LEGITIMATION IN THE ACADEMIC DISCOURSE OF TWO EARTH SCIENCE SUB-DISCIPLINES

PAVEL ADAMEK

NATIONAL INSTITUTE OF EDUCATION

NANYANG TECHNOLOGICAL UNIVERSITY

2022

LEGITIMATION IN THE ACADEMIC DISCOURSE OF TWO EARTH SCIENCE SUB-DISCIPLINES

PAVEL ADAMEK

A thesis submitted to the

National Institute of Education,

Nanyang Technological University

in partial fulfilment of the requirement for the degree of

Doctor of Philosophy

2022

Statement of Originality

I hereby certify that the work embodied in this thesis is the result of original research, is free of plagiarised materials, and has not been submitted for a higher degree to any other University or Institution. In addition, I declare that to the best of my knowledge, this thesis is free of plagiarism, and contains no material previously published or written by another person, except where due reference has been made in the text.

13/09/2022

Date

Supervisor Declaration Statement

I have reviewed the content and presentation style of this thesis and declare it is free of plagiarism and of sufficient grammatical clarity to be examined. To the best of my knowledge, the research and writing are those of the candidate except as acknowledged in the Author Attribution Statement. I confirm that the investigations were conducted in accord with the ethics policies and integrity standards of Nanyang Technological University and that the research data are presented honestly and without prejudice.

13/9/2022

. . . . . .

Date

. . . . . .

Authorship Attribution Statement

This thesis does not contain any materials from papers published in peer-reviewed journals or from papers accepted at conferences in which I am listed as an author.

13/09/2022

Date

Acknowledgements

I am profoundly grateful to the following people who have helped and supported me in the last seven years.

I wish to express a special thank you to my advisor, Associate Professor Csilla Weninger, who patiently and expertly guided me through my struggles with formulating my research framework, analysing my data, writing up my thesis, and understanding life at graduate school. I also wish to thank my original supervisor, Professor Hu Guangwei, who introduced me to the concept of knowledge-knower structures, thus propelling my interest in the underlying organising principles of disciplinary literacy.

I am grateful to other faculty members at the English Language and Literature academic group, especially to Associate Professor Ramona Tang and Dr Guo Libo, who taught me in my elective courses, and answered many a question about language and literacy.

Last but not least, I appreciate the unwavering support of my wife, Jade Hu, whose patience and understanding created the necessary conditions at home for me to finish my studies.

Table of contents

Acknowledgements I

Table of contents II

List of tables VI

List of figures VIII

Summary IX

Chapter I: Introduction 1

1.1 Institutional and personal context for the study 1

1.2 Earth sciences as an academic field 3

1.2.1 Major characteristics of earth science research 3

1.3 Academic discourse and perspectives on it 6

1.3.1 Corpus perspectives 9

1.3.2 Genre perspectives 13

1.3.3 Ethnographic perspectives 19

1.3.4 The three perspectives and this study 23

1.4 Purpose and the general research question 27

1.5 Significance 28

1.6 Thesis roadmap 31

1.7 Summary 33

Chapter II: Literature review 34

2.1 Academic literacies and English for academic purposes 34

2.1.1 Overview 34

2.1.2 The issue of specificity 38

2.2 Key genre: The research article 44

2.3 Disciplinarity and discourse community 51

2.3.1 Discourse community 53

2.3.2 Disciplines hard, soft, pure, and applied 58

2.3.3 Discourse communities and the nature of knowledge construction 61

2.3.4 Texts as discipline-specific social interactions 64

2.4 Disciplinary variation in the use of linguistic features 65

2.4.1 Themes 67

2.4.1.1 Academic and disciplinary vocabulary 72

2.4.2 Citations 79

2.4.3 Expressing certainty (hedging and boosting) 88

2.4.4 Personal pronouns 96

2.5 Appraisal of research into disciplinary discourse in the EAP tradition 104

2.6 Summary 106

Chapter III: Theoretical framework 109

3.1 Overview: Towards a social realist perspective on knowledge 109

3.2 Formative influences on Legitimation Code Theory 120

3.2.1 Bourdieu's thinking tools 121

3.2.2 Bernstein's pedagogic device and knowledge structures 132

3.3 Maton's Legitimation Code Theory (LCT) 139

3.3.1 Specialization 144

3.3.2 Semantics 161

3.3.3 Autonomy 164

3.3.4 Temporality 168

3.3.5 Density 169

3.4 Collaboration of LCT and SFL 171

3.5 Application of LCT in research into organising principles of academic fields of practice 172

3.6 EAP studies that have drawn on LCT 175

3.7 Specific research questions 182

3.8 Summary 184

Chapter IV: Methodology 185

4.1 Research design **185**

4.2 Collecting corpus data **185**

4.2.1 Constructing the sub-corpora 185

4.2.1.1 Selecting sub-disciplines of earth sciences 186

4.2.1.2 Selecting academic journals 190

4.2.1.3 Selecting RAs 191

4.3 Relating the concepts of LCT's Specialization to selected linguistic features **194**

4.3.1 Specialization involved in disciplinary practices 195

4.3.2 The first step: Identifying knowledge claims 197

4.3.3 Visibility of knowledge: Topical Themes 204

4.3.4 Visibility of knowers: Citation practices 218

4.3.5 Authoritativeness of knowledge: Epistemic certainty of propositions 224

4.3.6 Authoritativeness of knowers: Use of exclusive personal pronouns 234

4.4 Post-coding analysis **243**

4.5 Summary **245**

Chapter V: Results **247**

5.1 Visibility of knowledge and knowers **247**

5.1.1 Visibility in marine chemistry 248

5.1.2 Visibility in seismology 253

5.2 Authoritativeness of knowledge and knowers **259**

5.2.1 Authoritativeness in marine chemistry 260

5.2.2 Authoritativeness in seismology 267

5.3 Summary of quantitative results **274**

5.4 Qualitative results **275**

5.4.1 Topical Themes in marine chemistry and seismology 276

5.4.2 Citations as grammatical Subjects: The movers and shakers in marine chemistry and seismology 280

5.4.2.1 Influential individuals, institutions, and countries 280

5.4.2.2 Pace of knowledge shift 285

5.5 Summary of qualitative results **301**

Chapter VI: Discussion **303**

6.1 LCT perspective **303**

6.1.1 Knowledge and knowers in marine chemistry and seismology 303

6.1.2 Knowledge insights in marine chemistry and seismology 313

6.1.3 Potential born knower gaze 314

6.1.4 Age of knowledge and pace of knowledge change 316

6.1.5 Contribution to LCT 318

6.2 EAP perspective **326**

6.2.1 Discourse community 327

6.2.2 Specific vs. general EAP 329

6.3 Research into linguistic features **333**

6.3.1 Themes 334

6.3.2 Citation practices 336

6.3.3 Hedging and boosting 343

6.3.4 First-person pronouns 345

6.4 Summary **349**

Chapter VII: Conclusion **351**

7.1 Mapping disciplinarity in earth sciences **351**

7.2 Pedagogical implications **354**

7.3 Limitations **359**

7.4 Recommendations for further research **365**

References **368**

Appendices **411**

Appendix I: A list of journals from which the corpus research articles were sourced **411**

Appendix II: The coding sheet used to identify ranking clauses **413**

Appendix III: The coding sheet for hedges and boosters **424**

List of tables

Table 2.1 Disciplinary academic word lists 74

Table 2.2 Types of integral and non-integral citations modified from Swales (1990, p. 149) 82

Table 2.3 Hyland's (2005b) interpersonal model of metadiscourse 89

Table 2.4 Functional taxonomies of *we* 102

Table 4.1 Combinations of disciplinary names and keywords 192

Table 4.2 A rubric with types of Themes and their example realisations 208

Table 4.3 Theme constituents in the three Mood systems 209-210

Table 4.4 The coding table for Visibility of knowledge 214

Table 4.5 Types of integral and non-integral citations modified and combined from Hyland (1999), Swales (1990), and Thompson (2002) 221

Table 4.6 The coding table for Visibility of knowers 222

Table 4.7 In-text reference formats of the corpus articles 223

Table 4.8 The coding table for Authoritativeness of knowledge 227-228

Table 4.9 The coding table for Authoritativeness of knowers 239-240

Table 5.1 Descriptive statistics for Visibility of knowledge in marine chemistry 248

Table 5.2 Descriptive statistics for Visibility of knowers in marine chemistry 250

Table 5.3 Descriptive statistics for positions on ER and SR Visibility axes in marine chemistry 252

Table 5.4 Descriptive statistics for Visibility of knowledge in seismology 254

Table 5.5 Descriptive statistics for Visibility of knowers in seismology 256

Table 5.6 Descriptive statistics for positions on ER and SR Visibility axes in seismology 258

Table 5.7 Descriptive statistics for Authoritativeness of knowledge in marine chemistry 260

Table 5.8 Descriptive statistics for Authoritativeness of knowers in marine chemistry 262

Table 5.9 Usage frequency of first-person pronouns in marine chemistry 263

Table 5.10 Descriptive statistics for positions on ER and SR Authoritativeness axes in marine chemistry 265

Table 5.11 Descriptive statistics for Authoritativeness of knowledge in seismology 267

Table 5.12 Descriptive statistics for Authoritativeness of knowers in seismology 270

Table 5.13 Usage frequency of first-person pronouns in seismology 270

Table 5.14 Descriptive statistics for positions on ER and SR Authoritativeness axes in seismology 273

Table 5.15 The leading countries and institutions from which prominently cited authors hail 284

Table 5.16 Percentages of the corpus RAs published during different periods 286

List of figures

Figure 3.1 The specialization plane. 152

Figure 3.2 The epistemic plane - insights. 156

Figure 3.3 The social plane - gazes. 158

Figure 3.4 The semantic plane. 164

Figure 3.5 The autonomy plane. 167

Figure 3.6 The modalities of Temporality. 168

Figure 3.7 The modalities of Density. 170

Figure 5.1 Positions of all marine chemistry RAs in the Visibility plane. 252

Figure 5.2 Positions of all seismology RAs in the Visibility plane. 258

Figure 5.3 Positions of all marine chemistry RAs in the Authoritativeness plane. 266

Figure 5.4 Positions of all seismology RAs in the Authoritativeness plane. 273

Figure 5.5 Proposed positions of marine chemistry and seismology in the epistemic Visibility plane. 280

Figure 7.1 Scaled quantitative instrument (Maton & Howard, 2016, p. 58). 356

Summary

Researching (sub-)disciplinary literacy has a long tradition in English for Academic Purposes (EAP)/English for Specific Purposes (ESP). The main purpose of the research has been to illuminate how individual (sub-)disciplines vary in their communicative practices, mostly in the usage of individual linguistic features (e.g., stance nouns) or in information structuring of research article sections. The number of (sub-)disciplines studied this way is vast, although some fields have been on the margins of interest, such as earth sciences. Also, EAP/ESP research has focused little on larger principles underlying disciplinary knowledge building, and the linguistic features that reflect these principles, both at disciplinary and sub-disciplinary levels.

To partially fill these gaps, I select two representative earth science sub-disciplines, marine chemistry and seismology, and integrate corpus-based investigations of 120 discussion and conclusion sections of experimental articles from 2012 to 2017 from the two sub-disciplines with an approach from the sociology of knowledge, Legitimation Code Theory (LCT). Through the lens of LCT's Specialization dimension, I quantitatively show the degree to which knowledge and knowers are visible (prominent), and the level of authoritativeness (epistemic certainty) with which propositional content and knowers are presented. Specifically, I show that in the two sub-disciplines investigated, marine chemists present their specialist knowledge more visibly than seismologists, but seismology knowers figure more prominently than marine chemists. Conversely, knowledge in seismology is presented with a somewhat higher degree of epistemic certainty than in marine chemistry, but seismology knowers are overall less authoritative.

Further intriguing differences are revealed qualitatively. While disciplinary topical Themes concerned with objects of study strongly frame legitimate problems of both sub-disciplines, topical Themes concerned with methods and research products suggest that legitimate approaches in marine chemistry are framed less strongly than in seismology. Citation analyses identifying lead authors in both sub-fields show that few individuals are cited more than the rest. Types of knowledge that marine chemists are prominently cited for reveal that findings which are site-or area-specific and numerical may become obsolete faster, while comparable findings in seismology remain relevant throughout decades.

This research contributes to scholarship in multiple areas. First, I propose two new realisations of LCT's Specialization–whose main premise is that practices comprise knowledge-knower structures that specialise actors and discourses–by operationalising relations to knowledge structures through topical Themes and hedging and boosting, while relations to knower structures are operationalised through citations and personal pronouns. Second, I empirically investigate Specialization quantitatively, which is rare in LCT research. Third, I illustrate qualitative differences in what disciplinary knowers are cited for, where prominent knowers are based, and suggest reasons for differing age of cited knowledge. Fourth, I expand EAP/ESP research into academic literacy by focusing on little explored earth sciences. The results can raise junior scholars' awareness of the form of published (sub-)disciplinary knowledge and of the existing variability in the studied linguistic features, thus illuminating the potential for emphasis, highlighting, and self-promotion.

Chapter I: Introduction

This chapter begins with brief descriptions of the institutional and personal contexts from which the idea for this study originated, as well as a brief introduction to earth sciences–the academic field of focus in this study. Next, the study is situated within the tradition of academic discourse research, outlining the major perspectives and approaches, as well as gaps in extant research. The chapter then presents the overarching research question and the study's contributions.

1.1 Institutional and personal context for the study

The Earth Observatory of Singapore (EOS) at Nanyang Technological University is a research institute that focuses on geo-hazards and environmental challenges in Southeast Asia. Since its inception in 2008, the institute's main areas of interests have been earthquakes, volcanic eruptions, tsunamis, and climate change. The institute's postgraduate programme, a PhD in Earth Sciences, was later administratively transferred under a new interdisciplinary school–the Asian School of the Environment (ASE) in 2015, which also offers an undergraduate degree of Bachelor of Science in Environmental Earth Systems Science. The interdisciplinary nature of the institute's research is apparent from the integration of earth sciences, engineering and technology, environmental life science, and the social sciences.

I joined the EOS in January 2013, first on a part-time basis, later becoming a full-time Language and Communication Specialist. The majority of my work time is devoted to teaching writing and presentation skills classes to undergraduate and PhD students. Apart from teaching, I also extensively work with PhD students, post-doctoral

fellows, and faculty members on their texts, with the most common genre being the experimental research article, and the most common part-genre the conference abstract.

The people I work with come from a variety of countries, and a fair percentage of my students are proficient users of English. In the case of post-graduate students, their educational backgrounds provided them with at times completely different disciplinary induction: There are geologists, teachers, archaeologists, mathematicians, geophysicists, biologists, and computer engineers. Thus, for the majority, their PhD is a journey of socialisation into a new disciplinary culture (Casanave, 2008; Kwan, 2009)– here, I am referring to the Hallidayan notion of culture at the level of registers (Halliday, 2016). At a more refined level, the PhD students need to become familiar with the culture of two sub-disciplines, since some have been required to research two distinct areas, for example, microbiology of tsunami sediments and the role of GPS in tectonic plate movements. In my job, I try to sensitise my students to the myriad ways in which sub-disciplinary discourse practices may differ: I have taught lessons on, for example, the research article macrostructure, the moves and steps in the research article part-genres, citation practices, use of first-person pronouns (both inclusive and exclusive), metadiscourse, the use of active and passive structures, technical vocabulary, and nominalizations. At the same time, while exploring these topics and working with, among others, volcanologists, geodesists, seismologists, geochemists, climate scientists, and microbiologists, I realised that earth sciences comprise an intriguing amalgam of sub-disciplines, whose nature and larger organising principles are likely different from some related disciplines (e.g., physics, chemistry, or biology). Yet, the characteristics of the sub-fields or the entire field, as well as concomitant questions about disciplinarity,

have remained on the margin of research in English for Academic Purposes (EAP) and English for Specific Purposes (ESP).

1.2 Earth sciences as an academic field

Since this study focuses on a discipline that readers may be unfamiliar with, I believe it will be advantageous to provide some general background information about the field of earth sciences.

Earth sciences is an umbrella term for disciplines studying the Earth's processes taking place in the spheres of the land, oceans, and atmosphere (Reed, 2008). As a unified discipline, earth sciences can be considered relatively young. Reed (2008) states that while humans have observed and studied these processes for millennia, at the turn of the last century, geosciences were a compilation of independent fields. Then, during the International Geophysical Year of 1957-58, scientists introduced a new global perspective recognising that processes in the individual spheres interacted. The recognition eventually led to a merger of the spheres and their sub-disciplines under the field of earth sciences.

1.2.1 Major characteristics of earth science research

The specific issues that characterise earth science research have been pertinently summarised by Kleinhans et al. (2010). Since the discipline aims to both explain the past processes and determine (plus predict) the causal relationships of forces shaping the Earth, it is both a descriptive historical science and a natural science associated with causality and law formation. Disciplinary theories and hypotheses are oftentimes underdetermined, which means that direct causes and effects are difficult to establish.

This difficulty arises from several (at times simultaneous) sources: (a) Processes happen on time scales longer than human life or recorded history; (b) they cannot be instrumentally detected as they take place deep inside the planet's interior; and (c) processes could have been caused by a number of initial conditions. Thus, in the case of weak underdetermination, not enough evidence is available that would point unequivocally to one theory; in the case of strong underdetermination, no amount of evidence can clearly point to one theory. In the face of scarce or unavailable evidence, earth scientists follow two strategies: multiple working hypotheses and inference to the best explanation (abduction).

Also, as Kleinhans et al. (2010) point out, to deal with scanty evidence, earth scientists have found a solution in combining causal and narrative explanations. Narratives contribute most to explanation, because causal explanations cannot provide enough details. Narratives are in fact more than mere descriptions. Narratives reconstruct a sequence of events, from initial conditions to causal forces that initiated interactions. These reconstructions rely on robust process explanations (i.e., the underlying phenomena) and actual-sequence explanations (events step by step). However, narratives do not explain events by reducing or generalising them to principles of physics or chemistry. Rather, evidence for various phenomena is integrated. For example, glacial landforms, erratic boulders, and uneven rates of sea level through time indicate the same cause–ice ages. This kind of explanation can also be called common cause explanation, which rests on "the principle of the common cause" (Cleland, 2013, p. 7): Seemingly unrelated phenomena may be associated with one cause, because most events influence the environment in multiple ways, leaving behind diverse pieces of evidence that can point back to the same event.

The above-mentioned specific features of earth sciences have led to an ongoing debate among philosophers of science about the nature of the discipline. Various opinions have been put forward, for example:

- Distinct features of geology do not seem to conform to the traditional notions of how experimental science progresses (Raab & Frodeman, 2002);
- The field of earth sciences as a historical natural discipline is not inferior in its epistemology to experimental science (Cleland, 2001, 2013);
- Geology (a sub-discipline of earth sciences) is neither merely a derivative (Frodeman, 1995) nor an unautonomous discipline absolutely reducible to physics and chemistry, though some degree of reductionism may still apply (Kleinhans et al., 2010); and
- While reductionism has been rejected in fields like biology and psychology, earth sciences are still deemed reducible (Kleinhans et al., 2010).

Ultimately, though, philosophical debates about the nature of the field may be only a starting point. For as Kastens et al. (2009) have posited, understanding how geoscientists learn and think necessitates collaboration with scientists of thinking and learning. This collaboration is where my doctoral research comes in. To shed light on one facet of the nature of earth sciences, I investigate some of the larger organising principles underlying knowledge creation at the level of sub-disciplines. To this end, I bring together perspectives from the sociology of knowledge and applied linguistics. Sociologically, (sub-)disciplines can be seen as fields of practice, in which (sub-)disciplinary members, in competition with each other, are constantly trying to "shape what is defined as achievement to match their own practices" (Maton, 2014a, p. 17). On a deeper level then, (sub-)disciplinary conventions, albeit possibly but

imperceptibly fluid at all times, are a reflection of what practices constitute legitimate contributions to (sub-)disciplinary knowledge, and how (sub-)disciplinary knowledge is created. To reveal a facet of legitimate practices, (sub-)disciplinary conventions can be analysed linguistically, since language reflects social order and at the same time shapes it (Jaworski & Coupland, 2014; Rampton et al., 2004). Analysing (sub-)disciplinary discourse, especially writing, to which we turn next, can thus reveal not only what is important to a (sub-)disciplinary community, but also how the community is constructed through writing (Hyland, 2013a).

1.3 Academic discourse and perspectives on it

Put simply, academic discourse is the way academics think and use language (Hyland, 2009). The foremost function of academic discourse is knowledge creation (Basturkmen, 2009; Hyland, 2004), especially through publication of research articles (Canagarajah, 2002). Publication is not the end, though, since new knowledge needs to be negotiated with discourse community members–readers (Hyland, 2014). In order for the negotiation to be successful (i.e., for a writer's proposition to be accepted), the writer and the writing must be persuasive (Hyland, 2004).

To understand why academic writing can be considered knowledge creation, and why persuasion is crucial in this process, we need to inspect the authority that is often attached to academic writing, and how the authority is negotiated. As Hyland and Salager-Meyer (2008) and Hyland (2008a) state, academic writing holds a privileged position because it is perceived to consist of seemingly impartial descriptions, which can be verified, replicated and falsified, of the world around us and its phenomena. Thus, academic discourse has been seen as a channel for communicating objective facts

(Hyland & Salager-Meyer 2008), "a guarantee of reliable knowledge" (Hyland, 2008a, p. 2), and as representing "the language of truth" (Lemke, 1995, p. 151). However, Hyland (2008a) reminds us that this guarantee of reliability, which is perpetuated by the practice of writing as if from a detached, non-partisan and objective position, is based on our belief that what is observed in one case (or several cases) will be observed in all other cases. This belief has been criticised by philosophers of science (e.g., Cleland, 2001), because in reality, induction and falsification represent probabilities rather than certainties, and are contingent on a finite body of evidence, background assumptions, and experimental conditions. Moreover, observations need to be interpreted through the lens of assumptions of a particular theory, which is only good to the extent it can describe a whole category of observed phenomena and predict new phenomena (Hawking, 1993). This interdependence of observations and theory makes observations liable to being interpreted in at least one other (competing) way (Hyland, 2009, 2014), since "we do not know what reality is independent of a theory" (Hawking, 1993, p. 44), which also means that there is no impartial theory that can be used to test observations (Hawking, 1993). Since other competing theories may be used to interpret what has been observed, there is always the possibility that readers will reject our propositions (Hyland, 2008a). Then, "the persuasiveness of academic discourse does not depend upon the demonstration of absolute fact, empirical evidence or impeccable logic; it is the result of effective rhetorical practices, accepted by community members" (Hyland, 2004, p. 8). Such a state of affairs leads us to focus more on the ways that academics present their propositions rather than on how they demonstrate absolute truth (the proof).

According to Hyland (2006a), in order to communicate successfully, discourse community members need to present and organise their findings in ways that will be meaningful for their audience–disciplinary peers. Such organisational patterns draw on disciplinary conventions that represent what it means to be "doing" a particular discipline; in essence, the conventions represent the disciplinary view of the world and what the community considers as knowledge. In other words, Hyland (2004) notes that in order to be persuasive, a writer must meet two conditions. The first condition is that of adequacy, which means that propositions must be plausible and within the epistemological framework of the discipline. Adequacy can be demonstrated through the use of technical vocabulary and appropriate forms of argument, the choice of what is to be highlighted, what is expressed with caution, and what with conviction. The second condition is that of acceptability, which means that the writer has to "create a professionally acceptable persona and an appropriate attitude, both to readers [to show respect] and the information being discussed [the writer's authority and expertise]" (p. 13).

It might be interesting to note that whereas well-established discourse community members are at times allowed to breach established conventions of communication (Bizzell, 1992), the conditions of adequacy and acceptability must still be met. For example, during my time at the EOS, I have experienced a case in which a single-authored journal submission of the former EOS Director (an eminent researcher and a native English speaker) was rejected on language grounds. The former EOS Director mentioned to me that the editor had objected to the manuscript's narrative style, and suggested a "blander" writing style instead.[1] The differences between the

[1] The anecdote echoes Myers' (1985) comment that practically all researchers have been asked for "alteration that, in the writer's view, diminished the value of the article" (p. 593).

rejected and accepted versions would constitute an interesting case study of the bounds of adequacy and acceptability in earth sciences.

