example 18.

18 What is required is a discussion process at a broad social scale to define what the society wants future landscapes to look like, and what "services" they should provide us with. **In other words**, a vision is needed which includes both spatial and temporal scales, an inherent flexibility to deal with new opportunities and changes in public demands, and an acceptance that . . .

[AEE 2004 Abensperg]

Interestingly, this example might be said to address two audiences, a discipline-external audience first and the discipline-internal audience second. Rather than addressing the needs or expectations of a single community, then, the AEE writer appears to be aware of two distinct audiences and addresses both directly.

The other key use of *in other words* in AAEE is the summary of an argument that makes explicit reference to the need for interdisciplinary research, as in example 19.

19 To develop environmental protection strategies, it is necessary to have among other things, a clear understanding of the relationships between the environmental variables influencing and constraining the availability of natural resources and the human population pressure in the rural system. **In other words**, it is essential to understand the rural system in its total complexity by integrating the information related to its environmental system with the information related to its socio-economic system.

[AEE 2002 Feoli]

In example 19, the factors of *environmental variables* and *human population pressure* are rephrased in terms of natural science (*environmental system*) and social science (*socio-economic system*) disciplines. The necessity and complementarity of both approaches is a key point made about interdisciplinarity in AEE.

8.4.4 Summarising an argument: GEC

In other words is used most extensively in GEC, to paraphrase questions, summarise models, to articulate the conclusion from a set of results, and to add to or summarise an argument. It is used both to move from the technical to the less technical (example 20) and to move in the reverse direction, from the less to the more technical (example 21).

20 At times, however, major disruptions to the policy equilibrium occur and regulations fundamentally change. The interesting question to ask is what factors cause these shifts. **In other words**: why do policies change?

[GEC 2008 Brunner]

21 The more community members make use of this kind of support, the stronger are their safety nets. **In other words**, they enhance their social capital by nurturing horizontal social relations (reference).

[GEC 2008 Toni]

To illustrate the key uses in GEC, examples will be taken from one article, GEC 2008 Webber, which uses the phrase five times. The title of the article is 'Pricing China's irrigation water'. One use of the phrase introduces conclusions from research results, in this case obtained through observation (example 22).

22 For example, we observed farmers in Beidong village . . . paying an average of about [amount] per year for water; but they paid another [amount] per year for diesel to pump that water . . . and another [amount] for the pumps. . . . **In other words**, estimated price elasticities are not for resource

and infrastructure fees but for the very much larger resource plus infrastructure fees and pumping costs.

[GEC 2008 Webber]

Three of the other uses summarise background information. Example 23 is one example.

23 Even the most basic right to water is the product of de facto rather than de jure arrangements (references). Indeed, a measure of the distance between the reality in China and the ideal of a functions system of water rights is the recent argument that people should have a right to participate in water resources management, a right they do not now have (references). **In other words**, there are no transparent and secure rights to water in China.

[GEC 2008 Webber]

Here the sentence introduced by *in other words* frames the reported discussion of the issue in a form that is appropriate for the argument of this article. The final example of *in other words* in the article most explicitly introduces part of the writers' own argument (example 24).

24 [These farmers] do not have the option of being more efficient users of water. There is no water storage in the system. . . . Since farmers are charged for water on an areal basis, any reductions in their use of water . . . [fail to lead to] monetary savings. **In other words** the language of blame needs to be reframed: the inefficiency lies in the system of water management.

[GEC 2008 Webber]

In this article, then, *in other words* guides the reader through an argument rather than through a report of research. Previous literature and observation are part of that argument. They are shaped into a form that contributes to the writers' own perspective, and the phrase *in other words* is key in doing this.

8.5 Conclusion

In other words is very unevenly distributed through the corpus. It is more frequent in the social science journals REE and GEC than in the natural science ones. On the other hand, it is more frequent in the interdisciplinary AEE than the monodisciplinary PS, and by some measures it is more widespread in the interdisciplinary GEC than the monodisciplinary REE. The qualitative study of the phrase has suggested that although the meaning of the phrase is not different in the four journals, its typical use is slightly different in each. In PS the phrase addresses a discipline-internal audience and underlines what is important. In REE it is outward-looking towards policy makers. In AEE it is used to address two audiences of researchers. In GEC it guides the reader through an argument.

A key outcome of this chapter, then, is that *in other words* is used to negotiate the situation where two audiences are addressed that differ in terms of the kind of experience or expertise they bring to the article. Somewhat fancifully, one might say that the writer looks in one direction to address one audience and then turns round to address the other, with the turn signalled by *in other words*. Intriguingly, the rephrasing or explanation is by no means always from the more to less expert, from the more complex to the simpler. Writers are just as likely to address the less expert audience first and then, as it were, demonstrate their expertise by rephrasing in a more technical way.

This brings us back to the question of the individual writer and the community, as mentioned in the introduction to this chapter. Research into genre (e.g. Bazerman 1988; Swales 1990) makes complementary arguments: writing is not an individual activity alone but tends to adhere to community norms; at the same time, it is not a mechanical activity driven only by convention. The individual writer and the generic conventions of the community are in constant negotiation. Hyland (2005: 13) stresses the role of the community in discourse production, and comments that metadiscourse marks 'cultural or community knowledge' as well as social relations. He also makes the point that the degree of 'knowledge overlap' between writer and reader varies according to text type: writing for oneself can assume total knowledge overlap, and will probably include little or no metadiscourse; writing for an unknown audience, as in a mail-shot, 'needs to be far more explicit', with writers drawing on various devices 'to help readers understand the material and guide their responses to it' (Hyland 2005: 12). In other words, community is a key influence on writing and individual choice is heavily constrained by community membership.

In this study of the code gloss *in other words* we observe writers for interdisciplinary journals negotiating interaction with not one but several communities. Arguably, then, the role of the individual is particularly salient, as the individual is not so much constrained by the community but construes that community as they write. Use of the code gloss seems to be a key factor in this construction.

References

Bazerman, C. (1988). *Shaping written knowledge: The genre and activity of the experimental article in Science*. Madison: University of Wisconsin Press.

Crismore, A., Markkanen, R., & Steffenson, M. (1993). Metadiscourse in persuasive writing: A study of texts written by American and Finnish university students. *Written Communication*, 10: 39–71.

Harris, Z. (1959). *Linguistic transformations for information retrieval*. Washington, DC: National Academy of Sciences–National Research Council.

Hoey, M. (2001). *Textual interaction: An introduction to written discourse analysis*. London: Routledge.

Hyland, K. (2000). *Disciplinary discourses: Social interactions in academic writing*. London: Longman.

Hyland, K. (2005). *Metadiscourse: Exploring interaction in writing*. London: Continuum.

Hyland, K. (2010). Metadiscourse: Mapping interactions in academic writing. *Nordic Journal of English Studies*, 9(2): 125–143.

Hyland, K., & Tse, P. (2004). Metadiscourse in academic writing: A reappraisal. *Applied Linguistics*, 25(2): 156–177.

Sinclair, J. (1993). Written discourse structure. In M. Hoey, G. Fox & J. Sinclair (eds.), *Techniques of description*. London: Routledge.

Swales, J. (1990). *Genre analysis: English in academic and research settings*. Cambridge: Cambridge University Press.

Thompson, G. (2001). Interaction in academic writing: Learning to argue with the reader. *Applied Linguistics*, 22(1): 58–78.

Vande Kopple, W. (1985). Some explanatory discourse on metadiscourse. *College Composition and Communication*, 36: 82–93.

9 Multidimensional Analysis

Variation between and within journals

9.1 Introduction

We now move from predominantly qualititative data analytical methods to the use of primarily computational approaches to big data. In previous chapters, we have chiefly looked at evidence of how the four BEE4 journals differ from each other. In this chapter, we turn to the question of intra-journal variation – that is, how much linguistic variation we find within each journal, as well as how much linguistic variation there is between the journals. In Chapter 4, we observed that the section headings which are used in articles in each journal provide some evidence of structural and rhetorical variation within each journal, with PS distinguished by the almost exclusive use in each article of the rhetorical model of 'Introduction – Methods – Results – Discussion' (IMRD) for research article organisation, while in GEC, at the other end of the spectrum, only 7% of the first-level headings followed this pattern. It is important, however, to bear in mind that conventionalisation of the headings does not necessarily indicate that the running text that follows in each section is also highly conventionalized. As we have seen in Chapter 5, there can be differences between Introduction sections despite the shared form of the heading, and for that reason we need to extend our analysis to the linguistic features that are characteristic of the running text in articles as a whole.

BEE4 contains 4524 texts and approximately 19.2 million words. To handle such a large quantity of data, it is necessary to use a data reduction approach to analysis. The approach that we take here is to apply Multidimensional Analysis (Biber 1988, 1995, 2006) in order, first, to characterise all the texts in the four journals according to their linguistic features and then, second, to see what happens if we cluster the texts, journal by journal, in groups that share similarity in their linguistic profiles. This process allows us to make comparisons between the four journals (posing the question 'what is the overall linguistic profile of Journal A compared to Journal B?') and also to look at variation within each journal ('if we place the texts in Journal A into six different clusters, how similar/different are the six clusters to each other?'). In respect of the latter, our expectation is that there will be greater degrees of variation within the interdisciplinary journals, but that this variation will be greater in the journal where there is distance between the disciplines involved (that is, in GEC) as contrasted with the journal where the

disciplines are closer (that is, AEE). In addition to the consideration of variation in the four journals, though, this chapter is also an exploration of a corpus methodology, that of using MD analysis as a means to detect similarities of texts on the basis of shared linguistic features.

Multidimensional Analysis (MDA) is an approach that was developed by Biber (1988) to describe the characteristic features of different registers by quantifying the occurrences of a wide range of linguistic features within a corpus. In his 1988 study, Biber examined a variety of spoken and written registers of general English, such as 'telephone conversations' or 'planned speeches' for spoken discourse, or 'academic prose' or 'press reportage' for written, and he proposed six dimensions of variation:

Dimension 1: Involved vs. informational production
Dimension 2: Narrative vs. non-narrative concern
Dimension 3: Explicit vs. situation-dependent reference
Dimension 4: Overt expression of persuasion
Dimension 5: Abstract vs. non-abstract information
Dimension 6: On-line informational elaboration

Variation is captured by the location of registers at different points on a scale, running from positive to negative mean scores on each dimension, with telephone conversations, for example, very high on D1 (involved), negative on D2, D3, D5 and D6 (non-narrative, situation-dependent, non-abstract) while academic prose is strongly negative on D1 (therefore informational), slightly negative on D2 (tending to non-narrative), but positive on D3 and D5 (explicit in reference, and conveying abstract information).

Conrad and Biber (2001) observe that there are two categories of MD analysis: those that analyse register using Biber's (1988) dimensions and those that create new dimensions to explain specific registers or genres (Conrad & Biber 2001). In our work, we follow the second approach to MD analysis, and we have used MDA in a largely bottom-up manner rather than the top-down manner more commonly used in register analysis. The latter typically starts from distinctions between different categories of texts (e.g. spoken versus written, academic versus fiction and so on) and then uses the analyses to determine a set of language dimensions. Each register is then placed on every dimension in the set. By contrast, we use MD analysis here to place texts within a sub-corpus into groups, with the linguistic features used as the measures by which we cluster texts into groups rather than used as the means to say why one group differs from another. Once the groups have been formed, they then require interpretation in order to determine what each group represents.

A second difference between Biber's original study and our study here is that we are working with a specialised corpus, whereas Biber used a general corpus. His 1988 model was developed for the purpose of describing variation across speech and writing in general English language use, and it was derived from analysis of data taken from two general language corpora, the Brown corpus (written data)

and the London-Lund corpus (spoken language data). As his corpus contained a broad range of registers, the degrees of variation are quite marked – that is, the linguistic features of personal letters are markedly different from those of, say, scientific texts. In the BEE4 corpus, however, the degrees of variation are much narrower. All of the journal articles come under the heading of 'academic prose', which is a single register in Biber's study, and we aim to identify dimensions that are particular to academic prose in four environmental science journals, rather than using dimensions that span several registers. In this, we follow other researchers such as Friginal (2009) and Gray (2015) who have created MD models from their own specialised corpora to explore variation within, respectively, call-centers and journal articles in six disciplines.

Since Biber's original study, he, his colleagues and students and many others have applied and extended MD analysis in several spheres. In studies of academic discourse, Biber (2006) is the most comprehensive, examining language registers in university at large, using the TOEFL 2000 Spoken and Written Academic Language (T2K-SWAL) corpus sponsored by ETS and describing language use spoken and written university registers, both formal and informal, across the disciplines and levels. Studies that have focused on particular writers or genres include Hardy and Römer (2013), Hardy and Friginal (2016) and Gardner, Nesi and Biber (2018) employing MD analysis to describe university student genres of writing; Conrad (2014) comparing novice and professional engineering reports; Crosthwaite (2016) using Biber's original dimensions to assess student writing development longitudinally; Biber, Gray and Staples (2016) investigating ESL written exam data; Gray (2015) exploring disciplinary writing in six disciplines and three different types of research (theoretical, qualitative and quantitative).

The approach that we take presents challenges. Firstly, because there is likely to be less variation within the data than there would be in a general corpus, the dimensions will be less distinctive. If the dimensions are less distinctive, then it may be more difficult to explain the differences in profiles in meaningful ways. The differences may turn out to be quite trivial or they may not be easily explicable. Secondly, there are potentially large numbers of texts in each cluster. In AEE, for example, there are 1784 texts in total, which means that if, for example, six clusters are detected within the journal, the largest cluster must contain at least 298 articles. Such a large number of articles cannot be read, so the articles have to be sampled. By what criteria can the data be sampled? Thirdly, there is a problem of interpretation: we aim to group texts into clusters, according to their linguistic profiles, and then, by examining a number of texts within that group, arrive at an interpretation of why those texts are clustered. The groups of texts may be clearly distinct (for example, one group may present highly mathematical economic models while another contains ethnographic studies of rural farming practices), but other groups may not be easily distinguishable (for example, there may be two groups of articles that are reports of laboratory experiments, all using an IDRD organisation, which appear, on the surface to be the same). In the first case, it is possible to interpret the results of the clustering analysis quite confidently, but in the second case the interpretation is less obvious.

In the following sections, we briefly explain the MD model that we developed in previous research (Thompson, Hunston, Murakami & Vajn 2017), and then consider what the model reveals about variation between the journals in BEE4. The next stage of the chapter addresses the question of variation within each journal by looking at the results of clustering the texts by their MD linguistic profiles (what we term 'constellations'). This use of MD analysis, for the purpose of clustering texts in a corpus, is novel. In this sense, the chapter functions not only as a source of insight into variation between and within journals but also an evaluation of the potential of applying MD analysis to text categorization.

9.2 Dimensions in the BEE11 corpus

This approach was first used in Thompson et al. (2017), working with the BEE11 corpus, and technical details of the procedures that were followed in order to carry out the Multidimensional Analysis are provided there. In essence, a six-dimension model was developed from the analysis of the corpus; the labels for these dimensions are listed here and then each dimension is explained in some detail:

Dimension 1 (D1): System-oriented vs. action-oriented
Dimension 2 (D2): Explicit vs. implicit argumentation
Dimension 3 (D3): Informality
Dimension 4 (D4): Conceptual discourse
Dimension 5 (D5): Text-focused vs. site-focused
Dimension 6 (D6): Non-research world vs. research world

The first dimension in any MD model is the strongest, as it is the one that contains the greatest range of linguistic features and which accounts most strongly for variance in the data. Dimension 1 in our model is distinguished by, on the one hand, the use of present tense, determiners and abstract nouns, including stance nouns, process nouns and cognitive nouns, and, on the other hand, past tense and perfect aspect verbs. The former are, within the model, positive loading features on the dimension while the latter are negative. Our readings of articles that had either highly positive or low scores on this dimension led us to interpret the dimension as distinguishing between articles in which the focus is on systems, models or abstract concepts, as opposed to articles in which the focus is on actions (what the researcher, or human agents, did at particular times). An example of the former is shown as example 1, and the latter as example 2.

1 A respondent i will agree to pay any amount below EQSYM and refuse to pay any amount that exceeds EQSYM. When the first bid belongs to his interval of acceptable values, he may accept or refuse the bid offer. Here, we do not impose a precise rule: respondents can answer yes or no with any probability when the bid offer belongs to the interval.
[REE 2007 Flachaire]

2 Crop species may have also influenced soil C and N levels through production of residues of varying quality. The residue N content did not differ between tillage regimes, and averaged 6.2, 6.2, and 12.9mgNg .−1 for sorghum, wheat, and soybean residues, respectively (reference).
[AEE 2007 Wright]

The labels 'system-oriented' and 'action-oriented' thus capture the difference between, on the one hand, articles that are not time-specific and are about abstractions and ideas rather than about actions and, on the other, those that are clearly located in time and which report on actions that have been taken.

In Dimension 2, positively loaded features include modals (possibility, prediction and necessity) and adverbs of various kinds, including stance adverbs and conjuncts. Articles that score highly positively in this dimension are notable for the degree of explicitness about the relations between propositions, and about the author's stance. For example, clauses may be explicitly linked using conjunctive adverbs. Conversely, the negatively loading feature in this dimension is the use of passive voice, which is traditionally viewed as a means of suppressing human agency. We label this dimension as a contrast between 'explicit argumentation' and 'implicit argumentation', and illustrate high and low scoring texts in examples 3 and 4.

3 Without doubt, the Review has great merit. In fact, judged from the perspective of those calling for drastic and immediate action, the Review is probably as good as it currently gets with a CBA approach to climate change.
[GEC 2007 Neumayer]

4 The deleterious effect of free radicals are quantitated by the peroxidation of lipids and formation of the oxidative products; TBARS. TBARS formation in the cotyledons was enhanced by treatment with UV-B or AAPH (Fig. 3). When they were used in combination a slightly additive effect was observed (Fig. 3).
[PS 2003 Jain]

Dimension 3, surprisingly for a corpus of research articles, contained positively loaded features that are associated with spoken, conversational English: contractions, that-deletion and personal pronouns. There are no negatively loaded features, only the absence of those features. Journals in BEE11 with a relatively high score on this dimension have articles that report surveys and interviews with participants, and many include verbatim reports of spoken discourse. The label, 'Informality', is intended to indicate that transcriptions of others' voices, typically voices of non-researchers, are a characteristic of articles that score highly on this dimension. Example 5 illustrates a high scoring article on this dimension:

5 Interviewer: Why don't chemists do the same thing?
Young physicist: Why don't they have a similar arrogant attitude [small laugh]? I don't know!
[GEC 2008 Lahsen]

Dimension 4 too is characterised by positively loading features only – namely, word length, attributive adjectives, coordinating conjunctions (connecting phrases), topical adjectives, and to- complement clauses that are controlled by stance nouns. According to Biber (1988), higher average word length in a text indicates greater informational density. Attributive adjectives, on the other hand, relate to conceptual elaboration, and topical adjectives (which are also predominantly attributive) are used for classifying concepts (for example, political, public, social, national). Phrase connectors are used by writers to list, qualify, compare and contrast. An example of a text that scores highly on this dimension (Conceptual discourse) is given in example 6, which illustrates a high density of long words, and use of attributive and topical adjectives.

6 A growing theoretical literature discusses the principal components of resilient adaptation: some degree of overproduction or excess capacity; overlapping functions; rapid flow of materials, investment and information; responsive decision-making at an appropriate subsidiary level; diversification of inputs and of the economic base; alleviation of absolute poverty; learning from past events; mobilising systems to redistribute costs including insurance; and, active experimentation and support for innovation (references).

[GEC 2005 Pelling]

In Dimension 5, the positively loaded features are communication verbs (eg, *ask*, *say*, *claim*, *discuss*) and that complement clauses controlled by a communication verb, along with that complement clauses controlled by mental, factive, or likelihood verbs, while the single negatively loaded feature is 'place noun'. We interpreted this as a contrast between articles that contain a plurality of reported voices, or multiglossia, and are thus focused on other texts (see example 7), and texts that focus on places or sites, in the physical world (example 8).

7 Representatives of the new paradigm question the validity of the traditional perception of dryland ecosystems as equilibrium systems, allowing assessment of livestock carrying capacity and thus overgrazing. They claim that such systems should rather be seen as 'event driven' non-equilibrium systems (reference).

[GEC 2001 Rasmussen]

8 Although farm territories were generally fragmented (62% of the fields are smaller than 1ha) and scattered (30% of the fields are situated at more than 1km from the farmsteads), they all had a majority of fields in only one bocage area.

[AEE 2004 Thenail]

Dimension 6 is an oddity in that consists of only two features, one positively and one negatively loaded. Thompson et al. (2017) argue that as one of the two features is nominalisations, which are known to be a distinctive feature of academic

discourse, the dimension was worth including in the MD model (the other feature was 'common nouns'). In our analysis of the BEE4 corpus that follows, however, we have decided to exclude the sixth dimension as we found that differences in scores on this dimension were often accounted for by relatively minor aspects of the phenomena that they investigated, and in many instances articles had misleading scores on the dimension; some of the nouns treated by Biber (1988) as nominalisations would not have been viewed as such by Halliday (1994, 2004).

9.3 Inter-journal variation in BEE4

Thompson et al. (2017) used this model to profile all GEC articles in the period 1991–2010, using the scores per article on each of the six dimensions. They then clustered texts that share similar dimension profiles, to create what they term 'text constellations'. Again, technical details for this can be found in Thompson et al. (2017: 166–167).

Here we look at the results of applying the MD analysis to the four journals in BEE4. As mentioned in the previous section, we have removed D6 from the analyses, on the grounds that the scores on that dimension were not interpretable. In the first stage, the results of the Multidimensional Analysis of the texts in BEE4 are discussed in order to find out what they can reveal about the differences between the four journals, and, in the second stage, texts in each journal are placed into clusters (constellations) and those are examined, to see what variation exists between texts in each journal.

For the first stage we pose the questions:

1 Do the social science journals (GEC and REE) align with each other on dimension scores? And do the natural science journals (AEE and PS) align with each other?
2 Is it possible to differentiate the interdisciplinary journals (AEE, GEC) from the monodisciplinary journals (PS, REE) by their dimension score profiles?

The first question can be answered in a straightforward manner using quantitative results, while the second question will require a more nuanced answer, as the profiles may differ by dimension.

The mean dimension scores for the four journals on each of the five dimensions are shown in Table 9.1, in numerical form, with standard deviations also reported. The standard deviation scores are useful for identifying variance within a dimension where a low standard deviation is indicative of a low degree of internal variation. Figure 9.1 shows the closeness or distance of each journal to the others on each dimension.

On the first three dimensions, the two natural science journals (AEE and PS) are close to each other, most so on D1 and D3, and the two social science journals (GEC and REE) are also close, particularly on D2.

Our interpretation of D1 is that it primarily distinguishes between, on the one hand, research activities that involve the construction and evaluation of models

Table 9.1 The mean scores for the BEE4 journals on the five dimensions, with standard deviations.