We see, then, that academic discourse is intimately connected with disciplinary communities' social activities, the way they perceive and understand the world, and their epistemological beliefs (Hyland, 2009). The activities, ways of understanding, and beliefs have been studied by applied linguists from a variety of perspectives, among which are contrastive rhetoric, psycholinguistic processing, and critical discourse analysis (J. Flowerdew, 2002). However, linguists working within EAP and ESP paradigms have most notably focused on corpus-based work, genre analysis, and social context investigation (Charles, 2013)–perspectives that have probably impacted the teaching of academic discourse most directly (J. Flowerdew, 2002).

1.3.1 Corpus perspectives

Today, corpus linguistics is one of the most prominent approaches to investigating linguistic features (Anthony, 2013), and an approach that has contributed most to the description of academic discourse (Hyland, 2012a). At the heart of the approach is work with machine-readable collections of texts: corpora. A corpus can be defined as "a collection of pieces of language text in electronic form, selected according to external criteria to represent, as far as possible, a language or language variety as a source of data for linguistic research" (Sinclair, 2004, p. 23).

Since corpora usually consist of large quantities of textual data (hundreds of thousands to millions of words), although smaller, specialised corpora or corpora of video files might also be used (McEnery & Hardie, 2012), they allow us to see and describe repeated patterns of language use (Charles, 2013). As an approach, corpus

linguistics has been described as having these major characteristics (Biber et al.,1998, p. 4):

1) It is empirical, analyzing the actual patterns of language use in natural texts. It utilises a large and principled collection of natural texts, known as a "corpus", as the basis for analysis;

2) It makes extensive use of computers for analysis, using both automatic and interactive techniques; and

3) It depends on both quantitative and qualitative analytical techniques.

While the above-mentioned characteristics are generally agreed on in the corpus linguistics community (Anthony, 2013), an important point of debate has concerned the scope of the approach. According to Tognini-Bonelli (2001), computer corpora were first seen as a mere tool with which vast quantities of data could be processed. At the next stage, the corpus approach became associated with a distinctive methodological framework. Then, in a move beyond methodology, the quantity of data processed brought a radical change in theoretical and qualitative perspective, "trigger[ing] philosophical repositioning" (p. 48): The corpus approach "has offered insights into the language that have shaken the underlying assumptions behind many established theoretical positions in the field" (p. 48). Thus, Tognini-Bonelli (2001) argued that corpus linguistics "is much more than 'the methodological basis' for the enquiry, that it carries with it and becomes inseparable from the actual consequences of this stance; in this way, it becomes a discipline in its own right and acquires its own domain" (p. 49).

The concomitant of the proposed new theoretical status is a distinction between corpus-based and corpus-driven approaches (Biber, 2009, 2015; McEnery & Hardie, 2012; Tognini-Bonelli, 2001). Tognini-Bonelli (2001) states that an approach which is corpus based assumes that existing language features are well grounded in a linguistic theory, and the aim is to analyse how these given forms are used and how they vary. In contrast, a corpus-driven approach does not make assumptions about existing linguistic features. It does not separate data from theory building–"the commitment of the linguist is to the integrity of the data as a whole…the theory has no independent existence from the evidence" (p. 84)–and so the patterns and variations observed in corpora then become the basis for linguistic hypotheses and theoretical statements. The characteristics of the two approaches, it might be noted, should not be taken to imply any hierarchy, that is, the corpus-driven approach "is not in any way superior" (Biber, 2009, p. 279). Nevertheless, this binary distinction may be more theoretical than practical, since, for example, corpus-driven studies (even pattern grammar studies, which are often exemplified as the most developed kind of corpus-driven research) have often employed corpus-based methods as well, and could thus be viewed as hybrids (Biber, 2009). This hybridising, together with the facts that, firstly, corpus-based studies may also base most of their claims on corpus evidence rather than on pre-existing theories (McEnery & Hardie, 2012); secondly, that corpus-driven studies incorporate in research practice a plethora of methodologies (Biber, 2009); and thirdly, that Tognini-Bonelli's characteristics of the two approaches suggest the distinction is based on the extent to which each approach relies on corpus evidence, while "respect" (p.151) for corpus evidence is, in fact, a common ground (McEnery & Hardie, 2012), has led some scholars to question whether "the corpus itself has a theoretical status", and instead

posit that "*all* corpus linguistics can justly be described as corpus-based" (McEnery & Hardie, 2012, p. 6, original emphasis). The distinction has thus been viewed as "ultimately misleading" (McEnery & Hardie, 2012, p. 151).

In relation to language feature investigation, corpora of different sizes have been used. Some are large corpora built to offer a view of the national varieties of English, for example, the 100-million-word British National Corpus (BNC), which represents the complete range of varieties in spoken and written British English, or its American equivalent, the one-billion-word Corpus of Contemporary American English (COCA). Other corpora have been assembled to study specifically written academic discourse, for example, the British Academic Written English (BAWE) corpus, the Michigan Corpus of Upper-level Student Papers (MICUSP), or corpora of various disciplinary writings. With the help of these corpora, researchers have illuminated, for example, the differences between spoken and written expressions of stance in academic settings (e.g., Biber, 2006) and among disciplines (e.g., Hyland, 2005; Jiang & Hyland, 2015); evaluative use of reporting verbs in citations (e.g., Hyland, 2002a); the roles played by personal pronouns when employed by students and experienced writers (e.g., Harwood 2005a), across disciplines (e.g., Harwood, 2005c), and inclusively and exclusively (e.g., Harwood, 2005b); and use of relative clauses in journal descriptions (e.g., Tse & Hyland, 2010). In addition, corpora have been used to create word lists, such as the Academic Word List (Coxhead, 2000), and to critically appraise such lists (e.g., Hyland & Tse, 2007), or they have revealed multi-word lexical bundles across disciplines (e.g., Hyland, 2008b).

Corpus approaches have been criticised for focusing on texts extracted from their real-life contexts (Hyland, 2012a). This bottom-up focus on micro-features, as

opposed to the initial macro focus of genre analysts, then ushers in a "fragmented" view of the discoursal world (Swales, 2002, p. 152). Also, corpus linguists have been preoccupied with identifying sentence-level lexico-grammatical features instead of focusing on larger patterns in discourse, often interpreting their results post-hoc (i.e., without prior theoretical underpinning), though this imbalance is being increasingly addressed by researching extended discourse units (textlinguistic approaches) (L. Flowerdew, 1998, 2002).

1.3.2 Genre perspectives

Widening our perspective from lexico-grammatical features to include whole texts, the concept of genre has been extremely helpful in academic discourse analysis (Charles, 2013). While genre has been studied from multiple standpoints, such as folklore, literary studies, rhetoric, and linguistic, all the perspectives share several points in common (Swales, 1990). Among them are a recognition that genres exist within discourse communities; an interest in repeated structural patterns; and "an emphasis on communicative purpose and social action" (p. 45). The last point, I believe, is perhaps the most significant, because although genres need to encompass multiple factors (e.g., the message, format, audience), communicative purpose is paramount, and any major shift in purpose will possibly necessitate a change in genre (Bhatia, 1993).

Within applied linguistics, the concept of genre has been approached from three varied scholarly traditions: New Rhetoric, Systemic Functional Linguistics, and English for Specific Purposes, the last two being similar in their focus on language components of genres, but dissimilar in their target audiences (Hyon, 1996). The New Rhetoric tradition places more emphasis on social purposes under the premise that "a rhetorically

sound definition of a genre must be centred not on the substance or the form of discourse but on the action it is used to accomplish" (Miller, 1984, p. 151). Similarities in form are only the first and basic claim to a genre status, but genres need to include a pragmatic aim of accomplishing a social action (Miller, 1984). The emphasis on situational contexts and social actions then led rhetoricians to employ ethnographic rather than linguistic methods of analysis (Hyon, 1996).

For Systemic Functional Linguists, genres are "staged, goal oriented social processes": staged because they comprise multiple steps; goal-oriented because they are associated with achieving communicative functions; and social because of their link to specific audiences (Martin & Rose, 2008, p. 6). As social processes, genres cannot be considered in isolation: Individual genres exist within a network of genres, and since each one has a recurrent form, they together embody the social practices of a given culture (Martin & Rose, 2008). Scholars working in this tradition focus on explicating the sentence-level and text-level linguistic features in order to help mainly primary and secondary students write school genres (Hyon, 1996).

The ESP genre approach is mainly associated with work done by Swales (1990, 2004) and Bhatia (1993). Especially Swales' (1990) *Genre analysis* was such a seminal publication with arguably the biggest impact on academic discourse analysis and its applications in university writing courses (Cheng, 2015; Hyland, 2015) that the *Journal of English for Academic Purposes* recently devoted a special issue to commemorating its publication. In his book, Swales (1990) offers a famous paragraph-long definition of genre, the best-known part of which is perhaps the first two sentences:

> A genre comprises a class of communicative events, the members of which share some set of communicative purposes. These purposes are recognised by the

> expert members of the parent discourse community, and thereby constitute the rationale for the genre. (p. 58)

Swales (1990) acknowledges that the term genre is "extremely slippery" (p. 33), and this sentiment is further echoed in his later work (Swales, 2004), where he expresses his scepticism whether genres should be confined to definitions at all. The first reason he gives is that no definition of the term can accurately encompass all instances. The second reason is that definitions may blur our vision in the case of new genres. Instead, Swales (2004) offers a suite of six metaphors, such as genre as frame, as standard, as biological species, which could be variously called upon to contribute to our understanding of genres in specific contexts.

Following the 1990 publication, genre analysis within the field of academic writing research was mostly applied to the research article (RA), which is the most prestigious academic written genre (Hyland & Salager-Meyer, 2008). The trend became widespread in EAP/ESP/ERPP (English for Research and Publication Purposes), with research actively focusing on analysing the various rhetorical functional units, called moves and steps, of the Introduction-Methods-Results-Discussion (IMRD) part-genres. The numerous studies of disciplinary writing produced various part-genre models, some of which, for example, the CARS (Create-a-Research-Space) model (Swales, 1990; Swales & Feak, 2004), have proved their general applicability across a multitude of disciplines (Hyland & Salager-Meyer, 2008). Yet, further research has shown that even the CARS model is susceptible to broad disciplinary differences (Shehzad, 2010) and even to differences within individual sub-disciplines (Kanoksipatham, 2012; Samraj, 2002), which led some researchers to caution against indiscriminately applying models gleaned from one discipline to writing in another discipline (Brett, 1994). While

introductions are perhaps the most widely studied part-genre, other parts of the RA have also received their share of attention, for example, methods (Cotos et al., 2017; Lim, 2006; Peacock, 2011), results (Brett, 1994; Williams, 1999), and discussion (Basturkmen, 2012; Cotos et al., 2016; Holmes, 1997; Peacock, 2002). An interesting study in this regard is Kanoksilapatham's (2015) wide-ranging investigation of three (civil, software, and biomedical) engineering sub-fields for the rhetorical structures of all the major RA part-genres.

In addition to identifying the rhetorical structure of research article part-genres of various (sub-)disciplines in the form of moves and steps, genre-focused studies have also investigated whether moves and steps are associated with particular lexico-grammatical features. For example, Lim (2012) explored how writers of management RAs establish niche in introduction sections through contrastive discourse markers, negative verb phrases, or negative modifiers. Le and Harrington (2015) identified word clusters (e.g., *results suggest that*, *results are consistent with*, *a possible explanation for*) frequently used in the steps of the commenting on results move in the discussion sections of quantitative articles in applied linguistics, finding that different steps were realised by different clusters. Similarly, Cortes (2013), who mapped lexical bundles in the moves and steps in almost 1,400 RA introductions in 13 disciplines, found that certain lexical bundles strongly correlated with particular moves/steps, or even triggered the moves (i.e., they were at the beginning of the first sentence of the move, such as *little is known about the*). Another interesting finding was identification of long word clusters of more than six words, with some being complete units (e.g., *the rest of the paper is organised as follows*, *the purpose of the present study was to*). This form-function focus in genre-based research has also included investigation of potential

correlation between moves/steps and syntactic complexity. For example, Lu et al. (2020) found that social science writers from six disciplines employed syntactic complexity indices (sentence length, nominalisation, finite and non-finite dependent clauses, and left-embeddedness–the number of words preceding the main verb) variably in RA introductions to accomplish rhetorical goals (e.g., sentences realising a gap in move 2 of the CARS model were markedly more frequently left-embedded, whereas outlining the structure of the paper in move 3 was significantly less frequently expressed through long sentences, finite dependent clauses, or left-embeddedness).

The development of this research area from move/step identification to linguistic realisation of rhetorical goals testifies to the continued interest in and need for genre analysis studies. To support further research and improve its reliability and validity, some genre analysts have addressed relevant methodological issues, such as challenges and solutions when developing a move-and-step coding scheme (Moreno & Swales, 2018), suitability of various statistical measures of inter-coder agreement (Rau & Shih, 2021), and even operational criteria for defining what a research article actually is in order to move beyond common-sense understanding of genre in sampling, and in order to theoretically anchor genre analysis of research articles (van Enk & Power, 2017).

Another area of interest in genre analysis is associated with the broadening of perspective from text/language-focused analyses to a more dynamic view of genre as a developing phenomenon, shaped by the individual members of a discourse community (Flowerdew & Peacock, 2001). While genre analysts have always maintained that genres are situated in social contexts, they started focusing on how genres are constructed, interpreted and exploited, almost "reversing" the text-context emphases (Bhatia & Gotti, 2006, p. 10). One way genres may develop is under the influence of

digital communication. Digitisation, Xia (2020) notes, has afforded new communicative contexts in which texts can be multimodal or be interconnected with other texts by hyperlinks. This "hypertextuality further produces possible ways of constructing genres, and at the same time, it enables the readers to consume genres in their own customisable manners" (p. 146). The impact of digitisation has been investigated, for example, by Hefner (2018), who analysed a new genre: video methods articles from the *Journal of Visualised Experiments*, which aims to support researchers from a variety of physical and life sciences disciplines by visually demonstrating experimental protocols to enhance understanding of methodologies traditionally communicated in writing, and thus preempt/solve communication problems. The results show that the new hybrid genre draws on traditional written methods articles, lab reports, and conference presentations. Innovations, compared to written RAs, include a wide range and large quantity of visuals, which are argued to "make significant additional meanings possible" (p. 37), and, in the use of language, explicit interactional strategies, especially in the introduction and conclusion sections, reminiscent of presentations, with some of the direct audience engagement being also interpreted as humorous by viewers. In another study of digital media-led genre evolution, Perez-Llantada (2013) investigated a corpus of so-called Articles of the Future (AofF) prototypes, in which the content was enhanced by online research highlights, graphical abstracts, interactive graphs, embedded videos, and hyperlinks. The findings suggest that AofF prototypes reflect conventional RA structural practices; in other words, while multimodality and hypertextuality may enhance the content of the major academic genres like the research article or abstract, they may not necessarily result in innovation in terms of linguistic or rhetorical form and function (Perez-Llantada, 2016).

1.3.3 Ethnographic perspectives

The shift of focus to social structures, social identities, and social practices as an alternative to text-only focus (Hyland, 2013b) was a response to the "quest for a thicker description of language use" (Bhatia, 2002, p. 21). Long before these calls from EAP/ ESP scholars, the question of how language relates to its socio-cultural conditions of usage and production had been a central focal area of theory and research in a variety of approaches to language. Starting in the early 1960s, linguists who were dissatisfied with then-dominant formalist approaches influenced by Saussure and later Chomsky, which viewed language as an abstract or idealised system[2], turned to investigating mundane spoken (face-to-face) interactions to gain insights into the role of language in social organisation. This social focus did not always mean abandoning the formalist perspective, with some researchers (e.g., Labov (1972)) still selecting language rules (linguistic forms) as their point of departure, but examined their variation in the context of community interaction. Others (e.g., Hymes, 1974) argued that understanding people's linguistic behaviour had to start from social functions of communicative events, and then identifying the language structures employed to achieve the social functions. In other words, with the focus on communicative situational use, the goal was to understand how linguistic competence of individual community members to produce and interpret speech was regulated by speech community norms. Such norms could refer to categories bounded by gender, socio-economic class, geographical location, race, and so on (Weninger & Williams, in press). This focus motivated attempts to model communicative contexts, for example, in terms of setting, participants, ends (as

[2] For example, Labov (1972) noted that linguistics was "defined in such a way as to exclude the study of social behavior or the study of speech" (p.187) and that "it is questionable whether sentences communicating nothing to anyone are a part of language" (p. 183).

goals and as outcomes), act (message form and message content), key ("tone, manner, or spirit in which an act is done" (p. 57), instrumentalities (channels and forms of speech), norms (of interaction and of interpretation), and genres–Hymes' (1974) model for which he used the mnemonic "SPEAKING" (p. 62). Contexts were also seen as highly interactional, with spoken exchanges created and sustained by participants' recognition and interpretation of cues during their involvement (i.e., Gumperz's contextualisation cues) (Weninger & Williams, in press). Later, major research interest shifted from examining community norms to studying individual speakers (interactional participants), and to issues of identity and style (e.g., Bucholtz & Hall, 2005), although these had already been explored in pioneering studies, for example, in Goffman's virtual and actual identities (i.e., expectations of identities before interaction vs. identities formed during encounters) (Weninger & Williams, in press). This shift from community to individuals was motivated by the premise that speakers draw on language as a resource "in order to come across as particular types of people and to negotiate social relationships" (Weninger & Williams, in press, p. 9).

Interest in the social aspect of language use was also manifested in the changing conceptualisation of literacy. For example, in 1984, Street distinguished between autonomous and ideological models of literacy, with the former portraying literacy as acquisition of technical skills and strategies (such as decoding or sound/shape recognition/perception), and the latter linking literacy to cultural and power contexts in society (Street, 2013). Street also noted that, in fact, all literacy models arose out of ideological frameworks, since the autonomous model portrayed literacy acquisition as neutral, thus masking power relations between groups whose literacy practices varied. In 1990, James Gee introduced the term "new literacy studies" (p. 49), often referred to

as the NLS[3], to describe a body of research from various disciplines (e.g., anthropology, linguistics, education, social psychology) that opposed traditional notions of literacy, which viewed reading and writing as autonomous, asocial, and a reflection of an individual's cognitive skills. For NLS scholars, literacy had no meaning outside of cultural contexts of use, also implying that different contexts changed literacy's effects (Gee, 2015): Reading and writing were socially embedded within specific discourses (Gee, 1990), where discourse with a small *d* denoted language used in conversations, reports, stories, etc., while Discourse with a capital *D* denoted the specific ways of speaking/listening and reading/writing together with specific ways of effecting "recognisable social identities" through "acting, interacting, valuing, feeling, dressing, thinking, believing" (Gee, 2015, p. 171) as well as using tools and technologies.

Within EAP/ESP, focus on studying social aspects of learning and language use increased only in the past 20 years, as evidenced by the surge in research into, for instance, identity, discipline, and interaction, captured in the bibliometric study of Hyland and Jiang (2021). Alongside the field's traditional priority given to textual analyses (Lillis, 2008; Swales, 2019), linguistic ethnography, which originated from anthropological research, has aimed for a comprehensive description of complex contextual factors of language use (Blommaert, 2007). As opposed to many structural linguistic approaches, linguistic ethnography distinguishes itself by its focus on functions of language-in-use in specific situations (Blommaert, 2007), and includes both observation of text production practices and text authors' perspectives on their texts and

[3] Several decades after the term was introduced, Gee (2015) acknowledged that the word *new* meant that all ideas would inevitably age, and a better, though unwieldy, name would be "integrated social-cultural-political-historical literacy studies" (p. 167).

the practices (Lillis & Scott, 2007). The main premise of the approach has been aptly summarised by Rampton et al. (2004):

> Linguistic ethnography generally holds that to a considerable degree, language and the social world are mutually shaping, and that close analysis of situated language use can provide both fundamental and distinctive insights into the mechanisms and dynamics of social and cultural production in everyday activity. (p. 2)

Ethnography can be employed at three levels (Lillis, 2008): as a method, as methodology, and as "deep theorizing" (p. 355). Ethnography as a method is the most basic level. It is often operationalised as interviews or "talk around texts" (Lillis, 2008, p. 358). Either the text is the primary concern, in which case the information collected from text authors supplements the linguistic analyses, or the text writer takes the centre stage (*emic* perspective), and the viewpoints of the writer are probed to recognise what may be significant in a specific context (Lillis, 2008). Within the EAP tradition, ethnography is most commonly applied as (or, as Lillis (2008) notes, relegated to) a method, not a fully-fledged methodology, primarily associated with text-focused interviews. In this way, ethnography has been enacted in many studies of academic discourse (e.g., Dressen, 2014a; Harwood, 2006; Hyland, 2001, 2005; Kuteeva & McGrath, 2015; Lim, 2006).

Ethnography as methodology is characterised by a prolonged and continued engagement with text writers, and by collecting a range of background data in order to gain a fuller insight into data obtained from linguistic analyses (Lillis, 2008). This background data can be collected using observations, surveys, diaries, literacy history interviews, cyclical talks about texts, discussions, or correspondence between authors

and others (Hyland, 2013; Lillis, 2008; Lillis et al., 2010). Classic examples of studies where ethnography is used as specific methodology are Lillis et al. (2010) or Prior (1998). For example, Prior offers a sociohistoric account of disciplinary writing through rich descriptions of contexts and processes of writing in four seminars (language education, geography, American studies, and sociology) at the graduate level. His focus on graduate writing stems from his effort to understand the sociogenesis of disciplines. Thus, he asks not how disciplinary experts write (expert writing is "heavily abbreviated, presuppositional, and tacit" (p. xiii)), but how experts become experts.

At the highest or most abstract level, ethnography theorises the relationship between context and text through two key notions of indexicality and orientation (Lillis, 2008). Indexicality refers to "the specific ways in which bits of language (speech, writing) index, or point to aspects of social context"; orientation represents "how speakers/hearers orient to what is said and written" (p. 376)–that is, what stance speakers/hearers assume towards what is expressed or how they evaluate it. The notions are not fixed linguistic descriptors, but rather "intermediary categories" (p. 382) helping to connect contextual understanding with a textual analysis. This theorising attempts to bridge a gap between language and culture brought about by linguistic approaches emphasising the form of language in texts (Lillis, 2008).

1.3.4 The three perspectives and this study

Through linguistic analysis of legitimate practices, the three main perspectives just summarised above have the potential to shed light on the underlying organising principles of knowledge creation within earth sciences, a field that has so far remained underexplored in EAP/ESP research. Nevertheless, to the best of my knowledge, the

potential remains little realised. I am aware of only a handful of corpus-based studies of earth sciences. First, there are studies by Dressen (e.g., 2008, 2014a, 2014b). In Dressen (2014b), she investigated research articles from field geology, focusing on 13 language variables that "index the historically situated concerns of geological field practice" (p. 13), such as demonstrating the ability to think and interpret like a geologist through use of evaluative adjectives, references to geological time, or technical field descriptors. In Dressen (2008), she investigated field accounts from geochemistry, petrology, and structural geology for textual cues that point to geology's symbolic genres (e.g., seeing and professional vision: the "discipline's shared ways of seeing and interpreting field structures" or interpreting: "knowing how to interpret natural field facts within the discipline's current frames of interpretation" (p. 239)). Both of these studies describe some of the linguistic and genre features that novices need to master to write like disciplinary insiders. Then, there are two or three corpora that were collected by the same authors (e.g., Argamon et al., 2005, 2008; Argamon & Dodick, 2006; Dodick et al., 2009). These corpora were used to reveal methodological variations between historical and experimental sciences in order to empirically support the position of those philosophers of science who maintain that there is no single scientific method (e.g., Cleland, 2001). To compare the sciences, the studies drew on machine learning to identify topic-independent rhetorical features (conjunctions, comment, and modality) in published articles from paleontology, geology, and evolutionary biology (historical sciences) and physical chemistry, organic chemistry, and physics (experimental sciences). Importantly also, the representative sub-fields in these machine learning studies imply that earth sciences may incorporate distant specialisations, likely with differing legitimate practices for knowledge building, but were not distinguished.