	Dimension 1	*Dimension 2*	*Dimension 3*	*Dimension 4*	*Dimension 5*
AEE	145.93 SD 44.98	34.96 SD 13.04	4.51 SD 3.20	76.39 SD 15.22	–0.18 SD 10.65
GEC	216.77 SD 33.11	55.73 SD 16.33	10.54 SD 5.93	98.83 SD 19.34	9.44 SD 11.90
PS	124.00 SD 26.74	16.87 SD 12.65	3.31 SD 1.85	69.03 SD 13.34	18.08 SD 8.77
REE	274.76 SD 34.88	61.52 SD 17.90	9.06 SD 6.13	81.56 SD 14.08	12.55 SD 14.68

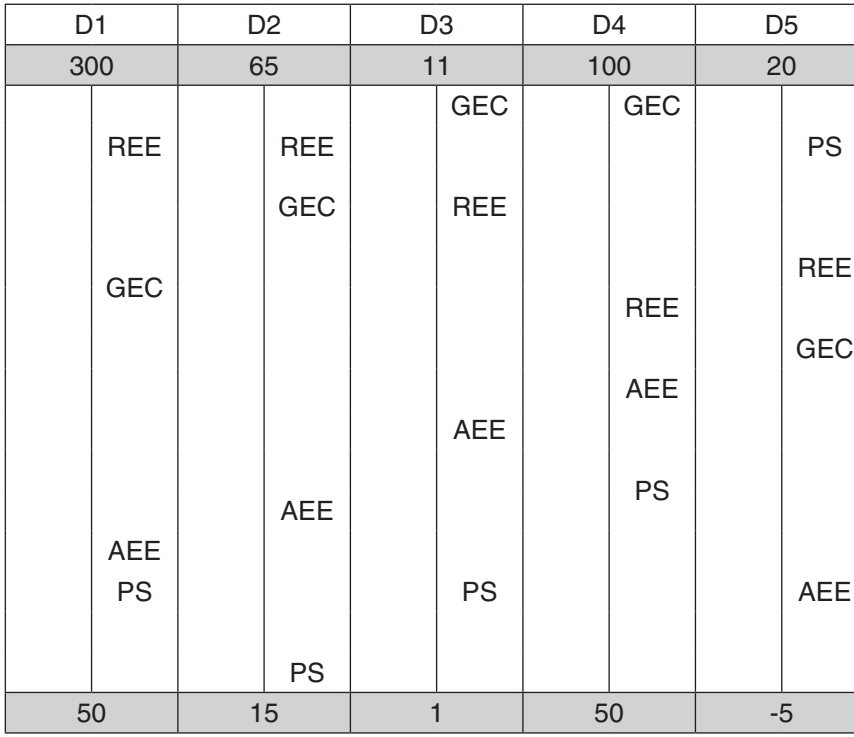

Figure 9.1 The spread of mean dimension scores for the four BEE4 journals on dimensional scales. (Note that each scale is different.)

and, on the other hand, research activities that are reported (for example, experimental field or laboratory research). AEE and PS sit together as journals in which experimental research is reported, and REE is characterised as model-oriented research. As we have observed in earlier chapters, REE is an economics journal and many of the articles in the journal are concerned with the development of

mathematical models. An example of this, characterised by the use of present tense, determiners and a process noun (approach) can be seen in example 9.

9 Note that if the discounted approach is adopted, the result can be associated with Chichilnisky's approach. According to Gerlagh and Keyser (ref), strict resource conservation induces Chichilnisky's non-dictatorship of the present. Then, maximizing the discounted sum of utility within the viability kernel should lead to a similar solution as Chichilnisky's criterion.

[REE 2007 Martinet]

GEC also contains model building articles and therefore scores highly on Dimension 1, but other types of article also appear in GEC, and thus the mean score is high but not as high as that for REE. PS by contrast is the lowest scoring of the four journals on both Dimension 1 and 2 and this can be accounted for by the observation that research articles in this journal typically are reports of experiments conducted by researchers, and this results in a higher use of past tense and also a lower use of argumentation markers (the emphasis is on report rather than discussion).

AEE is similar to PS on Dimension 1, which illustrates the prevalence of experimental research reports in the journal, but has higher scores on Dimension 2, which suggests that there may be more discussion and explicit argumentation in AEE than in PS.

As remarked previously, Dimension 3 contains features of informality. Articles which report individual human voices such as studies based on focus group or interview data tend to score highly on this dimension. As can be seen in Table 9.1, both GEC and REE are relatively high on this scale (in this case, GEC is more markedly so), while, not surprisingly, both PS and AEE are low.

The picture is not so clear on Dimension 4, and on Dimension 5 AEE and PS diverge. GEC is the highest scoring journal for Dimension 4, conceptual discourse, and a plausible explanation for this is that the journal contains many articles with discussions of national and international environmental policy. An example from GEC is example 10.

10 Global change and climate change are becoming issues of great public interest. In particular at the interface between science, policy and society new ways of inquiry and dialogue have to be developed. Traditionally, science has had great authority in defining what is a socially relevant problem and what approaches are appropriate in investigating them. Also in global change research and climate research political institutions for support and legitimacy play a significant role (reference).

[GEC 2006 Welp]

Dimension 5 contrasts a focus on places with a text focus and it is not surprising to find that AEE, which contains many experimental studies of ecosystems such as, for example, in grasslands, should score negatively on this dimension (that is, towards the site-specific end of the cline), but it is interesting that PS has the

highest mean positive score (18.08) of all four journals. This high score for PS on Dimension 5 can be accounted for by the very low scores that PS has on place nouns, the one negatively loading feature. Research in PS is very much laboratory-based and there are few references to geographical entities in the physical world such as forests, hills or streams. The lack of place nouns in the PS data thus exaggerates the effect of the positively loading features which, as can be seen in Table 9.1, are not exceptionally frequent overall.

On the basis of the dimension score profiles that are shown in Table 9.1, therefore, we can answer the two questions given previously by saying that the social science journals (GEC and REE) do align with each other on dimension scores, broadly speaking, although there is divergence on D4. The two natural science journals (AEE and PS) align with each other by comparison with GEC and REE but they diverge strongly on D5.

In answer to the second question, there is some evidence that the interdisciplinary journals (AEE, GEC) can be differentiated from the monodisciplinary journals (PS, REE) by their dimension score profiles (both GEC and AEE are less extreme than the monodisciplinary journals) but we need to look more closely at the variation within each journal in order to get a more finely detailed view.

9.4 Constellations

The dimension scores allow us to see broadly how the journals differ from each other, but in order to investigate variation within journals we need to look at similarities and differences between texts within each journal. Using the clustering approach developed in Thompson et al. (2017), we have created 'constellations' of texts for each journal that are built around clusters of texts which share similarities in their dimension score profiles. Details of the technical procedures for this can be found in Thompson et al. (2017). In simple terms, we group texts according to their scores on the five dimensions, so that, for example, articles that score high (in relation to other articles in that journal) on D1, high on D2, low on D3, and neither high nor low on the other two dimensions will cluster. It should be noted that these clusters do not create exclusive categories – while prototypical texts in each cluster may have profiles that are clearly different from the prototypes of other clusters, there are many texts that are less distinct. That said, however, the clusters can be seen to indicate tendencies in the data.

The number of clusters that are formed is not fixed; the choice of a number depends on the analyst's judgement. Conventionally, the analyst will make the judgement based on a reading of a dendrogram. However, in order to make the comparison possible between four different journals, we have chosen initially to set the number of constellations for each journal at six, as we did for the creation of constellations in Thompson et al. The decision to choose six was guided by a decision to have enough (but not too many) potentially meaningful groupings to make it possible to detect variation in each journal. It is probable that some groupings are actually similar to each other while in other cases they are clearly distinct. Having created six putative constellations, we then inspect each constellation in turn to determine whether or not the constellation is palpably distinct from the other constellations and we then make

decisions about whether to retain the six constellations or whether to merge one or more with others. This process is described in section 9.4.1. We assume here that interdisciplinary journals will have greater variation than will the monodisciplinary ones (that is, that they will have a larger number of distinct constellations), but that, given what we have observed in previous chapters, we need to compare the interdisciplinary social science journal (GEC) with the monodisciplinary social science journal (REE), and the same two journals with the two natural science journals (AEE and PS).

The creation of six putative constellations may create groupings that are notably distinct and potentially some which are not distinct. This leads us to formulate two questions about the four sets of constellations:

- What interpretations of the constellations are suggested by careful readings of examples of each constellation?
- Does the analysis suggest greater variation within the two interdisciplinary journals than within the monodisciplinary journals?

In the following sections we present a set of six box-and-whisker plots for each journal. Each box plot shown has been standardised to have the same lower and upper limits on the y-axis (–4 to 6), so that the scale is the same. The box plots show, from the upper left to the upper right, Constellations 1 to 3 (C1–C3), and on the second row, Constellations 4 to 6 (C4–C6). Each panel shows the box and whisker for five dimensions, going from D1 on the left to D5 on the right. The thick line in the middle of the box is the median, the top of the box is the third quartile, the bottom is the first quartile, and the line outside the box extends to the bottom and top values, except where these are outliers (represented by dots beyond the line).

An important point to make is that the zero point of each journal is particular to that journal, and so one cannot directly compare the scores on the plots from one journal to those for another journal. As we have seen in Table 9.1, the mean on D1 for PS is much lower than it is for REE, and so the zero point for D1 on PS constellations is a different point relative to the zero point for D1 on REE constellations. If one PS constellation has a high score on D1, this does not mean that the articles in that constellation are higher on D1 than articles in REE. In fact, the highest score on D1 for any article in PS is 155.9, and the lowest score on D1 for an article in REE is 171.1, with the highest score at 362.4. In other words, high scores on a given dimension in one journal are relative to an apparently high score on the same dimension in another journal. A constellation in PS that has a high mean on D1 will still be lower on the D1 scale than any of the REE constellations.

The value of the box-and-whisker plots is that they indicate linguistic profile variation **within** each journal. It is helpful to think of a constellation as having the potential to be a configuration for a particular kind of article. However, as the number of constellations has been arbitrarily set at six, for each journal, it is possible that the differences between some configurations (the constellations) are so small as to be of negligible significance. As a consequence, it will be necessary to merge two or more constellations to form a single constellation if examination of texts in those constellations reveals little difference between the constellations

9.4.1 Interpreting the constellations

To interpret the constellation labels, five texts from each cluster were closely read in order to determine the similarities between the texts in the constellation, and to create a label for that grouping of texts; with four journals and six constellations each, this came to a total of 120 articles. The five texts were selected on the basis of their scores, taking texts with score profiles as close to the mean score profiles for each dimension, on the grounds that they constituted prototypical members of the group.

The procedure followed was that the articles were read through once, without reference to the dimension scores, in order to establish what the article was about, and how it was structured. They were then read a second time, with specific attention paid to the linguistic features that distinguish that constellation from others. For example, if the constellation was distinguished by high scores on Dimension 2 (high compared to other article constellations in that journal) then attention was paid to the Dimension 2 linguistic features (modal auxiliary verbs, stance adverbs and conjuncts, among others).

In the following sections, we look at each journal in turn, and we begin by presenting the box-plots for the journal, and a set of initial observations that are based on comparing the box-plots. We then discuss the constellations, drawing on our close readings of the five-article sample.

9.4.2 The constellations, by journal: AEE

To illustrate how we interpret the six box-and-whisker plots in Figure 9.3, we first take the plot shown in Figure 9.2 which is the plot for Constellation

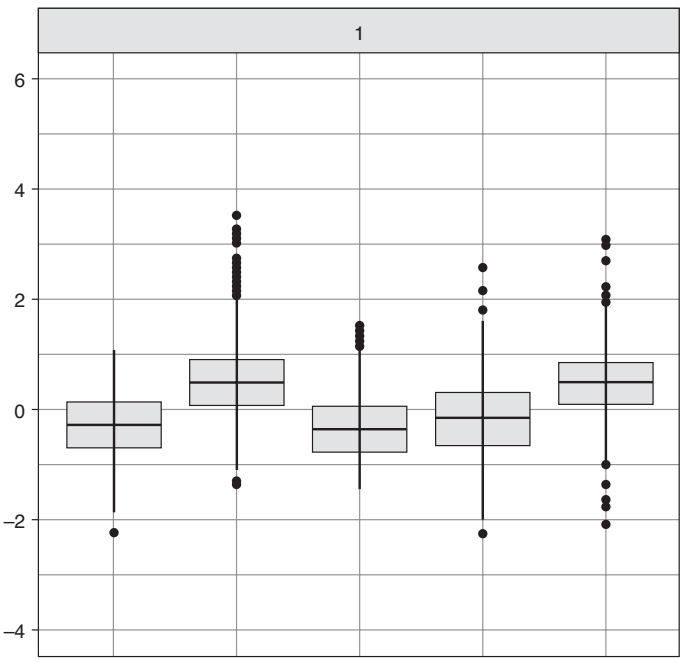

Figure 9.2 AEE six-constellation model.

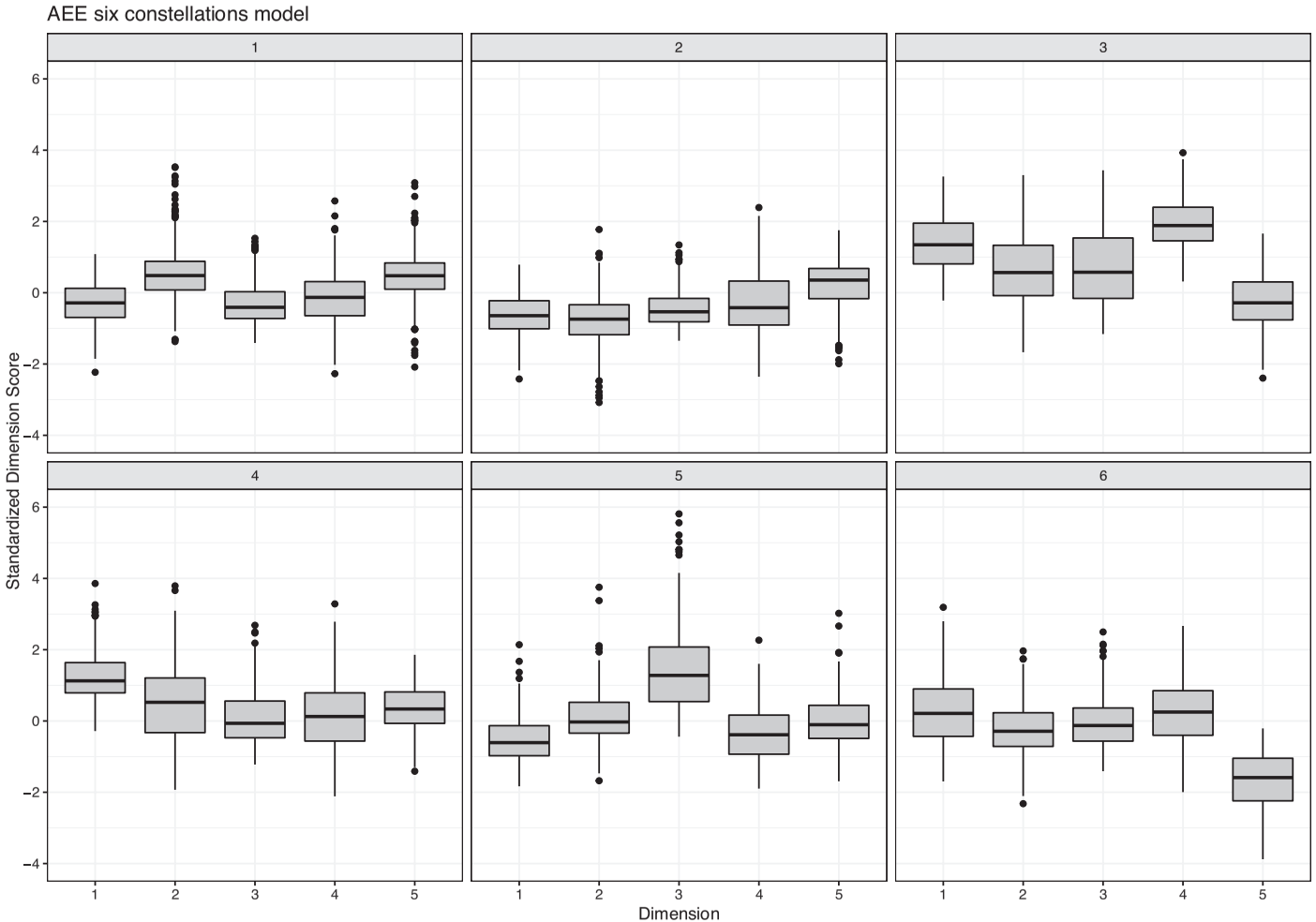

Figure 9.3 The first constellation for AEE in isolation.

1. In this figure we can see five 'boxes', with the first box on the left representing Dimension 1, the next to the right is Dimension 2, and so on, with the rightmost showing Dimension 5. Each box has a line in the middle which shows the median, with the top and bottom of the box showing the upper and lower quartile. The large dots above and below the straight vertical lines represent the outlier values. In this plot, D1 has only one outlier while D2 has several outliers above and one below. Comparing D3 and D4, we can see that the scores for D3 are more compressed with less distance between the upper and lower quartiles while D4 is more spread out and the vertical line is longer, which suggests greater variation. Across all the constellations, zero indicates the baseline and if the scores for a given dimension are high or low in any dimension, this indicates that the score on that dimension is high or low in relation to the standardized score for that journal. As we know that AEE has a low mean dimension score on D1 in comparison to REE and GEC (see Table 9.1) this means that in Constellation 1 the D1 scores are lower than normal within AEE.

Turning now to look at all the six constellations as shown in Figure 9.3, we observe that:

- C1 and C2 are similar except that C1 has a higher score on D2 (explicitness; C1 above zero, C2 below zero).
- C4 and C6 are similar except that C4 is slightly higher on D1 (system-orientedness) and C6 is lower on D5 (that implies that C6 is more site-focused).
- C3 resembles C4 on the first three dimensions, though the mean score for D3 is a little higher. Where it differs greatly is the high mean score for D4 (conceptual discourse) relative to C4.
- C5 is similar to C2 on the first two dimensions but has a markedly higher score on D3 (informality).

Our reading of the texts confirms our first observation, that Constellations 1 and 2 are broadly speaking similar. All the articles examined were field studies, which typically consist of the following elements: an introduction explaining the background and the problem, which is stated in the last sentences of the introduction; a materials and methods section that contains descriptions of the study area, the experimental design (where appropriate) and the analytical framework; results; discussion. The articles report field studies in a variety of contexts (savannah woodlands in Burkina Faso, coffee farms in Mexico, and so on), and the difference between them is that the articles in C1 have more extended discussions (hence the higher scores on explicitness). It is likely that this is usually due to the larger number of contributory or observed factors at play as can be seen in the speculative writing in example 11 or the discussion of variables in example 12.

11 Moreover, trampling by cattle and fertilisation by animal dung may have enhanced decomposition by compacting necromass and increasing its contact

with the soil, thus exposing it to a potentially more humid and microbial-rich environment.

[AEE 2005 Sawadogo]

12 Topsoil, humidity, vegetation and net profit/loss were measured on all farms, and are generally sensitive to one or more measures of agroecological resistance. However, fewer measurements make it difficult to statistically confirm clear patterns of resistance for rill and gully erosion. Landslides were a poor indicator of agroecological resistance, overall.

[AEE 2002 Holt-Gimenez]

Both Constellations 3 and 4 have strongly positive scores on D1, which indicates an orientation towards systems, and reading of the sample texts confirms this. In Constellation 4, four of the five articles construct models that can be used in assessments: AEE 2002 Arshad and AEE 2002 Nortcliff both focusing on soil quality assessment, AEE 2003 Sheldrick on soil nutrient audits, and AEE 2001 Esterhuizen, rather than building a model, reports on the use of a conceptual framework. It is interesting to note, in passing, the applied nature of these frameworks (example 13).

13 Whilst it is important to evaluate or "measure", it is also important to use this impact information in management decisions for better allocation of scarce resources.

[AEE 2001 Esterhuizen]

Constellation 3 has higher scores on D4, which suggests an orientation towards conceptual discourse, with greater use of attributive adjectives, particularly in relation to spatial and social categorization (which fits into Biber 2006's group of 'topical adjectives'). In the sample, there are articles that:

- Quantify nutrient balances in different regions of Ethiopia
  (AEE 2005 Haileslassie)

- Report a regional-scale study of faunal response to anthropogenic disturbance in Southern Africa
  (AEE 2004 Fairbanks)

- Review approaches to quantitative measurement of agricultural sustainability globally before testing one model on a dataset
  (AEE 2001 Nambiar)

These articles critique, or use, models at global, national and regional levels of analysis and are often aimed at stressing the importance of differentiating between these levels, as in example 14.

14 Our objectives were to illustrate the regional differences in nutrient depletion in Ethiopia. . . . This kind of information is critical for decision makers to

> plan and implement integrated nutrient management policies and strategies at a regional level.
>
> [AEE 2005 Haileslassie]

One article, AEE 2002 Doran, stands out in this constellation through its stylistic choices (expressions such as 'fraught with complexity'), its organisation (rather than employing an IMRD structure, it uses thematic headings such as 'Threats to global sustainability') and its detached observation of scientists in relation to farmers (example 15).

> 15 As mentioned earlier, use of a given approach for assessing or indexing soil quality is fraught with complexity and precludes its practical or meaningful use by land managers or policy makers. However, the use of simple indicators of soil quality and health which have meaning to farmers and other land managers will likely be the most fruitful means of linking science with practice in assessing the sustainability of management practices (references).
>
> [AEE 2002 Doran]

Constellation 6 resembles both C3 and C4 in its high scores on D1, and the articles in the sample are all related to modelling. C6 also has markedly low scores on D5, which indicates a focus on 'site-focused discourse'. Land is a major focus of study in AEE, and in these articles there is an even greater concern. Four of the five articles in the sample contain large numbers of 'Place nouns': AEE 2003 Sankhayan (186), AEE 2006 Rounsevell (230), AEE 2002 Smart (231), AEE 2009 Claessens (280), of which the majority are *land* and *area*. Constellation 6, therefore, is labelled 'Land use modelling', and we interpret the other two as 'Regional modelling' (Constellation 3) and 'Assessment frameworks' (Constellation 4).

Constellation 5 appears to be about 'Human voices' (the label given in Thompson et al. [2017] to the constellation distinguished by high scores on D3). Initially that interpretation is supported by readings of AEE 2005 Bardsley, which reports on open-ended interviews used to explore perceptions of in situ agrobiodiversity conservation, and AEE 2001 Goma, a participatory research study in three countries in which discussions with local peoples form part of the dataset. In these articles, human voices and human actors do play a clear role.

However, other articles do not contain evidence of human voices, and the reason for this is that there is an anomaly in D3 in relation to agricultural science. Dimensions bring together a range of linguistic features that typically co-occur in texts which load in the same orientation on that dimension. However, the relative proportions differ. Among the positively loading features on D3 are second- and third-person pronouns, nominal pronouns, contractions, that deletion, group nouns and wh- clauses. In some texts, the high incidence of one of these features can lead to a high positive score for the dimension, even though the other features

are not equally represented. An example of this for Constellation 5 is AEE 2010 Pfleeger, which contains 30 tokens in the group nouns category. The group nouns category includes words such as *bank*, *firm*, *household*, *community* and *government* which are intended to capture groups and institutions (Biber 2006). In AEE 2010 Pfleeger, however, the words *community* and *bank* are used in relation to plants ('plant communities' and 'seed banks'). The words have been treated in the MD analysis as terms that describe a group of humans but, in these instances, this is a term which can be used with a different sense in a specialised discourse. While there is evidence to support an interpretation that texts in this constellation present research that encompasses human voices, there are also some texts that are included by mistake, and not all the articles that appear in this constellation actually fit here.

In summary, for AEE, the six putative constellations have been examined and that number has been reduced to five groups:

1 Field studies (the original C1 and 2)
2 Regional modelling (C3)
3 Assessment frameworks (C4)
4 Research presenting human voices and human actors (C5)
5 Land use modelling (C6)

9.4.3 The constellations, by journal: GEC

We start with initial observations (see Figure 9.4):

- C1 and C5 are similar on D1 and D2 but differ greatly on D3 (informality) – C5 has the highest score for this dimension of all the constellations.
- C2 and C4 have similar scores for D1 and D2, but C2 has a higher score on D4 (conceptual discourse) and C4 is higher on D1.
- In C3 all the dimensions have a median below zero.
- C6 has markedly positive scores on both D1 and D2, by comparison with the other constellations.