From the genre perspective, Dressen's studies focused on the field account (as the discipline's so-called materialised genre) as well as symbolic genres. The corpora and machine learning studies looked at whole texts, not considering the varying rhetorical characteristics of individual research article sections (as acknowledged in the statement of limitations by, for example, Argamon and Dodick (2006)). Additionally, the latter studies were restricted to analyses of lexical features, which could be searched automatically, since complex syntactic structures could not be accurately identified by machine learning (Dodick et al., 2009).

To my knowledge, ethnography was utilised to investigate earth sciences only by Dressen, not only to confirm textual practices observed in her corpus (e.g., 2014b), but more notably, in a case study within her corpus-driven investigation, for example, of a field geologist's disciplinary voice changes over 10 years after her PhD as reflected in the use of 13 textual indexes (Dressen, 2014a).

While all these studies have contributed to describing aspects of the nature of earth sciences, what still remains unexplored are the larger organising principles underlying knowledge building in earth sciences, including any sub-disciplinary differences. These organising principles can reveal that (sub-)disciplinary knowledge is not homogeneous, but has its inner structure, forms, properties, and tendencies (Maton, 2014a), and that intellectual (sub-)fields do not differ merely owing to their accumulated bodies of information on a specialist subject matter, but also owing to their knowledge structures and knower structures. To investigate these issues, this study is driven by a theoretical framework from the sociology of knowledge, Legitimation Code Theory (LCT). The study also engages with the three perspectives on disciplinary discourse in the following ways. A corpus-based approach is engaged in a bottom-up

fashion in order to generate quantitative and qualitative data, with a focus on sub-disciplinary differences in the use of selected linguistic features. This lexico-grammatical focus is to a smaller degree underpinned by Systemic Functional Linguistics (SFL).

The genre perspective is represented by the focus on the research article, which perhaps best of all genres reflects the conventions and knowledge creation in academic fields, making it "the premier discourse of the academy" (Hyland, 2014, p. 6). The planned sociological interpretation of the patterns of use of the linguistic features also reflects EAP and SFL perspectives, for example, Swales' (1990) focus on social action within a discourse community as well as Martin and Rose's (2008) manifestation of the social practices of sub-disciplinary cultures. This study can then enrich the genre perspective by expanding the range of disciplines whose RAs have been researched, and by revealing the social practices of sub-disciplinary knowledge building.

Ethnography is engaged primarily to explore connections between prominently cited earth science authors and their institutional and country affiliations, which could hint at possible influencers in the sub-disciplines–a direction of investigation that is, to the best of my knowledge, rare in EAP/ESP. I have also consulted usage of selected linguistic features with my earth science colleagues (e.g., Section 6.4). Last but not least, while not part of this study, I have been able to draw on many hours of talking around texts during work with earth science authors. The general insight I have gained has provided me with invaluable, although often intangible, benefits when interpreting my corpus data.

1.4 Purpose and the general research question

In line with Kastens et al. (2009), I combine a linguistic approach with an approach from the sociology of knowledge, Legitimation Code Theory (LCT), which aims to reveal the underlying organising principles of disciplinary practices. Due to the scope of my research, I select two representative earth science sub-disciplines: marine chemistry and seismology[4]. Both utilise some principles of fundamental natural sciences (chemistry and physics, respectively), yet the sub-disciplinary characteristics (Section 4.2.1.1) can be expected to add a unique twist to their application. That is because many differences in knowledge-making and assumptions about how knowledge is built are reflected in disciplinary texts, and the textual differences should be visible also at the level of individual sentences (Love, 1993; MacDonald, 1992). To study sentence-level linguistic differences, I investigate how Themes, citations, hedging and boosting, and exclusive first-person pronouns are used in the sub-disciplines. These linguistic features were selected for their compatibility with the concepts that constitute LCT's dimension of Specialization (Section 3.3.1), my analytical tool, and for their potential to focus on both knowledge and knower practices in the sub-disciplines. At the same time, the selected features allow the dimension of Specialization to be operationalised in two novel ways in LCT research (Sections 4.3.3 - 4.3.6). Lastly, the selected features have been investigated in numerous EAP/ESP studies, and findings about the differences in the features' usage in marine chemistry and seismology can be compared with findings from previous research.

By investigating the four linguistic features in marine chemistry and seismology,

[4] I should note that the results of my research cannot be easily extrapolated to the whole discipline of earth sciences. Differences in writing–and hence in knowledge-making practices–have been found even within sub-disciplines of a field (e.g., Harwood, 2006), and my experience editing manuscripts from varied earth science sub-disciplines confirms their existence.

I intend to answer the following overarching research question:

> What are some of the legitimate knowledge-making practices and bases of achievement in the sub-disciplines of marine chemistry and seismology, as evident from the usage of selected linguistic features in published research articles, and how do these practices and bases of achievement differ?

To answer the overarching question, I also investigate separate questions for each of the four linguistic features. However, as the specific questions require more background information about my theoretical framework, I introduce them only at the end of Chapter III, in Section 3.7.

1.5 Significance

This study is significant at three levels. First, it improves our understanding of disciplinarity within earth sciences. This is significant given that earth sciences comprise an amalgam of sub-disciplines, have remained largely understudied, and, as mentioned in Section 1.2.1, earth sciences have been seen as lacking autonomy due to their alleged reducibility to more fundamental disciplines of physics and chemistry (Kleinhans et al., 2010). By focusing on the usage of selected linguistic features–Themes, citations, hedging and boosting, and exclusive first-person pronouns–I aim to uncover differences in the principles underlying knowledge and knower structures in two representative sub-disciplines, marine chemistry and seismology. Such differences may be typological and topological.

Second, this study contributes to broadening the EAP perspective on academic literacy by combining linguistic inquiry with an approach from the sociology of knowledge, Legitimation Code Theory (LCT). LCT aspires to reveal attributes of

legitimate disciplinary knowers and legitimate ways of referring to disciplinary knowledge (Maton, 2014a). This combination of approaches then contributes to moving EAP/ESP away from a narrow focus on textual practices to broader questions of (sub-)disciplinary knowledge building. In addition, this study extends extant EAP research into how linguistic features are used in published research articles in sub-disciplines (and in this case, by extension, in a discipline) that have so far been on the margins of interest. From this perspective, especially interesting can be (even partial) comparison of my results with results from already investigated disciplines.

Third, this study is directly applicable in my pedagogical practice, since I teach students who are either seeking to enter (through their PhD degrees or research associate positions) or have just become novice members of various earth science (not only marine chemistry and seismology) discourse communities by graduating or publishing during their degree programmes. Evidence from pedagogically motivated research into academic literacy has demonstrated that individual disciplines possess characteristic forms of acceptable communication (Hyland, 2006b), which are frequently beyond the skills of novice members (Luo & Hyland, 2017). As writing is still the main avenue through which students are assessed (Charles, 2013), and as publication is encouraged or expected during a PhD candidature at the ASE, textual analysis crucially contributes to our understanding of the nature of disciplinary communities (Hood, 2010). More discipline-specific descriptions that students will find directly helpful are thus clearly needed (Hyland, 2012a), with some authors encouraging the teasing out of differences even at the sub-disciplinary level, pointing out that "distinguishing between writing practices only at the disciplinary level is an oversimplification" (Harwood, 2006, p. 443). From a larger perspective on academic

writing in social context, this study then falls within the normative view, "which identifies academic practices in order to induct students into them, accommodating rather than problematising the status quo" (p. 145) rather than within the critical EAP stance (Charles, 2013).

Nevertheless, insights into what constitutes the English language features of publishable research are crucial. On an ideational level, language use (communication) is the enabling force behind disciplinary as well as scientific advancement (Penrose & Katz, 2010). This advancement is only possible once new research has been published, which in turn is only possible if it has been written in accordance with disciplinary conventions and expectations. On a practical level, successful English-medium publications represent capital in the form of social status and recognition (Lillis & Curry, 2006) both for individual academics and for universities (Hyland, 2016a), in addition to fulfilling the nowadays increasingly more common requirement of international publication during graduate study (Flowerdew, 2013). In the words of Hyland and Salager-Meyer (2008), published RAs are an integral part of the cycle which distributes rewards and maintains social stature: Publications build up researchers' credibility, which is instrumental in grant applications, whose award then enables researchers to purchase equipment and hire research assistants, which then helps in producing more data, which can support new scientific claims that can be published. Due to the existence of such a cycle, dissemination of insights into the features of published RAs then increases the chances of especially novice writers of their successful participation in their disciplinary communities. In fact, for Gosden (1993), "aware[ness] of the social dialogic nature of scientific writing and develop[ing]

the skills required in the appropriate use of linguistic resources which realise such social interaction" (p. 56) is pivotal for disciplinary community entrants.

1.6 Thesis roadmap

This Thesis consists of seven chapters. The introduction is followed by the literature review in Chapter II. There, I review in some detail the theoretical concepts and empirical studies most relevant to my research. The chapter starts with historical development of the EAP/ESP field and its flagship genre, the research article. Next, I move to consider formulations of disciplinarity and discourse community, as well as to critically review research to date into four linguistic features: topical Themes, citation practices, hedging and boosting, and exclusive first-person pronoun use. While the chapter identifies existing gaps in our understanding of how these linguistic features are used in academic writing and in earth sciences in particular, a more significant gap is pointed out in the predominant approach to studying academic literacy: application of purely linguistic methods of enquiry. The chapter concludes by supporting calls for broadening our research perspective.

In Chapter III, I propose that our perspective can be broadened by combining linguistic approaches with an approach from the sociology of knowledge. First, I devote space to selected topics and broad concerns of the sociology of education/sociology of knowledge, leading to social realism, which is an approach that Legitimation Code Theory (LCT), my theoretical framework, claims allegiance to. Second, I review two of the most fundamental influences on LCT formation, Pierre Bourdieu's (e.g., 1984, 1993) work on struggles in social fields of practice and Basil Bernstein's hierarchical and horizontal knowledge structures (1999) and classification and framing (2000).

Third, I provide the background to as well as some critical appraisal of LCT dimensions, predominantly the dimension of Specialization, which is drawn on in this thesis. Then, after summarising the collaboration of LCT and SFL, I review some of the research underpinned by LCT, highlighting its significant contribution to our understanding of knowledge creation as well as gaps in extant methodology and focus. Finally, I introduce my specific research questions.

Chapter IV outlines the methodological design of this study. First, I describe how I selected earth science sub-disciplines and collected my corpus. Next, I detail how LCT's dimension of Specialization was operationalised (i.e., how theoretical concepts were mapped on empirical data), including information on identifying knowledge claims and coding and analysis of four linguistic features: topical Themes, citations, hedging and boosting, and exclusive first-person pronouns. I also show how the linguistic features reveal Specialization in the Visibility of knowledge and knowers and Authoritativeness of knowledge and knowers.

Chapter V deals with this study's quantitative and qualitative results. The quantitative results are visualised with the help of Cartesian planes, showing not only the prevailing Specialization codes, but also a degree of overlap, and, more importantly, the differing concentrations and ranges of marine chemistry and seismology knowledge and knower Visibility and Authoritativeness. Further, qualitative exploration of specialist topical Themes suggests differing knowledge insights in the sub-disciplines, and investigation of prominently cited authors and what they are cited for reveals that several institutions are leaders in both sub-fields, and that the type of cited knowledge points towards faster obsolescence of claims in marine chemistry.

In Chapter VI, I return to my overarching question about the legitimate knowledge making practices in the two sub-disciplines. I take a broader view of how my results contribute first to LCT's theory and methodology, also in terms of pointing towards potential problems, and second to extant EAP research, methodology, and debates. I also suggest caution when applying the strong version of dichotomies (e.g., positivism vs. social constructionism) to linguistic characteristics of various academic disciplines.

Chapter VII concludes the study by offering a broad summary of the study's achievement and relevance, including the pedagogical significance, limitations, and recommendations for further research.

1.7 Summary

In this opening chapter, I have introduced my study. First, I have set it in my own institutional and personal contexts. Second, I have provided the reader with some general information about earth sciences, the field under focus, together with its characteristic explanatory strategies. Next, I have reviewed three main EAP/ESP approaches to studying academic discourse (corpus, genre, and ethnographic) in order to situate this study within the research tradition in my field, and tease out the gaps justifying this study, leading to the overarching research question and the study's contribution. Finally, I have provided a brief overview of the thesis organisation. In the following chapter, I turn to consider the major theoretical, historical, and empirical issues connected with research into literacy practices of academic disciplines.

Chapter II: Literature review

Academic discourse has attained such a prominent status in academia that a whole specialised field is devoted to studying its regularities and peculiarities, not only for the sake of theory building, but also for its pedagogical application in courses of academic writing for both native and non-native English speakers (for as Hyland (2016a) points out, there are in fact no native speakers of academic English). In this chapter, I will position the present study within this specialised field and the relevant research conducted to date. First, I will outline the beginnings of English for academic purposes (EAP), and situate the EAP tradition within the field of English for specific purposes (ESP). Next, I will turn to the principal genre of academia and of this study, the research article, and sketch its rise to prominence and its changing expression, aspects of which are my focus. Then, I will move to key concepts of disciplinarity and discourse community, before reviewing selected areas of research into discipline-specific research writing–areas that are relevant for the exploration in this study. Finally, I will devote space to some voices that critically appraise EAP research, and then highlight the need for the present study, in which I propose to broaden the view of (sub-)discipline-specific practices by drawing on a social realist framework theorising knowledge and knowledge-making practices–Legitimation Code Theory.

2.1 Academic literacies and English for academic purposes

2.1.1 Overview

In the previous chapter, I have related this study to three dominant approaches to studying academic discourse that I am going to draw upon. The three approaches belong to a much broader tradition of EAP research. In the following overview, I am going to

situate this study more firmly within this broader research. Let me begin by turning to the questions of what EAP actually is, how it originated, and where it intersects with the field of ESP.

Broadly defined, EAP is "concerned with researching and teaching the English needed by those who use the language to perform academic tasks" (Charles, 2013, p. 137). More specifically, ESP/EAP has been described as having four defining characteristics, all relevant to the characteristics of this study. The field is:

1) Designed to meet specific needs of the learner;
2) related in content (i.e., in its themes and topics) to particular disciplines, occupations and activities;
3) centred on the language appropriate to those activities, in syntax, lexis, discourse, semantics, etc.;
4) in contrast with 'General English' (Strevens, 1988, pp. 1-2).

The term EAP first appeared around the mid-1970s, and by the late 1970s was in more general circulation thanks to, in part, its use in published seminar proceedings and paper collections (Hyland, 2006b; Jordan, 1997). The field of EAP originated in the broader area of ESP (Charles, 2013), and while largely on the fringe of ESP at the beginning of the 1980s, EAP has risen to the position of prominence in today's language research, theory development, and research-based education (Hyland, 2006b). As a field of inquiry, EAP has turned into a significant research discipline in its own right, providing sophisticated descriptions of academic discourse that inform pedagogical materials and methods (Charles, 2013).

Developmentally, EAP is the reaction of the language teaching profession to several significant changes–changes that also constitute the rationale for studies like

mine: increased globalisation of tertiary education; rise of English as the lingua franca of the academic world; and the recognition that the communicative competence needs of those using English in discipline-specific settings differ from the needs of those who are learning English for other purposes (Hyland, 2006b). The last point dates back to the developments in applied linguistics in Great Britain in the 1960s, and is connected primarily with work done by Firth and Halliday, who saw language as a communication resource that needed to be adapted to the context and situation in which it was used (Flowerdew & Peacock, 2001). That was in a period dominated by the belief that scientific language was neutral, transparent, and universal (Swales, 2001). But as Halliday et al. (1965) recognised at that time, language is not a homogeneous entity, and approaches to language teaching need to account for differences in, for example, register (defined as a variety of language with a specific function, such as a register of science or of business). In their view, language teaching can be greatly supported by linguistics, whose primary contribution is to provide good descriptions of specialised languages and registers derived from analyses of large samples of such languages and registers used in specific situations. As Halliday et al. (1965) state,

> any description of a language which is at all delicate is bound to be specific in the sense that it relates only to a certain variety, or certain varieties, of the language; and institutional linguistics is needed to define and identify these varieties. For example, if we want an accurate account of the English of medical textbooks, we cannot expect the description given to be valid in detail also for the language of political speeches. (p. 170)

Some of the concomitant milestones in EAP development were then related to increased professionalism and differentiation. For example, they included the renaming

of The Group for Special English Language Materials for Overseas University Students to The British Association of Lecturers in English for Academic Purposes in 1989 (Jordan, 1997), and the establishment of the *Journal of English for Academic Purposes* in 2001 in response to a surge in articles dealing with academic texts and contexts (Johns, 2013).

One of the models of where EAP can be positioned in the broader context of English language teaching/learning is presented by Jordan (1997, p. 3). Starting from the four language skills of listening/speaking/reading/writing, which are applicable to all purposes, English can be learnt for social purposes (e.g., various daily communicative scenarios), general purposes (e.g., as a school subject for school exams), or specific purposes (ESP). ESP can then be further divided into EOP/EVP/EPP, English for occupational/vocational/professional purposes (e.g., English for doctors, hotel staff, or airline pilots), and EAP. EAP can then branch out into English for specific academic purposes (ESAP), such as medicine, economics or engineering, and English for general academic purposes (EGAP), which deals with areas such as listening and note-taking, academic writing, seminars and discussions, research and referencing skills, or general academic register. However, the distinction between EOP and EAP may not be as unambiguous, since many university courses, especially vocationally oriented ones, are meant to prepare students for their professional occupations, and thus EAP courses run at universities may also have EOP dimensions (Flowerdew & Peacock, 2001). Moreover, even the difference between ESP and EAP can get blurred, because "for most of its history, ESP has been dominated by English for academic purposes" and "EAP continues to dominate" (Johns & Dudley-Evans, 1991, p. 306)–a statement that was made some 30 years ago, and which I would venture to uphold even today.

Further distinctions can be made based on where EAP courses are taught and who the target audiences are. EAP is practised in a range of educational settings, for example, in secondary school courses, pre-university courses, university pre- and in-sessional part-time and full-time courses at undergraduate and post-graduate levels, courses for academic researchers writing for publication or preparing to present at conferences, native and non-native speaking contexts, distance programmes or computer-assisted learning (Charles, 2013; Hyland, 2006b; Jordan, 1997). While EAP also deals with issues related to needs analyses, syllabus and material design, or testing and evaluation, major research is conducted into written university genres, especially the research article. In effect, the description of academic language varieties is the primary enterprise of the discipline (Flowerdew & Peacock, 2001). Of relevance to this study will be areas focusing on specific academic purposes of disciplinary written research genres.

2.1.2 The issue of specificity

Initially conceptualised by Halliday et al. (1965), and later espoused by Strevens (1988) as the first of the defining characteristics of the discipline (mentioned in the previous section), the notion of specificity is central to EAP/ESP. Yet, it remains contested, both within and outside the field (Hyland, 2002b). The reasons put forward for non-specific EAP (EGAP) have encompassed a range of concerns and included, for example, the role of education at large, the conception of EAP's independence as a field of inquiry, the competencies of EAP teachers, the needs of learners of varying language proficiency, and transferability of language features and skills (Hyland, 2006b). Some of these concerns are worth mentioning in more detail.

According to one perspective outlined by Widdowson (1983), in the broadest sense of the role of language education, a specific focus is too narrow, limiting students' achievement, restricting their communicative competence, and especially disadvantaging students learning EAP. In other words, specificity "is in inverse proportion to educational value" (p. 105) because for Widdowson (1983), university education should prepare students for unpredictable situations, and not be confined to teaching formulas and patterns of prescribed linguistic behaviour.

Another strong position has been taken by some scholars towards the purpose of EAP. As a discipline, EAP/ESP is viewed not as a subject in its own right, but an endeavour subservient to other disciplines: It is a means to an end, which is to acquire a professional or vocational language competence (Hutchinson & Waters, 1987). Thus, learning ESP is a "dependent activity, a parasitic process" (Widdowson, 1983, p. 109) without its own purpose.[5]

Although some researchers may not consider EAP/ESP a fully-fledged subject, they still recognise the need to assist learners achieve professional or vocational language competence. In that respect, Spack (1988) suggested that disciplinary writing should be taught by disciplinary specialists, because only they are regarded as having the necessary knowledge background and the experience of becoming disciplinary members. According to Spack (1988), EAP teachers lacked competence in disciplines other than English, and were unable to teach specific disciplinary conventions, and thus they were better off focusing on general strategies and principles of writing transferrable to disciplinary writing.

[5] Despite the age of the citations in this paragraph, the issue of EAP's purpose is still current, as reflected in a recent book by Ding and Bruce (2017) titled *The English for Academic Purposes practitioner: Operating on the edge of academia*, in which the authors raise "the fundamental questions of this volume, whether EAP is a research-informed academic subject or a peripheral support service" (p. 3).

This position then brings us to concerns with generalizability and specificity of language and skills. Some generic skills among disciplines were unfortunately unproblematically assumed to exist (e.g., listening and note-taking, searching for sources in the library, or skimming and scanning for relevant information) (Hyland, 2006b), in addition to expressed conviction that differences in disciplinary writing were of little consequence. For example, according to Hutchinson and Waters (1987), "there is no grammatical structure, function or discourse structure that can be identified specifically with Biology or any particular subject. Such things are the product of the communicative situation" (p. 165). Subjects, in their view, differed only in their technical vocabulary and proportions of particular grammatical constructions. Others, such as Zamel (1998), acknowledged that academic discourse was not monolithic, and disciplines were in a constant state of change as they were continually reshaped by new disciplinary members. Yet, Zamel (1998) saw disciplinary differences as mere surface features, and focusing on them was reductionist. Instead, lessons should focus on "features that characterise all good writing" (p. 194).

The arguments for specificity largely mirror the arguments against it. For example, with respect to the nature of language education, the pro-specificity camp points out that students do not learn only isolated words, grammatical structures or genre formulas: All these elements carry disciplinary values, assumptions and meaning-making practices (Hyland, 2006b). In fact, teaching discipline-specific communication, persuasive argumentation, and interaction with readers cannot be separated from teaching the discipline itself (Hyland, 2017a). Furthermore, students do not learn academic English step by step, mastering its features in a pre-determined sequence.

Instead, they acquire new language features as and when they are ready (Hyland, 2006b).

The suggestion that disciplinary literacy should be taught by disciplinary specialists has also been rejected on the grounds that disciplinary specialists do not have the expertise to teach language skills (Lea & Street, 1999). Moreover, while subject specialists do tacitly acknowledge that writing is always embedded in disciplinary epistemology, they see language features influenced by disciplinary epistemology as self-evident, and therefore they assume that no explicit attention needs to be drawn to them (Lea & Street, 1999).

The main argument of the anti-specificity camp that language features and skills are generic and "transferrable across different disciplines and occupations" (Hyland, 2002b, p. 385) rests on the so-called Common Core Hypothesis, that is, the perception that there are lexico-grammatical features common to all varieties (Bloor & Bloor, 1986). Several viewpoints regarding this Common Core have been put forward. Bloor and Bloor (1986) conceded that Common Core lexico-grammar may exist, but the concept is necessarily an abstract one, for such a Common Core would consist of features such as plural morphemes, verb inflections, or word order; that is, elements that do not exist as a language variety or separately from any variety, but are always grounded and learned within some kind of linguistic context. Then, Bloor and Bloor (1986) concluded that while even specific language courses will enable students to acquire general linguistic competence, only specific language courses will enable students to acquire linguistic competence appropriate to specific situations.