Thompson et al. (2017) described six constellations for GEC. We have to look at the journal afresh, however, for two reasons: in this case, we are looking at the ten-year period 2001–2010 rather than the 20-year period examined in Thompson et al. (2017); secondly, D6 has been removed from the calculations of the constellations. In spite of this, we find that the six constellations that we identify here are broadly the same as those identified in the earlier publication.

Constellation 1, which we label 'Narrative', contains articles that deal with complex systems and situations with a range of factors, and that review past and present practice in order to build models for the future, or for use elsewhere. The low constellation scores for D1 indicate that past tense is used and that there is attention to past actions or events. GEC 2008 Urry, for example, develops a framework based on complex systems theory, and draws on past

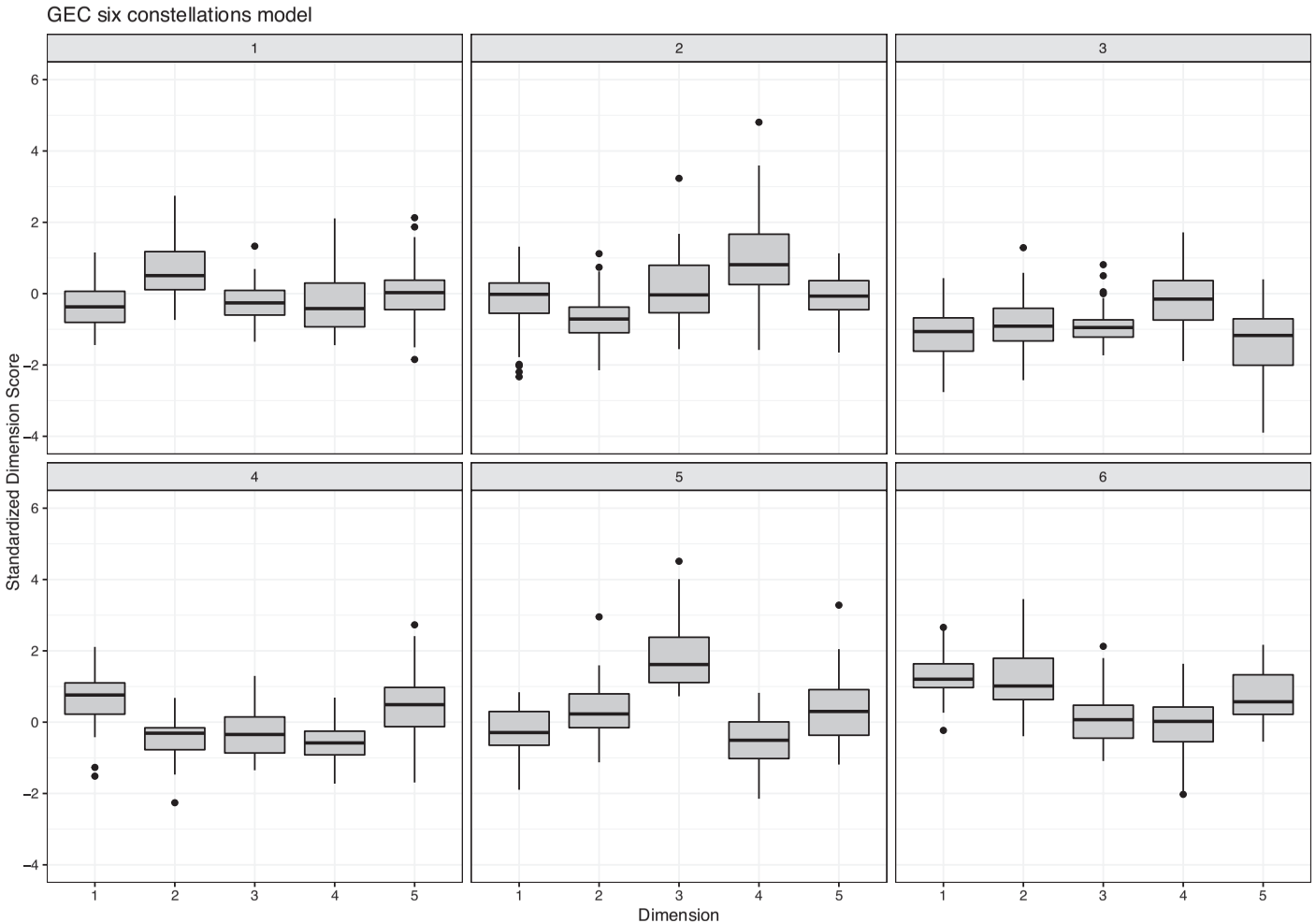

Figure 9.4 GEC six-constellation model.

and present in order to predict an automobility system of the future, as in example 16.

16 One billion petroleum-fueled cars were manufactured during the last century, with now over 600 million cars roaming the world. World car travel is predicted to triple between 1990 and 2050 (reference).

[GEC 2008 Urry]

GEC 2009 Fraser examines historic socio-economic and policy factors that have influenced land use in southern Romania before presenting a framework that can be applied to other areas. GEC 2007 Boyd reviews issues that environmental managers have to consider when adapting small-scale projects to the governance and livelihood conditions of low-income communities and uses historical examples to illustrate points (example 17).

17 This was shown in the experience of Fondo Bioclimatico in Mexico. The project started with the clear purpose of incorporating carbon as a development strategy, but as the project grew the objective moved away from development to a much more limited market aim.

[GEC 2007 Boyd]

The articles in Constellation 2 deal predominantly with social and contextual variables. Amongst the types of research presented are surveys and case studies and what is distinctive is the use of categories such as *social*, *regional*, *global*, and abstract concepts such as *vulnerability*, *responsibility*, *capability* as in example 18.

18 This paper has presented a robust assessment of the implications of climate change on international inequity. . . . To this end, two responsibility indicators, two capability indicators, and 33 sector-specific vulnerability indicators were selected based on their relevance, significance, scientific credibility, and public availability.

[GEC 2010 Fussel]

The articles may make comparisons between different levels (the local, the regional, the national, and so on). GEC 2008 Armitage constructs a theory of learning and management in a range of environmental contexts, and illustrates this with reference to five specific contexts. Of one, he states (example 19):

19 Perhaps the most important lessons are that not all resource management experiences can be transferred, and that learning may be very place and time-specific, especially in complex social – ecological systems. Expanding the spatial scope of activities in the lagoon also highlights how larger-scale resource systems are likely to be characterized by a greater degree of ecological complexity, exacerbating the challenges of a social learning process.

[GEC 2008 Armitage]

Constellation 3 is perhaps the most interesting of the six in that it resists easy categorization. The profile suggests articles that tend towards action-orientedness, and some conceptual discourse but there is no strong characteristic. The constellation is closest in profile to Constellation 4, 'Researching people', in Thompson et al. (2017), but the articles that were sampled here do not support this interpretation. GEC 2006 Ewers, for example, creates a model for measuring the relationship between forest cover and economic development, and GEC 2007 Nogues-Bravo uses various datasets to assess the expected change in surface temperature of 13 mountain systems in the world. Neither article has a focus on researching people. No clear similarity between the articles emerged from the reading, and we suggest that there is little justification for treating this as a distinct constellation, which in turn implies that a five-constellation model is appropriate for GEC.

Constellation 4 ('Modelling') concerns the development of conceptual or analytical frameworks and their application, in which attention is mainly focused on the approach/technique/theory and not on the human voice. GEC 2007 van Vuuren, for example, demonstrates approaches to downscaling and then applies the algorithms to a particular case. GEC 2009 Strassburg is an economics article that proposes a mechanism to guide policy on deforestation and then tests it through a simulation. This constellation has a high score on D1, and thus a greater concern with systems than actions.

Constellation 5 (Human voices) is distinguished by high scores on D3. Examples of articles in this constellation are a survey-based examination of social resilience to climate variability (GEC 2010 Marshall), a focus group study of climate mitigation measures denial (GEC 2001 Stoll-Kleemann), and a study of rural household demographics (GEC 2008 de Sherbinin) which features a list of research questions and quoted responses. These are not the only articles in GEC that report surveys; GEC 2010 Burch, for example, in Constellation 4 used semi-structured interviews, but the difference is that Burch does not use quotations and instead focuses on the issues raised and discussed, without bringing in human voices.

The final constellation, 6, is 'Review and theorisation'. It is high on D1 and D2, which indicates a concern with system and with explicitness. Examples of articles in this constellation are GEC 2004 Tol, which reviews assessments of impacts of climate change, asking what scenarios, data and methods are used, without applying any of these, and GEC 2006 Welp, which extensively reviews different theoretical frameworks before elaborating one of these, to construct a model of learning. Where Constellation 4 features articles that develop models and then applies them, in Constellation 6 the model may not be tested.

In summary, for GEC, six constellations have been examined and that number has been reduced to five groups, with the articles in C3 dispersed among the five remaining groups:

1 Narrative (C1)
2 Contextual variables (C2)
3 Modelling (C4)
4 Human voices (C5)
5 Review and theorisation (C6)

9.4.4 The constellations, by journal: PS

We start with initial observations (see Figure 9.5):

- C1 and C2 are similar in profile with C2 having slightly higher scores on D4 and D5.
- C4 and C6 are similar but C6 scores higher on D4 (conceptual discourse).
- C5 is the most notable with remarkably positive scores on D1 and D2. However, it should also be remembered that PS has generally low scores on D1 and D2 and so this profile is still low compared to the other journals.
- C3 has a potentially distinctive set of scores on D3 (informality).

The constellations in PS are far more difficult to interpret. Typically articles in PS consist of an introduction in which the 'real-world' value of the plant or plant process under study is explained, the current state of knowledge, or similar research, is presented and then the study is introduced, and then, optionally, the achievement of the study is explained. Next, the materials and methods used in the study are stated, with information about the sources given. The results are then presented in brief, followed by a discussion of the results. It is not uncommon for results and discussion to be collapsed into one section. The following are examples taken from the texts that were read for interpretation of the constellations:

Introduction (examples 20–22):

20 Plum is one of the most important fruit crops in China.
[PS 2009 Luo]

21 The effects of 1-MCP in delaying ripening have been studied on several fruits including avocado [4,5], apple [6], pear [7], plum [8,9], and guava [10,11].
[PS 2009 Luo]

22 This study provided basic information toward our understanding of the role of BRX-like genes in rice.
[PS 2010 Liu]

Materials and methods (examples 23–24):

23 In brief, the callus of V. vinifera L. cv. Cabernet Sauvignon was derived from anther tissue cultured on callus induction medium (MS, 0.25% (w/v) Gelrite, 3% (w/v) sucrose, 2μg/ml 2,4-Dichlorophenoxyacetic acid, 0.2μg/ml 6-benzyladenine, 5μg/ml AgNO3, pH 6.2) in Petri dishes at 25°C in the dark.
[PS 2009 Zhang]

24 Most of the purification and recombinant DNA procedures were performed according to the methods described in Sambrook et al. [17].
[PS 2002 Pandey]

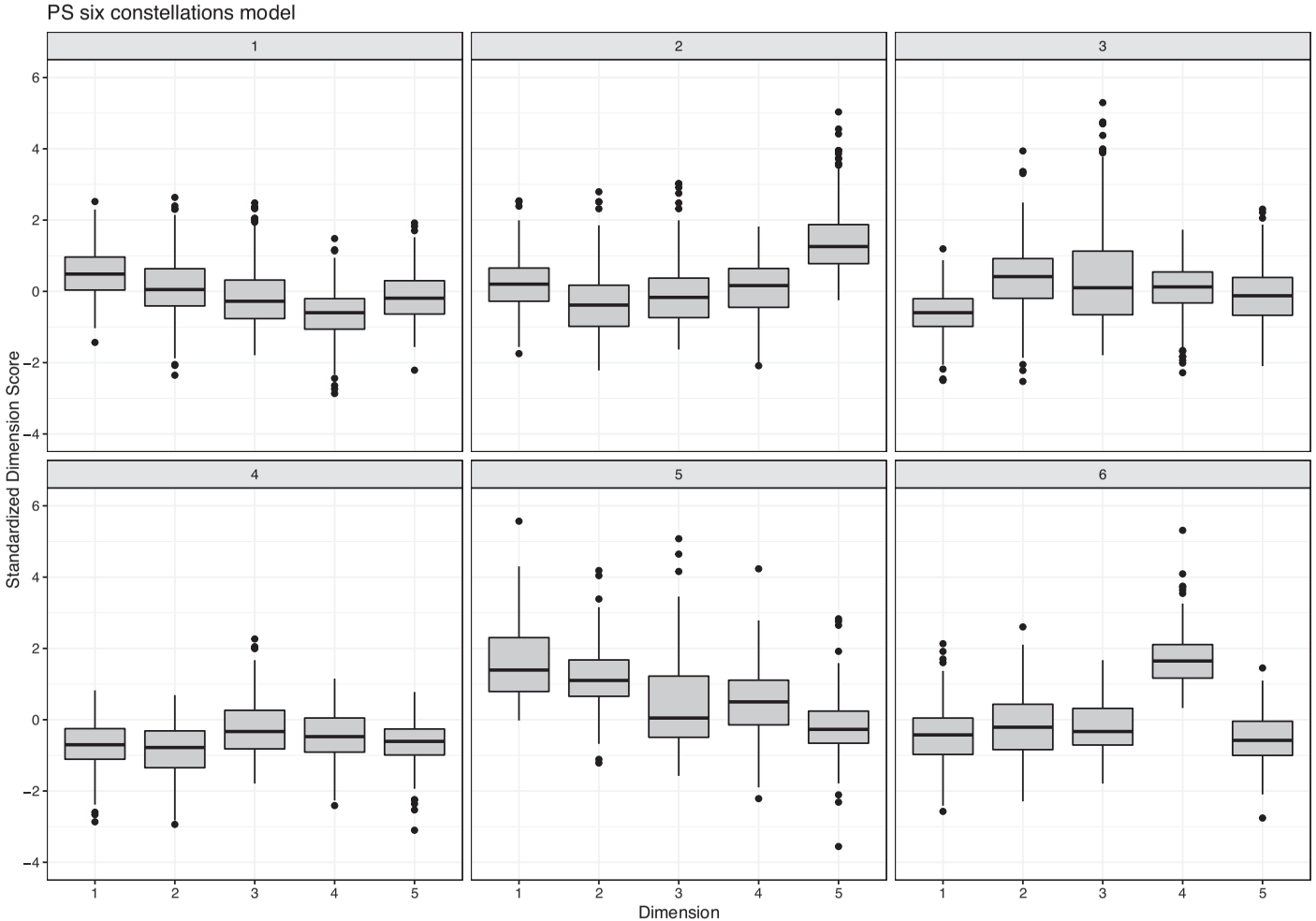

Figure 9.5 PS six-constellation model.

Results (example 25):

25 Compared to the control, berries from virus-infected clone 314 had significantly lower levels of 3-O-monoglucosides in the berry skin during berry maturation and ripening (Figs. 4–8) except at SV Site 2, where the levels of these anthocyanins in the berry skin were similar in virus-infected clone 314 and control. [PS 2008 Singh Brar]

Discussion (example 26);

26 [O]ur results show . . . [a] possible explanation could be that the phenylpropanoid pathway was repressed using ascorbic acid (5mg/L) and, after few days, this compound might be limited in the high-rate propagation using TIBs. [PS 2008 Arencibia]

Readings of the sample articles in each putative constellation resulted in the identification of only one grouping that was clearly different from the others. This was Constellation 5, which we have labelled 'Focus on systems'. The articles in this constellation are characterised by higher use of present tense (D1) where the authors posit statements about how things work, about processes and about interrelationships. The articles may be predominantly theory-oriented or they may feature lengthy discussions of theory (by comparison with other articles in PS) before testing the theory in practice. PS 2008 McClellan, for example, is a review article that looks at modes of regulation that control hormone responses in plant systems, while PS 2001 Shvetsova looks at interactions between plant cells and the environment.

Constellation 3 contained articles which were stylistically distinct in that some authors used a more elaborate and personalized style (PS 2008 Breton uses several adverbs such as 'indubitably', 'clearly'; PS 2008 Singh Brar uses 'we' frequently in the discussion, in addition to 'seems that', 'it is remarkable that'), but there is also a review article in this constellation (PS 2008 Fukao, on the mechanism 'underlying physiological and molecular regulation of ethylene-mediated acclimation responses to submergence in rice') which does not have much in common with the other articles. As there is no clear core around which the articles in the constellation cluster, it is not given a label.

In conclusion, only one distinct constellation in this journal has been identified (C5, 'Focus on systems'), and the other five will therefore be treated as forming the other constellation, which we label with the catch-all 'Experimental research'. It may seem strange that the most focused of our journals offers the most challenges in interpretation, but this is understandable in the light of the standard deviation scores for PS as shown in Table 9.1, which are the lowest on all five dimensions (26.74, 12.65, 1.85, 13.34, 5.74), indicating that there is less variation between articles within PS than is the case with the other three journals.

1 Focus on systems (the original C5)
2 Experimental research (the original C1–C4, C6)

9.4.5 The constellations, by journal: REE

We start with initial observations (see Figure 9.6):

- C1 and C2 are broadly similar in profile.
- C3 is distinguished by high scores on D5 (text-focused discourse) and a slightly high level of informality.
- C4 scores do not deviate much from the zero line although it is notable that the D1 score is similar to that for C2.
- C5 has remarkably positive scores on D1, D2 and D3 – a profile that is distinct from the other five.
- C6 is distinguished by high scores on D4 (conceptual discourse).

As was shown earlier in the chapter, REE, as a journal, has high scores for all the five dimensions but particularly so for D1 and D2. In interpreting the six putative constellations, then, it is assumed that each D1 score will begin from a position of strong positive loading on that dimension: a predominance of verbs in present tense (as opposed to past tense), uses of abstract, process and stance nouns, articles, first-person pronouns. Consequently, the scores for D1 in C4 and C6 do not indicate that the texts eschew the use of positively loading features in that dimension, nor that they are primarily in the past tense.

The readings of sample texts for Constellations 1, 2 and 4 revealed little difference between the groups and we have merged them together to form a single constellation, labelled 'model building'. Typically, these articles introduce a set of problems that may be in the 'real world' or may be 'research-world' problems, and then they develop a model for addressing the problem(s), usually through criticism of previous models, and test it on a dataset. In these articles, human voices are usually absent and where humans are introduced they are constructs within the world of theory, as with 'producers' in example 27:

27 In our model, the adjustment costs of investments make it profitable to start investing early in non-polluting technologies, and due to the low (operating) costs of production, producers find it profitable to produce once the capacity is there.

[REE 2007 Kverndokk]

Two constellations have a high score on D3 (informal language): Constellations 3 and 5. C3 also has a high score on D5 (text-focused) while C5 has a high score on C2 (explicitness). C3 is a small constellation (13 texts) and we label it 'Human voices'. Examples are REE 2007 Flachaire, which creates a model for the analysis of respondent uncertainty on willingness-to-pay surveys, and REE 2007 Whitehead, which is a survey-based study of respondent uncertainty. Both of these articles contain second-person pronoun use (in the survey questions).

Constellation 5 can be merged with C3 as it contains articles that similarly focus on human voices, individual decisions and actions. The D2 scores are higher, and this suggest a more discursive form of writing but the attention to human voices

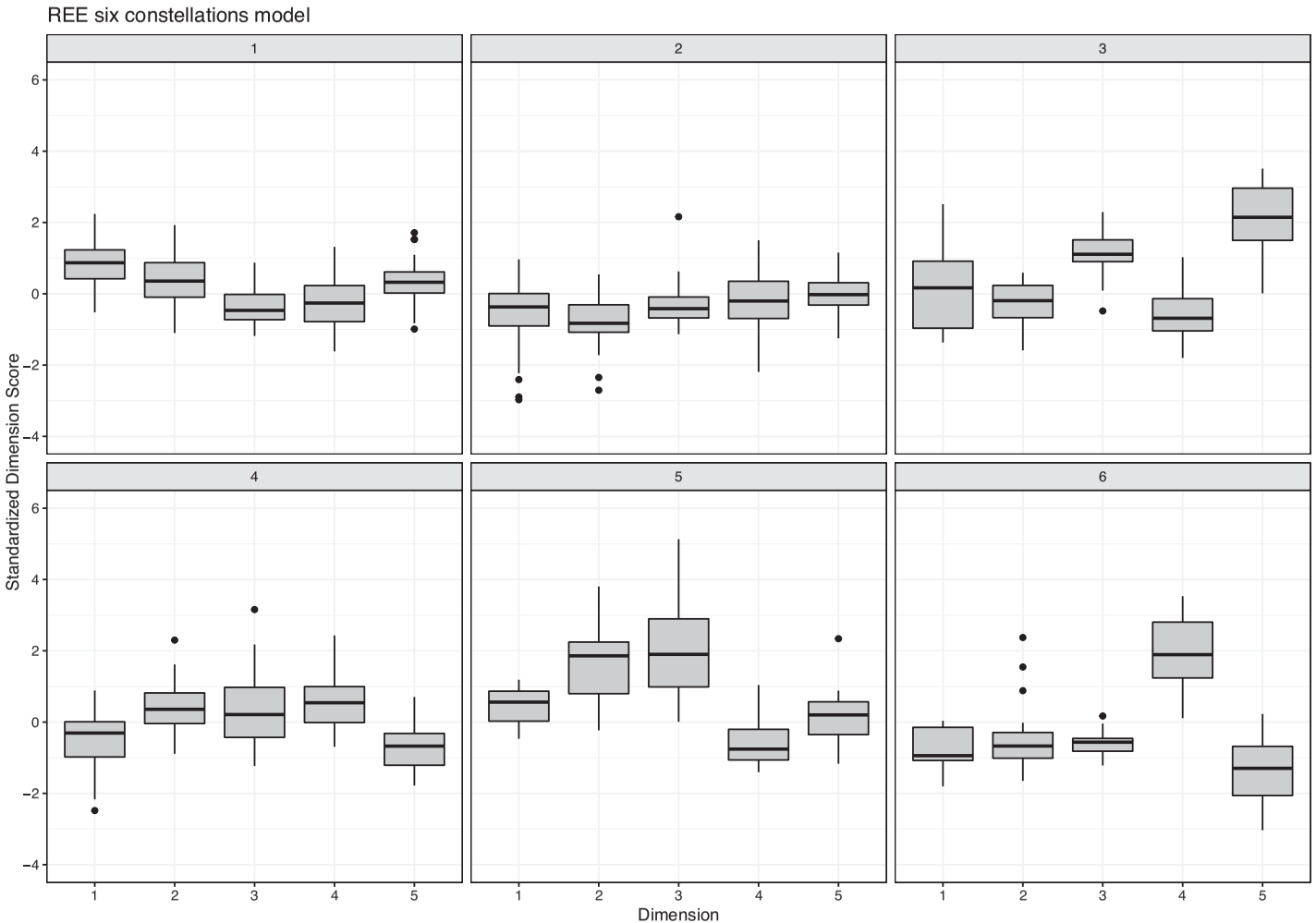

Figure 9.6 REE six-constellation model.

and actors is similar. Three of the articles (REE 2007 Dijkstra, REE 2008 Brekke, REE 2008 Konishi) contain high use of second- and third-person pronouns (84, 74 and 69 respectively) as illustrated in example 28.

28 Given her choice, the change from the high to the low concentration level has a positive welfare value (of saving her from the adverse health effect). If the individual is perfectly informed, then she would not protect at all at the high concentration level because she knows it is ineffective.
[REE 2007 Dijkstra]

As was the case in AEE Constellation 5, there are some texts that do not fit into that category because of problems in the MD analysis. One problem was the misidentification of some uses of possessive 's' ('S' preceded by an apostrophe) which had been incorrectly tagged as contractions of third person singular verb BE, which led to false counts, and the second problem was the high counts of 'group nouns' for some articles. REE 2009 Antelo, for example, is an article about systems for allocation of permits to firms which contains 184 instances of the lemma FIRM. As this word is in the list of group nouns, the article scores highly on D3 but it has very low to zero scores for the other features in D3. The word 'firm' is used repeatedly because that is the topic of the article, and the article is better placed within the 'Modelling' constellation than this one.

Finally, Constellation 6 is distinguished only by a high score on D4 (conceptual discourse). A feature of this dimension, as we have observed with Constellation 3 of AEE, is the frequent use of attributive adjectives, a positive loading feature on D4, as in example 29.

29 This study explores the **economic** potential of an **ecological** approach to **invasive agricultural** pest control that relies on habitat management to enhance the population of **natural** enemies.
[REE 2010 Zhang] (emphasis added)

The five sample texts are varied in their foci of interest: REE 2007 Jaccard addresses the question of what constitutes what is a good CO_2 reduction strategy; REE 2005 Johannesen develops a bio-economic model to analyse the impact of the integrated conservation and development projects (ICDPs) benefit-sharing components; REE 2010 Zhang creates a spatial bioeconomic model for managing the habitat of natural enemies of crop pests; REE 2010 Lewis combines econometric models of land-use change with spatially explicit simulations and ecological models; REE 2010 van der Ploeg examines the question of why many resource-rich countries save less than their marginal Hotelling rents on natural resource reserves. This diversity (other than the perhaps coincidental pair of bioeconomic models) suggests that the articles do not cohere around particular approaches or concerns. The high use of attributive adjectives may be worthy of deeper analysis in a separate study but within the context of attempting to interpret constellations, we choose to merge C6 with C1, C2 and C4.

Consequently, for REE, there are two constellations:

1 Modelling (the original C1, C2, C4 and C6)
2 Human voices and decisions (C3 and C5)

9.5 Conclusion

In this chapter we have used Multidimensional Dnalysis to explore variation both between and within the journals. The initial dimensional profiles of the four journals showed that the two natural science journals are on the whole close to each other in their linguistic profiles, and there is evidence to suggest that PS, a monodisciplinary journal, contains far less variation than the other three journals (on the basis of the SD scores). REE, by comparison with GEC (the monodisciplinary and interdisciplinary social science journals) scores highly on Dimensions 1 and 2, which suggests that many of the articles involve model development and testing (the predominant use of present tense, the explicit argumentation), and if GEC does not score quite so highly on these dimensions it may be that this is indicative of its higher levels of internal variation (the journal contains articles that involve model testing, but also contains articles that report experimental research).

An important point to note is that the dimension scores for the journals establish that, for example, use of present tense and more stance features is characteristic of GEC and REE, with human voices also being distinctive of some articles in GEC, while past tense, passive voice and fewer adjectives are typical of PS and AEE. In the second half of this chapter, a possible six constellations were inspected for each journal. For each journal five articles were read closely to see whether the constellation could be interpreted as constituting a clear, functional grouping. The analysis showed that each journal had fewer than six distinct constellations. Of the four journals, PS and REE showed the least intradisciplinary variation, with only two constellations, and GEC and AEE, the two interdisciplinary journals, have the most (five). The answer to our second question, therefore, is that there is more variation within each of the two interdisciplinary journals.

There are riders to these conclusions which stem from the limitations of the methods employed. Multidimensional Analysis was developed in order to identify differences between spoken and written registers. The set of linguistic features that are tagged and quantified for the purposes of such analyses was originally assembled by Biber in the 1980s for the investigation of general language use, and the features may not be the most appropriate index of linguistic differences within specialised discourses rather than within general language use. Within written academic language uses, for example, the use of noun phrases consisting of a head noun premodified by a noun, or postmodified noun phrases consisting of a head noun followed by *of*, are known to be distinctive. Another acknowledged feature of academic writing is the range of reporting verbs used, either followed by a *that* + clause or not. It is likely that a Multidimensional Analysis that quantifies the uses of features such as these in addition to the standard features such as the use of present or past tense, the use of perfective aspect or passive voice, would

provide a better linguistic profile of research journal texts than the set of linguistic features used here. Thompson and Gray (forthcoming) have begun work on such an approach, building a new set of linguistic features for use with a 50-journal corpus that has been built in collaboration with Elsevier.

A second problem with Multidimensional Analysis in this instance is that the model does not appear to be strong enough to distinguish between all different types of text. As stated earlier in the chapter, MDA is at its strongest when there is greater variance in the data because the registers under analysis are more diverse. In some instances in our analysis, we found that the dimension scores differentiated between texts according to topic as was the case when some articles were distinguished by being primarily about sites rather than non-sites (Dimension 5). This suggested that articles that were about lakes or forests were different from articles that were about risk or adaptation and, though this was plausible in some cases, it was certainly not so in others. In addition, where a journal had a low dimension score, particularly on Dimensions 3, 4 and 5 (which accounted for only small amounts of variance anyway), there was a danger that an increased score on that dimension in one constellation was of little meaning, as it was actually caused by a small number of linguistic feature instances. In one case, an AEE article appears to be about social groupings but the cause of this was the use of the word 'community' where this was initially treated by the analysis as a collective term for humans, but in that article it actually referred to non-human groupings.

MD Analysis was developed by Biber as a method for describing language registers, and in this chapter we have applied it for a different purpose: linguistic profiling of large numbers of research articles in order to explore variation within a corpus, firstly as a means of establishing variation between journals and secondly a means to identify groupings of articles within journals. We have observed a number of limitations to this approach, some of which can be addressed by changing the sets of linguistic and semantic features captured in the analysis so that they are more reflective of research writing.

What MD analysis has achieved in this study is (1) to distinguish between the journals, (2) to offer hypotheses about what will be important in distinguishing 'registers' in this corpus, (3) to generate hypotheses about potential intra-journal variation. The hypotheses can then be tested by the qualitative research – the 'sanity check'. Unsurprisingly, perhaps, the hypotheses work best in those instances where there are sharp differences between article types in a journal (i.e. in interdisciplinary journals with distance between the disciplines). The qualitative analysis, or reading, in turn inspires reflection on and potentially revision of the categories underpinning the original hypothesis, leading to an iterative process.

References

Biber, D. (1988). *Variation across speech and writing*. Cambridge: Cambridge University Press.

Biber, D. (1995). *Dimensions of register variation*. Cambridge: Cambridge University Press.

Biber, D. (2006). *University language: A corpus-based study of spoken and written registers*. Amsterdam: John Benjamins.

Biber, D., Gray, B., & Staples, S. (2016). Predicting patterns of grammatical complexity across language exam task types and proficiency levels. *Applied Linguistics*, 37: 639–668.

Conrad, S. (2014). Expanding multi-dimensional analysis with qualitative research techniques. In T. Berber Sardinha & M. Veirano Pinto (Eds.), *Multi-dimensional analysis, 25 years on: A tribute to Douglas Biber* (pp. 273–295). Amsterdam: John Benjamins.

Conrad, S., & Biber, D. (2001). *Variation in English: Multi-dimensional studies*. London: Longman.

Crosthwaite, P. (2016). A longitudinal multidimensional analysis of EAP writing: Determining EAP course effectiveness. *Journal of English for Academic Purposes*, 22: 166–178.

Friginal, E. (2009). *The language of outsourced call centers: A corpus-based study of cross cultural interaction*. Amsterdam: John Benjamins.

Gardner, S., Nesi, H., & Biber, D. (2018). Discipline, level, genre: Integrating situational perspectives in a new MD analysis of university student writing. *Applied Linguistics*: 1–30.

Gray, B. (2015). *Linguistic variation in research articles: When discipline tells only part of the story*. Amsterdam: John Benjamins.

Halliday, M. A. K. (1994). *An introduction to functional grammar* (2nd ed.). London: Routledge.

Halliday, M. A. K. (2004). *The language of science*. London: Continuum.

Hardy, J., & Friginal, E. (2016). Genre variation in student writing: A multi-dimensional analysis. *Journal of English for Academic Purposes*, 22: 119–131.

Hardy, J., & Römer, U. (2013). Revealing disciplinary variation in student writing: A multi-dimensional analysis of the Michigan Corpus of Upper-level Student Papers (MICUSP). *Corpora*, 8(2): 183–207.

Thompson, P. & Gray, B. (forthcoming) Disciplinary variation in published academic writing.

Thompson, P., Hunston, S., Murakami, A., & Vajn, D. (2017). Multi-dimensional analysis, text constellations and interdisciplinary discourse. *International Journal of Corpus Linguistics*, 22(2): 153–186.

10 Topic Modelling

What a journal is about

10.1 Introduction

This chapter is the second to focus on quantitative, bottom-up methods of corpus investigation. It discusses the approach to corpus analysis known as 'Topic Modelling' and asks what it can tell us about the journals in the BEE4 corpus, in particular whether interdisciplinary journals have any distinguishing characteristics when Topic Modelling (TM) is applied. The chapter begins with an account of Topic Modelling, and what it has previously shown about one of our interdisciplinary journals: GEC.

Topic Modelling is a statistical approach to corpora that can be used to answer the question 'what is this corpus about?' (Murakami, Thompson, Hunston & Vajn 2017). It works by calculating the extent to which words co-occur in a text of a given length. A 'topic', using this approach, is defined as a list of words that have a high probability of co-occurrence within a text. For the purposes of the research behind this chapter, each journal article is divided into 'texts' of about 300 words. This is carried out via an automated process that counts 300 words and then continues to the next paragraph break. As a result, a 'text' in this study means 300+ words, beginning and ending at a paragraph break. As an example of a topic, here are the top 20 words in a 'topic' identified from the journal GEC:

- popul, growth, increas, rate, per, incom, capita, china, econom, gdp, fig, level, high, decreas, trend, higher, rapid, low, declin, less

[GEC50 topic 13]

(Some of these 'words' are actually word-stems obtained using the Porter stemming algorithm [Porter 1980]. For example, 'popul' could include one or more of *population*, *populations*, *populate*, *populating* etc.) Looking at the list gives us a sense of what the texts it comes from are about, though it may be difficult to put that into words. Possible names for the topic could be 'changes in population levels' or 'links between population levels and economic factors' or 'changes in population levels and economic factors in China'. On a more abstract level we might say that the topic relates to the object of the research, rather than to, for example, the research method used.

Once a number of topics have been identified in a corpus, they can be the focus of further investigation. Murakami et al. (2017) did this in relation to a corpus of articles from the journal GEC covering a 20-year span (1990 to 2010). In that study, 60 topics were identified (for the implications of this, see the next section). As a proof of concept of the method, one topic was selected and the four articles containing that topic were identified. The topic contained the words *water*, *river*, *basin*, *suppli*, *flow*, *irrig*, *resourc*, *avail*, *use*, *stress*. The titles of the four articles confirm that all are about water security and water scarcity.

Next, topics were explored as a means of gaining information about the structure of research articles that is independent of the section titles used. It was shown, for example, that topics that tend to occur in the first 25% of each article include those associated with citations, such as dates (*1995*, *2000* etc.) and formulations such as *et al*., also topics suggesting article overviews, including words such as *studi*, *paper*, *approach* and *discuss*.

Of particular interest in the study of GEC as a journal responding to developments in a socially oriented research field was how topic frequency changes over time. Murakami et al. (2017) note that topics that are more frequent in the early years of GEC than in the later years relate to specific environmental problems such as pollution, global warming and energy scarcity. Emerging in the later years of the corpus are topics related to people, their vulnerability to environmental change and the ways they tackle environmental issues. In other words, between 1990 and 2010 there was a shift from singling out problems to humanising the solution to those problems.

Turning to how Topic Modelling can be used to distinguish between types of article, a profile of each individual article in GEC was obtained showing the number of occurrences of each of the 60 topics in it. Contrasting profiles were noted. Using this method it could be seen that some articles focus on a single topic, whereas in others there is a more equal spread of topics, with no one topic dominating. It is speculated that multi-topic articles that take a broad view of a particular problem, or that combine insights from different domains, will be more prevalent in interdisciplinary journals. An article exemplified in Murakami et al. (2017) (GEC 1992 Dahlberg) recommends rethinking aspects of use of renewable resources using contextual analysis to capture environmental, social and technological dimensions of how systems work. It is interdisciplinary in the sense that it recommends the adoption of methods from one discipline to a problem usually addressed through other means; it also spans a number of topics that the Topic Modelling approach identifies as distinct, such as cattle-grazing, fisheries and coastal wetlands. Each of a large number of topics occurs with roughly equal frequency.

A final point of interest is that Topic Modelling gives a perspective on 'what words mean' by putting each word in the context of the words it is likely to co-occur within a 300+ word text. One consequence of this is that polysemous words are distinguished. For example, in GEC, *level* means 'height' in the topic '*flood*, *level*, *sea*, *rise*, *coastal*, *loss* etc.' [GEC50 topic 40] , 'degree of gas concentration' in the topic '*emiss*, *energy*, *CO2*, *greenhouse*, *reduct*, *gas . . . level* etc.' [GEC50 topic 3], and 'point on a scale' in the topics '*popul*, *growth*, *increas*, *rate*, *per*,

income . . . level etc.' [GEC50 topic 13] or '*variabl*, *model*, *effect*, *signific*, *result*, *posit . . . level* etc.' [GEC50 topic 24]. More nuanced distinctions can be made. For example, *agriculture* in GEC can be linked to the ways land is used – '*land*, *use*, *area*, *agricultur*, *cattl*, *graze* etc.' [GEC50 topic 36] – or to the ways crops are grown – '*crop*, *product*, *food*, *soil*, *agricultur*, *yield* etc.' [GEC50 topic 30] – or to the ways communities make their living – '*household*, *farmer*, *farm*, *villag*, *incom*, *livelihood . . . agriculture* etc.' [GEC50 topic 25]. These are not different meanings, but there are distinguishable nuances of meaning between them.

In this chapter, we first present an accessible account of the method used in this chapter (section 10.2), referring the reader to the paper by Murakami et al. (2017) for further details. We then offer a number of observations of the differences between the journals based on broad-based information (sections 10.3 and 10.4). The next studies are more detailed comparisons between the two natural science journals (section 10.5) and the two social science ones (section 10.6). The chapter ends with a diachronic study (section 10.7) and a conclusion.

Note: Throughout this chapter we refer to each topic by a reference number, such as [GEC50 topic 3]. This indicates that the journal is GEC, the topic lists being used are those derived from identifying 50 (not 40 or 60 etc.) topics, and within that set of 50 topics the one being identified is number 3.

10.2 Topic Modelling: the method

Topic Models establish the degree to which words will co-occur within a text of specified length. From the output the keywords of each topic can be derived based on the probability distribution of each topic over word types. This results in a pre-specified number of wordlists, each one of which comprises a topic. The approach to Topic Modelling that we use here is based on Latent Dirichlet Allocation (Blei, Ng & Jordan 2003). We used the topicmodels package (Grün & Hornik 2011) in R (R Core Team 2015) to build the Topic Models. Further details of the method are as in Murakami et al. (2017). Here we consider the implications of what might be called the Topic Modelling methodology.

It is important to remember that although the methodology is rigorous and automatic, it is entirely dependent on some arbitrary human decisions. Firstly, what constitutes a 'text of a given length' has to be decided. It is possible to select whole articles as the unit of measurement, but that is problematic because it assumes that an article is 'about' a single thing. Following Murakami et al. (2017), and as explained previously, we specify chunks of texts of 300 words or slightly more; this is approximately a paragraph or two. This compensates for articles of different lengths. It also means that the pattern of occurrence of topics across a single article can be traced. As noted, topics that tend to occur at a given position in articles can be identified, and articles that have many constituent texts that include the same topic can be distinguished from those whose chunks are spread across many topics, thus identifying articles that are sharply focused and those whose discussion is more wide-ranging (see section 10.1).

The second arbitrary decision is how many topics to obtain. The script finds the number of topics it is instructed to find; it does not determine the best number of topics for a given corpus. It is not the case, then, that we can say 'this journal includes 60 topics but that one includes only 40 topics', or something similar. Murakami et al. (2017), working on GEC articles published over 20 years (1990 to 2010), looked at the topic wordlists produced by asking for 20, 30, 40 etc. topics and took the decision that 60 topics offered the right degree of granularity. For the work in this chapter, which considers more journals over a shorter time span, we decided to take 50 topics. In other words, the script was set to produce 50 topic wordlists for each journal. It must be remembered that this is an arbitrary number, not the 'right answer'. Similarly, when we talk about the topic lists we refer to the top 20 words in each list (that is, the 20 words that most strongly co-occur within the 300 word chunks). Each list consists of more than 20 words and, again, the choice of 20 as a way of representing the topic is arbitrary.

Finally, any name or label given to a topic is a subjective choice, though not an entirely arbitrary one, as illustrated with the 'population levels' topic. It is important to note that a topic is not an off-the-shelf entity but a construct formed by the word-list. It is often helpful to give a topic a name or mnemonic, and this process will be used extensively, but that name does not define the topic.

Using Topic Modelling, the question 'what is this corpus about?' might be answered thus: 'It is about the 50 clusters of ideas indicated by these lists of words'. Note that it is beyond expectation that two corpora would share any topic wordlists. Assessing the degree of overlap between journals, then, is a matter of assessing qualitatively how similar the topic wordlists are.

With these points in mind, we now locate Topic Modelling, as a methodology, in relation to other concepts in the field. The idea of co-occurrence is, of course, prevalent in Corpus Linguistics. The frequent co-occurrence of two words in a short span (typically, 9 or 11 words long, that is ±4 or ±5) is known as collocation. Collocation is a major contributor to word meaning, as famously memorialised in Firth's dictum: 'you shall know a word by the company it keeps' (Firth 1957: 11). The co-occurrence of strings of words has been studied under the heading of 'lexical bundles' (e.g. Biber & Barbieri 2007) and of 'p-frames' (Fletcher 2002–2007). Such strings are often classified in terms of their grammatical form and/or discourse function, with comparisons made between different corpora in terms of the frequency of the strings and their function. From a more entirely qualitative perspective, Sinclair's concept of the 'unit of meaning' stresses the importance of recurring phrases that are identified in terms of lexical, grammatical and semantic consistency (Sinclair 1991). Sinclair's approach requires close reading of concordance lines and has been largely responsible for our understanding that the linguistic unit to which meaning belongs is the variable phrase rather than the word.

Co-occurrence as understood in Topic Modelling is both similar to and different from the approaches mentioned previously. Whereas collocation measurements are based on close proximity, co-occurrence as defined in Topic Modelling takes a 'span' of a much larger number of words, at least 300 in our case, because they are about the meanings to be found in a text rather than about phrases. Whereas units of meaning

are identified essentially by the human eye, an approach which prioritises sequential patterning, Topic Modelling takes a computational approach and treats texts as 'bags of words'. Lists are produced that depend entirely on the presence or absence of a word in an unsequenced collection of 300 words. 'Bag of words' approaches are often (understandably) treated with suspicion by linguists, because they ignore the contribution made by word order or by syntax to meaning. Consider, for example, the difference between *the* ***icing*** *on the* ***cake*** and *covered in* ***cake icing*** . Both phrases include the words *cake* and *icing*, but the first is a metaphorical use and the second is a literal one. A 'bag of words' approach loses this distinction. Corpus investigation, however, is largely about rearranging naturally occurring data so that it is viewed in a way other than that obtaining during the normal reading process. Both the qualitative, meaning-based concordance line approach and the more abstract, quantitative approaches such as Topic Modelling contribute to this rearrangement. (See Chapter 11 for further discussion.)

Another familiar and productive notion in corpus investigation is that of the 'keyword', that is, a word that is relatively more frequent in a given text or collection of texts than another (Scott 1997). Scott and Tribble (2006: 59–60) demonstrate that identifying the keywords in a text gives information as to what the text was about. Keywords have been extensively used, in particular to identify on quantitative grounds words which are then subjected to more qualitative scrutiny. There is some debate about how keywords are most effectively identified (Baker 2004; Gabrielatos 2018), but the essence of the methodology is to eliminate words which are frequent in the language overall in favour of those words that are distinctive to the text or corpus in hand. Topic Modelling offers an alternative way of suggesting what a corpus 'is about'. It could be said to offer a more nuanced, if complex, approach to topic and has the advantage of not discriminating against words which are important in a corpus though also frequent in the language overall. As explained previously, topics are designated based on degree of co-occurrence rather than on the grounds of salience or distinctiveness. It also has the benefit that it is not dependent on a comparison between corpora: the topic lists in each journal are identified without recourse to a reference corpus, avoiding the difficulty of identifying a suitable reference corpus. However, there is a further stage in our methodology that leads us towards the concept of keyness. The 200 lists of 20 words gives a combined list of 1000 tokens. There is some repetition, with some types occurring once only and some occurring in several lists. It is then possible to identify how many lists each type occurs in. For example, in PS the word *use* occurs in 11 out of the 50 topic lists, *differ* occurs in eight, and *transport*, *travel*, *tourism* and *transit* occur in one only. Arguably, the words that occur in three or more topic lists are 'key' to the journal, though this keyness does not lie in a contrast with the language norm. The next section takes this principle further.

10.3 Studying the broad-range words

Application of the Topic Modelling methodology that we have used to the BEE4 corpus journals yields 50 topic lists in each of the four journals. That is, there are 200 objects of study, and no established method for approaching them. Our task,

then, has been to explore various possibilities for comparing the journals. We begin by finding ways of processing and reducing the amount of data.

As noted in the previous section, in each of the four journals, there are many words that occur in more than one topic-list. Setting an arbitrary cut-off point of five topic-lists, we can identify those words in each journal that have this degree of broad range. Table 10.1 shows these words and the number of topic-lists they occur in for each journal. Table 10.2 shows (1) the words from Table 10.1 that are

Table 10.1 Words occurring in multiple topic lists in four journals.

AEE		*GEC*		*PS*		*REE*	
Word	*No. of lists*	*Word*	*No. of lists*	*Word*	*No. of lists*	*Word*	*No. of lists*
use	12	use	11	use	13	level	14
differ	9	also	10	plant	12	use	12
year	8	can	10	analysi	9	can	10
product	7	import	9	differ	9	differ	10
studi	7	increase	9	contain	8	chang	9
agricultur	6	area	8	effect	7	one	8
effect	6	differ	8	growth	7	result	8
level	6	effect	8	high	7	case	7
plant	6	level	8	also	6	effect	7
relat	6	develop	7	concentr	6	margin	7
tabl	6	economy	7	gene	6	optim	7
type	6	relat	7	increase	6	product	7
also	5	time	7	result	6	two	7
area	5	base	6	sequenc	6	valu	7
base	5	chang	6	show	6	will	7
content	5	global	6	studi	6	first	6
density	5	issu	6	tissu	6	function	6
high	5	may	6	two	6	high	6
land	5	respons	6	accumul	5	market	6
mean	5	state	6	detect	5	may	6
measur	5	estim	5	observ	5	rate	6
observ	5	high	5	product	5	also	5
per	5	impact	5			assum	5
result	5	inform	5			benefit	5
soil	5	intern	5			control	5
system	5	market	5			reduc	5
time	5	peopl	5			relat	5
total	5	potenti	5			set	5
veget	5	region	5			time	5
29	170	29	196	22	152	29	187

Table 10.2 Unique and shared broad-range words in four journals.