Others, for example Johns (1988), have stated that some rough commonalities can be found among disciplinary language conventions, "but the differences among

them may be greater than the similarities" (p. 55). One such example is writing summaries, which, Johns (1988) reminds us, is purported to be a general academic skill, yet several significantly different types of summary have been found. Another example comes from McCarthy's (1987) 21-month-long ethnographic study of one college student. The study demonstrated how essays, papers, and reports in three different freshman and sophomore subjects left the student feeling that every time the writing he was expected to do was completely unlike what he had done before. He was unable to automatically transfer skills gained in previous subjects to new contexts with different knowledge problems, and in each successive subject he was ultimately like a stranger in a strange land. Moving beyond specific genres, not even what constitutes a coherent organisation or a persuasive argument is really transferrable from one discipline to another (Lea & Street, 1999). Thus, while the existence of general similarities may be tempting, the identification of language features common across disciplines has proved challenging (Hyland, 2011a). It is true that lists of core features include items like metadiscourse and hedging (Hyland, 2002b), but research into the usage of various features has shown considerable disciplinary (e.g., Hu, 2018; Hu & Cao, 2015; Hyland, 2005; Khedri et al., 2013) and sub-disciplinary (e.g., Harwood, 2005a; Harwood, 2006) differences. Additionally, academic writers, when interviewed in mixed-methods studies as disciplinary insiders, show that they are acutely aware of what language features best suit their fields (e.g., Hyland, 2001, 2004), again confirming that academic literacy is subject specific (Hyland, 2002b). Other specificity proponents, for example, Bhatia (2002), have taken a more decisive stance: A Common Core may not exist–words and structures have many meanings which change depending on the context. On their way to disciplinary membership, students need to acquire specific communication skills, not

just extend their general literacy to help them participate in disciplinary discourse. They need "a range of literacies to handle disciplinary variation in academic discourse" (p. 27).

In summary then, I should point out that the above-mentioned debate surrounding the tension between general and specific EAP is ongoing. So far, it has taken place in four stages, as described by Anthony (2011). In the first stage, in the 1960s, the move to specificity was inspired by research into the lexico-grammatical differences between disciplines–a focus that was in the second stage expanded to paragraphs and whole texts. The next stage was characterised by debates about the necessity for specialised disciplinary knowledge in order to teach disciplinary writing, and the corresponding lack of it on the part of EAP teachers. The acknowledgement of this lack then led to a wider-angle perspective in the form of proposals to focus on core elements transcending disciplinary differences. In the last and current stage, the return to narrow-angle perspectives was influenced by, among others, findings enabled by corpora.

In my view, this latest shift is perhaps best epitomised by the following quotation from Hyland (2016c):

> The take-home message here is that the discourses of the academy do not form an undifferentiated, unitary mass but a variety of subject-specific literacies. Disciplines have different views of knowledge, different research practices, and different ways of seeing the world, and as a result, investigating the practices of those disciplines will take us to greater specificity (p. 27).

I fully agree with Hyland's characterisation above, and also with Tarone et al. (1981) that language feature use across genres and disciplines needs investigating and

"should not simply be assumed" (p. 202). But I also argue that such investigation should take us to even greater specificity of sub-disciplines. This is then one of the aims of the present study: to illuminate some of the variations at the level of sub-disciplines and in earth sciences, that is, two EAP areas with many blanks. Perhaps no other genre has been more in the centre of research endeavours to describe (sub-)disciplinary differences than the research article–the "most prestigious genre of scholarly writing" (Hyland & Salager-Meyer, 2008, p. 305).

2.2 Key genre: The research article

If we agree with Hyland (2014) that the research article (RA) is "the genre where knowledge is principally negotiated, agreed on and archived" (p. 6), and with Canagarajah (2002) that a publication in the form of an RA is an indication that a piece of knowledge has been recognised and accepted, then we can safely conclude that in the recent decades, we have witnessed an explosion of knowledge. At the beginning of the 21st century, the number of journal articles published every day already reached 6,000, after decades of doubling publication numbers every five years, effectively adding up to an almost 600% growth since the 1950s or 1960s (Budin, 2002). Not long ago, the boom engaged an estimated 5.5 million researchers, 2,000 publishers, and 17,500 research or higher education institutes (Lillis & Curry, 2010). More recent numbers estimate that journal publishing is worth billions of dollars, employs in excess of 100,000 individuals who support between 6.7 and 8.9 million active researchers. Peer-reviewed English-language journals number more than 28,000 (with the number having grown steadily by 3.5% annually over the last 300 years), with the roughly 8,700 SCI and 3,000 SSCI journals publishing about 1.5 million RAs annually, accounting for

most citations due to their higher impact factors (Ware & Mabe, 2015). Nevertheless, despite the huge numbers of published RAs, what gets published is just the tip of an iceberg of what is submitted, especially by English as an Additional Language (EAL) scholars (Hyland, 2016a and references therein). For example in 2013, *Nature* published only 7.8% of almost 11,000 submitted articles (Gould, 2014).

The sheer number of publications in this prestigious genre reflects how publishing has become obligatory for many academics and institutions, whose reputation, credibility and indeed productivity are measured by the lengths of their publication lists (Canagarajah, 2002; Hyland, 2007; Hyland, 2016a). In fact, it is not only the number of publications. In some disciplines (e.g., marketing), publication targets revolve around getting research published in high impact journals (Tusting et al., 2019). Moreover, the increased pressure on academics to publish their research has been compounded by increased pressure to publish in English language journals (Hyland, 2011b; Starfield, 2014). It is no wonder, then, that in the last decades, discipline-specific research in EAP/ESP has been increasingly focused on the RA genre (Lin & Evans, 2012). In fact, the RA is perhaps the most studied of all academic genres (Samraj, 2014, 2016). To understand how the genre has climbed to its present position of prominence, we need to understand how it came to symbolise the "truths" about nature and our world. Let us, then, take a brief historical excursion through its development.

The genre is about 350 years old, and its inception was bound with the establishment of the first scientific journals. According to Hyland and Salager-Meyer (2008), the RA developed in two countries almost simultaneously: as a result of the endeavours of Denis de Sallo in France and Henry Oldenburgh in England. The initial form of journals appeared with the founding of the *Journal des Scavans* in January 1665

in Paris, and the *Philosophical Transactions of the Royal Society* in March 1665 in London. These first scientific journals were platforms for scientists to regularly debate, and recurrent social situations eventually shaped the form that the discussions and the articles took.

How the form of the experimental report metamorphosed in the first 140 years, or the first 90 volumes, since the establishment of the *Philosophical Transactions of the Royal Society* has been studied by Bazerman (1988), whose description of the development of scientific discourse traces the appearance of many of the now-familiar experimental report features. Bazerman's recount focuses on several points:

a) The experimental format

Perhaps surprisingly, a mere one twentieth to one fifth of the articles published in the first century and a half were experimental reports, with the majority being observations and accounts of natural events. By 1800, experimental accounts comprised 39% of the journal contents. In the first volume, experimental reports were "cookbook recipes for creating marvellous effects or effects of practical use" (p. 66). Soon after, experiments became synonymous with deliberate testing of natural phenomena. In further developments, by volume 30, experiments assumed an adjudicating function as they were being designed to test hypotheses and to solve debates and disputes about natural phenomena. From this adjudicating function, their next role was to establish general truths about nature. By volume 80, the role that experiments played slowly started to resemble the role they have today: To demonstrate how authors arrived at their conclusions, with hypotheses presented first, and experiments designed to verify them. In volume 90, experiments served to test beliefs and expand knowledge. Experimental results thus

justified general knowledge claims, which were themselves considered parts of even broader general claims.

b) Methodology

In the early volumes, experimental methods were given limited space to briefly inform the reader what type of experiment was done. Soon afterwards, authors started to describe their methodology and experimental conditions with increasing care. Two trends become apparent. First, authors engaged in disputes designed their experiments to address or prevent objections of opponents. Procedures were described in detail to demonstrate how proper steps were followed, while the procedures of the opponents were portrayed as being less careful and at times erroneous. Second, authors not engaged in disputes devoted less space to their methods, but since conducting multiple experiments under varying conditions became part of discoveries of natural phenomena, distinguishing which features and conditions of individual trials potentially affected results became crucial, and such features were therefore specified. In volumes 80 and 90, as experiments functioned to test hypotheses and put forward general claims, methodologies were detailed, often accompanied by illustrations of apparatuses in the interest of exactness, disqualification of alternative explanations, and persuasion of the reader.

c) Reporting of results

Early articles, especially those in the cookbook recipe form, did not report any particular results, because following the recipe was assumed to always lead to the same results. However, similarly to methodology, results began to be more elaborate with emerging scientific disputes. When, by volume 60, experiments were being designed to test authors' hypotheses, their own results were reported to

confirm the hypotheses. In addition, results reported by other authors could be mentioned to provide further support, even though the other experiments were not necessarily designed to test the same hypotheses. Results also became increasingly quantitative, which facilitated comparison, and could serve as direct proof.

d) Validation of claims

Early experiments were conducted in meeting rooms before assembled scientists, whose communal experience carried sufficient validity. However, as experiments became more complex and results more contested, methodologies were described in more detail, and extraneous variables were eliminated as much as possible, experiments moved to laboratories and could no longer be witnessed by large assemblies. Instead, designated witnesses, individuals of prestige and high social standing, including in some cases the king and the queen, were key to attesting to the verity of the findings. However, as opposing claims could also be accompanied by lists of prestigious witnesses, detailed methodology allowing replication and detailed results allowing critical assessment gained importance. By volume 40, when scientists cogitated over complex phenomena and put forward theories, experiments were conducted in private, being witnessed only by the experimenters. Scientists then had to tell a persuasive story. Thus, the detail and plausibility of experimental accounts became the means of vicarious witnessing, which was crucial to earning the readers' trust in the conclusions reported and in the experiment as the proof: Accounts turned into "a retrospective guided tour of conclusions and experimental evidence" (p. 75).

e) Article organisation

Throughout the period, experimental articles grew longer. First articles were short

texts in which journal editors reported what they had found out through various channels, including hearsay. Soon after, one- or two-page authored articles were published, mainly in the narrative form of what was done and what results were obtained. When articles became forums for disputes, the format changed to open with the contested phenomenon, followed by the opponent's work, the author's experimental design, his results, and sometimes conclusions. By volume 40, hypotheses appeared at the onset of reports. When natural phenomena began to be viewed as more complex, authors began their articles by introducing the complex phenomena, often accompanied by detailing failed experiments. Authors then described a series of their own experiments that should shed light on the phenomena, drawing conclusions after each individual experiment, which usually necessitated further experimenting. At the end, when the phenomena were adequately understood, all results were synthesised. In the last volumes examined by Bazerman, towards the end of the 18th century, scientific articles morphed into a format not dissimilar to the typical experimental article of the last 100 years. As articles became more focused on general knowledge claims, introductions tended to include philosophical statements about general knowledge, followed by a problem, which could also include a statement of a gap in scientific knowledge. The problem was followed by claims intended to offer resolution, supported by experiments, and conclusions discussing the implications of the findings.

While Bazerman's historical excursion has focused on significant selected aspects of the history of the RA, it by no means encompassed all of them. Other developments included more frequent use of phenomena as the subjects of sentences rather than author-centred writing; more nominalised forms and abstractions; greater

numbers of references in all RA sections; more prominent figures, tables, and graphs; shorter methods sections; and multiple authorships (Swales, 2004). Yet, the historical changes mentioned above[6] sufficiently illustrate how communication conventions varied.

The changing format of the RA throughout its existence has reflected the creative responses of writers to arising situational needs (Bazerman, 1988). But should the additions and alterations be viewed as improvements? According to Hyland and Salager-Meyer (2008), a more appropriate view is that the RA is the result of "selective pressures and adaptive mechanisms, increased cognitive complexity, higher standard of proof, and greater volume of data coupled with a dramatic increase in the number of scientific articles" (p. 305). From this perspective, the short historical review underscores that the RA is a conventionalised product whose adaptations and alterations are contingent on the priorities and norms of a community (disciplinary or sub-disciplinary) at a given point in time, bearing in mind that the norms and priorities reflect knowledge-making practices. At the same time, the adaptations are also, although to a lesser degree, a result of external driving factors, for example, as seen today in government and institutional policies. These external drivers impact disciplines to different extents, which for instance Tusting et al. (2019) illustrate by comparing how academics' writing practices in marketing, history, and mathematics have been variously shaped by UK's 2014 Research Excellence Framework. Another example of external selective pressure comes from Hyland and Jiang (2016), who report a general (in some disciplines dramatic) increase in the frequency of self-mention, most notably in the use

[6] I should mention that the above pages dealt with the experimental RA, since this type is the focus of the present study. However, the experimental RA is by no means the only kind of scientific article that exists. Other types include review articles, theoretical articles, science letters, and methods papers (Swales, 2004).

of personal pronouns, in both hard and soft fields in the last 50 years, presumably in order to highlight personal contributions, and to underscore one's impact in the disciplinary community to readers belonging to grant and tenure-awarding committees. Simultaneously, the reported increase is an example of shifting values and norms in how visible researchers are when building their disciplinary knowledge.

As a genre, today's RA strives to present a balance of factual information and social interaction with disciplinary peers (Hyland, 1999). While in some disciplines facts are supposed to be reported neutrally so that they speak for themselves, "practically all commentators have concluded that these RA texts are richly persuasive rather than flatly expository" (Swales, 2004, p. 218). In fact, persuasion has been a long-standing integral part of scientific experiments. The reader will remember that even 17th and 18th century scientists aspired to persuade their peers of the correctness of their research by, for example, multiplying witnesses through public experiments, detailed experimental protocols that facilitated replication and thus distance witnessing, and virtual witnessing–imagining with the help of words and illustrations: "production in the reader's mind of such an image of an experimental scene as obviates the necessity for either its direct witness or its replication" (Shapin, 1984, p. 491). However, what is considered persuasive today depends on what discipline an RA has been written in, or perhaps more precisely, for what discourse community it is meant. It is then the concept of a discourse community to which we turn next.

2.3 Disciplinarity and discourse community

As Christie and Maton (2011) report, proponents of disciplinary demise would have us believe that disciplinarity is largely irrelevant, because the accelerating rates of

technological change and globalisation apparently render it a reflection of past social norms, currently "out-moded and quickly becoming supplanted by 'inter-', 'cross-', 'multi-', 'trans-' or 'postdisciplinarity' " (p.1). While the authors assure us that these proclamations have been all too hasty, and disciplinarity is alive and well, they conclude that there is still much to learn about what disciplinarity actually entails, especially what its underlying structural principles are, despite extensive and intensive debates about the nature of disciplines in the past decades.

One of the early works touching on disciplinarity is Toulmin's (2003) *The uses of argument*, originally published in the late 1950s, in which the author recognises that argumentation and reasoning patterns in one field are "distinct from patterns to be found in other fields" (p. 236). While rough similarities in argument presentation do exist, stretching the similarities to the point of claiming that some ways of arguing are universally valid or invalid means that we expect some forms of argument to "achieve something which it is no business of theirs to achieve" (p. 236). Toulmin's discipline-specific ways of arguing draw on the notion that effective communication requires knowledge of /proficiency in community-specific norms of communication in addition to linguistic proficiency. The same notion was later advocated by Hymes (e.g., 1974); that is, that competence in language should be seen as communicative rather than simply linguistic (knowledge of structural elements), because it is needed for appropriate participating "as not only a speaking, but also a communicating member" (p. 75). The existence of disciplinarity and concomitant discipline-specific ways of communicating is the underlying premise of this study: a premise that also includes the idea that discipline-specific ways of communicating can be used to delineate discourse communities, not only at the level of a discipline, but also at the finer level of a sub-

discipline.

2.3.1 Discourse community

The concept of a discourse community is at the heart of disciplinary communication because disciplinary knowledge is generated, organised, and maintained in the disciplinary community through writing as a social activity (Faigley & Hansen, 1985; Hyland, 2009). As Hyland (2006b) reminds us, according to social constructivism, the intellectual milieu of a discipline in which academics reside dictates what problems are investigated, what methods are employed, how results are interpreted, and in what ways the interpretations will be written about. Through writing, then, academics affirm shared viewpoints and their community membership.

What actually is this discourse community that its members uphold? The origins of the concept can be traced to linguists who saw writing as a social process (Swales, 1990). Swales quotes from a conference paper presented by Bruce Herzberg in which a discourse community is conceptualised as representing:

> a cluster of ideas: that language use in a group is a form of social behaviour, that discourse is a means of maintaining and extending the group's knowledge and of initiating new members into the group, and that discourse is epistemic or constitutive of the group's knowledge. (p. 21)

The above characterisation is suggestive, but does little to help us define or identify a discourse community (Swales, 1990). The term itself is still imperfectly defined, and various criteria could be used (Bizzell, 1992; Starfield, 2001). One such collection of criteria, Swales' (1990) six discourse community characteristics, has been particularly influential:

1) Discourse community members share public goals. These goals do not simply equate studying the same object or subject, since it is goal sharing which is criterial.
2) Discourse community members communicate through agreed channels (e.g., meetings, journals, and newsletters).
3) Discourse community members exchange information and feedback. Community membership implies that the information and/or feedback is made use of, or at least its existence is known.
4) Discourse community members communicate via genres that represent what the community values and how it operates. The genres bespeak appropriate topics, format, as well as ways of arguing and negotiating.
5) Discourse community members use specific (technical) lexis to communicate with each other.
6) A discourse community includes both members who possess a high level of relevant expertise and novices in a ratio that allows the community to continue operating.

The criteria are useful in that they are specific enough to apply, so that we can start determining whether a discourse community exists in a specific case. In the context of this study[7], the criteria then tell us that academics researching seismology and academics researching marine chemistry belong to two discourse communities: For example, they have their own journals and meetings where they exchange information on agreed topics, and they have their technical lexis. The criteria (specifically criterion 2) were applied by Swales (1990) himself to the solution of the "café owner problem",

[7] To clarify, I here apply Swales' six criteria to academic contexts, but Swales himself notes that discourse communities exist in other contexts as well.

that is, whether café owners could be regarded as forming a discourse community. His answer is that discourse community members need to share communication channels, which café owners do not, and be knowingly part of a community, which they are not. Yet, the criteria do not erect an impenetrable wall isolating community members from the world. In fact, individuals may be members of several discourse communities at the same time (Swales, 1990). Again, in the case of the two earth sciences sub-disciplines, seismologists and marine chemists belong to at least one more common discourse community: that of earth scientists. Neither are the criteria a mental straitjacket. The fact that a discourse community has agreed common goals does not make it completely monolithic, because any discourse community comprises members who differ in their background, experience and beliefs (Hyland, 2004). Such diversity necessarily means that not all members will identify themselves to the same extent with the community's worldview–its values, methods, or conventions (Swales, 1990); in addition, the norms that communities use may range on a continuum from stringent to developing (Swales, 1990). Indeed, the continuum allows for conflicting epistemological perspectives within communities, despite shared goals of what is worthwhile to study and which methods to use. For example, in an interview with the former Director of the Earth Observatory, I found out that the Director believes that earth scientists should not hide behind passive constructions in their writing (*It was found that…*), because it is people who do science. In fact, the Director called the use of the passive in writing in his field his pet peeve (personal communication, 9th July, 2014). The Director's comment echoes what Bizzell (1992) remarked of her field: "Each scholar tends to favour the elements that seem to lead composition studies in the direction he or she wants the field to go" (p. 223).

Swales' (1990) definitional endeavours have stimulated further debate on disciplinarity and have been critically evaluated. Bizzell (1992) critiques Swales' six criteria on the grounds that they depict only stylistic features. Swales' concept implies that joining a discourse community represents voluntarily signing up for its cause, its extensive project. This element of choice, however, may disregard the socio-economic realities of who will be in the position to choose to participate (e.g., because of having the time or the means), or who will likely become (willingly or not) a community member. In addition, Bizell points out that Swales' stylistic definition of a discourse community does not account for the community's power to shape world views– discourse communities encourage certain kind of thinking (e.g., analyzing or classifying), which then becomes the habit of the members' minds. At the same time, community members may be socially and culturally predisposed to this way of thinking based on the kind of thinking that their prior experiences exposed them to. To what degree sub-disciplinary writers adhere to and diverge from sanctioned conventions, and indeed what some of the conventions are, should become apparent in my analysis of published RAs.

Another type of critical evaluation concerned what is missing from the current conceptions of a discourse community: power and power relationships. According to Starfield (2001), just like all social relations, relations within a discourse community are not equal. Stylistic definitions do not illuminate how power is exercised, how it is resisted, or how this struggle affects the community. Unproblematic conceptions of a discourse community imply that all members can communicate with the same authority, and opinions of all members will be equally regarded as worthy of being paid attention to. In contrast, the fact of the matter is that those who do not command the norms of

appropriate disciplinary language, for example, by dint of the social groups they belong to or based on their dominant language (or sociolect), may not be deemed to make legitimate contributions or even have the right to make such a contribution, and they may be excluded or silenced.

Adherence to appropriate norms then reflects power relations among discourse community members in at least two ways. First, successful community members can discern the sanctioned ways of thinking, the language, the structures, and the way to present evidence; those who are unable to do so will be viewed as communicatively incompetent in the community context (McCarthy, 1987). Discourse community members share internal standards about what is appropriate and what is worth writing about, and it is exactly the degree of match between the readers' and writers' internal standards which determines how a piece of writing will be judged (Johnson et al., 1994). Second, adherence to community discourse conventions also differentiates novice members from established ones. New members are expected to abide by the conventions, whereas well-established members are not bound to the same extent and are even expected to breach them at times to stimulate convention development (Bizzell, 1992). The issues of power, adherence to approved norms, and the permeability of conventional boundaries for junior and senior discourse community members are undoubtedly valid, and their investigation would be fascinating and illuminating.

The last point of criticism to be mentioned in this brief overview of the vast literature on the notion of discourse community and its practices is that from Prior (1998). He notes that since the early 1980s, the central principle in the research into disciplinary writing has been that writing is situated and acquired within discourse

communities, which are imagined as "unified social territories or abstract systems of codified knowledge" (p. xii), and authoritative entities exerting normative but consensual pressure on all their members. Such a portrayal is too static, and it is becoming increasingly untenable. Ethnographic research has highlighted how disciplinary writing is a heterogeneous, laminated process, comprising personal, institutional, sociocultural, and disciplinary histories, which may be forging alliances, conflicting and/or competing. Disciplinary writing then, is a dialogic interpretation of community and personal perspectives: a process through which community members produce and reproduce their community every time they write.[8] According to this sociohistoric view, then, a disciplinary community is constituted through doing, participating, and negotiating, perhaps reaching an imperfect mutual understanding or approximating a consensus, rather than through merely occupying a social territory and sharing an abstract knowledge system. Communities have "potentials for alignment that emerge from any common experiences" (p. 31).

2.3.2 Disciplines hard, soft, pure, and applied

What is becoming obvious from the previous pages is that the notion of discourse community has been dissected along various lines, and numerous considerations have been proposed. To those, I would like to add two other categories, this time disciplinary: hard vs. soft and pure vs. applied. Both proposed in the early 1970s, they characterise disciplines based on whether they orientate towards a paradigm (hard vs. soft), and whether they are concerned primarily with theory building or application of knowledge to real-world problems (pure vs. applied) (Becher, 1989).

[8] In fact, a two-way process is at play here, since the writer's style may also influence community norms (McCarthy, 1987).

Combining the two categories then yields a type of quadrant that helps classify disciplines into hard pure/hard applied and soft pure/soft applied, even though some established disciplines are not straightforwardly pigeonholed, and sub-disciplines within a discipline may straddle two categories (Becher, 1989).

According to Becher's (1989) characterisation, acknowledged as broad brush, somewhat imprecise, and hardly permanent due to changes in the structure of disciplines and nature of knowledge, hard pure disciplines can be described as accumulating knowledge in a linear fashion, starting with basic issues.[9] Criteria for accepting or refuting knowledge claims are relatively clear and agreed on. In contrast, the soft pure fields revisit and re-explore topics already researched by others. Basic issues are never completely (or satisfactorily) resolved, and knowledge is built in a reiterative and recursive manner. Criteria for making disciplinary contributions are not unified, and new knowledge mostly takes the form of interpretations.[10] The topics studied by hard pure disciplines seem to be clearly defined and bordered from others. New research picks up where previous has left off. The purpose is to dissect, analyse, calculate, and model the interactions between several controlled variables in order to look for general principles. In contrast, soft knowledge fields allow for certain permeability between research issues, which can be either addressed for the first time or already studied in the past, and any part of the issue can be re-examined.[11] Issues are considered in light of the complex interlay among multiple variables less amenable to complete control, and this aspect is acknowledged in weaker claims of findings. Having

[9] The familiar metaphor used by Isaac Newton of standing on the shoulders of giants aptly exemplifies this type of disciplinary progression.

[10] However, some researchers (e.g., Hyland, 2004, 2008a, 2009, 2014) argue that interpretation is equally an inherent feature of knowledge contribution in all disciplines.