Unique broad-range words				*Shared broad-range words*
AEE	*GEC*	*PS*	*REE*	
year	import	analysi	one	use
tabl	develop	contain	case	differ
type	economy	growth	margin	effect
content	global	concentr	optim	also
density	issu	gene	valu	high
land	repons	sequenc	will	
mean	state	show	first	
measur	estim	tissu	function	
per	impact	accumul	rate	
soil	inform	detect	assum	
system	intern		benefit	
total	peopl		control	
veget	potenti		reduc	
	region		set	

unique to each journal and (2) the words that are shared across all four lists. As can be seen from Table 10.2, all the journals have a handful of words that are very broad, occurring in over ten topic lists, and about 20 other words that occur in five or more topic lists. For convenience the term 'broad-range words' is used to refer to the words in Table 10.1.

As might be expected, the 'unique' words reflect the content of the journals. AEE is about land, soil and measurements. GEC is about the economy, development, people and the future. PS is about analysis of tissue and gene sequencing. REE is about modelling economic functions. The shared words are of a high level of generality and reflect an interest in comparison and cause–effect.

The words in Table 10.2 that are neither unique nor shared among all journals are shared between three journals. AEE, GEC and REE share three words (*level*, *relat*, *time*). AEE, PS and REE share two (*product*, *result*). This means there are connections between all the journals except between GEC and PS. Looking at the broad-range words that are shared between two journals, AEE is connected to both PS (*studi*, *plant*, *observ*) and to GEC (*area*, *base*), GEC is also connected to REE (*can*, *chang*, *may*, *market*), but PS is only slightly connected, by one word each, to GEC (*increase*) and REE (*two*). Combining these observations confirms that most pairs of journals are connected by at least five broad-range words, but GEC and REE are connected by seven, PS and REE by only three, and GEC and PS by only one. If sharing these words in this way indicates a closeness of content, GEC and REE are the most close and GEC and PS the least. AEE is the journal with the most even degree of closeness with each of the other journals.

Taking all this information together, it would seem from this evidence that three of the journals, AEE, GEC and REE, are relatively closely linked, whereas the fourth journal, PS, is linked only to the other natural science journal, AEE. This is consistent with what has been discovered about the journals in other chapters,

especially Chapter 9. The words that link three journals are connected to research activity itself (*level*, *relat*, *time*, *product*, *result*). Those that uniquely link AEE and GEC are about place (*area*, *base*); those that uniquely link GEC and REE are about possibility and markets (*can*, *may*, *chang*, *market*); those that uniquely link AEE and PS relate to experimentation and plants (*studi*, *plant*, *observ*). The other pairs of journals (AEE-REE; GEC-PS; PS-REE) are either not linked uniquely or are linked by one unique word only.

Thus we have used Topic Modelling to characterise the content focus of each journal and to map the relationships between journals.

10.4 BEE4 key articles

The next approach to the topic word-lists exploits the fact that each list is derived from texts of 300+ words and that each article is composed of a number of such texts – about ten if the article is 3000 words long and about 20 if it is 6000 words long. The articles that contain most of any one topic can then be identified, and each occurrence of that topic noted with its article code and position. For example, AEE50 topic 1 consists of the following words:

- CO2, plant, elev, stress, respons, leaf, effect, O3, temperatur, concentr, growth, leav, cultivar, light, level, ambient, condit, experi, radiat, photosynthesis

The topic might be labelled 'effects of elements, temperature and light on plant growth'. If the 20 (300+ word) texts are extracted that have the highest probability of occurrence of the topic words – that is, the 20 most typical paragraphs for that topic – we find that five articles account for 15 of those texts, with another five articles accounting for the other five texts. The list of 20 typical paragraphs has led us to ten articles in total. This distribution might be contrasted with REE50 topic 3 (*region*, *field*, *forest*, *land*, *agriculture*, *differ*, *log* etc.), where the 20 most typical text-chunks are spread over only two articles, one including 15 of the typical paragraphs and the other for the remaining five. Another contrasting situation is PS50 topic 1 (*protein*, *bind*, *function*, *kinas*, *interact*, *involve*, *identifi* etc.), where the 20 most typical paragraphs are spread across 19 articles, with only one article including two such text-chunks. We therefore have two extreme situations, one where a topic appears to be restricted to a small number of articles only, and one where a topic is spread among many articles, and various points between the two.

It is then possible to look at the topics that tend to be 'niche', appearing in only a few articles, and also at the journals that tend to include such niche topics and those that do not. For example, REE50 topic 3, which relates to forests, has been established as 'niche' in REE. In GEC, the only topic that includes the word *forest* is GEC50 topic 45 (*forest*, *deforest*, *area*, *cover*, *land*, *tree*, *timber* etc.). The 20 most typical text-chunks for this topic in GEC occur in ten articles. A similar topic in AEE – AEE50 topic 7 (*forest*, *tree*, *plantat*, *burn*, *clear*, *coffee* etc.) – occurs in eight articles. The word *forest* does not occur in the PS topic lists. In other words, it seems that 'forestry' as a topic does not appear in PS, does appear in REE but only in two articles that each focus intensively on the topic, and is more widespread in both AEE and GEC.

What is more striking is the difference between the journals. The methodology for establishing this is to take arbitrary cut-off points of five and 15. If the typical paragraphs for a topic occur in five or fewer articles, that topic will be described as 'niche'. If the typical paragraphs for a topic occur in 15 or more articles, that topic will be described as 'spread'. The journals can then be compared according to how many of their 50 topics are 'niche', 'spread' or 'neither'. Table 10.3 gives for each journal a list of topics classified in these three ways. The number of topics in each list is also shown for convenience. In most journals the 'neither' list is larger than any other, but the proportion of topics in the 'niche' and 'spread' lists do distinguish between the journals, placing REE at one end of a continuum with

Table 10.3 'Niche' and 'spread' topics in four journals.

	AEE		*GEC*		*PS*		*REE*	
	Topics	*No. of topics*	*Topics*	*No. of topics*	*Topics*	*No. of topics*	*Topics*	*No. of topics*
'niche' topics = ≤ 5 articles	9, 28	2	2, 4, 5, 6, 9, 17, 19, 20, 26, 31, 35, 37, 41, 48, 49, 50	16		0	1, 2, 3, 4, 6, 9, 11, 12, 13, 15, 18, 19, 20, 21, 22, 23, 24, 25, 26, 28, 29, 31, 32, 33, 34, 35, 36, 38, 42, 43, 44, 45, 46, 47, 48, 49, 50	37
'spread' topics = ≥ 15 articles	3, 11, 12, 15, 16, 17, 20, 21, 22, 24, 29, 33, 34, 36, 40, 41, 44, 45, 46, 48, 49, 50	22	33, 38, 42	3	1, 2, 3, 5, 8, 9, 12, 18, 20, 21, 23, 24, 26, 27, 30, 31, 33, 35, 38, 42, 43, 45, 46, 47, 49, 50	26	14, 16, 30, 37, 39	5
'neither' topics = 6–14 articles	1, 2, 4, 5, 6, 7, 8, 10, 13, 14, 18, 19, 23, 25, 26, 27, 30, 31, 32, 35, 37, 38, 39, 42, 43, 47	26	1, 3, 7, 8, 10, 11, 12, 13, 14, 15, 16, 18, 21, 22, 23, 24, 25, 27, 28, 29, 30, 32, 34, 36, 39, 40, 43, 44, 45, 46, 47	31	4, 6, 7, 10, 11, 13, 14, 15, 16, 17, 19, 22, 25, 28, 29, 32, 34, 36, 37, 39, 40, 41, 44, 48	24	5, 7, 8, 10, 17, 27, 40, 41	8
Total topics		50		50		50		50

PS at the other end. GEC sits closer to REE and AEE sits closer to PS. This may be a result of the smaller number of articles in GEC and REE compared with AEE and PS, but it may also reflect a greater tendency in social sciences to 'break new ground' and tackle new areas of research rather than consolidate existing ones. It would appear that the 'niche topic' versus 'spread topic' distinction differentiates between natural science and social science. The two interdisciplinary journals lie between the natural science and social science extremes.

10.5 Comparing AEE and PS

Having pursued approaches that reduce the data, we consider it necessary to deal directly with all 200 of the topic lists produced by our application of Topic Modelling. We thus now turn to a more detailed study of the topic wordlists, and compare pairs of journals. We first compare the two natural science journals: AEE and PS. We seek to answer the question: can the interdisciplinary journal AEE be shown to have more internal variation than the monodisciplinary PS? Each of the 50 topics in each journal is annotated in two ways. First a distinction is drawn between 'research process' and 'content'. An example of a 'research process' topic is:

- studi result may high higher found suggest low also howev observ lower differ relat probabl due present like import influenc

[AEE50 topic 12]

The words in this list indicate that research results have been identified and interpreted, leading to an assessment of probability, but the topic of that research cannot be identified.

An example of a 'content' topic is:

- sucrose enzyme sugar metabol syntas glucose synthesi carbohydr phosphate pathway accumul biosynthesis fructose dehydrogenase active tissue carbon atp ec

[PS50 topic 13]

This topic is about the processing of sugars, although from a simple list of words it is not possible to be any more precise about the research process being used.

Table 10.4 gives a rough count of the topics in each category in all four journals, with the figures for AEE and PS shown in bold. The numbers are approximate

Table 10.4 'Process' and 'content' topics in four journals.

	AEE	*GEC*	*PS*	*REE*
Process	**16**	11	**12**	19
Content	**32**	39	**38**	31
Total	**50**	50	**50**	50

because distinguishing between 'process' and 'content' is to some extent subjective. The table does give an idea of the proportion of the categories, however.

The next stage in this study is to annotate the 'content' topics with an informal summary of what they are about. This is inevitably a subjective process and is as much a heuristic as it is a classification. That is, it is a way of forcing attention on to the topic lists so that generalisations within each journal and distinctions between the journals can be proposed. The results of this exercise for each journal are shown in Tables 10.5 and 10.6. In each table, the first interpretative label is given in the first column, with a generalised label in the final column. Table 10.5a shows the results for PS.

Table 10.5a Interpretation and generalisation of content topics in PS.

Interpretation	*Topics*	*Generalisation*
experiments on proteins	1	internal to plants
plant growth conditions	2, 16, 19, 44, 45, 46	external to plants
actions of enzymes	3, 13	internal to plants
actions of chemicals	4, 10, 28	external to plants
studies of genes	6, 7, 9, 17, 18, 22, 23, 25, 43, 49	internal to plants
effects on plants	8	external to plants
plant mutations	11	internal to plants
genes and amino acids, proteins	14, 34	internal to plants
cell structure and changes	15	internal to plants
plant resistance to disease/insects	27	external to plants
improving plant species as crops	29	crops
plant lifecycle	32, 36, 47	lifecycle
plant species as crops	37, 39	crops
genes and bacteria	38	internal to plants
actions of metals	40	external to plants
plant nutrition	41	external to plants
plants and light	48	external to plants

Table 10.5b Content topic generalisations in PS.

Generalisation	*Topics*
Things internal to plants: genes, proteins, enzymes, cell structure	1, 3, 6, 7, 9, 11, 13, 14, 15, 17, 18, 22, 23, 25, 34, 38, 43, 49
Things external to plants: metals, chemicals, light, nutrition	2, 4, 8, 10, 16, 19, 27, 28, 40, 41, 44, 45, 46, 48
Plant lifecycles	32, 36, 47
Plants as crops	29, 37, 39

Once the initial interpretation had been carried out on the PS content topics (in Table 10.5a), it became apparent that a distinction could be drawn between studies of entities internal to plants, such as proteins and genes, and those relating to entities external to plants, such as chemicals, metals and light. Only a few topics cannot be fitted into these categories: topics relating to plant lifecycles, and those that specify plants used as crops. Table 10.5b summarises these generalisations.

Table 10.6 repeats this exercise for AEE. Table 10.6a focuses on the entities in each topic. The interpretation in the right-most column mirrors the 'internal–external' split observed in PS. Table 10.6b summarises this distinction and adds the category of 'people'.

Table 10.6a Interpretation and generalisation of content topics in AEE.

Interpretation	*Topics*	*Generalisation*
plants conditions of growth	1	internal to agriculture
systems of crop growth	2, 19, 40	internal to agriculture
insects and crops	4	internal to agriculture
gases	20, 25	internal to agriculture
insects	3, 43	internal to agriculture
farm layout	5, 13	internal to agriculture
water	8, 26	internal to agriculture
land use	9, 10, 23	external to agriculture
energy	6	internal to agriculture
forests	7	external to agriculture
birds, wildlife, ecosystems	14, 33, 35	external to agriculture
soil	18, 25, 28, 29	internal to agriculture
weeds and herbicides	37	internal to agriculture
metals and soil	39	internal to agriculture
people – farmers	30	people, farmers
economic policies	31	people, politics
grass and livestock	32, 38	internal to agriculture
agricultural systems	35, 42	internal to agriculture
places	45	places
weather and seasons	49	external to agriculture

Table 10.6b Content topic generalisations in AEE.

Generalisation	*Topics*
Internal to agriculture: soil, systems, insects, weeds	1, 2, 19, 40, 4, 20, 25, 3, 43, 5, 13, 8, 26, 6, 18, 25, 28, 29, 37, 39, 32, 35, 42
External to agriculture: agriculture and the environment	9, 10, 23, 7, 14, 33, 35, 49
People: farmers, politics, places	30, 31, 45

Two key interpretative findings emerge from this part of the comparison between the two journals. The first is the 'internal–external' split. Many topics in PS are about either what is internal to the plant (genes, enzymes, proteins etc.) or what is external to it and interacting with it (metals, chemicals, light etc.). In AEE, many topics are about entities internal to agriculture (agricultural systems, soil quality, livestock etc.) and others are about the interaction between agriculture and aspects of the environment (forests, birds, water systems etc.). Drawing out these parallels also draws attention to the exceptions. In PS, there are three topics that are about plants as crops. The presence of these three topics is a reminder that in the other PS topics the commercial and nutritional value of plants is effectively ignored. This does not necessarily mean that it is not mentioned at all in the research articles, but it emerges as a component of a topic only in three cases. In AEE there are a few topics that are about human engagement, and these topics are unlike anything in PS. One topic relates to farmers, villages and households. Another relates to development, economics and policies.

Having looked at the 'content' topics, we can now turn attention to the 'process' topics, that is, those that are about the process of doing research – establishing a method, processing data and drawing conclusions. Because there are fewer of these it is possible to show the interpretation process in more detail. Table 10.7

Table 10.7 'Process' topics in PS.

Topic	*Items*	*Interpretation*	*# Articles*
12	treatment day control treat seedl time period experi 24h follow respect reach	experiment carried out on seedlings	18
20	use valu data test analysi differ experim calcul mean three per signific	analysis of experiment results	19
21	use sampl extract determin dri measure water standard analysi content acid weight column hplc methanol temperature	analysis of sample measurements	20
24	may suggest result observ differ studi howev also possibl can present due might sinc indic high report this therefor although	interpretation of results	19
26	use section fluresc protoplast stain gfp microscop incub particl sampl tissue fix ethanl	preparing tissue samples for analysis	20
30	mm 10 min -1 ml mg 50 100 20 30	measurement	17
31	32 33 35 34 tabl 36 show 31 40 similar respect result	measurement	15
33	fig show observ shown result detect data indic similar 2a 1a 3a compar 1b also 2b	presenting results	19
35	protein gel antibody use fraction extract anti sds band page purify blot molecular western detect mass separ stain analysi electrophoresi	preparing samples for analysis	19
42	increas level signific content higher decreas tabl concert compar lower observ differ total show	interpretation of results	16
50	ph buffer extract centrifuge assay contain incub ad reaction supernat 1mm phosphate homgen	preparing samples for analysis	16

shows the relevant topics in PS. The topic number is shown, followed by the top words in the wordlist, then the interpretation. The right-most column in the table gives the number of articles (out of a maximum of 20) from which the most typical texts with that topic come. Omitted from this table is a topic that consists of dates and other citation 'words' (such as *et al*.).

Table 10.8 gives the same information for AEE.

Table 10.8 'Process' topics in AEE.

Topic	*Items*	*Interpretation*	*# Articles*
3	sampl number within record locat per point survey select distanc	sampling and measuring	19
11	-1 year rate average tabl per annual rang total respect -1year amount estim ha mean 10	measurements	19
12	studi result may high higher found suggest low also howev observ lower differ relat probabl	interpretation of results	20
15	total respect tabl 10 mean average 12 20 rang	measurements	16
16	use data map class inform base classif cell analysi land image grid	identifying and classifying	17
17	sampl use measure determin dri collect method extract taken chamber analysi analyze	sampling, measuring and analysing	19
21	site two first differ three studi year second tabl one four ii type repress group similar third	presentation of results	15
22	variabl use analysi test data statist analys regress varianc mean effect perform linear signific	statistics	19
24	valu calcul eqsym estim use measure mean assum base obtain equat index data	statistics	16
27	indic assess use evalu impact method environment risk base pesticide relat quality differ level consid object approach score effect defin	assess environmental impact and risk.	6
34	signific differ treatment higher tabl compar effect lower 5 control similar greater show among observ fig interact result howev	interpretation of results	18
36	can may will howev need potenti require limit like improve term strategi examp possibl even provid also conclus time import	prediction of future	16
41	plot treatment experi experiment two three appli design block control four week	experimental treatment with a control set	20
46	increas decreas effect reduc result chang redut declin due hoev caus less also affect negat lead	interpretation of results	19
47	model simul predict data paramet use estim result measure base uncertainty	use of model	10
48	fig show correl posit tabl valu relat high relationship observ indic variat negat low highest similar shown trend result higher	interpretation of results	19

The two tables show considerable overlap: process topics in both journals are about methods, the production of data and interpretation of results. The range in PS is more restricted, however. AEE includes topics relating to the assessment of risk (27), future prediction (36) and the use of models (47). In both journals, most process topics are spread across a large number of articles. In AEE, however, there are a few relatively 'niche' process topics, occurring most significantly in only six or ten (out of 20) articles.

From this study, what can we conclude about these two journals? Both are about the natural world. Both treat their subjects – plants and agriculture – as entities with internal and external influences, but only AEE includes topics that take a specifically environmental approach. Only AEE includes topics related to social issues as well as natural ones. There is little evidence that either journal includes articles that are niche in terms of research process, but AEE does include two slightly more niche process topics, and there is a small amount of evidence that it is more focused on research processes in general. From this we might conclude that there is indeed evidence that AEE as an interdisciplinary journal has more internal variation than PS.

10.6 Comparing GEC and REE

We now turn to the two journals selected for study because they both have a social science focus. Again, we ask whether the interdisciplinary journal GEC demonstrates more internal variation. As might be expected, the topic wordlists in these journals reflect a human dimension. For example, the list of words in GEC50 topic 45 includes words indicating natural entities or processes such as *forest*, *deforest*, *area*, *cover*, *land* and *tree*. The same topic wordlist also includes words indicating human institutions, such as *properti*, *nation*, *use*, *plantat* and *road*. Similarly, REE50 topic 3 includes natural-entity words such as *field*, *forest*, *land*, *log* and *carbon* but also human-institution words such as *agricultur*, *trade*, *use*, *cost* and *margin*.

In order to compare the journals with each other, a procedure is followed that is in principle the same as that was followed previous section, in that the topics are interpreted and scrutinised to identify generalisations. The process is not identical, however, simply because each analysis is a response to the specifics of each set of topics. Comparing GEC and REE involved the following procedure:

- We first distinguished 'content' and 'process' topics. This distinction is shown in Table 10.4, repeated here for convenience, with these two journals shown in bold.
- Each content topic was then summarised in no more than two words.
- For each journal we then made a table consisting of an intuitive topic label and the topics that belong to it. These are shown in the first two columns of Tables 10.9 and 10.10. In the majority of cases, it was not possible to assign more than one topic to each label, but we did, for example, assign six topics to the category 'risk /damage levels and future predictions' in GEC and seven topics to the category 'business finances' in REE.
- As a further step, we assigned each of the column 1 labels to one of 12 broader categories, shown in the third column of Tables 10.9 and 10.10. This allows a comparison between the two journals, shown in Table 10.11.

Table 10.4 (repeated) 'Process' and 'content' topics in four journals.

	AEE	*GEC*	*PS*	*REE*
Process	16	**11**	12	**19**
Content	32	**39**	38	**31**
Total	50	**50**	50	**50**

Table 10.9 'Content' topics in GEC.

Topic label	*Topics*	*Category*
community development	1	People, local
fisheries	2	People, local
greenhouse gas emissions	3, 19	Physical
government	4, 10	People, political
social justice	6	Abstract
water resources	7	Physical
ecosystem conservation	8	Physical
risk/damage levels, future predictions	11, 17, 20, 32, 43, 44	Future
international politics	12, 50	People, political
population growth	13	People, system
agricultural trade	14	People, system
problems and attitudes	15	People, perception
discourse and media	16	Discourse
history	18	History
global development	22, 27	People, system
farmers	25	People, local
scientific knowledge	28	Knowledge
food production	30	People, system
environment and migration	31	People, system
technology and development	35	People, system
land use	36	Physical
people's perceptions	37	People, perception
environmental systems	39	Physical
seas and flooding	40	Physical
forests	45	Physical
climate change	46	Physical
models	47	People, system
social practices	48	People, system
indigenous knowledge	49	Knowledge
29 distinct topic labels		

Table 10.10 'Content' topics in REE.

Topic label	*Topics*	*Category*
business and government	1	People, business
forests	3	Physical
taxation and climate	4	People, political
fisheries	5	People, local
production and technology	7	Technology
conservation and biodiversity	8, 21	Physical
land use and development	9	People, system
technology and climate	10	Technology
systems of choices	11	People, system
oil and fuel – energy	12, 45, 47	Physical
health and risk	13	People, system
international trade	15	People, business
consumption and technology	19	Technology
business finances	24, 27, 29, 33, 46, 49, 50	People, business
households	25	People, local
waste disposal	31	People, system
government and social welfare	34	People, political
future	38	Future
gases and emissions	42	Physical
water	43	Physical
pollution and regulation	44	People, political
agriculture	48	People, system
22 distinct topic labels		

Table 10.11 Comparison of GEC and REE content topics.

Content topic category	*GEC: number of topics*	*REE: number of topics*
Physical	9	8
People, local	3	2
People, political	4	3
People, system	9	5
People, business	0	9
People, perception	2	0
Technology	0	3
Future	6	1
Discourse	1	0
History	1	0
Knowledge	2	0
Abstract	1	0

As before, it is important to note that because of the highly subjective nature of this exercise, not too much significance should be attached to the precise ratios suggested by the numbers in Table 10.11. It is worth noting, however, that GEC has more topics categorised as 'future' while REE has more categorised as 'People, business'. The most important finding, however, is that GEC simply includes more categories (10 out of the total 12). REE includes 7 categories. Not only that, but GEC includes more different types of topic. Discourse and history, as topic areas, come from very different knowledge domains from those dealing with politics or the natural world. GEC includes topics dealing with people's perceptions of environmental issues, as well as topics around (media) discourse and history. Importantly, it also includes two topics here labelled 'knowledge', which deal with how knowledge is constructed. These are shown here:

- research, scienc, knowledg, scientif, scientist, earth, question, understand, ipcc, work, institut, make, poli, expert, author, problem, view, domain, public, part [GEC50 topic 28]

- health, peopl, communiti, disease, mani, tradit, indigen, import, generat, effect, knowledge, arctic, access, also, popul, culture, 2008, includ, heat, inuit [GEC50 topic 49]

These are the topics that question the unique value of the scientific paradigm. Topic 28 includes words that were discussed in detail in Chapter 6, around the concept of 'science'. In that chapter it was noted that some articles in GEC present science in antagonistic terms. Topic 49 suggests an alternative world view, with knowledge located in traditional communities.

In short, although GEC and REE both include a range of topics, there are more topic categories, and more types of topic, in GEC than in REE. A small number indicate an antagonism towards scientific research.

Turning to the 'process' topics, as we previously listed each topic, its interpretation and the number of articles in which that topic is highly probable, from a total of 20. The results are shown in Tables 10.12 and 10.13. As before, topic wordlists that consist of dates are excluded.