[11] I would add that soft fields are not immune to trends in research topics, so the choice of what to study is not completely random.

raised the last point, I believe it needs to be mentioned that Becher's characterisation of disciplines is now 30 years old, and recent research has in fact found that the rhetoric of claims of findings in hard and soft fields is converging (Hyland & Jiang, 2016).

Becher (1989) acknowledges that hard applied and soft applied disciplines are less easily characterised. Nevertheless, he suggests that hard applied fields combine quantitative approaches with qualitative judgements, for example, in engineering designs or medical diagnoses, and are directed towards utility, namely, effective products and techniques for the daily life. Soft applied disciplines, such as education, grapple with the intricacies of human experiences, aiming to improve our personal and social lives. Their outcomes are protocols and procedures.

Becher's division of disciplines into hard vs. soft and pure vs. applied summarised above together with their characteristic ways of approaching their objects of study raises an intriguing question: How do earth scientists communicate compared to academics in hard and soft fields? If we invoke the major characteristics of earth science research from Chapter I, we have a discipline which exhibits attributes of both hard and soft fields: It is both descriptively historical and a natural science associated with causality and law formation (Kleinhans et al., 2010); it is a field which is linear and interpretive; it both models selected variables to arrive at general principles and acknowledges that earth processes are results of complex variables not amenable to laboratory control.

Disciplines and discourse communities as concepts do not have definitions that everyone agrees upon. Numerous criteria for defining disciplines have been proposed, such as the existence of a domain of specialist knowledge, independent academic departments, an international community, professional associations, specialist journals,

or methods of enquiry (Becher, 1989). However, disciplines do not neatly fit into predefined categories. Despite the continuing debate, the concept of a discipline (and a discourse community) remains a useful descriptor, not least because the various classifications and criteria help us look for continuities and interrelatedness between and among knowledge fields (Becher, 1989). Indeed, according to Hyland (2006b), it would be difficult if we did not have the concept of a discourse community: Although the boundaries may be drawn imperfectly, the concept allows us to identify how disciplinary members "communicate in a way that others can see as 'doing biology' or 'doing sociology'" (p. 41). In other words, the term helps us explain the social base of writing conventions and of interpreting writing (Hyland, 2003a).

2.3.3 Discourse communities and the nature of knowledge construction

As Bizzell (1992), Hyland (2006b), and McCarthy (1987) point out, disciplinary members use language in ways approved of by the disciplines. This language use in turn (re)affirms what is disciplinarily-appropriate communication (McCarthy, 1987). In the case of writing, appropriate communication comprises accepted forms of reasoning, negotiating knowledge claims, describing research procedures, referring to key concepts, using technical vocabulary, and judging what is likely (un)known (Hyland, 2006b). To understand exactly why appropriate language use assumes this discipline-maintaining function (Faigley & Hansen, 1985), we need to acknowledge that disciplinary rhetorical and discourse strategies reflect disciplinary knowledge practices, which in turn reflect fundamental assumptions (ontological, epistemological, and methodological) about research (Hu, 2018). These assumptions have in the recent decades been largely moulded by one of two prominent perspectives.

One perspective is offered by social constructivism, often associated with soft disciplines[12]. In this view, knowledge does not exist independently of us: It is constructed in social communities through their discourse (Hyland, 2006b). Since knowledge of the world is not purely objective but "emerges in part through our perceptions of that world during our interactions" (Hyland, 2006b, p. 39), it may change from culture to culture, and from period to period, and is indeed discipline specific: "By entering a discipline, one commits oneself to looking at experience in the particular way established by that discipline"[13] (Bizzell, 1992, p. 148). The metamorphic potential of knowledge necessitates that claims are justified through discourse (hence language appropriateness). Claims are one's interpretations, which need to compete with other potential interpretations, and if they are convincingly argued for, then they are agreed on by the disciplinary community and accepted as knowledge (Hyland, 2006b). Engaging in interpretation is predicated on discourse community membership (Bizzell, 1992). "These competing interpretations shift attention from what happens in the lab or the library to what happens on the page" (Hyland, 2006b, p. 39). This position can also be summarised as constructivist relativism (Maton, 2014c), where knowledge is a reflection of power and interests.

In contrast, many hard disciplines subscribe to positivism, summarised, for example, in Hu and Wang (2014). Positivism posits that objective reality exists independently of subjective human perceptions and experiences. This objective reality is governed by causes and effects of physical laws, which can be measured or otherwise

[12] Although a well documented case of how knowledge is socially constructed in a harder discipline (biology) is presented by Myers (1985).

[13] This commitment applies, of course, also at a finer level of sub-fields or schools of thought, where it may in some cases be a non-negotiable requirement for one's career. For example, on the cases of orthodox economics, Chomskyan transformation grammar, and theoretical physics, Maton (2014a) illustrates more "extreme" cases of looking at experience or the world through a specific lens even at the cost of decreasing relevance of the research to the real world or observed data.

studied inductively and deductively. Since objective reality is believed to be describable, hard-discipline scientists use standardised procedures to devise experiments that rigorously control intervening variables. The resulting knowledge is considered an impartial, empirical, and reliable basis for our understanding of and interactions with the world. This position can also be summarised as positivist absolutism (Maton, 2014c), where knowledge is decontextualised and value free.

Social constructivism and positivism have implications for language use in academic discourse. Echoing Toulmin (2003), who first recognised disciplinary differences in argumentation (see above), Firestone (1987) asserts that disparate assumptions about what constitutes reality are mirrored in diverging persuasive strategies. He explains that social constructivism is often realised by qualitative methods, entailing rich descriptions of contexts by an immersed researcher in order to depict the situation in as much detail as possible, so that readers can orientate themselves better in the situations and be convinced that the writer's conclusions are sensible. Positivism is expressed through quantitative methods, in which a detached and unbiased researcher tries to reduce error in order to statistically determine the role of randomness and causality. Even the bareness of the positivistic scientific style, from which ornamentation and human agency are purposely missing in an attempt to let the data tell their own story, is in fact a rhetorical device with considerable persuasive power. Put differently, Hood (2011) asserts that in the soft fields, voices of researchers and participants are visible as vehicles of knowledge and observations. In contrast, in the hard fields, the researcher is invisible, and instead phenomena and abstract entities are brought to the fore–Geertz's (1988) "author-evacuated" (p. 9) scientific writing. Another linguistic difference is mirrored in the use of citations and adjoining reporting

verbs, with soft fields tending to employ more subjective propositions and dialogically expansive citations, and hard fields more factual information and dialogically contractive citations (Hu & Wang, 2014).

The opposing conceptions of social constructivists and positivists of how knowledge is generated and how it is linguistically described seem hard to reconcile. We may feel that in this dichotomy a side has to be taken. But perhaps what we need is to take a more middle-of-the-road approach. For example, we can acknowledge, following Firestone (1987), that findings gleaned by quantitative and qualitative approaches will provide different information, and so we can complement one method by the other in order to triangulate findings to strengthen their reliability. Alternatively, we can heed Bizzell's (1992) advice: Since knowledge is not merely communicated by academic discourse but constituted in the process of communication, then "insofar as we can achieve a critical distance on our discourse, we achieve a critical distance on our knowledge." (p. 138).

2.3.4 Texts as discipline-specific social interactions

To complete my account of disciplinarity, I need to at least briefly mention the role of disciplinary writing in accruing benefits and building community hierarchies. If we adopt the social view of writing as an activity that constitutes and maintains a disciplinary community (Faigley & Hansen, 1985), we can perceive RAs not only as scholarly publications, but also as interactions between members of an academic disciplinary community (Hyland, 2004). The social interaction role of RAs becomes more apparent if we realise how writing can be beneficial to researchers socially in a discourse community. First, invoking two of Swales' (1990) definitional criteria of a

discourse community–the criteria of information exchange and use of specific representative genres–discourse community members have to acquire certain accepted community discourse practices in order to earn the right to have their opinions heard and received (Hyland, 2004). That is the initial condition to meet and the first hurdle to clear. Second, having earned the right to communicate with discourse peers through published pieces of work, researchers gain access to rewards, both social (e.g., status, respect, credibility) and material (e.g., grants). As I have already mentioned in Chapter I, research publications are part of a cycle of credibility, research grants, and new studies leading to even more publications (Hyland & Salager-Meyer, 2008). Therefore, such a socially instituted system of benefits controlled through writing ensures that writers are often willing to communicate through institutionalised forms and use established norms (Hyland, 2004).

This normative function of writing is why studies like this PhD thesis are useful, because they can shed light on what forms of (sub-)disciplinary communication are sanctioned, beginning with the differing usage of linguistic features in RAs, and from there teasing out the larger organising principles underlying knowledge creation which are indicated by the linguistic differences, by combining an EAP approach with an approach from the sociology of knowledge. It is these underlying principles that can then be made explicit to aspiring disciplinary members, who can apply them and hence acquire access to the material and social benefits that published research can bestow.

2.4 Disciplinary variation in the use of linguistic features

Investigation of the variations in language feature usage across academic fields has been a very active research area in EAP. The list of (often comparatively)

investigated features is extensive, including, among others, various aspects of citation practices (e.g., Halevi, 2013; Harwood, 2009; Hu & Wang, 2014; Hyland, 1999), use of the passive voice (e.g., Banks, 2017; Leong, 2014; Tarone et al., 1998), self-mention (e.g., Cheung & Lau, 2020; McGrath, 2016; Starfield & Ravelli, 2006), demonstratives this/these (e.g., Gray, 2010; Gray & Cortes, 2011), hedging and boosting (e.g., Hyland, 1998; Peacock, 2006; Skelton, 1997), stance and reader engagement (Hu & Cao, 2015; Hyland, 2005a), shell (signalling, carrier, metadiscursive) nouns (e.g., Aktas & Cortes, 2008; Flowerdew, 2003; Jiang & Hyland, 2015, 2018; Mousavi & Mouni, 2014), types of grammatical Subjects in thematic positions (e.g., Gosden, 1993; MacDonald, 1992; Martinez, 2003), academic vocabulary (e.g., Liu & Han, 2015; Valipouri & Nassaji, 2013; Wang et al., 2008), lexical bundles (e.g., Hyland, 2008; Ren, 2021), phrasal verbs (Alangari et al., 2020), or evaluative *that* constructions (e.g., Hyland & Tse, 2005; Kim & Crosthwaite, 2019). These studies have largely confirmed that the observed variations reflect differences in knowledge construction. Out of all the linguistic features listed above, in this section, I will review Themes, citations, hedging and boosting, and first-person pronouns. While these four features have been the subject of multiple EAP/ESP studies of (sub-)disciplinary differences in academic writing, their usage in earth sciences has received little to no attention. Findings from this research can potentially be compared with a wealth of findings regarding the usage of these features in other academic fields, and this comparison can thus contribute to the debate about the nature of earth sciences and the field's potential reducibility to other natural science disciplines (Section 1.2.1). In addition, the relevance of these four features to this study lies in their compatibility with the analytical tool used here, that is, with the concepts that constitute LCT's dimension of Specialization (Section 3.3.1). Importantly also, the selected

features allow operationalising the dimension of Specialization in two novel ways in LCT research (Sections 4.3.3 - 4.3.6).

2.4.1 Themes

The concept of Theme has spawned a vast literature analysing both spoken (e.g., Backlund, 1992; Forey & Sampson, 2017) and written texts in non-academic and academic contexts. Within the latter contexts, thematisation has been of interest because, as North (2005) points out, it likely reflects disciplinary epistemological values. A Theme can be thought of as information that appears in the initial clause position (e.g., the grammatical Subject, Predicator, Complement, or circumstantial Adjunct of time/place/manner/etc.).[14] This initial position carries significance in the message construction since the writer/speaker has selected it as the starting point (Halliday & Matthiessen, 2014). The Theme then describes the focus ("aboutness") of the message, although the Theme is better thought of as a point of departure of a message (in the case where it does not equate the grammatical Subject) since the Subject can also signal the "aboutness" of a clause (Thompson, 2013). From there, the Theme functions to organise and carry forward the discourse (Halliday & Matthiessen, 2014). This means that a thematic change shifts "the focus of the clause in significant functional ways" (Bloor & Bloor, 2013, p. 75).

One interesting strand of research focused on whether various configurations of Themes can distinguish text types or genres. For example, a text's discursive purpose and subject matter significantly correlate with its thematic organisation (Lavid, 2000): expository texts (encyclopaedias) tend to employ topical Themes; descriptive texts

[14] For more technical information about Themes, see the methodology (Section 4.3.3).

(geography manuals and travel guides) abound in location Themes (also, see El-Issa (2016) below); narrative texts (historical passages and biographies) frequently feature temporal Themes; instructive texts (manuals and recipes) commonly use process and temporal Themes; and argumentative texts (essays and editorials) predominantly select textual Themes (conjunctives) to introduce the next argument stage. In an example of a genre-focused study, tourist guides, when compared to biographies and geography textbooks, have been found to characteristically contain longer unmarked Themes (mostly denoting places), marked Themes realised by circumstantial Adjuncts (again denoting places and time), existential (*there*) Themes, and interactional Themes (El-Issa, 2016). Another example of a genre-focused study was Ghadessy's (1999) pioneering investigation of grammatical and lexico-semantic features of Themes in research article abstracts. Ghadessy compared several features identified in 150 abstracts from 30 disciplines to the same features from studies of obituaries, programmes, narratives, expository texts, and football commentaries, finding that regarding Theme selection (Thematic references to speaker/hearer/writer/reader, major animate text participants, abstract concepts, process, time, and location) abstracts were negatively correlated with all but expository texts. Abstracts are thematically more writer than reader oriented (e.g., the writer is used as a Theme, though infrequently, in *I/ we/the author(s)*, but the reader is not); processes or temporal/location Themes are used infrequently or negligibly; and crucially, Themes are mostly abstract entities owing to the frequent use of grammatical metaphors (grammatical metaphor is "a noticeable feature of the abstracts", p. 151), mostly through nominalisation of adjectives, verbs, and adverbs.

Another research strand focused on uncovering the disciplinary differences or characteristics of academic RAs, either as a whole or its various part-genres. For example, Gosden (1993) divides grammatical Subjects into four domains (*Participant*, *Discourse*, *Hypothesized and objectivised*, and *Real world*), and from their frequency in the four traditional RA sections (IMRD) in physics, chemistry, and biology shows that texts (especially in the discussion sections) become progressively more rhetorically persuasive and interactional, with declining *Real-world* Subjects (objects of study, mental processes, procedures, *it*, *there*) but higher use of visible *Participant* first-person *we* and integral citations as Subjects, and *Hypothesized and objectivised viewpoints* (*it is clear that…*, *it is interesting to note that…*). Contrary to this increasing interactivity, civil engineers completely ignore the *Participant* domain in their results and discussions, instead selecting to foreground the more objective and impersonal *Real-world* domain Subjects in almost 80% of sentences (Ebrahimi et al., 2014). In another classification of grammatical Subjects into *Phenomenal classes* (disciplinary study material) and *Epistemic classes* (methods, conceptual tools, researchers, and previous research), MacDonald (1992) shows that, overall, psychology and history are concerned with the *Phenomenal class* of groups as objects of study (27% and 44% of Subjects, respectively) rather than individuals (1%) as their study material (compared to literature's 30% focus on individuals and 10% on groups), with psychologists most frequently framing their sentences as being about evidence and reasons (49% of Subject Themes) or previous research (12%), thus foregrounding findings, methods, and inferences as the dominant way of knowledge building, as compared to 15% and 6% in history, respectively, and 7% and 5%, respectively, in literary studies.

Thematic studies of abstracts also found interesting inter- and intra-disciplinary variations. For example, Lores (2004) observes two major rhetorical organisation structures of linguistics abstracts: IMRD-structured informative abstracts and CARS-structured indicative abstracts. The IMRD abstracts tend to hide authorial presence, instead foregrounding *Discourse* (textual) or impersonal *Real-world* domains, whereas CARS-structured abstracts reveal it, showing that either overt or hidden personal agency in knowledge construction is acceptable in the discipline. The tendency of linguists to de-emphasise personal agency in IMRD abstracts was confirmed by Ebrahimi and Chan (2015). In their corpus, only 7% of applied linguistics IMRD abstracts utilised the *Participant* domain as the grammatical Subject (mostly in the methods section to highlight procedural novelty), compared to almost a third of the sentences in IMRD abstracts in economics, with first-person pronouns used in all four IMRD sections, possibly to boost the texts' interactivity.

The fact that Themes reflect differences in disciplinary knowledge making and the nature of knowledge becomes visible rather early: in student writing, that is, in the writing of those still learning the conventions of their fields (North, 2005). Comparison of essays written for the history of science class and authored by science majors and arts/social science/education majors revealed that the variation in thematic structures was underlain by diverging conceptions of knowledge construction in soft and hard fields. While the science majors used markedly fewer textual and interpersonal Themes, letting bare historical facts speak for themselves, arts/social science/education majors' use of textual Themes indicated more explicitly the rhetorical structure and persuasive nature of their argument, and their use of interpersonal Themes revealed either their epistemic commitment (stance) to the propositions or overt attribution of claims to

others. The two groups thus used Themes to depict knowledge either as consisting of plain facts or as constructed and negotiated by writers/readers. In addition, in student writing in at least some disciplines (whose writing has been analysed), there seems to be a general progression from Themes representing individuals and objects (real world phenomena) to Themes highlighting knowledge making (theories, scholars, and their perspectives or analyses as abstract entities) (Gardner, 2008). In some fields (wildlife behaviour vs. conservation biology), a greater use of epistemic Themes related to research, researchers, and knowledge making correlates with more positive evaluation by expert writers (Samraj, 2004).

While the above findings strongly suggest that students' use of Themes in writing progresses in the general direction of Thematic preferences in their fields (Gardner, 2008), unsurprisingly, differences between student and experienced writers in the use of Themes remain evident. In Montemayor-Borsinger's (2003) comparative study, a research paper drafted by a PhD student features half the instances of self-mention as Subjects but three times more Themes pointing to the text itself (e.g., *this section*) compared to the final (published) version rewritten by a head of the research laboratory. More interesting, though, are the quantitative (more frequent in the expert version) and especially qualitative differences in the writers' use of Instantial Subject Themes–Subjects that are disciplinary in nature, but allow packaging complex new ideas, rather than comprising only established technical terms. Instantial Subjects (underlined) such as "*The best available data from Small Scale Neutron Scattering is consistent with…", "The type of defects formed, how they evolve with increasing current and how their number and type are related to the magnitude of critical current and sample history are still open to questions",* or "*A careful look at the data shows…"*

(p. 47) used by the expert writer reveal that abstract concepts are treated as objective entities which can be discreetly evaluated or commented on, and thus serve the purpose of rhetorical positioning. In contrast, the novice writer's Instantial Subjects (underlined) either miss the opportunity to evaluate and or influence the reader subtly ("*The Fourier transform of the vortex structure in tilted fields is anisotropic…*") or are parts of rather straightforward evaluations ("*Thus it is not clear which is the type of defects…*").

Investigating the proportion of disciplinary topical Themes in published texts can give us a hint to what degree the legitimate contribution in a (sub-)discipline necessitates disciplinary specificity. Disciplines differ in their level of technicality. For example, according to Becher (1989), physics and mathematics are less intelligible because they employ a special, compressed, quantitative code. The humanities use some technical terms employed in otherwise familiar, everyday discourse. Disciplines like sociology, law or geography, employ varying specialist words with meanings likely to be alien to disciplinary outsiders. It is vocabulary, then, in thematic positions in this study, that "often provides a way of identifying which words best distinguish the texts of a particular author or group of authors from another" (Hyland, 2012b, p. 68). Let us now turn to selected issues in identifying technical and academic vocabulary.

2.4.1.1 Academic and disciplinary vocabulary

The earliest descriptive study of technical/scientific vocabulary that I am aware of dates to the early 1960s: Barber (1962, cited in Coxhead & Nation, 2001) looked at three academic texts of varying lengths (but not more than 10, 000 words each), and found clear evidence that certain words were much more common in academic texts than in non-academic ones. The study was influential for ESP in general, and for

research into academic vocabulary in particular (Coxhead & Nation, 2001). The research trend culminated in studies whose aim was to establish lists of disciplinarily non-specific academic words "common to a wide range of texts, and not so common in non-academic texts" (Nation, 2001, p. 189), that is, words falling between technical disciplinary terms and general vocabulary. The lists included Xue and Nation's (1984) University Word List and perhaps the best known of all such lists–Coxhead's (2000) Academic Word List (AWL). The latest iteration of an academic word list is Gardner and Davies' (2013) Academic Vocabulary List (AVL).

Compared to general academic words, which evidence differing frequencies, ranges, collocations, and meanings across disciplines (Hyland & Tse, 2007), disciplinary vocabulary is seen as perhaps the most easily recognisable hallmark separating academic disciplines (e.g., Becher, 1989; Hutchinson & Waters, 1987; Swales, 1990). Discipline-specific academic word lists (not to be confused with lists of technical vocabulary) were developed largely in the last decade, for example, in medicine (Wang, Liang, & Ge, 2008), agricultural science (Martinez, Beck, & Panza, 2009), applied linguistics (Vongpumivitch, Huang, & Chang, 2009), chemistry (Valipoury & Nassaji, 2013), and environmental science (Liu & Han, 2015). Comparing the lists (Table 2.1 below), we can see that each contains varying numbers of word families (vis-à-vis the AWL and other disciplines), which overlap with the AWL only partially. Granted, the lists are based on corpora of disciplinary RAs, while the AWL corpus included RAs, course books, texts from other corpora, textbook chapters, and laboratory manuals. Also, the criteria for inclusion into individual lists are not identical.

However, the surface-level comparison shows that the numbers of discipline-specific expressions can differ substantially[15].

Table 2.1

Disciplinary academic word lists

	No. of families	**Shared families**	**Tokens in corpus**	**Corpus contents**	**Inclusion criteria**
AWL	570	–	3.5 million	28 subject areas from art, commerce, law and science	At least 10x in at least 15 subject areas
Disciplinary word lists					
Environmental science	458	318	862,242	200 RAs in 10 subject areas	At least 30x in at least 8 subject areas
Medical science	623	342	1,093,011	288 RAs in 32 subject areas	At least 30x in at least 16 subject areas
Agricultural science	92	33 (Sub-list 1)	826,416	218 RAs	Items above the mean frequency for all items in the corpus
Chemistry	1,400 (including 683 GSL families)	327	4 million	1,185 RAs in 4 subject areas	At least 10x in each subject area
Applied linguistics	603	475	1.5 million	200 RAs from 5 journals	At least 5x in each journal

Apart from the disciplinary differences shown in Table 2.1, even sub-disciplines of the same discipline show considerable variance when it comes to vocabulary (Maher, 2016). Evidence for such variance comes from a research study by Hsu (2011) into 12 subject areas of business studies. The vocabulary knowledge thresholds for optimal comprehension, defined as knowledge of 98% of the running words in a text plus proper nouns (Laufer & Ravenhorst-Kalovski, 2010), have been found to diverge in the sub-

[15] Word families from the General Service List (GSL) were eliminated from the count, except for chemistry. The GSL was compiled by Michael West in 1953 as a list of 2,000 high-frequency word families (Nation, 2001).

disciplines of business by several thousand words!

Another type of vocabulary research focused on purely technical words, which can be defined as terms that are "recognizably specific to a particular topic, field or discipline" (Nation, 2001, p. 198). While the definition sounds simple, we need to realise that in fact technical words form a cline. For example, Nation (2001, p. 198) creates four bands:

- Category 1: The word is rarely used outside a specialised field (e.g., *morpheme, hapax legomena, lemma*).
- Category 2: The word can appear both in and beyond a specialised field, but the meaning changes (e.g., *reference, type, token*).
- Category 3: The word's technical meaning can be easily understood outside of a specialised field (e.g., *range, frequency*).
- Category 4: The word may be more commonly used in a specialised field, but it has no technical meaning, though background knowledge would provide a reader with a more precise meaning (e.g., *word, meaning*).

A similar cline comprising five levels appears in a more recent categorisation proposed by Ha and Hyland (2017), who argue that "any sense of a word implies a certain degree of technicality" (p. 46). Their TAM (Technicality Analysis Model) categorises words from least technical (TAM1), slightly technical (TAM2), moderately technical (TAM3), very technical (TAM4) to most technical (TAM5) based on their general/specialised meanings, frequencies in reference word lists, and monosemy/ polysemy.

Due to this complexity, various estimates of how many words in academic texts are technical have been proposed. According to Nation (2013), dictionaries of

disciplinary technical vocabulary likely contain at least 1,000 to 2,000 words, with those for disciplines like medicine or botany easily exceeding 6,000 words. Hutchinson and Waters (1987) note that technical vocabulary has been reported to cover only 9% of words in a large corpus of technical and scientific writing. Schmitt (2010) suggests that technical vocabulary comprises about 10% of a text. Hu and Nation (2000) estimate that in an academic text, low frequency words (many of which would be technical vocabulary) and proper names account for about 14%. However, these figures represent only rough averages. In actual empirical studies, Valipouri and Nassaji (2013) find that around 14% of their corpus of RAs in chemistry is composed of technical words; Nation and Chung (2003) report that out of the total number of words in an anatomy text, 37.6% (or one in every three running words) are technical words, while in an applied linguistics text, technical words make up only 16.3% (or one in every five running words) of the total. The lower proportion in the applied linguistics corpus is explained by the discipline's more accessible nature.

We can see from the above that empirical studies have also arrived at different proportions of technical vocabulary, showing that some disciplines use larger numbers of technical words. Yet, a point to consider is how the proportion of technical vocabulary in a corpus can be determined. There have been several approaches to identifying technical vocabulary. In their paper on the subject, Chung and Nation (2004) point out that none of the approaches is prevalent or can be considered standard, and studies comparing the reliability of the various approaches are practically non-existent. They propose and evaluate four approaches:

1. Using a rating scale

In this approach, presented as the most direct, disciplinary insiders rely on their judgement and field knowledge to identify technical words. They use a rating scale (a four-point rating scale is used by Chung and Nation) distinguishing the relatedness of individual words to the disciplinary subject matter. To counter potential discrepancies in how the scale is used, inter-rater reliability should be calculated.

2. Using a technical dictionary

While this approach seems straightforward, two caveats have to be taken into account. First, various disciplinary dictionaries may exist, and they may have been compiled following different inclusion criteria. (Also, as Nation (2013) notes, disciplinary dictionaries may often be single-authored, potentially further increasing inclusion inconsistency.) Second, decisions have to be made whether words appearing as main-entry headwords, as sub-entries and/or in definitions will be included as technical; whether members of the same word family will be counted as one word; and whether technical terms must be single words or can be part of multi-word expressions.

3. Using clues in the text

This approach relies on writers' direct indications that words are technical. The indications can be in the forms of definitions, bolding/underlining/bracketing, labels, and illustrations. However, Chung and Nation note that definitions can be hard to identify at times; typographical clues may serve a variety of purposes; and non-technical words may also be labelled in diagrams. The major issue, however, is that not all technical words are usually indicated, which is a consequence of a narrow focus on a specific issue in a text.

4. Using a computer to automatically extract technical terms

Computers can identify technical terms based on either linguistic or statistical cues. Linguistic information taken into account may include parts of speech, inflections, and grammatical structure. Often, linguistic cue analyses are followed by statistical analyses. Statistical formulas rely on the fact that technical terms occur in disciplinary texts more frequently than in non-disciplinary ones (roughly 50 times more often) (indeed, frequency is an oft-invoked criterion in studies of specialised vocabulary as evidenced, for example, by studies on disciplinary academic word lists in Table 2.1), but the formulas may also take into account other information, such as frequencies of different senses of the same word. Comparing the four approaches, Chung and Nation (2004) conclude that the rating scale is the most accurate measure but also the most laborious, whereas relying on authors' clues is the least reliable way. Yet, no single method may prove at the same time reliable (although Chung and Nation report a 100% success rate for the rating scale) and practical for dealing with large datasets, and thus combined approaches have been proposed, for example, Kwary's (2011) hybrid method combining keyword analysis and systematic classification, and the already mentioned Ha and Hyland's (2017) TAM.

The various approaches to identifying specialist vocabulary make it possible to investigate the degree of (sub-)disciplinary technicality. This investigation of technicality can then be combined with investigating topical Themes, thus revealing the proportions of disciplinarily specific topical Themes vs. other topical Themes in published texts. These findings, interpreted in the light of an approach from the sociology of knowledge, can make visible not only the legitimate contributions in (sub-)disciplines, but also potentially some of the larger organising principles underlying knowledge creation, specifically the strengths of relations of

(sub-)disciplinary discourses to (sub-)disciplinary knowledge structures. To the best of my knowledge, such an investigation has not been conducted, neither in earth sciences or in other fields. In addition, earth science texts have been thematically explored in only a handful of studies: a geology textbook (Love, 1993); conclusions of 30 geology RAs (Rashidi & Ghaedi, 2013), and student writing in an environmental science course (Samraj, 2004). Finally, this study can contribute to examination of disciplinarity by drawing on lenses from EAP, SFL, and sociology of knowledge.

2.4.2 Citations

Citation is one of the defining features of academic writing (Feak & Swales, 2009; Hyland, 2002a). It is also one of the most highly visible disciplinary conventions, reflecting disciplinary beliefs about appropriate uses of testimony and appeals to authority (Dowdey, 1992). As such, differing citation usage across (sub-)disciplines has attracted much research attention, not only in applied linguistics, but also in sociology, and information science (White, 2004).

Citations can be defined as "attribution of propositional content to another source" (Hyland, 2002a, p. 115). Since this attribution is commonly identifiable by the name of the source and year of publication, citations have also been defined "as a research report which has a specific reference point that is clearly identifiable" (Charles, 2006, p. 314). Of course, some citations may not have a particular identifiable reference point, yet they may refer to positions of other researchers–Charles herself (2006) terms such cases general reference–such as accepted positions in the sciences (in the case of earth sciences, they may, for example, be references to Newton's laws of physics).

However, I contend that the proportion of such general references is comparatively small, and thus the above definitions can be considered satisfactory.

Citations do not represent a "functionally empty" practice of displaying familiarity with sources, as some students may believe (Coffin, 2009, p.164), nor are their functions exhausted by simply delineating the context of knowledge or problem to which new research is going to contribute. Citations serve multiple functions–Harwood's (2009) author informants commenting on their use of citations named 11 purposes, for example, signposting, engaging, or advertising.[16] Some citations, Becher and Trowler's (2001) informants revealed, function as solidarity markers ("If we don't read them, perhaps they won't read us." (p. 115)) or as markers of "keeping good intellectual company" (p. 114). Other citation functions are explicitly argument building (e.g., positioning, supporting, building), and thus are tightly integrated as a means of persuasion into writers' arguments (Coffin, 2009; Hyland, 1999). In this way, citations show that a disciplinary conversation is ongoing (Charles, 2006), with the style of the conversation subject to social, intellectual, and epistemological disciplinary considerations (Becher, 1989).

From another perspective, citations are a measure of success in academia. They are evidence that a published paper has attracted attention and interest of the disciplinary community (Paul, 2000)–which is not insignificant when put into perspective. Large numbers of published articles, in fact, remain uncited for years. In some disciplines, the percentage of uncited papers is under 40% (e.g., physics and chemistry); in other fields such as engineering or language and linguistics, the proportion is close to 80% (Hamilton, 1991). While finer differences exist at sub-

[16] For other classifications of citing functions and for an overview of past studies on citers' motivations for citing see White (2004).

disciplinary levels (e.g., 51.9% of biomedical engineering RAs are uncited vs. 76.8% in mechanical engineering), and the rates are now almost three decades old, the proportion of uncited papers has likely not diminished given the recent surge in the number of published papers (summarised in Hyland, 2016a). Thus, garnered citations are also a form of reward in the form of monetary awards, tenure, and grants (Paul, 2000), that is, citations propel the academic cycle of rewards and status mentioned in Section 1.5 (Hyland & Salager-Meyer, 2008).

Citations can be divided into integral and non-integral (Swales, 1990). As Table 2.2 below shows, integral citations feature the researcher's name as a sentence element, making the researcher more prominent (e.g., Hyland, 1999; Thompson, 2002). Non-integral citations include parenthetical or superscript references, rendering authors less important than the reported message (e.g., Coffin, 2009; Thompson, 2002). The prominence of integral citations can be further distinguished based on the grammatical position the cited source occupies, with cited authors in Subject positions (authors' names as head nouns or pre-modifiers) being notably acknowledged for their contribution, while less so if authors appear in non-Subject positions (e.g., passives or other clause constituents) or in phrases (e.g., adjunct agent structures, such as *according to*…) (Hyland, 1999).

Table 2.2

Types of integral and non-integral citations (modified from Swales, 1990, p. 149)

Integral	Non-integral
a) Subject *Brie (1988) showed that the moon is made of cheese.* b) Passive agent *The moon's cheesy composition was established by Brie (1988).* c) Part of a noun phrase *Brie's theory (1988) claims that the moon is made of cheese.* d) Adjunct structure *According to Brie (1988), the moon is made of cheese.*	a) Parenthetical citation *Previous research has shown that the moon is made of cheese (Brie, 1988).* b) Super-scripted citation *It has been established that the moon is made of cheese.* [1-3] c) Contrastive citation *The moon may be made of cheese (but cf. Rock, 1989).*

Integral and non-integral citations have been classified further, based on functions performed rather than structural integration, allowing more fine-grained distinctions. For example, Thompson (2002) and Thompson and Tribble (2001) further distinguish Swales' integral citations into verb controlling, naming, and non-citations, and non-integral citations into source, identification, reference, and origin (the examples, which are modelled on Swales' (1990) example sentences, have been added by me):

a) Verb controlling: the cited agent controls the verb

   Brie (1988) has observed the cheesy structure of the moon.

b) Naming: the cited source is a noun phrase or a part of a noun phrase (e.g., as a method, formulation, or someone's work)

   The cheesiness scale of Brie (1988) has been influential.

c) Non-citation: a name given without the year of publication, usually when already provided earlier

The cheesy structure of the moon, first observed by Brie, has led to searching for similar structures on other planets.

d) Source: an attribution of a proposition or a statement (which can be used either as a support for one's argument or to launch a counterargument) to another source

The moon seems to be made of cheese (Brie, 1988).

e) Identification: either an explicit or an implicit identification of an agent in a sentence

It has been suggested that the moon is made of cheese (Brie, 1988).

f) Origin: similar to Source, but the attribution is focused on a concept, technique, or product, and the citation is in the form of a noun phrase.

The cheesiness scale of the moon (Brie, 1988) can be applied to other planets as well.

The same fine-grained division of non-integral citations (d-f) could be applied to super-scripted citations. The last type of non-integral citation in the classification of Thompson (2002) and Thompson and Tribble (2001) is similar to Swales' (1990) contrastive citation (the example, which is modelled on Swales' (1990) example sentence, has been added by me):

g) Reference: a reference to another text for further detail (often accompanied by the directive "see")

The moon may be made of cheese (see Rock, 1989).

A citation's prominence also depends on its textual integration, namely, the extent to which a reference has been quoted directly or assimilated into a text in a summarised or re-worded form (Coffin, 2009). Coffin distinguishes three options (the examples, which are modelled on Swales' (1990) example sentences, have been added by me):

a) Insertion: a directly quoted source

The discovery that "cheese seems to be the primary material in the moon's planetary make-up" (Brie, 1988, p.22) has stunned the research community.

b) Assimilation: a reworded, summarised or paraphrased proposition

The discovery that the moon seems to be made of cheese (Brie, 1988) has stunned the research community.

c) Insertion + assimilation: a combination of a direct quotation and rewording

When Brie (1988) described cheese as the main component of the moon's "planetary make-up" (p. 22), our understanding of how planets were created changed forever.

The above-mentioned classifications are applicable to all disciplines, but that is where similarities end. Research into disciplinary citation practices has uncovered clear differences in the use of integral and non-integral citations. For example, regarding the prominence of cited sources, Hood (2011) finds that in the natural sciences, the source authors are mainly rendered invisible by superscript notations, which implies that the projected knowledge is more important.[17] In contrast, in the humanities, she discovers that the source authors are visible, often appearing in integral citations, indicating that the kind of knower is important for legitimising knowledge claims. On Hood's (2011)

[17] Hood (2010) additionally distinguishes superscript references by the size of the superscript numbers, positing that small superscripts further reduce the visibility of origins of cited propositions.

cline of authorial source visibility, the social sciences occupy the middle space, although, as Cao and Hu (2014) find, various social science disciplines can occupy different positions within this middle ground. Those social science disciplines that use integral citations more frequently are more likely knower-oriented disciplines (Cao & Hu, 2014). Similar findings are presented by Hyland (1999), who shows that while non-integral citations are generally preferred across both hard and soft disciplines, writers in the humanities and social sciences are more than two times as likely to utilise integral citations than natural scientists. Integral citations do occur in the natural sciences, but they are comparatively infrequent (Hyland, 1999; Okamura, 2008).

If we focus on the grammatical positions of cited authors in integral citations, we can also find disciplinary preferences. In the natural sciences, authors are placed more frequently in passive or adjunct grammatical structures (Hyland, 1999). Also, various procedures are named after researchers, but that does not necessarily acknowledge the contributions of those researchers: Rather, the names are used in a technical sense (Hyland, 1999). This usage is also noted in the social sciences by Hood (2011), who mentions that authors' names may at times function as mere pre-modifiers of a product of a research process.

Yet another disciplinary difference is the extent to which researchers cite previous work. The extent is reflected mainly in two ways. First, higher numbers of citations have been found in the humanities and social sciences RAs than in the natural sciences texts (Halevi, 2013; Hyland, 1999). The discrepancy can be striking. For example, an average of 104 citations were used in sociology RAs vs. 25 in physics RAs (Hyland, 1999). However, this conclusion does not apply across all hard fields. For example, it is not borne out by citation numbers in molecular biology (Hyland, 1999),

presumably because biology is a relatively new field with more descriptive methods, and "a disciplinary ethos that emphasises proprietary rights to claims" (Hyland, 1999, p. 357)[18], and because medicine is a composite field with connections to multiple disciplines, and its RAs are intended for diverse audiences (Hu & Wang, 2014). The general numerical difference can be explained, according to Hyland (1999), by the fact that natural scientists tend to research well-defined and narrowly specific problems, where certain shared background is expected. In these hard fields, a knowledge claim eventually loses its reference as it is incorporated into the canon of disciplinary knowledge and becomes a given "fact". In addition, in some fast moving fields (e.g., biology), focusing on developments within one's own specialization is the only feasible strategy due to lack of time and difficulties understanding technical vocabulary or argument structure of other fields (Becher & Trowler, 2001). In contrast, humanities and social sciences researchers cannot presuppose an extensive shared background since they research much broader problems and topically diverse issues. Therefore, while natural scientists integrate their findings into specific contexts of previous research, and at the same time draw on it for support, in the humanities and social sciences researchers draw on several related fields and historical studies to negotiate answers to problems that cannot be answered by following a single well-defined path, which means that older research cannot be simply discarded as outmoded (Hyland, 1999). In other words, "some research topics benefit from a historical grounding, and reference to some older sources speaks to the writer having taken an appropriately broad time perspective" (Pecorari, 2006, p. 19).

[18] Becher and Trowler (2001) note that "some would take issue with this" (p. 115) interpretation.

The second difference in the extent of reference to previous work is that the age of cited references tends to be somewhat lower (i.e., the references are more recent) in the natural sciences and engineering. Studies have found that that the median age of all types of engineering references is only nine years (Young, 2014), while in the humanities and social sciences (theses and dissertations), the average reported age ranges from 19 to 22 years (Pecorari, 2006). Another study has attributed about 70-80% of references in the hard disciplines to more recently published sources (within the last 15 years), whereas in the humanities and the social sciences newer references typically make up between 40% and 70% of the reference lists, with large differences between disciplines (e.g., 45% in social sciences and 68% in health professions) (Halevi, 2013).

To the above-mentioned differences between the natural sciences, the social sciences, and the humanities, we can add other aspects of disciplinary citation practices that have been investigated: the choice of reporting verbs (e.g., Hyland, 2002a; Thompson & Ye, 1991), phraseological patterns in reporting clauses (e.g., Charles, 2006), writer stance, and nature of source (e.g., Coffin, 2009). In addition to these areas, I propose that more research could focus on how citation practices reflect the principles underlying (sub-)disciplinary knowledge building, where knowledge needs to be understood not only as a canon of specialist findings, but also as an object having its form, properties, and tendencies (Maton, 2014a). To date, only initial steps in that direction have been taken by Hood (2011) and Cao and Hu (2014), who have suggested links between the visibility of cited sources and types of knowledge and knower structures theorised by sociologists of education and sociologists of knowledge. Especially interesting may be investigation into what citation practices reveal about knowledge creation in earth sciences, a field combining features of historical and

natural sciences (Cleland, 2001; Cleland, 2013) and displaying characteristics that do not seem to conform to the traditional notions of how experimental sciences progress (Raab & Frodeman, 2002).

2.4.3 Expressing certainty (hedging and boosting)

From the previous literature review sections it is evident that academic writing is hardly, as Hyland (2002d) points out, the objective, faceless, and impersonal style that traditional views depict it to be. Instead, we could say that writing in "even the most "rhetorically innocent" style (Hyland, 2005a) is itself a way of persuading one's audience. Persuading one's audience then shifts our focus towards writer-reader interaction.

Writer-reader interaction has been investigated from several perspectives. For example, Ho and Li (2020) list the following approaches: metadiscourse; appraisal; interacting via politeness strategies; providing explanation, exemplification, and reformulation; interacting with the reader-in-the-text; and creating and including a fictionalised reader in the text. Most research that I am aware of utilised the construct of metadiscourse[19]. Metadiscourse has been defined and understood in different ways, which can be put on a cline from a narrow to an extended conceptualisation (Hyland, 2017b): at one end, metadiscourse refers to textual features that help organise texts (e.g., direction, purpose, internal structure); then, metadiscourse can in addition include references to writers, readers, and texts; and finally at the other end, metadiscourse extends to encompass linguistic devices that signal the writer's stance towards the text content and the reader, as well as the preferred way the text should be interpreted by the

[19] Indeed, Hyland (2017b) notes that metadiscourse features extensively both as a term and a research topic in various databases.

reader. This end of the continuum is represented by Hyland's (2005b) interpersonal model of metadiscourse (reproduced below), which comprises an interactive dimension (resources that organise texts and assist readers' comprehension) and an interactional dimension (a writer's evaluative textual voice in dialogue with readers).

Table 2.3

Hyland's (2005b) interpersonal model of metadiscourse (p. 49)

Category	**Function**	**Examples**
Interactive	**Help to guide the reader through the text**	**Resources**
Transitions	express relations between main clauses	in addition; but; thus; and
Frame markers	refer to discourse acts, sequences or stages	finally; to conclude; my purpose is
Endophoric markers	refer to information in other parts of the text	noted above; see Fig; in section 2
Evidentials	refer to information from other texts	according to X; Z states
Code glosses	elaborate propositional meanings	namely; e.g.; such as; in other words
Interactional	**Involve the reader in the text**	**Resources**
Hedges	withhold commitment and open dialogue	might; perhaps; possible; about
Boosters	emphasize certainty or close dialogue	in fact; definitely; it is clear that
Attitude markers	express writer's attitude to proposition	unfortunately; I agree; surprisingly
Self mentions	explicit reference to author(s)	I; we; my; me; our
Engagement markers	explicitly build relationship with reader	consider; note; you can see that

The interactional dimension of this influential model was subsequently presented in finer-grain detail by Hyland himself (Hyland 2005a), to elaborate on components of stance and engagement. Through stance, writers communicate their opinions, evaluations, level of certainty, and place themselves more or less authoritatively in the text, using hedging, boosting, attitude markers, and self-mention. Through engagement (reader pronouns, directives, questions, shared knowledge, and personal asides), writers connect with their readers, explicitly acknowledging their presence, in order to pull readers into co-constructing the meaning of the text in the way that guides readers to align their interpretation with writers.

Most recently, Hyland's interpersonal model of metadiscourse has served as a basis for the Metadiscourse-Stance-Participation (MPS) framework (Ho & Li, 2020), which is an attempt to "plug conceptual loopholes" (p. 61) and thus offer a theoretically more robust model of reader-writer interaction that excludes from Metadiscourse categories that are not exclusively text internal: hedging, boosting, and attitude markers, because they refer to a writer's attitudes to real-world phenomena or other texts; and evidentials (e.g., *according to*...), because they refer to other texts. The excluded categories of hedging, boosting, and attitude markers instead form the Stance part of the MPS framework.

Of the various elements of writer-reader interaction, hedging and boosting (also called evidentiality; Hyland and Jiang (2016)) are directly relevant to this thesis, as the two concepts are used to operationalise Authoritativeness of knowledge (Section 4.3.5). For the purpose of the operationalisation, it is of little relevance whether hedging and boosting are conceptually subsumed under interactional metadiscourse - stance or directly under stance. Hedges, according to Hyland (2005a), are a set of linguistic devices (e.g., modals such as *might*, adverbs such as *possibly*) that signal lack of certainty in or of complete commitment towards a proposition, instead communicating plausible reasoning. When researchers advance new propositions, they do not only build on existing knowledge, but possibly also contradict some established claims, which may create tension with others in the disciplinary community (Cheng & Unsworth, 2016). Hedging allows writers to avoid direct invalidation and confrontation by creating a dialogic space that acknowledges the readers' role in negotiating or assessing claims before they become accepted (Hyland, 2005a). This dialogic space then expresses the writer's deference to the disciplinary community (e.g., Hyland, 1996a).

Where hedges open dialogic space, boosters close it (Hyland, 2005b). Boosters are linguistic devices that express commitment and certainty (Hyland, 2005a) through, for example, lexical verbs such as *show* or *demonstrate*, or adverbs such as *always* or *obviously*. Expressing conviction is important since authors need to persuade journal editors and readers that the claims that are being advanced in article manuscripts are robust (Peacock, 2006). Boosters, especially if they express evidential (i.e., unambiguous) truth (Skelton, 1997), such as the above examples *show* and *demonstrate*, and are coupled with words such as *figures*, *data*, or statistical terms communicate that findings are facts that emanate from a study's results rather than from its writer(s)' judgements. Boosting further helps writers to engage readers by communicating "solidarity with an audience, stressing shared information, group membership" (Hyland, 1998, p. 350) when boosters (e.g., *it is a well-known fact*) express that the claims/facts put forward are considered as familiar to the disciplinary reader (Vassileva, 2001). In general, then, both hedges and boosters reveal the writer's positioning towards one's findings and issues examined in own texts, as well as evaluation of findings and positions of other disciplinary members in other texts (Hyland, 2005a).

Hedging and boosting have been the subject of numerous studies aiming to tease out differences in their usage underlaid by the writers' first language background (e.g., Farokhi & Emami, 2008; Vassileva, 2001), discipline (e.g., Hyland, 1998; Peacock, 2006), research paradigm (e.g., Hu & Cao, 2015), register (e.g., Biber, 2006), or ability to identify such expressions in texts (e.g., Hyland, 2000; Lewin, 2005). The research strand focusing on disciplinary differences is the most relevant to my thesis, and I shall summarise some of its findings below.

Hyland (1998) investigated textual practices in published RAs from philosophy, marketing, applied linguistics, and sociology (representing the humanities and social sciences) and microbiology, physics, electrical engineering, and mechanical engineering (representing the sciences). Textual analyses, which involved searching the corpus against a list of 180 hedges and boosters, were complemented by interviews of authors from the disciplinary communities, both about the rhetorical practices and intentions of their texts and texts written by others. The findings show that in all disciplines, hedges are more frequent than boosters, with the humanities/social science disciplines employing more markers of evidentiality than the sciences, especially hedges, by about two and a half times. This quantitative difference can be interpreted as a reflection of the ontological and methodological assumptions and discourse conventions of the humanities/social sciences and natural sciences. The former disciplines inquire into areas which are more impacted by contextual vagaries. Variables are not easily controlled for, and results are more diverse. Less can then be claimed unequivocally, and more interpretive effort and personal projection are needed. The latter disciplines work on a narrower set of clearly defined problems (within each discipline), and draw on common background knowledge and methodology. Procedures and findings are portrayed more impersonally, as if data was speaking for itself, which is achieved by less frequent use of hedges and boosters. Another interesting finding is that both categories of evidentiality are largely represented by a relatively small number of commonly occurring expressions. The most common hedges are the modal verbs *may*, *would*, *could*, and *might* (all within the top five together with the adjective/adverb *possible(y)*), with the lexical verbs *suggest*, *indicate*, *seem*, and *assume*. Among the most common boosters are *will*, *the fact that*, *show*, *it is clear/clearly*, and *indeed*. Even

here, however, disciplinary preferences are noticeable. Humanities and social sciences texts include more cognition verbs (e.g., *think*, *believe*, or *suspect*), which carry subjective connotations, while natural science texts employ less-subjective discourse verbs (e.g., *indicate*, *suggest*, *show*) and also prefer modal verbs, which are more easily connected with inanimate objects and phenomena.