Tables 10.12 and 10.13 suggest that these journals are roughly equivalent in terms of the research process topics. REE has more in the way of calculations, and GEC has arguably more different approaches to research. Both have a mixture of topics that are spread across many articles and those that appear to be concentrated in a few. The prevalence of topics which are focused on a few articles is a distinguishing feature of these two journals in comparison to AEE and PS.

This section has explored in more detail the comparisons between AEE and PS and between GEC and REE. It shows that the interdisciplinary journals are distinguished, not by having more topics, but by having a wider range of types of topic.

Table 10.12 'Process' topics in REE.

Topic	*Items*	*Interpretation*	*# Articles*
2	probabl, expect, uncertainti, inform, will, case, decis, distribut, instrument, can, assum, make, correl, result	measuring probabilities	4
6	use, data, estim, studi, approach, result, differ, method, base, adjust, include, analysi, number, avail	obtaining data and results	5
14	may, import, issu, problem, paper, howev, mani, exampl, particular, like, even, need, might, larg, reason	interpreting results	16
16	eq, term, condit, equal, first, can, margin, hand, sinc, order, respect, express, two, equat, side	performing calculations	17
17	eqsym, function, given, follow, condit, denot, assum, let, proposit, optim, maxim, defin, paramet, constant	performing calculations	12
18	public, measur, good, number, chang, characterist, hous, one, increas, us, articl, privat, sale, also, addit, control, include, provid, level, result	interpreting results	2
20	wtp, respons, question, respond, bid, bound, amount, valu, answer, mean, prefer, individu, first, state, rang, initi	obtaining results through question/answer	4
22	tabl, year, average, per, unit, valu, 10, paramet, total, mean, report, million, simul, three, annual, rang, present, calcul, sampl, number	performing calculations	5
23	one, model, set, function, attribute, can, follow, divers, given, relev, speci, property, general	building model	1
26	equilibrium, game, solute, constraint, nash, non, function, agent, cooper, can equilibria, problem, strategi, player, social, case, payoff	applying game theory	5
28	program, particip, pay, survy, payment, group, respond, design, mechan, hypothet, respons	using survey and interview	4
32	quality, inform, chang, pay, valu, willing, improve, use, valuat, sever, effect, recreat	interpreting results	2
35	fig, level, curv, per, relat, differ, capita, show, line, world, two, distribut, illustr, point, equal	interpreting results	3
36	test, result, data, panel, hypothesi, unit, estim, time, use, cross, restrict, empir, factor, find	interpreting results	4
37	model, section, paper, result, use, present, analysi, assumpt, discuss, consid, general, show, two, follow, differ, allow	interpreting results	16
39	increase, effect, reduc, higher, decreas, result, lower, case, negat, less, posit, impact, will	interpreting results	15
41	run, long, time, short, effect, use, gdp, urban, lag, studi, term, day, air, concentr, impact, relationship, level, high, mortal, associ	describing method	6

Table 10.13 'Process' topics in GEC.

Topic	*Items*	*Interpretation*	*# Articles*
5	studi, case, factor, analysi, result, find, chang, research, examin, compar, import, differ	interpreting results	4
21	data, use, avail, includ, base, method, estim, type, provid, sourc, observ, inform, time	describing method	12
23	group, interview, respond, particip, survey, question, respons, two, inform, three, ask	using survey and interview	6
24	variabl, model, effect, signific, result, posit, differ, correl, relationship, measur	describing model	8
26	assess, process, inform, stakehold, polici, particip, user, decis, issu, tool, make	describing method	6
29	indic, valu, use, differ, eqsym, index, categori, fig, measur, weight, determin, repres, relat	performing measurements	9
33	approach, section, paper, discuss, analysi, use, present, framework, method, differ identify	interpreting results	17
34	year, total, per, average, tabl, 10, million, estim, 20, show, annual, fig, 0, 50	presenting figures	14
41	scale, local, level, region, spatial, can, across, map, effect, impact, may, area, larg, relat	interpreting results	4

10.7 Diachronic study

Topic Modelling can be used to study changes in the frequency of topics over time. For example, as noted in the introduction to this chapter, Murakami et al. (2017) point out that in GEC over a period of 20 years topics relating to the identification of specific environmental problems such as pollution, global warming and energy use were more frequent in the early years of the journal, whereas alternative approaches such as discourse studies or studies of local adaptations to environmental change were more prominent in the later years. Many other topics would remain constant across time, or would have a peak at some point during the time span.

For the present study, graphs showing the frequency of each of the 50 topics in each journal were produced (i.e. 200 graphs in all), and these were scrutinised to observe trends within each journal. The results are given for each journal individually here.

PS (See Table 10.14 for examples of the topics referred to.)

The topic graphs for the journal PS show some clear diachronic changes across the ten years studied. Of the 50 topics, 16 can be observed to demonstrate a movement from high frequency to low or vice versa. Topics 1, 20 and 43 are instances of an increase in frequency, while topics 3, 30 and 46 are instances of a decrease in frequency. Other 'rise' and 'fall' topics have a less consistent pattern of change, but overall finish the decade markedly more or less frequent than they started it. The other 34 topics demonstrate no clear rise or fall.

Table 10.14 Examples of topics over time in PS.

Topic number	*Topic words*	*Example of . . .*
1	protein, bind, function, kinas, interact, involv, identifi, factor, arabidopsi, like, encod, regul, domain, spot, signal, includ, also, specif, phosphoryl, relat	rise in frequency
46	cultur, medium regener, shoot, callus, explant, ms, -1, media, calli, transfer, vitro, week, somat, plantlet, supplement, contain, embryogen, suspens, select	fall in frequency
29	use, plant, product, can, crop, speci, improve, studi, genet, system, import, high, develop, introduct, mani, effici, research, yield, larg, produc	'gene' topic: rise in frequency
7	dna, genom, hybrid, fragment, two, band, one, analysi, copi, singl, probe, sequenc, size, digest, restrict, detect, southern, site, methyl, insert	'gene' topic: fall in frequency

Among the topics that increase in frequency are several that include words related to genetics, such as *gene*, *(c)dna* and *genome* such as 'gene'. For this reason, a more specific study has been carried out on all these topics, 13 in all. Eight of these topics (6, 9, 17, 22, 23, 29, 43 and 49) are more frequent at the end of the decade than the beginning. The other five (7, 14, 18, 25 and 38) are the same or less frequent at the end of the decade. Those topics that include references to gene sequencing (7, 14, 25, 38, 43 and 49) are more likely to finish lower. Topics that include references to crop improvements (6, 23 and 29) are more likely to finish higher.

In conclusion, then, over ten years the journal PS shows an overall rise in the frequency of topics relating to genetics. Within that general trend, there is a decrease in topics relating to the basic research connected to gene sequencing and an increase in topics relating to applications of that research to improving crops. This is consistent with the change in editorial policy for the journal discussed in Chapter 3 (section 3.4.3).

AEE (See Table 10.15 for examples of the topics referred to.)

AEE is another journal that shows some clear trends, with about 20 topic graphs out of the 50 showing a clear rise (e.g. 15, 29 and 44) or fall (e.g. 3, 4 and 36) in frequency. Among the rising topics are those that relate to gases and emissions (topics 1, 20 and 25) and soil quality (topics 18, 25 and 29). Among the falling topics, perhaps surprisingly, are some that relate to people (topic 30: farmers, topic 31: development and economics) as well as those relating to the control of insects (topic 4) and weeds (topic 37). It would appear from this that at the end of the studied decade there is a greater focus on how broader environmental changes impinge on agriculture and a reduced focus on what might be called routine interventions against pests. An excursion into a more human perspective appears not to have been maintained over time, though it must be remembered that the ten-year span

Table 10.15 Examples of topics over time in AEE.

Topic number	*Topic words*	*Example of . . .*
29	soil, depth, content, layer, surfac, clay, moistur, densiti, properti, profil, bulk, textur, sandi, sand, loam, top, physic, horizon, capac, fine	rise in frequency
36	can, may, will, howev, need, potenti, requir, limit, like, improv, term, strategi, exampl, possibl, even, provid, also, conclus, time, import	fall in frequency
20	emiss, n2o, flux, ch4, measur, co2, gas, carbon, soil, methan, oxid, greenhous, ghg, net, respire, atmospher, estim, product, factor, sourc	'emissions' topic: rise in frequency
18	soil, organ, carbon, soc, microbi, matter, activ, residu, decomposit, pool, sequestr, tillag, stock, content, aggreg, biomass, ratio, earthworm, ct, fraction	'soil' topic: rise in frequency
30	farmer, land, use, popul, villag, household, cultiv, local, resourc, product, market, livestock, area, access, tradit, incom, major, import, rural, sourc	'farmer' topic: fall in frequency
4	plant, damag, orchard, pest, cotton, use, control, pesticid, fruit, diseas, insect, resist, insecticid, pollin, effect, popul, wild, non, sever, pollen	'insects' topic: fall in frequency
37	weed, seed, grass, herbic, year, plant, perenni, sown, establish, seedl, annual, control, cover, emerg, mixtur, bank, clover, speci, densiti, flower	'weeds' topic: fall in frequency

studied is only a segment from the life history of the journal and does not bring us up to the present day.

Another noticeable feature of the AEE topics is that 'process' topics seem to have increased in frequency in the ten years studied. Of the 16 process topics shown in Table 10.7, eight (11, 12, 15, 16, 17, 34, 41 and 48) show a rise in frequency and only four (3, 24, 27 and 36) show a fall. This suggests a general increase in focus on the methods of arriving at results and on their explicit interpretation.

REE (See Table 10.16 for examples of the topics referred to.)

A feature of REE is that few topics (topic 7, 14, 35) show a clear upward or downward trend in the decade being studied, and it is not possible to generalise from them. Instead, the topics typically show peaks of frequency at various times over the ten years. This suggests a consistent variation of approach without specific identifiable trends.

GEC (See Table 10.17 for examples of the topics referred to.)

GEC is similar to REE in that relatively few topics show a clear upward or downward trend, with only seven topics clearly following this pattern. There are, however, nine topics that finish higher than they started and nine that finish lower. Two specific overall themes appear to have reduced in frequency over the ten years: 'future risk' and 'development'. There are eight topics that include words such as 'future' and 'risk'. Of these, two (32 and 43) finish the decade more

Table 10.16 Examples of topics over time in REE.

Topic number	*Topic words*	*Example of . . .*
35	fig, level, curv, per, relat, differ, capita, show, line, world, two, distribut, illustr, point, equal, share, see, one larg, scenario	rise in frequency
7	output, product, effici, input, measur, function, produc, index, good, technolog, bad, set, desir, process, distanc, observ, regul, given, frontier, direct	fall in frequency
23	one, model, set, function, attribut, can, follow, divers, given, relev, speci, properti, general, refer, consid, weight, featur, note, requir, restrict	mid-period spike in frequency
11	util, set, function, choic, consider, form, can, prefer, demand, specif, altern, system, given, welfar, expenditur, two, incom, structur, repres, condit	mid-period dip in frequency

Table 10.17 Examples of topics over time in GEC.

Topic number	*Topic words*	*Example of . . .*
32	adapt, capac, respons, measur, option, strategi, can, cope, chang, exampl, abil, determin, potenti, depend, effect, adger, econom, condit, institut, specif	'future' topic: rise in frequency
20	decis, uncertainti, make, inform, can, may, one, choic, forecast, probabl, differ, process, exampl, strategi, risk, maker, particular, also, outcom, predict	'future' topic: fall in frequency
35	technolog, resourc, develop, effici, improv, access, new, low, import, like, base, countri, also, transfer, innov, high, industri, adopt, pollut, may	'development' topic: rise in frequency
47	scenario, model, use, project, futur, region, differ, result, global, sres, simul, rang, base, estim, world, assumpt, assum, a2, econom, develop	'development' topic: fall in frequency
1	manag, communiti, develop, learn, resourc, project, program, need, effort, sustain, involv, plan, local, strategi, approach, support, success, implement, address, activ	'society' topic: ends at high frequency
11	vulner, risk, disast, hazard, exposur, natur, assess, event, social, sensit, factor, communiti, relat, econom, research, condit, stress, capac, physic, affect	'society' topic: ends at low frequency

frequent than at the start, but the other six (11, 20, 37, 40, 44 and 47) finish the decade at the same or a lower frequency. Similarly, of the seven topics including the word 'development', two (1 and 35) finish more frequent and the other five (22, 26, 27, 47 and 50) finish the decade the same or less frequent.

A more detailed study has been carried out on topics that include the words 'society/social' or 'community' (or both). There are six such topics: 1, 11, 27, 31,

48 and 49. The graphs show mixed results, with three topics (1, 48, 49) ending the decade at relatively high frequency and the other three (11, 27, 31) having one or more peaks but then ending at relatively low frequency. If the two sets of three are compared, then the 'diminishing' topics are seen to be those that focus on problems, including words such as *problem*, *challenge*, *conflict*, *threat*, *risk* and *disaster*, while the 'increasing' topics are those focusing on plan and solution, including words such as *program*, *strategi*, *success*, *structur* and *knowledge*.

Overall, in GEC as in REE it is less easy to spot chronological trends than it is in PS or AEE. A more detailed approach is needed and with this it is possible to identify a slight trend towards construing communities as acting on environmental change as well as being affected by it.

10.8 Conclusion

This chapter has used one form of Topic Modelling to explore the content of the BEE4 journals. The chapter has combined a quantitative, objectiv, and 'linguistically naïve' approach to processing language data with a highly subjective, interpretative and 'statistically naïve' approach to interpreting the outcome (Hunston 2016). For the purposes of this chapter, the outcome comprises the 200 topic wordlists, 50 for each journal.

The first study in the chapter (section 10.3) looked at the 'broad-range' words, that is, those that occur in the most number of topic lists. These words identify the themes that are characteristic of each journal. They also suggest a topological view of journal connections, shown in, suggesting that GEC combines themes also found in AEE and REE, and that AEE combines themes also found in GEC, REE and PS. It also suggests that GEC and PS share relatively little. This in turn implies that GEC and PS are at opposite ends of a spectrum, but also that the two interdisciplinary journals, AEE and GEC, act as pivot points between the other journals. Similar observations about the relationships between journals were made in relation to Dimensions 1 and 2 in Chapter 9.

The second study (section 10.4) exploited the distinction between 'text' as used in this Topic Modelling study, where each text is about 300 words long, and 'text' as the word is usually used, where each journal article comprises a single unit. The use of 300+ word 'texts' in this version of Topic Modelling gave us the opportunity to see how spread topics were across different journal articles. This proved instructive in distinguishing between the natural science journals AEE and PS on the one hand, and the social science journal REE on the other. Both AEE and PS had few 'niche' topics and many 'spread' topics, while REE was the reverse. GEC emerged as different from both, sitting midway between the two extremes. More significantly for our purposes, we have demonstrated greater variation in the two interdisciplinary journals than in their monodisciplinary counterparts. In other words, AEE has more variation than PS, and GEC has more variation than REE.

The two previous studies treated the topic wordlists at a high level of abstraction. The next study (sections 10.5 and 10.6) took a more interpretative approach to the topics and compared AEE with PS and GEC with REE. The AEE/PS comparison showed that AEE located research within a broader environmental approach

and to a very limited extent included a social perspective as well as a natural science one. There is also a small amount of evidence that AEE includes more 'niche' research approaches and focuses a little more on research processes themselves.

The range of topics in GEC is more demonstrably more diverse than that in REE. It includes topics that are antagonistic to other research paradigms. Both REE and GEC are more likely than the natural science journals to have research process topics occurring in only a few articles each, suggesting a relative lack of conformity in these journals compared to the natural science ones.

In the final, diachronic study no attempt was made to compare journals. However, it emerged that PS and AEE shared the feature of having topics that had a clear upward or downward trend, while GEC did so to a lesser extent and REE hardly at all. AEE is noticeable for taking a larger-scale perspective on environmental issues towards the end of the studied decade. GEC is noticeable for introducing a focus on the response of individual communities to environmental issues. A longer time span than ten years might be necessary to show clearer trends in all the journals.

The chapter has shown some of the potential for using a Topic Modelling approach in studies of this kind. Differences between the journals have been revealed, with some evidence pointing to the distinctive characteristics of the two interdisciplinary journals. One point about Topic Modelling that perhaps needs to be stressed is that of all the approaches to text analysis used in this book, this is the one that is most beyond the conscious control of the researcher writers. Writers may be very conscious of their strategy in writing introductions (see Chapter 5). They may consciously choose whether to write about 'observations' or 'arguments' (see Chapter 7), and they are aware that the words they use may have different meanings in other disciplines (see Chapter 6). Writing a topic as defined in this chapter, however, is entirely beyond human control, and indeed can never be the outcome of only one person writing.

From the studies in this chapter it would appear that interdisciplinary journals contain a broader range of topics than their monodisciplinary counterparts, even when the number of topics is held constant.

References

Baker, P. (2004). Querying keywords: Questions of difference, frequency, and sense in keywords analysis. *Journal of English Linguistics*, 32(4): 346–359.

Biber, D., & Barbieri, F. (2007). Lexical bundles in university spoken and written registers. *English for Specific Purposes*, 26(3): 263–286.

Blei, D., Ng, A., & Jordan, M. (2003). Latent Dirichlet allocation. *Journal of Machine Learning Research*, 3(1): 993–1022.

Firth, J. R. (1957). *Papers in linguistics 1934–1951*. Oxford: Oxford University Press.

Fletcher, W. (2002–2007). *KfNgram*. Annapolis, MD: USNA. Retrieved April 2019, from www.kwicfinder.com/kfNgram/kfNgramHelp.html

Gabrielatos, C. (2018). Keyness analysis: Nature, metrics and techniques. In C. Taylor & A. Marchi (Eds.), *Corpus approaches to discourse: A critical review* (pp. 225–258). Oxford: Routledge.

Grün, B., & Hornik, K. (2011). Topicmodels: An R package for fitting topic models. *Journal of Statistical Software*, 40(13). Retrieved from www.jstatsoft.org/v40/i13

Hunston, S. (2016). Interpreting quantitative data in Corpus linguistics. In D. Mayaffre, C. Poudat, L. Vanni, V. Magri & P. Follette (Eds.), *Proceedings of the 13th international conference on statistical analysis of textual data* (pp. 37–48). Nice: University Nice Sophia Antipolis.

Murakami, A., Thompson, P., Hunston, S., & Vajn, D. (2017). "What is this corpus about?" Using topic modeling to explore a specialized corpus. *Corpora*, 12(2): 243–277.

Porter, M. (1980). An algorithm for suffix stripping. *Program*, 14(3): 130–137.

R Core Team. (2015). *R: A language and environment for statistical computing*. Vienna: R Foundation for Statistical Computing. Retrieved from www.R-project.org/

Scott, M. (1997). PC analysis of key words–and key key words. *System*, 25(2): 233–245.

Scott, M., & Tribble, C. (2006). *Textual patterns: Key words and corpus analysis in language education*. Amsterdam: Benjamins.

Sinclair, J. (1991). *Corpus concordance collocation*. Oxford: Oxford University Press.

11 Conclusion

11.1 The research in this book

The aim of the research reported in this book has been to use Corpus Linguistics methods, combined with some discourse analysis methods, to investigate the language of journals centred around an interdisciplinary field. The data chosen for investigation comprised articles from journals devoted to environmental issues. Four journals were selected as the main data for the investigation. They were chosen to form a number of contrasts (see Chapter 3 for detail). Primarily, two of the journals are identified as monodisciplinary and two as interdisciplinary, based on citation data. Of the two interdisciplinary journals, one (*Agriculture, Ecosystems & Environment*) incorporates disciplines that are proximal, or closely related, while the other (*Global Environmental Change*) includes disciplines that are distal, or less closely related.

There were two main reasons for selecting interdisciplinary research discourse as the focus of the research. One was that interdisciplinary research is widely regarded as essential for solving the world's problems. As was shown in Chapters 1 and 2, research funding bodies and organisations such as universities generally see narrow disciplinarity as a barrier to carrying out really effective research, and they are keen to encourage, facilitate and fund interdisciplinary research. At the same time, interdisciplinary research is seen as problematic, taking people outside their intellectual comfort zones and requiring them to engage in collaborations that can be challenging. There is, therefore, a need for research of all kinds that addresses or even simply unpacks this difficulty. Our research has focused on research outputs: the articles that are published in journals (AEE and GEC) with an interdisciplinary readership. They have been compared with articles from journals (PS and REE) with a monodisciplinary readership. This is intended to complement research of a more ethnographic kind that focuses on aspects of interdisciplinary work such as meaning negotiation or social relations in meetings (e.g. Choi & Richards 2017).

The other reason for selecting the topic is to break new ground in focusing on journals that traditional approaches to academic discourse tend to avoid. In Chapter 4, for example, it was pointed out that much disciplinary-focused research avoids analysing journal articles that do not conform to an IMRD structure. It is

increasingly acknowledged that 'discipline' is insufficient as an identifier of the components of a corpus (Gray 2015 and Chapter 2), but alternative studies of interdisciplinary research discourse are rare (for exceptions, see, for example, Teich & Holtz 2009; Muguiro 2019). Breaking away from the 'comparing disciplines' model of corpus research, as this book has shown, requires some ingenuity in re-working accepted corpus methods (see Chapter 9) or adopting new ones (see Chapter 10).

The methods we have used in our study have been wide ranging (see Figure 1.1). They include some text analysis procedures, such as the study of article headings (Chapter 4); frequency comparisons based on groups of words, such as the study of status markers (Chapter 7); processing of concordance lines, such as the studies of *environment* or *important* (Chapter 6); an adaptation of Multidimensional Analysis (Chapter 9); and an exploration into Topic Modelling (Chapter 10). Of these, the quantitative methods (Chapters 9 and 10) have been the most innovative, but we have needed all the studies we have done to draw the conclusions presented in this chapter.

In this final chapter, we summarise our findings about the nature of interdisciplinary research discourse (section 11.2) and consider what the implications of this might for teaching writing skills (section 11.3). We then draw some further conclusions about the implications of our research for Corpus Linguistics more generally (section 11.4). We finish by considering where future research might lead (section 11.5).

11.2 Conclusions about interdisciplinary research discourse

Although we have used a variety of methods in our research, the lessons learned from each method have reinforced rather than contradicted each other. It is perhaps worth saying initially that we have of course learned something about the nature of research into the environment. The articles in our corpus have been written by researchers in a very large number of countries – this is clearly an international field – but perhaps surprisingly they have not been written by very large teams. Research conducted prior to the present studies (Hunston & Thompson 2016) shows that overall the number of authors per paper in the BEE11 corpus journals increased between 2000 and 2010, but that even in 2010 only around half the papers in AEE and PS have five authors or more, and that in the same year over half the papers in GEC have only one or two authors, with this proportion even higher in REE. This suggests that particularly in the social science areas researchers may work in large teams, but they do not write in large teams.

We have made the point throughout this book that interdisciplinarity in writing is discoursally constructed. This emerged particularly in the study of introductions (Chapter 5) and the study of the code gloss *in other words* (Chapter 8), where it was apparent that writers construe their readership as interdisciplinary and position themselves in relation to that readership as either 'insiders' or 'outsiders'. Our study of article introductions shows that a minority of writers choose to identify themselves as standing outside a mainstream of the research community, thereby

construing an explicitly interdisciplinary stance. Other writers are less overt, but still clearly exercise choice over writer and reader positioning. The monodisciplinary and interdisciplinary article introductions diverge in this respect. Similarly, our study of *in other words* shows some writers explicitly addressing two audiences, one inside and one outside the writer's own discipline. What became apparent also was that in a journal such as AEE it is only a minority of articles that exercise this choice.

This recognition that 'being interdisciplinary' is a matter of choice, and that an interdisciplinary identity is construed through writing, underpins our other findings. It has particular relevance for addressing variation within journals (that is, identifying why articles are not consistent within in a single journal) and for using this research in the teaching of academic writing.

The issue of heterogeneity is another of our key findings. Briefly put, the interdisciplinary journals we have investigated in detail (AEE and GEC) are not the same as each other. For many of the features we have studied, AEE is more similar to its monodisciplinary counterpart PS than it is to GEC. It is only when AEE and PS are compared that the unique features of AEE, those that identify it as interdisciplinary, are apparent. It will come as a disappointment to some that we have identified no linguistic features that could be said to be diagnostic of interdisciplinary writing, in the sense that an interdisciplinary journal or article of necessity includes those features and others do not. This means that interdisciplinarity in writing is less forensically identifiable than 'academic-ness' or 'science-ness' are. For the teacher of writing there is no infallible recipe for interdisciplinary writing that might be offered to the student. Nonetheless, we have identified the kinds of choices that writers make when construing their interdisciplinary identities.

For our third finding we turn away from consideration of individual articles and their writers to the investigations of variation within whole journals. One expectation in researching interdisciplinary discourse is that it will evince internal variation: the articles in an interdisciplinary journal will be more different from each other than the articles in a monodisciplinary journal are. This expectation has indeed been met in a number of studies we have undertaken. Key to these is the Multidimensional Analysis study (Chapter 9), which used a quantitative, bottom-up process adapted from Biber's (1988) methodology and concluded that there are more derived 'constellations' of articles in AEE than in PS and in GEC than in REE. The equally quantitative Topic Modelling study (Chapter 10) similarly suggests greater variation in types of topic in the interdisciplinary journals than in their monodisciplinary counterparts. This finding is confirmed by many of the other studies. For example, the study of article structures (Chapter 4) found greater variation in the interdisciplinary journals and the study of individual words, including status markers (Chapters 6 and 7), did so too.

Our final major finding is that the tripartite distinction proposed by Barry and Born (2013) in relation to the nature of interdisciplinary fields can be adapted to relate to interdisciplinary writing. We have noted that writers may present their research as 'integrative' to another discipline, as 'supportive' of it, or as 'antagonistic' to it. These distinctions can be observed in the interdisciplinary journals

AEE and GEC. We have further generalised from these observations to propose a discourse activity that we term 'disciplinary reflexivity'. This occurs when a writer names or otherwise refers to their own discipline or approach or that of others. Further investigation of these instances has often led to the identification of either an 'integrative' (or 'conciliatory') or an 'antagonistic' stance towards other disciplines.

This theme has emerged in a number of our studies. It was an outcome from the study of status markers (Chapter 7), of the words *science*/*scientist*/*scientific* (Chapter 6), and *inadequate* (Chapter 5), and of Topic Modelling (Chapter 10), as well as comprising a key finding from the study of introductions (Chapter 5). An important difference between the interdisciplinary natural science journal, AEE, and the monodisciplinary one, PS, is the amount of disciplinary reflexivity in AEE and its striking absence from PS. Furthermore, a key difference between the two interdisciplinary journals is that writers in AEE typically construe their relationship to other disciplines as integrative whereas those in GEC are just as likely to take an antagonistic stance. Almost all the instances of antagonism that we have encountered are in GEC.

To conclude this section, we might point out the difference between talking of an 'interdisciplinary journal' and talking of an 'interdisciplinary article'. The distinguishing characteristic of an interdisciplinary journal does not lie in any one set of language features but rather in the degree of variation when compared with monodisciplinary counterparts. Each article in an interdisciplinary journal could be resolutely monodisciplinary (though in practice few are), and the journal as a whole would still represent this degree of variation. For the individual article, it is clear as noted previously that writers exercise choice in construing their positioning in relation to their readership. They may choose to write in a 'monodisciplinary' way or may choose to position themselves as outsiders; they also select the degree of compliance or opposition they will adopt towards other disciplines – the degree to which they represent themselves as contributing to an interdisciplinary field or as challenging that field. Each focus – text or corpus – prioritises different results.

11.3 Teaching writing for an interdisciplinary readership

In considering the relevance of our research for the teaching of writing, we bear in mind a specific context: teaching early-career researchers (ECRs) who have completed PhD theses how to write for journals. This may be in the context of English for Academic Purposes, where the students are assumed not to be expert speakers of English, but it is not necessarily so. What advice would we have for a teacher in that situation?

The first point to make is that writing for an interdisciplinary journal is an important skill for ECRs in many subjects to acquire, and in the future it is likely to be more so. Writing for an interdisciplinary journal may therefore take its place in the curriculum alongside writing for peers within a discipline and presenting research to a popular or lay audience. Just as special training may be needed to

enable the brilliant biochemist to give a TED talk or give a lecture at the Hay Festival, so awareness needs to be raised of the particular requirements of an interdisciplinary journal.

As noted previously, there is no 'recipe' for writing for an interdisciplinary audience. However, many of the themes of an ECR writing course may simply be extended to take interdisciplinary writing into account. All student writers are trained to be aware of their audience and to shape their discourse to meet audience expectations, assumptions and needs. Just as expert and non-expert audiences may be envisaged, the audience that is academic but beyond the discipline might be addressed specifically. The ECR writer may be exposed to, say, introductions in journals such as GEC, such as those studied in Chapter 5, and invited to consider issues such as:

- How do writers from a completely 'outside' discipline (such as the discourse analysts who present a study of newspaper discourse for the *Global Environmental Change* journal [GEC 2004 Boykoff]) make their work relevant to environmental scientists? How do they frame the problem to which their work is a solution? How might this be done differently if the work were to be presented in an Applied Linguistics journal?
- What kind of detail might an environmental scientist (such as the scientists writing about grasslands in China [GEC 2006 Piao]) include when writing for GEC that they might omit if writing for a more specific journal?
- Where known antagonism exists, for example between natural scientists and social scientists who study climate change, how do various writers negotiate their own position? What options appear to be open for saying that other people are wrong, or that one's own work is essential?
- What values are attached to word choice, for example using the pronoun *we* or selecting an evaluative word such as *inadequate*? What alternative formulations exist?
- How might discourse markers such as *in other words* be used to switch focus from one audience to another? When might such expressions be useful or essential?

One important message here is that writing for an interdisciplinary audience is not the same as writing for a public or lay audience. The ECR who hopes to talk to the general public might learn to simplify information and to find a practical relevance for their research. Writing for an interdisciplinary audience requires more subtle adjustments and a more clear understanding of the intellectual expectations of that audience. It is an increasingly significant aspect of an ECR's career training.

11.4 Conclusions about Corpus Linguistics

We now turn to conclusions that might be drawn from our approach to Corpus Linguistics in this book. As noted in the introduction to this chapter, we have both made use of very traditional corpus investigation techniques, notably

carrying out word-based studies in which we identify comparative frequencies and scrutinise sample concordance lines, and introduced innovation, specifically in our approach to Multidimensional Analysis (Chapter 9, see also Thompson, Hunston, Murakami & Vajn 2017) and in our adoption of Topic Modelling (Chapter 10, see also Murakami, Thompson, Hunston & Vajn 2017). We now set these methods in the context of Corpus Linguistics traditions and practices.

We have, as noted, both adhered to and departed from standard Corpus Linguistics practices. Two concepts in particular are relevant here. The first is the practice in Corpus Linguistics of 'making the familiar strange'. The Key Word in Context format of concordance lines presents very ordinary language in such a way that its latent patterning (Sinclair & Coulthard 1975: 127), normally obscured, becomes apparent. Measures of collocation (Church & Hanks 1990), identification of keywords (Scott 1997), and observations of grammatical patterning (Hunston & Francis 2000) all reformat corpus data so that the output is new and yet relates to what any user of the language intuitively knows. A common reaction to any corpus-based observation is 'oh, of course!' – 'oh' because the information is new and unexpected, and 'of course' because once attention is drawn to the phenomenon it appears to have been known at some level all along.

Perhaps surprisingly, we regard Topic Modelling (Chapter 10) as a contribution to the tradition of 'making the familiar strange'. The topic wordlists are derived from an automated process and are presented to the researcher as an object for interpretation, rather in the way that concordance lines, or lists of collocates etc., are so presented. Seeing the BEE4 corpus as a series of 200 wordlists certainly presents a challenge of interpretation but allows insights to be drawn that are not apparent from other studies.

The second concept is the phenomenon of comparison. Comparison lies behind most Corpus Linguistic work. For the most part this involves comparing two (or more) corpora, representing discourse communities separated by context, geography or time, and characterising each against the background of the other. Other Corpus Linguistics work compares items within a single corpus, for example when measures of collocation are obtained or recurring phraseologies identified. Although corpus comparison has proven to be extremely valuable and even indispensable, it does present drawbacks. In particular, there is a risk that the contexts that are compared are essentialised through the comparison, similar to the ways that influential theorisations such as Becher's (1989) *Academic Tribes and Territories* run the risk of essentialising disciplines (Trowler 2014).

Although we would have liked to have broken away from the practice of comparison, it is so embedded in Corpus Linguistics practice that we have been unable to do so. Much of the research in this book involves comparing the journals in the BEE4 corpus and drawing conclusions. We have, however, attempted to acknowledge and even prioritise intra- and well as inter-corpus variation. In particular, our approach to Multidimensional Analysis (Chapter 9) represents an attempt to apply

a methodology that does not depend on comparing corpora. Instead of dividing texts a priori into corpora, which are then compared, we have operated a bottom-up method that identifies clusters or 'constellations' of texts. This means that the text groupings are the output from the method rather than its input. This avoids preconceived notions of difference and hopefully goes some way to avoid essentialising text categories.

In both the Multidimensional Analysis and Topic Modelling studies we have responded to the challenge of interpreting data derived from evidence that is internal to the corpus. There have been no precedents for how this might be done (other than our own previous work). Deciding how to identify text constellations, or how to compare topic wordlists, has been in each case an iterative and heuristic process. Future work may formalise this somewhat, but at the moment these studies remain a combination of the highly quantitative and algorithmic and the interpretative: the objective and the subjective. This, again, is in keeping with an extensive Corpus Linguistic tradition.

Returning to the notion of comparison: one reason that (sub-)corpora are often compared is that dividing available texts into sub-corpora imposes a welcome order upon data that is often messy. By retaining interdisciplinary journals as single entities we have worked with, and even embraced, this messiness. As with all Corpus Linguistics, we have been charged with reducing the amount of information we are working with at any one time, to make the task manageable. The challenge is to find ways of achieving that reduction while retaining 'messiness'. Selecting words or categories of word, and sampling concordance lines, is a traditional way of reducing information. Multidimensional Analysis and Topic Modelling offer alternatives to these. Word selection is essentially a top-down method, in that what is to be searched for is specified in advance. MDA and TM are bottom-up methods; not only do we not know what will be found, but we need to impose meaning on what is found.

Our research, then, has been a combination of the top-down and the bottom-up. It is important to retain that combination. The danger in top-down research is that only what is searched for will be found, and existing knowledge confirmed. The danger in bottom-up research is that it is so exploratory that it is difficult to interpret the results and to make sense of the data generated. In our own work we observe a productive tension between being open-ended and not limiting ourselves to predictable hypotheses, on the one hand, and finding meaningful things to say, on the other. This tension lies at the heart of Corpus Linguistics.

11.5 Towards the future

Research projects open doors rather than close them. Our research presented in this book paves the way for further investigations of discourse that is inherently messy because it cannot easily be divided into homogeneous sub-corpora. In addition we have suggested new ways of undertaking Multidimensional Analysis and applying Topic Modelling to corpora whose content is unknown.

Both these approaches have proved powerful tools for our investigations but we have found some limitations in the procedures that we followed. In the case of MD analysis, there is a case for identifying different linguistic features in the corpus either in addition to or in place of the general linguistic features that the Biber Tagger currently works with. These features could be ones that research has identified as characteristic of academic discourse, such as the uses of reporting verbs, or the range of noun phrase structures used. In relation to Topic Modelling, there is room for corpus linguists to evaluate different TM methods; in this book, we have used a method that works with single words that have been stemmed using the Porter stemmer, and other techniques (incorporation of bigram measures, uses of other stemmers/lemmatisers, etc.) can be tested in future.

In this research we have moved between text and corpus but we have not engaged with writers, or readers, at all. Future research might replicate Myers' (1990) work with biologists, where individuals are followed through the process of writing texts of different kinds. Explorations of how individual researchers approach the task of presenting their work for different journals would provide a useful corollary to the studies reported here.

References

Barry, A., & Born, G. (Eds.). (2013). *Interdisciplinarity: Reconfigurations of the social and natural sciences*. London: Routledge.

Becher, T. (1989). *Academic tribes and territories: Intellectual enquiry and the culture of disciplines*. Buckingham: Open University Press.

Biber, D. (1988). *Variation across speech and writing*. Cambridge: Cambridge University Press.

Choi, S., & Richards, K. (2017). *Interdisciplinary discourse: Communicating across disciplines*. London: Palgrave.

Church, K., & Hanks, P. (1990). Word association norms, mutual information, and lexicography. *Computational Linguistics*, 16: 22–29.

Gray, B. (2015). *Linguistic Variation in research articles: When discipline tells only part of the story*. Amsterdam: John Benjamins.

Hunston, S., & Francis, G. (2000). *Pattern Grammar: A corpus-driven approach to the lexical grammar of English*. Amsterdam: Benjamins.

Hunston, S., & Thompson, P. (2016). *Convergence and divergence in an interdisciplinary research field: The discourse of "global environmental change"*. Karlstadt Seminar, Germany.

Muguiro, N. (2019). *Interdisciplinarity and academic writing: A corpus-based study of three interdisciplines*. Unpublished PhD thesis, University of Birmingham.

Murakami, A., Thompson, P., Hunston, S., & Vajn, D. (2017). What is this corpus about? Using topic modelling to explore a specialised corpus. *Corpora*, 12(2): 243–277.

Myers, G. (1990). *Writing biology: Texts in the social construction of scientific knowledge*. Madison, WI: University of Wisconsin Press.

Scott, M. (1997). PC analysis of key words – and key key words. *System*, 25(1): 1–13.

Sinclair, J., & Coulthard, R. (1975). *Towards an analysis of discourse: The English used by teachers and pupils*. Oxford: Oxford University Press.

Teich, E., & Holtz, M. (2009). Scientific registers in contact: An exploration of the lexico-grammatical properties of interdisciplinary discourses. *International Journal of Corpus Linguistics*, 14(4): 524–548.

Thompson, P., Hunston, S., Murakami, A., & Vajn, D. (2017). Multi-dimensional analysis, text constellations, and interdisciplinary discourse. *International Journal of Corpus Linguistics*, 22(2): 153–186.

Trowler, P. (2014). Depicting and researching disciplines: Strong and moderate essentialist approaches. *Studies in Higher Education*, 39(10): 1720–1731.

Appendix A

Articles referred to in the body of the book

Agriculture, Ecosystems & Environment articles

Maul, J., Cooper, C. (2000). Water quality of seasonally flooded agricultural fields in Mississippi, USA 81/3:171–178

Smallwood, K.S., Geng, S., Zhang, M. (2001). Comparing pocket gopher (*Thomomys bottae*) density in alfalfa stands to assess management and conservation goals in northern California 87/1:93–109

Kumagai, T., Hidema, J., Kang, H., Sato, T. (2001). Effects of supplemental UV-B radiation on the growth and yield of two cultivars of Japanese lowland rice (Oryza sativa L.) under the field in a cool rice-growing region of Japan 83/1–2:201–208

Hoffmann, I., Gerling, D., Kyiogwom,U., Mané-Bielfeldt, A. (2001). Farmers' management strategies to maintain soil fertility in a remote area in northwest Nigeria 86/3:263–275

Goma, H., Rahim, K., Nangendo, G., Riley, J., Stein, A. (2001). Participatory studies for agro-ecosystem evaluation 87/2:179–190

Laloë, F., Lauckner, B., Piron, M., Rahim, K. (2001). Surveys and decisions in the context of multidisciplinary programmes: estimators and indicators 87/2:129–140

Esterhuizen, J., Liebenberg, G. (2001). The use of indicators within a comprehensive impact assessment approach in three South African research programmes 87/2:233–244

Morales, H., Perfecto, I., Ferguson, B. (2001). Traditional fertilization and its effect on corn insect populations in the Guatemalan highlands 84/2:145–155

Zurayk, R., el-Awar, F., Hamadeh, S., Talhouk, S., Sayegh, C., Chehab, A., al Shab, K. (2001). Using indigenous knowledge in land use investigations: a participatory study in a semi-arid mountainous region of Lebanon 86/3:247–262

Gerbens-Leenes,P., Nonhebel, S., Ivens, W. (2002). A method to determine land requirements relating to food consumption patterns 90/1:47–58

Peroni, N., Hanazaki, N. (2002). Current and lost diversity of cultivated varieties, especially cassava, under swidden cultivation systems in the Brazilian Atlantic Forest 92/2–3:171–183

Feoli, E., Vuerich, L., Zerihun, W. (2002). Evaluation of environmental degradation in northern Ethiopia using GIS to integrate vegetation, geomorphological, erosion and socio- economic factors 91/1–3:313–325

Holt-Giménez, E. (2002). Measuring farmers' agroecological resistance after Hurricane Mitch in Nicaragua: a case study in participatory, sustainable land management impact monitoring 93/1–3:87–105

Kujawa, K. (2002). Population density and species composition changes for breeding bird species in farmland woodlots in western Poland between 1964 and 1994 91/1–3:261–271

Doran, J. (2002). Soil health and global sustainability: translating science into practice 88/2:119–127

van den Berg, M., Driessen, P. (2002). Water uptake in crop growth models for land use systems analysis I. A review of approaches and their pedigrees 92/1:21–36

Nyffeler, M., Sunderland, K. (2003). Composition, abundance and pest control potential of spider communities in agroecosystems: a comparison of European and US studies 95/2–3:579–612

Cowell, S., Parkinson, S. (2003). Localisation of UK food production: an analysis using land area and energy as indicators 94/2:221–236

Dendoncker, N., Van Wesemael, B., Rounsevell, M., Roelandt, C., Lettens, S. (2004). Belgium's CO2 mitigation potential under improved cropland management 103/1:101–116

Pontius, R., Shusas, E., McEachern, M. (2004). Detecting important categorical land changes while accounting for persistence 101/2–3:251–268

Abensperg-Traun, M., Wrbka, T., Bieringer, G., Hobbs, R., Deininger, F., Main, B., Milasowszky, N., Sauberer, N., Zulka, K. (2004). Ecological restoration in the slipstream of agricultural policy in the old and new world 103/3:601–611

de Paz, J., Ramos, C. (2004). Simulation of nitrate leaching for different nitrogen fertilization rates in a region of Valencia (Spain) using a GIS – GLEAMS system 103/1:59–73

Thenail, C., Baudry, J. (2004). Variation of farm spatial land use pattern according to the structure of the hedgerow network (bocage) landscape: a case study in northeast Brittany 101/1:53–72

Manson, S. (2005). Agent-based modeling and genetic programming for modeling land change in the Southern Yucatán Peninsular Region of Mexico 111/1–4:47–62

Haileslassie, A., Priess, J., Veldkamp, E., Teketay, D., Lesschen, J.P. (2005). Assessment of soil nutrient depletion and its spatial variability on smallholders' mixed farming systems in Ethiopia using partial versus full nutrient balances 108/1:42370

Sawadogo, L., Tiveau, D., Nygård, R. (2005). Influence of selective tree cutting, livestock and prescribed fire on herbaceous biomass in the savannah woodlands of Burkina Faso, West Africa 105/1–2:335–345

Paranychianakis, N., Chartzoulakis, K. (2005). Irrigation of Mediterranean crops with saline water: from physiology to management practices 106/2–3:171–187

Vromant, N., Nguyen T., Chau, T. (2005). Overall effect of rice biomass and fish on the aquatic ecology of experimental rice plots 111/1–4:153–165

Parikesit,K., Takeuchi,A., Tsunekawa,O.S., Abdoellah (2005). Resource analysis of small-scale dairy production system in an Indonesian village – a case study 105/3:541–554

Fédoroff, E., Ponge, J., Dubs, F., Fernández-González, F., Lavelle, P. (2005). Small-scale response of plant species to land-use intensification 105/1–2:283–290

Mottet, A., Ladet, S., Coqué, N., Gibon, A. (2006). Agricultural land-use change and its drivers in mountain landscapes: A case study in the Pyrenees 114/2–4:296–310

Arriaga, L., Huerta, E., Lira-Saade, R., Moreno, E., Alarcón, J. (2006). Assessing the risk of releasing transgenic Cucurbita spp. in Mexico 112/4:291–299

Raiesi, F. (2006). Carbon and N mineralization as affected by soil cultivation and crop residue in a calcareous wetland ecosystem in Central Iran 112/1:13–20

Brady, J., Wallender, W., Werner, I., Fard, B.M., Zalom, F., Oliver, M., Wilson, B., Mata, M., Henderson, J., Deanovic, L., Upadhaya, S. (2006). Pesticide runoff from orchard floors in Davis, California, USA: A comparative analysis of diazinon and esfenvalerate 115/1–4:56–68

Brush, S., Perales H. (2007). A maize landscape: Ethnicity and agro-biodiversity in Chiapas, Mexico 121/3:211–221
Pascual, U., Perrings, E. (2007). Developing incentives and economic mechanisms for in situ biodiversity conservation in agricultural landscapes 121/3:256–268
Tomich, T., Timmer, D., Velarde, S., Alegre, J., Areskoug, V., Cash, D., Cattaneo, A., Ericksen, P., Joshi, L. et al. (2007). Integrative science in practice: Process perspectives from ASB, the Partnership for the Tropical Forest Margins 121/3:269–286
Zander, P., Knierim, A., Groot, J., Rossing, W. (2007). Multifunctionality of agriculture: Tools and methods for impact assessment and valuation 120/1:43–56
Wright, A., Dou, F., Hons, F. (2007). Soil organic C and N distribution for wheat cropping systems after 20 years of conservation tillage in central Texas 121/4:376–382
Jackson, L., Pascual, U., Hodgkin, T. (2007). Utilizing and conserving agrobiodiversity in agricultural landscapes 121/3:196–210
Farahbakhshazad, N., Dinnes, D., Li, C., B. Jaynes, D., Salas, W. (2008). Modeling biogeochemical impacts of alternative management practices for a row-crop field in Iowa 123/1–3:30–48
Yadav, V., Malanson, G. (2008). Spatially explicit historical land use land cover and soil organic carbon transformations in Southern Illinois 123/4:280–292
Pollini, J. (2009). Agroforestry and the search for alternatives to slash-and-burn cultivation: From technological optimism to a political economy of deforestation 133/1–2:48–60
Muthuri, C., Ong, C., Craigon, J., Mati, B., Ngumi, V., Black, C. (2009). Gas exchange and water use efficiency of trees and maize in agroforestry systems in semi-arid Kenya 129/4:497–507
Yan, H., Fu, Y., Xiao, X., Huang, H. He, H., Ediger, L. (2009). Modeling gross primary productivity for winter wheat – maize double cropping system using MODIS time series and CO2 eddy flux tower data 129/4:391–400
Manandhar, R., Odeh, I., Pontius, R. (2010). Analysis of twenty years of categorical land transitions in the Lower Hunter of New South Wales, Australia 135/4:336–346
Oom, S., Hester, A., Legg, C. (2010). Grazing across grass:shrub boundaries: Can spatial heterogeneity increase resistance? 139/1–2:159–166
Acosta-Martínez, V., Bell, C., Morris, B.,Zak, J., Allen, V. (2010). Long-term soil microbial community and enzyme activity responses to an integrated cropping-livestock system in a semi-arid region 137/3–4:231–240