Disciplinary differences in the usage of evidentiality (only boosters) have been found also by Peacock (2006). The natural sciences' (represented by physics and environmental science) preference for less frequent boosting and a narrower range of expressions, especially *demonstrate*, *show*, *find*, or *establish* points to a tendency to minimise personal involvement, and to create the impression that knowledge claims emanate from data, evidence, and observations. In comparison, humanities and social sciences (language and linguistics, law, public and social administration) writers rely to a larger extent on personal persuasion and presentation, which is reflected in boosting noticeably more frequently, using a range of expressions. Common boosters (apart from those listed above) are, for example, *especially*, *apparently*, *particularly*, or *considerably.* Peacock's study has also confirmed that a handful of commonly-occurring expressions (e.g., *show*, *will/will not*, *clear(ly)*, *establish*, and *must*) account for a large percentage of the total number of boosters (at least 40% in the corpus, though in some disciplines close to 60%, and *show* represents 20% of all boosters in physics). The overlap in preferred boosters between this study and Hyland's research mentioned above is striking, given that the studies investigated mostly different disciplines, and Peacock drew on a twice-longer list of boosting expressions, which also explains why overall higher frequencies of boosters are reported there.

The two studies summarised above provided a snapshot of disciplinary usage of hedging and boosting relevant at and around the time of their publication. A complementary perspective has been provided in a diachronic study into how usage of evidentiality expressions changed during a half century from 1965 to 2015 in applied linguistics and sociology (representing the social sciences) on the one hand and electrical engineering and biology (representing the sciences) on the other hand (Hyland & Jiang, 2016). The findings reveal a surprising trend towards usage convergence as the frequency of hedging and boosting in published RAs gradually decreased in the social sciences by between 21% and 38% (per 10,000 words of text), while at the same time increasing in the sciences (with the exception of boosting in biology), albeit less dramatically. A possible explanation for this trend may lie in the rise of quantitative studies in the social sciences, accompanied by less subjective expressions of evidentiality in favour of data-driven claims. For example, cognition verbs like *believe* or *think* have declined, while research-oriented verbs such as *find*, *demonstrate*, or *show* and hedges like *suggest* and *likely* have increased. Simultaneously, researchers in the sciences have come to use stance markers more (particularly boosting in engineering), possibly in a bid to promote their role in knowledge creation in response to external factors, such as tenure and grant award decisions. Engineers have also started using a wider range of expressions, with *establish*, *prove* and *clearly* becoming more common, underscoring the strength of claims.

In the last study reviewed here, Hu and Cao (2015) investigated the use of hedging and boosting in post-methods sections in RAs from applied linguistics (specifically the sub-field language teaching and learning), education (the sub-field learning and instructional science), and psychology (the sub-field counselling and

clinical research). Their findings yet again confirm preferences for certain expressions, both discipline based and paradigm based (quantitative vs. qualitative studies), such as for the epistemic *show* in quantitative studies (especially in applied linguistics and education) and for *will* in qualitative research. While *show* appears in qualitative studies as well, it accompanies interpretations of qualitative evidence, unlike *show* in quantitative studies, where it is often used to support findings of statistical nature. There were no notable differences in the rhetorical functions that hedges and boosters perform.

A new angle Hu and Cao (2015) bring is in the interpretations of at least some findings through a theoretical and analytical framework from the sociology of knowledge–knowledge and knower structures, later subsumed into legitimation code theory (LCT), which is also the framework used in this thesis (Section 3.3). A higher count of markers of evidentiality, especially boosters, in applied linguistics compared to the other two disciplines is interpreted as "a function of the knower code at work" (p. 20). Applied linguistics represents a field where knowledge claims are legitimated through insights of disciplinary authors (knowers). Boosters assist in positioning authors as privileged disciplinary knowers by strengthening authorial voice and emphasising conviction, which helps in persuading readers. In comparison, the lower number of boosters in psychology is interpreted as a preference for a less author-centred expression, with legitimate knowledge underscored by adherence to agreed-on methodologies and procedural adequacy, which then help align readers with new knowledge claims put forward.

Hu and Cao's study is the only one I am aware of that has attempted to link an EAP approach to investigation of hedging and boosting with an interpretation following a sociological framework. While their study can be methodologically refined to more

robustly characterise disciplinary knowledge-making practices, the suggested link has the potential to further illuminate some of the complexities of disciplinary discourse beyond what corpus, ethnographic, and genre perspectives have or can achieve. In this thesis, I aim to exploit this potential.

2.4.4 Personal pronouns

One of the linguistic features that research on disciplinary specificity in EAP/ ESP has long focused on is the use of personal pronouns. Personal pronouns overtly place the author inside the text and highlight authorial identity (Hyland, 2014). This overt presence has a long tradition, since at the onset of academic publishing, what later became RAs took the form of letters that scientists sent to editors (Bazerman, 1988). In them, Bazerman (1988) notes, author scientists/contributors described first-hand accounts, employing first-person pronouns, of observed phenomena to awe their readers, but also to promote their discoveries and gain publicity and support for their explorations and experiments. But when the letters gradually changed from private communication to be more public and publishable-as-they-were, personal pronouns started to disappear as researchers took a backstage position from the previous emphasis on individual skill and personal/communal witnessing. As an alternative, the passive voice tended to appear more frequently and came to dominate scientific writing in the 20th century (Leong, 2014). At the same time, the first-person narrative was kept alive, perhaps as Swales (1990) notes, because early researchers' use of first-person pronouns projected their sense of honesty and modesty, avoiding the presupposition inherent in the passive that anybody could (or even worse could not) replicate the experiment.

Today, personal pronouns mark a stylistic feature representing a philosophical[20] "choice of scientific paradigm" (Starfield, 2015, p. 258) between positivistic objectivity with elided researcher presence (Starfield, 2015) and linguistically constructed reality and enacted social relations (Halliday & Martin, 1996). Within the traditionally positivist orientation, "a hallowed concept for many" (Hyland, 2001, p. 208), personal pronouns can then be regarded as a challenge to objectivity (Starfield, 2015), making their use marked (Harwood, 2005a), highly controversial (Hyland, 2001), and thus "the object of intense scrutiny and judgment" (Starfield, 2015, p. 254). However, interestingly, the traditional axiom of "author-evacuated" scientific writing (Geertz, 1988, p. 9) may be weakening, as the tendency for authors to reveal their presence is growing (Basal & Bada, 2012; Cheung & Lau, 2020; Hyland & Jiang, 2016).

The increase in authorial presence mirrors changes in the use of the passive voice. For example, Leong (2014) shows a decrease in the number of passive constructions used in his corpus of RAs from six journals (*Nature* and *Science* among them) covering a variety of research fields. He finds that while in the early 1990s, about half of all the clauses were passive, today the proportion stands at around 30%, with most of the instances concentrated in the methods sections. Leong (2014) acknowledges that despite a generally more widespread encouragement for active-voice writing, the proportion of the passive voice in scientific writing is unlikely to drop any further. In addition, he shows that the drop in contemporary disciplinary writing does not apply universally to all fields: In RAs from chemistry, physics, and biology, the passives

[20] The reason I call the choice philosophical is because Starfield notes that using first-person pronouns entails questions "of authorial self-representation, of choice of scientific paradigm, the role of the researcher, and the nature of research itself" (p. 258).

account for almost 40% of sentences; in RAs from psychology, the proportion is below 30%.

Leong's findings indirectly confirm findings of others that the frequency with which personal pronouns appear varies across disciplines. Generally, they are less common in the natural sciences–hard fields, and more common in the humanities and social sciences–soft fields (Hyland, 2005a), which are more interpretive since their study subjects (humans) allow less control over variables, making the researchers work harder to persuade their readers (Hyland, 2014). This generalisation does not always hold, though, given that disciplinary writing conventions are dynamic, and studies show that some hard disciplines can feature vastly more first person pronouns than traditional soft disciplines such as literature (e.g., Cheung & Lau, 2020). Yet, even a disciplinary lens may allow us to discern only rough tendencies, since personal pronouns have been shown to vary within disciplines as well (e.g., Harwood, 2006; McGrath, 2016).

Across disciplines, personal pronouns perform similar pragmatic functions. In a study of eight selected disciplines from natural/social sciences and the humanities, Hyland (2003b) shows that personal pronouns are most commonly used to explain a procedure, state results or claims, elaborate an argument, and state goals. Similarly, in biology, personal pronouns are most common in the discussion to restate results to claim responsibility for novel findings, and in results sections to state results, and restate methodology that led to the novel results (Martinez, 2005). In literary studies, the majority of first-person pronouns relate to explicating the reason for conducting a particular study, expressing a viewpoint or attitude towards previously published knowledge, and presenting finding and claims (Cheung & Lau, 2020). In electronic engineering, computer science, and physics, personal pronouns also highlight

commitment and contribution to research, give reason, and indicate necessity (Kuo, 1999). In business management RAs, pronouns appear by far most frequently in the methods section followed by half that frequently in results and discussions (Duenas, 2007), and their most common functions are basically the same as mentioned above.

However, we need to remember that personal pronouns are not monolithic (Kuo, 1999; Starfield, 2015). Firstly, they may refer not only exclusively to writers themselves, but also inclusively to the disciplinary reader (Hyland, 2014; Kuo, 1999), and thus may appear in first-, second-, third-person or indefinite forms (Kuo, 1999).

Secondly, according to Harwood (2005a) and Martinez (2005), personal pronouns emphasise the writers' roles to varying degrees, making some roles low risk (e.g., explaining a procedure) and some high risk (e.g., stating results and making claims). These roles have been summarised in functional frameworks, which highlight functions such as stating a purpose/result/claim/opinion; explaining a procedure; developing an argument; or guiding the reader. For example, one of the earlier and influential functional pronoun taxonomies (based on undergraduate essays, published academic articles and books in linguistics) was that of Tang and John (1999), which recognises six roles behind first-person pronominal uses, clearly showing the heterogeneous nature of this linguistic feature. The identities are presented below in the authors' order from the least powerful to the most powerful authorial presence (p. S27-S29):

- *'I' as the representative* often refers either to human beings in general or to disciplinary members.
- *'I' as the guide* includes both the writers and the readers sharing the experience of reading through the text.

- *'I' as the architect* positions the writers in the role of organisers and commentators on what is going to come. Tang and John acknowledge that this role may be similar to the one of a guide, but they point out that *'I' as the architect* emphasises the creative and organising part of the writers as opposed to merely guiding and accompanying the readers on their journey through a text.
- *'I' as the recounter of the research process* emphasises what the writers did in order to devise their experiment.
- *'I' as the opinion-holder* communicates the writers' perspective or standpoint on known and established issues in the discipline, often through reference to previously published research.
- *'I' as the originator* is the strongest authorial identity in the functional framework. It positions the writers as "perceiv[ing] themselves as people who have the *right* and the *ability* to originate new ideas" (p. S29; original emphases).

Similar functional categories have been proposed by other authors. For example, Harwood (2005a) identifies seven uses of the pronoun *I* referring to methods sections in the writings of student and expert authors. He finds that this methodological use of the personal pronoun can appear as a muted, low-risk strategy, but it can in fact highlight the authors' contribution, promote their acumen as researchers, display accurate judgment, and emphasise the uniqueness of their methodology. Cheung and Lau (2020) adopted Tang and John's (1999) framework, and identified two new roles: '*I' as researcher* (justifying the need for a study and the approach used by providing an overview of a research landscape, referencing previous studies, explaining terms, and stating limitations) and '*I' as explainer of the research process* (largely in the methods

section; explaining diagrammatic and tabulated information about the research process, and guiding through models, calculations, and formulas). Hyland's (2002c) characterisation also displays similarities with Tang and John's (1999) regarding the cline of authority. Nevertheless, Hyland purposely includes only exclusive uses of personal pronouns, and his typology also differs in its focus on "clear discoursal functions" (p. 1099) as opposed to Tang and John's (1999) authorial roles, which in Hyland's words possesses "the advantage of more clearly highlighting only cases of self-mention and in avoiding the kinds of discoursal overlaps which occur when employing labels such as 'guide' and 'architect'" (p. 1099). I do not necessarily agree with Hyland, because from my perspective *'I' as the guide* and *'I' as the architect* fulfil clearly distinct functions: In the former "the writer as guide is always implicitly or explicitly accompanied by the reader" (Tang & John, 1999, p. S28) and presents the text and/or argument from a common vantage point, whereas the latter is more of a metadiscoursal organisational signpost reserved only for the author(s). Their separation then allows a more detailed analysis. For comparison, I have summarised Tang and John's, Hyland's, Harwood's, and Cheung and Lau's frameworks in Table 2.4 below. In the table, I have attempted to match the functions according to their approximate similarities, which was easier for Tang and John's and Cheung and Lau's on the one hand, and for Hyland's and Harwood's taxonomies on the other hand. However, the functions in the former two taxonomies roughly correspond to only several initially listed functions of the latter two taxonomies.

Table 2.4

Functional taxonomies of Tang and John (1999), Hyland (2002c) and Harwood (2005a)

Tang and John (1999)	**Cheung and Lau (2020)**	**Hyland (2002c)**	**Harwood (2005a)**
'I' as the architect of the essay	*'I' as architect*	*Stating a purpose*	*'I' to state a purpose*
'I' as the recounter of the research process	*'I' as recounter of the research process*	*Explaining a procedure*	*Methodological 'I'*
'I' as the opinion holder	*'I' as opinion holder*	*Elaborating an argument*	*'I' to elaborate an argument*
'I' as the originator	*'I' as originator*	*Stating results/ claims*	*'I' to state results and claims*
'I' as a representative	*'I' as representative*	*Expressing self-benefits*	*'I' to express self-benefits*
'I' as the guide through the essay	*'I' as guide*	*First person in acknowledgements*	*'I' as a disciplinary servant*
	'I' as researcher		*'I' to define terms*
	'I' as explainer of the research process		

Thirdly, all the pragmatic functions can also be seen through other lenses. Thus, personal pronouns can be relational (Starfield, 2015), dialogic (Hyland, 2014), interactional (Kuo, 1999), political (Pennycook, 1994), and promotional (Harwood, 2005c; Hyland, 2003b). Inclusive *we*, used as a marker of solidarity, can be used to mitigate one's criticism of claims advanced by other researchers in published RAs, while exclusive *we* can soften (the imposition on the disciplinary community of) one's own claims (Myers, 1989). In fact, a single instance of a personal pronoun can function in multiple ways at the same time, and so categorising personal pronouns according to only one taxonomy of functions may be problematic (McGrath, 2016; Vladimirou, 2007). Fourthly, patterns of personal pronoun use change with time (Harwood, 2006).

Compared to the range of options identified above, how personal pronouns are used in earth sciences has been little investigated. In fact, to the best of my knowledge, the topic has been explored by only one or two researchers. The research studies conducted by Dressen (e.g., 2002, 2003, 2008, and 2014a) focus primarily on textual silences–or what remains unsaid–in geologists' field accounts, and on the development of geologists' disciplinary voice in a 10-year period following the completion of their PhDs. For example, conventional authorial distancing between the writer and the research account (Dressen, 2002) can be seen in the (non-)use of exclusive first-person pronouns, which Dressen (2014b) includes under personalisation cues, one of the variables that index geological field practice. While geologists still have the option of using exclusive personal (and possessive) pronouns to overtly signal their presence in field accounts, this type of signal is the least common marker of persuasion and community membership (Dressen, 2003). Dressen (2014b) interviewed experienced geologists who have confirmed that a field description shows to the reader that the writer has field experience and is able to observe and interpret like a field geologist. While having the field experience and the ability to observe and interpret are necessary prerequisites for disciplinary practices and culture in geology–indeed, the experience "makes geology *geology*" (p. 74; original emphasis)–experienced practitioners still prefer to let the observed facts come to the fore. Geological field accounts may thus be notable for their complete absence of human characters (Dressen & Swales, 2000). Geological novices use personal pronouns abundantly because they want to emphasise what they did (Dressen, 2008), unlike more experienced geologists who let rocks tell the story (Dressen & Swales, 2000). However, if, on the rare occasion, an exclusive personal pronoun is used by experienced geologists, then the claim which is thus

personalised "must be taken to be true" (Dressen, 2008, p. 248). In this way, pronouns strengthen claims by showing personal investment.

Clearly, the specific but necessarily limited focus of Dressen on first-person pronouns in one part-genre within the RA in geology is unlikely to represent the full range of uses. However, in the absence of other similar studies of earth sciences, Dressen's studies raise an interesting question of how these pronouns used in other RA part-genres and in other sub-fields contribute to authorial prominence in the discipline. More importantly, the prominence that first-person pronouns might afford to authors in turn raises questions about the existence of an underlying (sub-)disciplinary knower structure, underlying relations to the knower structure, and knower role in legitimate knowledge creation. These are the larger issues I aim to explore in my thesis.

2.5 Appraisal of research into disciplinary discourse in the EAP tradition

Academic discourse is a social accomplishment and a cultural activity shaped by epistemology, practices, and reader-writer interactions (Bazerman, 1988; Hyland, 1999). As such, studying academic discourse means that EAP/ESP researchers have to come to grips with complexities of human behaviour and human perceptions. I hope that through the selective review of the literature in this chapter and the EAP/ESP research traditions sketched in Chapter I, I have demonstrated to the reader that EAP/ESP researchers have achieved notable progress in accounting for language use across academic disciplines.

Despite our endeavours to attend to the multitude of social, contextual, and ideological factors by following various paths of enquiry, challenges remain, not least among them the question of what we should focus on more: the various dichotomies of the bottom-up or top-down approaches to textual analysis; the linguistic features at the

sentence level or the discoursal features at the textual level; the text or the social context. Our approaches to studying academic discourse have not filled the various gaps still evident today. For example, Lillis (2008) pointed out a gap between language and culture, text and context brought about by theoretical stances based on linguistic/ formalist approaches to language study. Our currently used lenses may simply not be powerful enough to reveal how we can bridge the gap. The gaping hole was pertinently summarised by Swales (1997):

> The EAP movement tends to rely on a rather narrow range of canonical texts in such areas as systemic linguistics, needs analysis and corpus studies, and thus is prone to ignore important work in areas such as anthropology, sociolinguistics, rhetoric and social theory. The wider body of work usefully raises as many questions as it usefully provides answers and insights. (p. 373)

Similar voices have echoed from other quarters. For example, Atkinson (1998) stated that studies of the discourse of science have overly relied on one approach (rhetorical analysis), and that in the case of such a complex phenomenon as scientific discourse, a number of integrated methodological approaches should be used: "basically the more the better" (p. 169); Blommaert (2007) asserted that if we wish to understand how language functions in complex social settings, "it's about using all there is to use" (p. 687); and in the same vein, Hyland (2016b) concluded his overview of approaches to writing research noting that combining a range of methods is inevitable when dealing with the complexity of writing. In fact, methodological eclecticism could be a guiding principle for social scientific research more generally. As Bourdieu noted, as long as our selected methodology is consonant with the question we ask, "social research is something much too serious and much too difficult… to deprive ourselves of this or that

resource available in the full panoply of traditions of our discipline-and of the sister disciplines of anthropology, economics, history, etc." (Wacquant, 1989, p. 54). Despite these calls, some decades old, a lot of EAP research remains narrowly focused on areas that have been over-explored, methodologically circumscribed, and inward looking (Swales, 2019).

Therefore, the overarching objective in this study is to contribute to unearthing how academic discourse functions by employing more of what is out there, more of what is available beyond the confines of EAP approaches. I do not only answer the call of Bhatia (2002) for a broader view of academic discourse, from common core perceptions of academic literacy to more discipline-specific writing, but I also advocate an even broader view that encompasses how sub-discipline-specific language use reflects the underlying knowledge practices, and shows what counts as a legitimate sub-disciplinary contribution, namely, what Hyland (2006b) calls the nature of [sub-]disciplinary knowledge itself. I thus attempt to bridge the gap pointed out by some EAP authors by combining insights from research into academic discourse in the various EAP traditions with an approach from the sociology of knowledge, Legitimation Code Theory (LCT), which allows exploration of various facets of the same practices, thus aiming to reveal the organising principles behind knowledge building.

2.6 Summary

In this chapter, I have reviewed some of the theoretical concepts and empirical research related to studying disciplinary variation in written academic discourse. In the first three Sections (2.1 to 2.3), I have reviewed important historical and theoretical developments pertaining to EAP/ESP, which reveal the complexities, controversies, and

the changing nature of academic literacy practices. First, I have shown how the study of academic literacy has become a field in its own right, and how the field has gradually zoomed in on practices in specific academic disciplines, recognising the limits of generalising practices across academia. Second, I have sketched how academic literacy practices, as exemplified by the most prominent genre, the RA, have always been shaped by creative responses to situational needs–be they persuasion, competition, or disciplinary beliefs–rather than resulting from blind adherence to rigid rules. Third, I have described how a peer discourse community exerts social pressure on academic writers, and represents the ultimate influence on the shape that disciplinary discourse takes.

In Section 2.4, I have concentrated on four selected linguistic features most relevant to this study: topical Themes, citation practices, hedging and boosting, and exclusive first-person pronouns. All these features have been staples of EAP/ESP research, producing a wealth of disciplinary-specific findings. However, the range of disciplines is simply too vast for EAP/ESP to have encompassed all of them to date. Such is the case of earth sciences, a discipline that seems to have been largely bypassed, despite its unique features originating from both the natural sciences and the humanities. More importantly, I have pointed out that the limitations inherent in current EAP/ESP approaches do not allow us to see how the selected linguistic features reflect larger organising principles of knowledge building.

Finally, in Section 2.5, I have acknowledged previous calls from within the EAP/ESP community to broaden the purely linguistic methods of enquiry, and to push the limits of our understanding of academic writing by including perspectives from other fields. In this study, I will answer the calls by combining approaches from various

EAP traditions with Legitimation Code Theory (LCT), an approach which aligns itself with a social realist perspective on knowledge.

Chapter III: Theoretical framework

In this chapter, I shall elucidate the theoretical framework, Legitimation Code Theory (LCT), underpinning this study. The chapter will start with a summary of some of the academic history and the broad concerns of the sociology of education/sociology of knowledge, to which LCT claims allegiance. Next, space will be devoted to the principal influences on LCT, before LCT's five dimensions and some critiques of LCT are described. Finally, I will outline an established collaboration between LCT and SFL, and review several studies utilising LCT, including studies that have combined LCT and EAP.

3.1 Overview: Towards a social realist perspective on knowledge

This section will summarise some of the developments in the sociology of education that led to the social realist position, to which LCT claims allegiance.

While not attempting to define the sociology of education, Young (1971, 2020) states that the obvious focus of the field is knowledge–what it is, what is considered educational knowledge, and how it is transmitted to students[21]; that is, problems which have, barring several exceptions, "been largely neglected" (Young, 2020, p. 11). The sociology of education can, on the one hand, be thought of as a separate specialisation from the sociology of knowledge, for example on account of the latter focusing on the production of knowledge through research and scholarship in academic fields (the sociology of science, the sociology of history, etc.) and the former focusing on knowledge recontextualisation and transmission in schools (Young, 2020). On the other

[21] For Bernstein (1971), educational knowledge can be dissected into curriculum (determines what is selected as valid knowledge—most often knowledge unrelated to everyday knowledge of the pupils and their communities), pedagogy (determines its valid transmission), and evaluation (determines its valid realisation/reproduction by students).

hand, the two areas have also been interconnected in works on how knowledge is socially organised in educational contexts (e.g., Young' (1971) *Knowledge and control* or Bourdieu and Passeron's (1977) *Reproduction in education, society and culture*), implying that the "sociology of education is no longer conceived as the area of enquiry distinct from the sociology of knowledge" (Young, 1971, p. 3).