Global Environmental Change articles

Kempton, W. (1991). Lay perspectives on global climate change 1/3:183–208
Wescoat, J. (1991). Managing the Indus River basin in light of climate change Four conceptual approaches 1/5:381–395
Boxer, B. (1991). Societal contexts of ocean pollution science Cross-national comparisons 1/2:139–156
Andresen, S., Wettestad, J. (1992). International resource cooperation and the greenhouse problem 2/4:277–291
Wynne, B. (1992). Uncertainty and environmental learning Reconceiving science and policy in the preventive paradigm 2/2:111–127
Dovers, S., Handmer, J. (1992). Uncertainty, sustainability and change 2/4:262–276
Jasanoff, S. (1993). India at the crossroads in global environmental policy 3/1:32–52
Yarnal, B. (1994). Agricultural decollectivization and vulnerability to environmental change A Bulgarian case study 4/3:229–243

Shackley, S., Skodvin, T. (1995). IPCC gazing and the interpretative social sciences A comment on Sonja Boehmer-Christiansen's: 'Global climate protection policy: the limits of scientific advice' 5/3:175–180

Munson, A. (1995). Should a biosafety protocol be negotiated as part of the Biodiversity Convention? 5/1:46204

van Asselt, M., Rotmans, J. (1996). Uncertainty in perspective 6/2:121–157

Sanderson, S. (1997). The hopeful state of human dimensions research 7/4:299–301

Martens, P., Kovats, R., Nijhof, S., de Vries, P., Livermore, M., Bradley, D., Cox, J., McMichael, A. (1999). Climate change and future populations at risk of malaria 9:S89-S107

Koop, G., Tole, L. (2001). Deforestation, distribution and development 11/3:193–202

Rasmussen, K., Fog, B., Madsen, J. (2001). Desertification in reverse? Observations from northern Burkina Faso 11/4:271–282

Farrell, A., VanDeveer, S., Jäger, J. (2001). Environmental assessments: four under-appreciated elements of design 11/4:311–333

Lambin, E., Turner, B., Geist, H., Agbola, S., Angelsen, A., Bruce, J. et al. (2001). The causes of land-use and land-cover change: moving beyond the myths 11/4:261–269

van Asselt Marjolein, B., Rijkens-Klomp, N. (2002). A look in the mirror: reflection on participation in Integrated Assessment from a methodological perspective 12/3:167–184

Harrington, L., Lu, M. (2002). Beef feedlots in southwestern Kansas: local change, perceptions, and the global change context 12/4:273–282

Patt, A., Gwata, C. (2002). Effective seasonal climate forecast applications: examining constraints for subsistence farmers in Zimbabwe 12/3:185–195

Jones, S. (2002). Social constructionism and the environment: through the quagmire 12/4:247–251

Nelson, K., de Jong, B. (2003). Making global initiatives local realities: carbon mitigation projects in Chiapas, Mexico 13/1:19–30

Brown, D. (2003). The importance of expressly examining global warming policy issues through an ethical prism 13/4:229–234

Carmichael, J., Tansey, J., Robinson, J. (2004). An integrated assessment modeling tool 14/2:171–183

Boykoff, M., Boykoff, J. (2004). Balance as bias: global warming and the US prestige press 14/2:125–136

Newell, B., Crumley, C., Hassan, N., Lambin, E., Pahl-Wostl, C., Underdal, A., Wasson, R. (2005). A conceptual template for integrative human – environment research 15/4:299–307

Antilla, L. (2005). Climate of scepticism: US newspaper coverage of the science of climate change 15/4:338–352

Adger, N., Arnell, N., Tompkins, E. (2005). Successful adaptation to climate change across scales 15/2:77–86

Pelling, M., High, C. (2005). Understanding adaptation: What can social capital offer assessments of adaptive capacity? 15/4:308–319

Bakkenes, M., Eickhout, B., Alkemade, R. (2006). Impacts of different climate stabilisation scenarios on plant species in Europe 16/1:19–28

Piao, S., Mohammat, A., Fang, J., Cai, Q., Feng, J. (2006). NDVI-based increase in growth of temperate grasslands and its responses to climate changes in China 16/4:340–348

Welp, M., de la Vega-Leinert, A., Stoll-Kleemann, S., Jaeger, C. (2006). Science-based stakeholder dialogues: Theories and tools 16/2:170–181

Adger, N. (2006). Vulnerability 16/3:268–281

Turner, R., Burgess, D., Hadley, D., Coombes, E., Jackson, N. (2007). A cost – benefit appraisal of coastal managed realignment policy 17/3–4:397–407
Neumayer, E. (2007). A missed opportunity: The Stern Review on climate change fails to tackle the issue of non-substitutable loss of natural capital 17/3–4:297–301
Armitage, D., Marschke, M., Plummer, R. (2008). Adaptive co-management and the paradox of learning 18/1:86–98
van Aalst, M., Cannon, T., Burton, I. (2008). Community level adaptation to climate change: The potential role of participatory community risk assessment 18/1:165–179
Enfors, E., Gordon, L. (2008). Dealing with drought: The challenge of using water system technologies to break dryland poverty traps 18/4:607–616
Lahsen, M. (2008). Experiences of modernity in the greenhouse: A cultural analysis of a physicist "trio" supporting the backlash against global warming 18/1:204–219
Urry, J. (2008). Governance, flows, and the end of the car system? 18/3:343–349
Webber, M., Barnett, J., Finlayson, B., Wang, M. (2008). Pricing China's irrigation water 18/4:617–625
Toni, F., Holanda, E. (2008). The effects of land tenure on vulnerability to droughts in Northeastern Brazil 18/4:575–582
Risbey, J. (2008). The new climate discourse: Alarmist or alarming? 18/1:26–37
Weinhold, D., Reis, E. (2008). Transportation costs and the spatial distribution of land use in the Brazilian Amazon 18/1:54–68
Brunner, S. (2008). Understanding policy change: Multiple streams and emissions trading in Germany 18/3:501–507
Pahl-Wostl, C. (2009). A conceptual framework for analysing adaptive capacity and multi-level learning processes in resource governance regimes 19/3:354–365
Grabherr, G. (2009). Biodiversity in the high ranges of the Alps: Ethnobotanical and climate change perspectives 19/2:167–172
Lövbrand, E., Stripple, J., Wiman, B. (2009). Earth System governmentality Reflections on science in the Anthropocene 19/1:41456
Salick, J., Fang, Z., Byg, A. (2009). Eastern Himalayan alpine plant ecology, Tibetan ethnobotany, and climate change 19/2:147–155
Bridges, K., McClatchey, W. (2009). Living on the margin: Ethnoecological insights from Marshall Islanders at Rongelap atoll 19/2:140–146
Lawrence, A. (2009). The first cuckoo in winter: Phenology, recording, credibility and meaning in Britain 19/2:173–179
Grasso, M. (2010). An ethical approach to climate adaptation finance 20/1:74–81
Weichselgartner, J., Kasperson, R. (2010). Barriers in the science-policy-practice interface: Toward a knowledge-action-system in research 20/2:266–277
Boykoff, M., Frame, D., Randalls, S. (2010). Discursive stability meets climate instability: A critical exploration of the concept of 'climate stabilization' in contemporary climate policy 20/1:53–64
Biesbroek, R., Swart, R., Carter, T., Cowan, C., Henrichs, T., Mela, H., Morecroft, M., Rey, D. (2010). Europe adapts to climate change: Comparing National Adaptation Strategies 20/3:440–450
Füssel, H. (2010). How inequitable is the global distribution of responsibility, capability, and vulnerability to climate change: A comprehensive indicator-based assessment 20/4:597–611
Uusi-Rauva, C., Tienari, J. (2010). On the relative nature of adequate measures: Media representations of the EU energy and climate package 20/3:492–501

Marin, A. (2010). Riders under storms: Contributions of nomadic herders' observations to analysing climate change in Mongolia 20/1:162–176

Barthel, S., Folke, C., Colding, J. (2010). Social – ecological memory in urban gardens – Retaining the capacity for management of ecosystem services 20/2:255–265

Zeitoun, M., Allan, J.A., Mohieldeen, Y. (2010). Virtual water 'flows' of the Nile Basin, 1998–2004: A first approximation and implications for water security 20/2:229–242

Plant Science articles

Jeleńska, J., Deckert, J., Kondorosi, E., Legocki, A. (2000). Mitotic B-type cyclins are differentially regulated by phytohormones and during yellow lupine nodule development 150/1:29–39

Refshauge, S.J., Nayudu, M. (2001). Invasion by Pseudomonas syringae pv. coriandricola is responsible for bacterial blight of coriander 161/3:621–625

Shin, R., Lee, G., Park, C., Kim, T., You, J., Nam, Y., Paek, K. (2001). Isolation of pepper mRNAs differentially expressed during the hypersensitive response to tobacco mosaic virus and characterization of a proteinase inhibitor gene 161/4:727–737

Magioli, C., Barrôco, R., Benıć io Rocha, C., Santiago-Fernandes, L., Mansur, E., Engler, G., Margis-Pinheiro, M., Sachetto-Martins, G. (2001). Somatic embryo formation in Arabidopsis and eggplant is associated with expression of a glycine-rich protein gene (Atgrp-5) 161/3:559–567

Sharma, R., Sharma, A., Fujimura, T., Machii, H. (2001). Stress-responsive enhanced expression of mitochondrial processing peptidase gene after UV exposure in mulberry 161/4:807–812

Kraepiel, Y., Agnès, C., Thiery, L., Maldiney, R., Miginiac, E., Delarue, M. (2001). The growth of tomato (Lycopersicon esculentum Mill.) hypocotyls in the light and in darkness differentially involves auxin 161/6:1067–1074

Fukuhara, M. (2002). Acoustic characteristics of botanical leaves using ultrasonic transmission waves 162/4:521–528

Dickstein, R., Hu, X., Yang, J., Ba, L., Coque, L., Kim, D., Cook, D., Yeung, A. (2002). Differential expression of tandemly duplicated Enod8 genes in Medicago 163/2:333–343

Nandi, S., Suzuki, Y., Huang, J., Yalda, D., Pham, P., Wu, L., Bartley, G., Huang, N., Lönnerdal, B. (2002). Expression of human lactoferrin in transgenic rice grains for the application in infant formula 163/4:713–722

Pandey, G., Reddy, V., Reddy, M., Deswal, R., Bhattacharya, A., Sopory, S. (2002). Transgenic tobacco expressing Entamoeba histolytica calcium binding protein exhibits enhanced growth and tolerance to salt stress 162/1:41–47

Jain, K., Kataria, S., Guruprasad. K.N. (2003). Changes in antioxidant defenses of cucumber cotyledons in response to UV-B and to the free radical generating compound AAPH 165/3:551–557

Baldan, B., Bertoldo, A., Navazio, L., Mariani, P. (2003). Oligogalacturonide-induced changes in the developmental pattern of Daucus carota L. somatic embryos 165/2:337–348

Zoriniants, S., Nosov, A., Monforte-Gonzalez, M., Mendes-Zeel, M., Loyola-Vargas, V. (2003). Variation of nuclear DNA content during somatic embryogenesis and plant regeneration of Coffea arabica L. using cytophotometry 164/2:141–146

Ristic, Z., Wilson, K., Nelsen, C., Momcilovic, I., Kobayashi, S., Meeley, R., Muszynski, M., Habben, J. (2004). A maize mutant with decreased capacity to accumulate chloroplast protein synthesis elongation factor (EF-Tu) displays reduced tolerance to heat stress 167/6:1367–1374

Balconi, C., Cavallini, A., Natali, L., Motto, M. (2004). Evaluation of in vitro methods to define the role of the cob and pedicel-placento-chalazal tissues in the amino acid supply to the developing maize endosperm 166/5:1313–1320

Zhang, Z-H., Li, P., Wang, L-X., Hu, Z-L., Li-Huang Zhu, L-H., Zhu, Y-G. (2004). Genetic dissection of the relationships of biomass production and partitioning with yield and yield related traits in rice 167/1:43678

Lanfermeijer, F., Jiang, G., Ferwerda, M., Dijkhuis,J., de Haan, P., Yang, R., Hille, J. (2004). The durable resistance gene Tm-22 from tomato confers resistance against ToMV in tobacco and preserves its viral specificity 167/4:687–692

Courdavault, V., Burlat, V., St-Pierre, B., Giglioli-Guivarc'h, N. (2005). Characterisation of CaaX-prenyltransferases in Catharanthus roseus: relationships with the expression of genes involved in the early stages of monoterpenoid biosynthetic pathway 168/4:1097–1107

Pandey, D.M., Kang, K-H, Yeo, U-D. (2005). Effects of excessive photon on the photosynthetic pigments and violaxanthin de-epoxidase activity in the xanthophyll cycle of spinach leaf 168/1:161–166

Caravaca, F. Alguacil, M., Hernández, J., Roldán, A. (2005). Involvement of antioxidant enzyme and nitrate reductase activities during water stress and recovery of mycorrhizal Myrtus communis and Phillyrea angustifolia plants 169/1:191–197

Shinogi, T., Hamanishi, Y., Otsu, Y., Wang, Y., Nonomura, T., Matsuda, Y., Toyoda, H., Narusaka, Y., Tosa, Y., Mayama, S. (2005). Role of induced resistance in interactions of Epilachna vigintioctopunctata with host and non-host plant species 168/6:1477–1485

Wu, Q.-S., Wan, X.-Y., Su, N., Cheng, Z.-J. , Wang, J.-K., Lei, C.-L., Zhang, X., Jiang, L., Ma, J.-F., Wan, J.-M. (2006). Genetic dissection of silicon uptake ability in rice (Oryza sativa L.) 171/4:441–448

Tallón, C., Quiles, M.J. (2007). Acclimation to heat and high light intensity during the development of oat leaves increases the NADH DH complex and PTOX levels in chloroplasts 173/4:438–445

Singh Brar, H., Singh, Z., Swinny, E., Cameron, I. (2008). Girdling and grapevine leafroll associated viruses affect berry weight, colour development and accumulation of anthocyanins in 'Crimson Seedless' grapes during maturation and ripening 175/6:885–897

Arencibia, A., Bernal, A., Yang, L., Cortegaza, L., Carmona, E., Pérez, A.,Hu, C., Li, Y., Zayas, C., Santana, I. (2008). New role of phenylpropanoid compounds during sugarcane micropropagation in Temporary Immersion Bioreactors (TIBs) 175/4:487–496

Luo, Z., Xie, J., Xu, T., Zhang, L. (2009). Delay ripening of 'Qingnai' plum (Prunus salicina Lindl.) with 1- methylcyclopropene 177/6:705–709

Zhang, J., Ma, H., Chen, S., Ji, M., Perl, A., Kovacs, L., Chen, S. (2009). Stress response proteins' differential expression in embryogenic and non-embryogenic callus of Vitis vinifera L. cv. Cabernet Sauvignon – A proteomic approach 177/2:103–113

Huang, N., Li, C., Lee, J., Yen, H. (2010). Cytosine methylation changes in the ice plant Ppc1 promoter during transition from C3 to Crassulacean acid metabolism 178/1:41–46

Saidi, H., Ladouce, N., Hadhri, R., Grima-Pettenati, J., Drira, N., Gargouri-Bouzid, R. (2010). Identification and characterization of differentially expressed ESTs in date palm leaves affected by brittle leaf disease 179/4:325–332

Zhu,Y., Zeng, H., Dong, C., Yin, X., Shen, Q., Yang, Z. (2010). microRNA expression profiles associated with phosphorus deficiency in white lupin (Lupinus albus L.) 178/1:23–29

Ye, H., Foley, M., Gu, X. (2010). New seed dormancy loci detected from weedy rice-derived advanced populations with major QTL alleles removed from the background 179/6:612–619

Liu, J., Liang, D., Song, Y., Xiong, L. (2010). Systematic identification and expression analysis of BREVIS RADIX- like homologous genes in rice 178/2:183–191

Resource and Energy Economics articles

Thompson, A. (2001). The Hotelling Principle, backwardation of futures prices and the values of developed petroleum reserves – the production constraint hypothesis 23/2:133–156

Dupont, D., Grafton, R.Q., Kirkley, J., Squires, D. (2002). Capacity utilization measures and excess capacity in multi-product privatized fisheries 24/3:193–210

Bjørner, T., Jensen, H. (2002). Energy taxes, voluntary agreements and investment subsidies – a micro-panel analysis of the effect on Danish industrial companies' energy demand 24/3:229–249

Aiken, D.V., Pasurka, C. (2003). Adjusting the measurement of US manufacturing productivity for air pollution emissions control 25/4:329–351

Mulder, P., de Groot, H., Hofkes, M. (2003). Explaining slow diffusion of energy-saving technologies; a vintage model with returns to diversity and learning-by-using 25/1:105–126

Boxall, P., Rollins, K., Englin, J. (2003). Heterogeneous preferences for congestion during a wilderness experience 25/2:177–195

Maddison, D. (2003). The amenity value of the climate: the household production function approach 25/2:155–175

Armsworth, P., Kendall, B., Davis, F. (2004). An introduction to biodiversity concepts for environmental economists 26/2:115–136

Feng, H., Zhao, J. (2006). Alternative intertemporal permit trading regimes with stochastic abatement costs 28/1:24–40

Dijkstra, B. (2007). An investment contest to influence environmental policy 29/4:300–324

Naghavi, A. (2007). Can R&D-inducing green tariffs replace international environmental regulations? 29/4:284–299

Kverndokk, S., Rosendahl, K. (2007). Climate policies and learning by doing: Impacts and timing of technology subsidies 29/1:58–82

Flachaire, E., Hollard, G. (2007). Starting point bias and respondent uncertainty in dichotomous choice contingent valuation surveys 29/3:183–194

Martinet, V., Doyen, L. (2007). Sustainability of an economy with an exhaustible resource: A viable control approach 29/1:17–39

Hénin, P., Schubert, K. (2008). How long can excess pollution persist? The non-cooperative case 30/2:277–293

Antelo, M., Bru, L. (2009). Permit markets, market power, and the trade-off between efficiency and revenue raising 31/4:320–333

Golub, A., Hertel,T., Lee, H-L., Rose, S., Sohngen, B. (2009). The opportunity cost of land use and the global potential for greenhouse gas mitigation in agriculture and forestry 31/4:299–319

Chattopadhyay, S. (2009). The random expenditure function approach to welfare in RUM: The case of hazardous waste clean-up 31/1:58–74

Drechsler, M., Wätzold, F., Johst, K., Shogren, J. (2010). An agglomeration payment for cost-effective biodiversity conservation in spatially structured landscapes 32/2:261–275

Jiang, S. (2010). Aquaculture, capture fisheries, and wild fish stocks 32/1:65–77

Botzen, W., Bouwer, L., van den Bergh, J. (2010). Climate change and hailstorm damage: Empirical evidence and implications for agriculture and insurance 32/3:341–362

Nishitani, K. (2010). Demand for ISO 14001 adoption in the global supply chain: An empirical analysis focusing on environmentally conscious markets 32/3:395–407

Zhang, W., van der Werf, W., Swinton, S. (2010). Spatially optimal habitat management for enhancing natural control of an invasive agricultural pest: Soybean aphid 32/4:551–565

Aronsson, T., Backlund, K., Sahlén, L. (2010). Technology transfers and the clean development mechanism in a North – South general equilibrium model 32/3:292–309

Appendix B

The table below shows the first 40 items on the list of lemmas investigated for Chapter 7, in order of total normalised frequency. The table is presented without editing. It illustrates the variation in frequency of many of the lemmas among the four journals. It also shows that this method of identifying status markers introduces some 'noise'. For example, the word *case* is included here because *case* can be followed by an appositive that-clause and can mark status. Many instances of *case*, however, occur in the phrase *in case* (*of*), which does not mark status, so the overall frequency is inflated. The verb *PROVIDE* occurs in the list because it is sometimes followed by *that*. It does not indicate status, however, and most instances in this corpus have the meaning 'give'. The two most frequent nouns in the list, *effect* and *result*, mark connections between propositions rather than their status. Nouns such as *problem* and *risk* indicate both status and a judgement about the possible state of affairs. The adjectives *good* and *positive*, in this corpus, rarely indicate status. The list could be edited to exclude such words, but given the methodology in the chapter, which for the most part is not affected by the 'noise', and the possibility of introducing further error by such editing, this course of action has been rejected.

The table shows each lemma, its part of speech, its frequency per million words in each journal corpus, and the total frequency per million words.

	Lemma	*POS*	*AEE*	*GEC*	*PS*	*REE*	*Total*
1	Effect	Noun	2015.24	1267.75	1295.92	3104.23	7683.14
2	Result	Noun	1649.32	1275.18	1822.06	2602.15	7348.70
3	Show	Verb	1646.64	903.07	3182.19	1242.34	6974.24
4	Analysis	Noun	1479.73	1188.71	2175.19	1194.86	6038.50
5	Condition	Noun	1050.11	798.58	1328.30	1189.23	4366.22
6	Case	Noun	560.60	1200.65	364.80	2231.22	4357.26
7	Find	Verb	1111.01	617.70	1111.54	1044.40	3884.64
8	Provide	Verb	723.43	1281.01	427.36	1021.07	3452.86
9	Consider	Verb	724.96	819.53	293.87	1215.78	3054.14
10	Suggest	Verb	588.76	738.90	1092.05	601.05	3020.77

	Lemma	*POS*	*AEE*	*GEC*	*PS*	*REE*	*Total*
11	Scenario	Noun	513.46	1850.44	7.41	630.82	3002.13
12	See	Verb	440.84	1067.24	298.94	1156.24	2963.27
13	Information	Noun	482.50	953.14	162.57	1188.43	2785.64
14	Indicate	Verb	732.86	467.58	996.02	538.29	2734.75
15	Problem	Noun	268.83	1263.77	53.23	1082.22	2668.05
16	Risk	Noun	333.30	1265.89	17.97	1024.28	2641.45
17	Approach	Noun	476.64	1169.62	159.97	770.02	2576.25
18	Assume	Verb	345.66	482.96	74.08	1598.78	2501.50
19	Observation	Noun	697.82	185.92	1261.35	315.41	2460.50
20	Good	Adjective	475.00	592.50	229.94	894.74	2193.18
21	Report	Verb	573.98	305.53	951.57	310.58	2141.67
22	Estimate	Verb	636.54	392.52	196.32	885.89	2111.28
23	Expect	Verb	329.48	510.38	227.60	1040.38	2107.74
24	Require	Verb	388.60	747.92	323.36	617.15	2077.03
25	Point	Noun	395.22	540.78	221.69	734.62	1891.92
26	Line	Noun	153.27	149.05	1289.89	265.53	1857.74
27	Allow	Verb	356.36	437.35	275.48	774.05	1843.24
28	Remain	Verb	442.11	475.80	359.72	469.90	1747.53
29	Decision	Noun	160.03	746.06	359.72	469.90	1747.53
30	Calculate	Verb	815.04	249.04	278.09	340.36	1682.53
31	Positive	Adjective	316.87	271.85	230.48	841.64	1660.84
32	Assessment	Noun	280.68	1066.98	32.10	126.33	1506.09
33	Know	Verb	219.91	346.38	467.83	358.06	1392.17
34	Indicator	Noun	656.92	508.69	38.83	141.61	1346.05
35	Knowledge	Noun	152.89	738.90	76.97	370.13	1338.88
36	Assumption	Noun	133.91	363.62	21.81	792.55	1311.89
37	Appear	Verb	262.21	331.26	395.94	276.79	1266.20
38	Implication	Noun	99.12	208.46	69.56	873.02	1250.16
39	Fact	Noun	196.47	359.90	197.01	401.51	1154.88
40	Probability	Noun	145.76	220.40	25.11	741.86	1133.12

Index