Suitable starting points of exploration into knowledge in education are, Young (2008) suggests, the relationship of humans to the world and the symbols humans use to understand the world. "It is the the symbolic nature of our relationship to the world that is the basis of our knowledge, and it is that relationship that distinguishes us as human beings" (Young, 2008, p. xvi). Such an investigation of symbols and relations to the world was undertaken by Durkheim, one of the "founding fathers of sociology" (Jobling, 1973, p. 203) whose work significantly contributed to establishing the sociology of education as a discipline (Young, 2020) in his anthropological/sociological study *The elementary forms of religious life* (Durkheim, 1995). Investigating "the simplest and most primitive religion that is known at present" (p. 1), which he found mainly at Australian tribes (though examples from totemic beliefs elsewhere were also provided), Durkheim aimed to explain the human reality (for him, nearly all social institutions arose out of religion: "the idea of society is the soul of religion" (p. 421)). He proposed a distinction, which he argued applied to all societies and all religions, between the sacred and the profane. These two terms denote "a classification of the real or ideal things that men conceive of" (p. 34); that is, the sacred and the profane delineate fundamental categories of thought–boundaries between "two worlds with nothing in common" (p. 36), incompatible, even "hostile and jealous rivals" (p. 37). The sacred domain represents things ideal and transcendental, related to religion; the profane

domain represents the everyday material world. Importantly, objects of religious thought (nature, society, human beings) were the same objects that later scientists reflected on. Indeed, for Durkheim, "the essential notions of scientific logic are of religious origin" (p. 431), with both religion and later science "attempt[ing] to connect things to one another, establish internal relationships between those things, classify them, and systematise them" (p. 431). The classification and systematisation of thought was made possible by the use of concepts–relatively unchangeable (unlike one's perceptions or sensations), collective (rather than personal) representations of categories and classes (rather than particular objects), as it is generalised experience, together with lasting trends, that impacts societies' existence.

Durkheim's boundary between categories of thought (everyday vs. religious/ scientific) can also be traced in Vygotsky's (1986) differentiation between two conceptual systems: spontaneous and scientific. Spontaneous concepts are "situational, empirical, and practical" (p. 194), and develop from experience in everyday life, but "relations of generality" (p.197) between/among spontaneous concepts are not at first strongly developed in a child's mind. Scientific concepts, not "reflecting mere appearances of objects, as empirical concepts do" (p. 173), are beyond a child's everyday experience. The two types of concepts were seen as interdependent, since the learning in the classroom of the relationship between a scientific concept and the object it denoted, Vygotsky argued, was mediated by the existence and readiness in children's minds of everyday concepts (e.g., the concept of history can be facilitated by a child's understanding of the difference between past and now, here and elsewhere). Thus the very notion of scientific concepts implied a system of positions and relationships with other concepts; the acquisition of a scientific concept necessitated acquisition of "its

system of relations that determine its measure of generality" (p. 173). This systematisation that children were exposed to was in turn transferred to spontaneous concepts, which, while obviously not completely unrelated–such a state would render impossible "intellectual operation requiring coordination of thoughts" (p. 197)–were "cardinal[ly]" (p. 205) differentiated from scientific concepts by the absence of a system (i.e., undeveloped relations of generality).

The two systems of knowledge structures, namely everyday and academic knowledge, also appeared in the work of Bernstein (1999) as horizontal and vertical discourses. Briefly (an expanded description is provided below in Section 3.2.2), horizontal discourse is everyday/common-sense knowledge that all people can potentially acquire as it stems from "common problems of living and dying" (p. 159). In comparison, vertical discourse is either systematically and hierarchically structured knowledge, such as in the sciences (hierarchical knowledge structures), or it is knowledge in the form of specialised languages, specialised modes of enquiry and specialised criteria for text production and dissemination, such as in the social sciences and the humanities (horizontal knowledge structures). Access to and transmission of vertical discourse is regulated via education.

Bernstein's hierarchical and horizontal knowledge structures were further taken up by Maton (e.g., 2014a), who subsumed them into LCT's dimension of Specialization (more detail provided in Section 3.3.1) when seeking for a way to analyse within a single framework concepts related to intellectual fields of production (the above knowledge structures) and concepts related to recontextualisation and reproduction of knowledge (classification and framing, described in Section 3.2.2).

Systems of knowledge, and more broadly systems of thought, with their relationship to education were also topics in the works of Pierre Bourdieu. For example, in *Systems of education and systems of thought* (1971), Bourdieu argued that our mental processes may be governed by "unconscious master plans" (p. 194) acquired within the standing school system. These unconscious patterns can be thought of as culture, or "a programme of perception, thought and action" (p. 199), which, during differing periods in history, determines how reality is perceived, what theme/topic is thinkable or of legitimate concern, and "the particular form in which it is thought" (p. 195). In societies where culture is dominantly passed on during schooling, it is the role of academic training to "produce individuals equipped" with "a common individual unconscious" (p. 194). To fulfil its function of passing on culture, scholastic training has to be methodically organised: It classifies (e.g., by various -isms), establishes hierarchies, and selects works and ideas worthy of attention and preservation. In other words, the education system makes certain aspects of thinking and of reality more prominent, thus mapping boundaries of an area that can be travelled or explored, "with compulsory turnings and one-way streets" (p. 196), usually in a particular sequence that, in Bourdieu's view, is equally "alien to considerations of logic and teaching" (p. 196) as following an alphabetically ordered syllabus. Academic training then teaches how to differentiate these ideas, making sure that culture appears as coherent, differentiated, and determined.

For what purpose, we may ask, does the education system function in such a way? In Bourdieu's view, the acquisition via education of unconscious patterns of viewing the world (i.e., acquisition of a specific type of reality) legitimates and perpetuates distinction (meaning separation) based on social class and distinction

(meaning preeminence) of those whose "system of categories of perception, language, thought and appreciation" acquired via scholastic training "sets them apart from those whose only training has been through their work and their social contacts with people of their own kind" (p. 200).[22] To establish and sanction distinction, the education system became segregated, with higher levels of education and education in the classics reserved mostly for economically and culturally favoured classes. For Bourdieu, this duality of education created duality (or rift) in culture (academic vs. popular). However, only academic culture was objectified by being organised around the works that embodied it; popular culture was "deprived of" (p. 201) objectification, since defining popular culture on the same principle as academic culture (i.e., around patterns of thought, language, and works that embody the reality and experience of people excluded from academic culture) was foreign to the working classes in intention and expression.

Questions and issues raised by Bourdieu regarding differing outcomes of education for social classes had also previously appeared in demographic and sociological studies of the 1950s in the UK, which, for example, revealed how important a determinant social class was for student selection into grammar schools despite the professed objectivity of standardised and IQ tests (Young, 2020). In fact, according to Young (2020), "it was the earlier revelations of the social class basis of educational inequality that established the sociology of education as a distinct field" (p. 12). The revelations and culturally-based distinction of what counted as mere experience versus knowledge also inspired a particular direction, starting in the early

[22] Similarly, Bernstein (1971) noted that educational knowledge, through strong control (strong framing) of content selection, organisation, and pacing of delivery, "gives special significance to those who possess it" (p. 58), as compared to the mundane knowledge derived from everyday experience. This differentiation in the type of knowledge possessed is further maintained by the fact that mundane knowledge is allowed into the curriculum only for lower-achieving pupils.

1970s, in the sociology of education which became known as the New Sociology of Education (NSOE), developed from Young's (1971) book *Knowledge and control*. This sociological approach to knowledge challenged the knowledge base of the dominant liberal curriculum (Young, 2008). Seeing that "how a society selects, classifies, distributes, transmits and evaluates the educational knowledge it considers to be public, reflects both distribution of power and the principles of social control" (Bernstein, 1971, p. 47), proponents of the NSOE interpreted the continued lower academic achievement or failure of working-class students as a result of academic curricula having been historically constructed to preserve the class division of society (Young, 2008).

The main ideas, as well as problems not recognised at the time, in *Knowledge and control* and of the NSOE were aptly summarised by Young (2010). The main premise of the book was that "all knowledge is no more or less than someone's or some group's experience of making sense of the world" (p. 10); that is, all knowledge was socially constructed. When experience belonged to a privileged group such as experts or professionals, it was called knowledge, which conferred power and prestige on that experience. And if knowledge was what was transmitted through education, the focus of the sociology of education had to be what knowledge actually was and how it was transmitted; in other words, what getting education actually meant had to be investigated simultaneously with the more frequently asked question at that time of who was getting education.

Four consequences followed. First, knowledge (or curriculum) was a political instrument since it was constructed through inclusions and exclusions determined by distribution of power in society, and served to preserve the social status quo. Second, given the link between internal knowledge structure and distribution of social privilege/

social power, knowledge was necessarily relative, because there was no objective measure to distinguish whether some knowledge afforded more insights, was more reliable, or brought us closer to truth. The focus was then on the "knowledge of the powerful", the question of power in society, rather than on "powerful knowledge" (Young, 2010, p. 11). Third, by distinguishing educational/curricular knowledge from everyday experience, some social groups legitimated their views on knowledge and experience, thus masking power relations. This premise was consonant with the concept of symbolic violence in education in Bourdieu and Passeron (1977):

> PA [pedagogic action] is, objectively, symbolic violence first insofar as the power relations between the groups or classes making up a social formation are the basis of the arbitrary power which is the precondition for the establishment of a relation of pedagogic communication, i.e. for the imposition and inculcation of a cultural arbitrary by an arbitrary mode of imposition and inculcation (education). (p. 6)

Four, in a political struggle about the nature of education, the stance that a difference between knowledge and experience was ideological could serve as a foundation for criticism of any curricularised or institutionalised knowledge. Indeed, it led in some radical instances to uncritical privileging of experience of the dominated groups (however they were defined) or calls for abolishing schools.

The thesis that knowledge was purely a reflection of power relations in society as laid out in the 1971 book missed three problems (Young, 2010). First, it missed a contradiction that the thesis itself could be rejected based on its own rationale: "Why should anyone accept it when it claims to be the basis for rejecting the idea that some knowledge is more reliable than others" (p. 12). Second, it missed the fact that some

types of knowledge are more powerful not because of their social origin. Third, it missed that its criticism did not offer either an alternative to the curriculum it rejected or any grounds (save for grounds of social power) for including/excluding what a curriculum should contain.

Of relevance here may also be Bernstein's view of the NSOE, since it was Bernstein, after all, who, as Young's tutor and later senior colleague, encouraged him to write a dissertation on curriculum, which later became the basis of the first chapter of *Knowledge and control* and to edit the book (Young, 2008; 2010). Despite Bernstein's earlier support for some of the issues raised by the NSOE (his paper on classification and framing appeared in *Knowledge and control*) and his shared concern about unequal distribution of education opportunities, he critiqued the endeavour, along with theories of cultural reproduction, stating the programme "whatever else it produced, did not produce what it called for" (2003b, p. 144). The chief reason for this failure according to Bernstein (2003b) was excessive attention to what was reproduced in education and by whom, as if education (pedagogic discourse) itself was just a neutral medium for transmission of external power relations. This focus then left unanalysed the medium for reproduction, the nature of educational discourse. Bernstein's critique was, in hindsight, acknowledged by Young (2010, 2020), who described his earlier work both as emancipatory in questioning what knowledge was and the role it played in education, leading to his further intellectual development (Young, 2008), and as idealistic, assuming that "criticising a social phenomenon would lead to its inevitable collapse" (Young, 2020, p. 13). Instead, the idealism ended when confronted at the end of the 1970s and the beginning of the 1980s with neo-liberal economic policies, eventually fragmenting the sociology of education (Young, 2020).

Young's earlier position of knowledge of the powerful, which can be described as social constructivism, was later transformed to a position of powerful knowledge (e.g., his 2008 book), which can be described as social realism. Social realism is a relatively new school of thought within the sociology of education. It is also a school with which LCT, the framework underpinning the methodology of this thesis, aligns itself intellectually. Social realism can be viewed as "a coalition of minds" rather than a doctrine-based school of thought (Maton, 2014a, p. 9). According to Van Krieken et al. (2014), social realism posits that firstly, education is a distinct social field of practice, differentiated from other fields by producing, curricularising, teaching and learning knowledge; and secondly, social realism rejects as reductionist the assumption that knowledge is merely a reflection of social power. Social realism acknowledges that knowledge is created and shaped by struggles among social groups. In fact, Young (2008) emphasised repeatedly (e.g., pp. xviii, 60, 220) Durkheim's idea that the social character of knowledge was a precondition for objectivity. At the same time, some knowledge claims and knowledge forms hold more power than others. In other words, some knowledge "transcend[s] the historical conditions of its production (e.g. Euclid's geometry and Newton's physics)" (Young, 2008, p. 28). As Maton (2014a) adds, social realists assert that knowledge is both social and real; that is, it possesses properties and tendencies that affect the real world, for example, social inclusion and educational achievement.

Social realists maintain that a sound theory of knowledge (what knowledge is) is absent not only from discussions about the implications of the so-called knowledge society for education (e.g., Young, 2008), but also that the lack of a sound theory of knowledge has long rendered education research blind to the internal structure of

knowledge (e.g., Maton, 2014a). Knowledge has commonly been reduced to "knowing and nothing but, or to power and nothing but" (Maton, 2014a, p. 7). Social realists decry the exclusive focus of *relations to knowledge,* and reject as false the dichotomic choice between knowledge being only either objective and value free (knowledge essentialism) or knowledge being only socially constructed and thus reflecting power in society (knowledge relativism) (e.g., Young (2008), who refers to it, quoting others, as the epistemological dilemma).[23] Meanwhile, one of the key factors of education, the forms taken by knowledge (or *relations within knowledge*), has remained obscure (Maton, 2014a; Van Krieken et al., 2014). Social realism then brings the forms of knowledge into focus (Van Krieken et al., 2014).

LCT's originator and principal researcher, Karl Maton (e.g., 2016a), characterises LCT as allowing us to look beyond appearances of practices in social fields since the theory provides us with the tools to describe the organising principles underlying rules, ways of doing, and the bases of achievement. This character, Maton states, offers LCT the advantage of avoiding many false dichotomies of education research: free-floating theoretical concepts unable to explain real-life data vs. empirical findings bogged down in the particulars of their contexts so much that they cannot be generalised; qualitative vs. quantitative approaches; humanism vs. science, etc.

LCT concepts are meant to be enacted both in diverse empirical research (see Section 3.3 for a review of relevant studies into academic disciplines) and in generating or transforming practice (e.g., Macnaught, Maton, Martin, & Matruglio, 2013; Martin, 2013; Maton, 2014b). Maton (2016a) emphasises that LCT's universality has enabled those who have used its conceptual toolkit to investigate:

---

[23] These two stances could be linked to positivism and social constructionism (Section 2.3.2).

a) diverse practices (e.g., research, teaching, and beliefs),

b) diverse disciplines (e.g., physics, ballet, and jazz),

c) diverse institutional levels (e.g., secondary schools, vocational colleges, and universities),

d) diverse levels of analysis (e.g., disciplines, single texts, and individual words),

e) diverse national settings (e.g., Asian, European, and Scandinavian), and

f) its synergy with other approaches and methods (e.g., systemic functional linguistics, qualitative interviews, and textual analysis).

Maton (2016a) also posits that LCT does not only describe the inner workings of social fields and their practices, but also "promote[s] social justice" (p. 3). He asserts that by describing the oftentimes tacit organising principles underlying practice and the bases of achievement in social fields, LCT ensures that the rules of the game are explicitly available to everyone, not only to select actors in the field.

3.2 Formative influences on Legitimation Code Theory

While LCT is "allied" (Maton, 2014a, p. 8) to social realism, according to Maton (2016a), LCT also draws on concepts and ideas from a plethora of disciplines, such as philosophy, linguistics, physics, sociology, and cultural studies. For example, in very general terms, some of the influences on LCT were Bourdieu's field theory, which showed how relations between structures in society and practices were mediated by effects of various fields of practice; Bhaskar's critical realism and Popper's critical rationalism, which underscored that explanatory power varied with types of knowledge claims (some being considered more trustworthy and/or closer to the truth); or

Bernstein's code theory, which brought to the fore the impact of knowledge structures themselves on explanatory power and positioning within fields of practice (Maton, 2014a). These approaches highlighted that achievement, status, and identity were shaped by both epistemic and social forms of power. LCT aims to integrate the above ideas around the concept of legitimation (Maton, 2014a). The two most profound influences on LCT were the sociological theories and concepts of Pierre Bourdieu and, "above all", Basil Bernstein (Maton & Chen, 2020, p. 37)[24]. The extensive oeuvre of the two scholars inspired research traditions active decades after their deaths. The concepts introduced by them were drawn on in educational and linguistic research, and would constitute potential alternative approaches to LCT in researching the main question investigated in this thesis.

3.2.1 Bourdieu's thinking tools

Pierre Bourdieu was a French sociologist interested in the power dynamics in society, in how power structures are preserved, and in the relationship between social class, education, taste, and culture. To illustrate how these concerns are interrelated, we can consider Bourdieu and Passeron's (1979) essay *The Inheritors*. On the role of higher education qualifications and the preservation of social power in favour of those already dominating in the social hierarchy, Bourdieu and Passeron showed, using statistics, the extent to which working-class students were underrepresented in higher education. The authors also argued that the underrepresentation and often lower educational achievement of working-class students reflected (apart from the obvious gap in financial

[24] Speculatively, the influence of Bernstein and Bourdieu on the creation of LCT might have reached beyond familiarity with their work, since Maton (2005b) acknowledged in his PhD thesis, where he proposed the LCT framework, personal (likely repeated) contact with Bernstein and discussion of ideas underpinning his thesis with Bourdieu, among others. The LCT website (www.legitimationcodetheory.com) also includes a mention that Maton learned from Bernstein and Bourdieu during his doctoral studies.

resources) the differences in cultural models and values between the working class and educational institutions. For instance, to succeed in higher education, in some study programmes, students would need to possess values and models acquired from "extracurricular culture" (p. 17), such as from visits to museums, galleries, concerts, or theatres, and even scientific culture requires "first-degree experiences" (p. 21), such as books found in home libraries, family holidays spent doing culture-focused visits, and being used to particular mode of communication. These values and models are expected and presupposed by educational institutions, yet reflect habits and tastes of the privileged classes. Thus, the main function of higher education institutions is to perpetuate social privilege rather than transmit specialist knowledge.

Bourdieu's arguably most influential concepts are three interrelated notions of habitus, capital, and field, which are at the heart of his field theory. In the following paragraphs, I will briefly characterise each notion in turn.

Habitus is a system of generative schemes (Bourdieu, 1984). It is created by social processes and can be described as "not only a structuring structure" but also "a structured structure" (p. 170). Habitus thus represents schemes that both organise social practices and influence our perceptions of such practices on the one hand, and that reflect how our perceptions of the social world are shaped by the "internalization of the division into social classes" (p. 170) on the other hand. A different habitus arises from different conditions of existence, and the distinguishing features are given not only internally, but also by relations with other conditions in the social system (Bourdieu, 1984). Habitus can be applied to disciplines and discourse communities, since Bourdieu (1990a) points out that those agents who occupy similar social positions experience

similar conditionings, are very likely to have similar interests and tendencies, and thus produce similar practices.

The second notion, that of capital, has been elucidated in some detail in Bourdieu's (1986) essay. In this work, capital is seen, in a somewhat circular or cyclical fashion, as accumulated labour (when materialised) which enables those who possess it to appropriate social energy in the form of materialised or living labour. The structure and distribution of capital in all its various forms mirrors and determines the structure of society. But capital, far from being purely a fiscal matter, known as economic capital (convertible to money or property rights), exists also in the forms of cultural capital (e.g., educational qualifications), and social capital (e.g., social connections).

Cultural capital consists of three sub-types:

a) The embodied state, which is linked to an individual's body (i.e., various forms of culture or cultivation of body and/or mind), has to be acquired first hand, presupposes an investment of time, is incorporated into the individual's habitus, declines with and is limited by the individual's biological and cognitive capacity (and also by the time the individual can devote to acquiring it when free from economic obligations), and cannot be immediately transferred. In the context of disciplinary communities, the embodied state refers to, for example, gaining knowledge of the tacit norms and values of discourse community writing. With this knowledge, disciplinary members can, for example, recognise and competently produce what Swales (2004) calls genre hierarchies (e.g., the relative importance of an RA as compared to a monograph) and genre chains (e.g. a cover letter, an RA manuscript, a response to reviewers' comments). For

young researchers, acquiring the embodied state represents learning to participate in the disciplinary community (Casanave, 2008; Kwan, 2009).

b) The objectified state, which can be represented by cultural artefacts (films, paintings, books, machines, etc.), can be transferred immediately in its material form if an individual has the necessary economic capital, but its consumption requires embodied state of cultural capital. As Bourdieu puts it, an individual can buy a machine or a painting, but needs access to embodied capital (either personally possessed or possessed by others) to use any of the objectified capital in accordance with its intended purpose. Capital in its objectivised form is materially or symbolically effective (or recognised as valuable) only if it is possessed, employed as a weapon in the struggle in the social field of practice, and it accrues benefits to its possessor. In the context of a disciplinary discourse community, the objectified state refers to, for example, production of and access to knowledge in the forms of scientific journals and published RAs. Such access or lack of it can privilege centre scholars and seriously disadvantage those on the periphery (Canagarajah, 2002).

c) The institutionalised state is represented most clearly by academic qualifications. Such qualifications are seen as objectified certificates of cultural competence, because they are officially guaranteed as independent of their bearer, as opposed to the educational level of a self-educated individual, whose value of capital may be questioned at any time. The institutional recognition also allows for qualifications and their holders to be ranked, evaluated, and to establish their economic worth on the labour market, for instance, based on the principle of scarcity. In a disciplinary discourse community, the institutionalised

state refers to, for example, the number of invited plenary talks, the impact factor of a journal one has published in, and, as Hyland (2007) notes, the number of publications on a CV that thus attests to an individual's reputation and productivity.

Social capital is linked to group membership, that is, an individual's social network and social relationships, which support the individual with the weight of the collective capital. To be more exact, the amount of an individual's social capital is determined by the number of connections that could be called on in the social network, and the value of the capital each connection possesses. The existence of group membership and valuable relationships, once established, does not automatically continue but has to be continuously produced and reproduced by investment of time and effort, both on the part of the individual and the group.

I believe that one aspect of social capital mentioned in Bourdieu's (1986) essay is especially pertinent to disciplinarity: the limits of a group. Each group shares a symbolic constitution which defines and limits what kind of exchange (e.g., actions or words) is recognised as constitutive of the group. Each group member then is a custodian of the legitimate exchanges, and each group regulates the right of anyone to freely declare themselves group members, because every new member may potentially redefine the limits of exchange and therefore the group's identity. Once a new member is admitted, s/he must expend effort to acquire and maintain competencies that enable the group to perpetuate recognised exchanges and reproduce its social capital. In the case of written communication, the regulation of access and a stamp of approval are represented in the process of an RA publication (Canagarajah, 2002).

The last, but not the least, of the three concepts is that of field. Bourdieu (1993) describes fields as "structured spaces of positions (or posts) whose properties depend on their position within these spaces and which can be analysed independently of the characteristics of their occupants" (p. 72). Fields (e.g., academic disciplines or institutions) differ from one another in their specific interests and stakes which are perceived and coveted only by the field's occupants. Fields run on the principle of competition, whereby field occupants/agents recognise the specific stakes and interests, and are prepared to "play the game" (p. 72).

The notion of competition (or struggle) means that the structure of the field reflects "a state of the power relations among agents or institutions engaged in the struggle" (p. 73). In his essay on capital, Bourdieu (1986) refers to this state as capital distribution reflecting the social structure. Thus, the structure itself becomes one of the stakes in the game and determines the choice of strategies: Agents who control the power relations and thus the legitimate characteristics of the field tend to employ conservation strategies ("orthodoxy") to maintain what constitutes legitimate contribution and hence valued capital, while those with smaller amounts of field-specific (legitimate) capital (newcomers, younger agents) tend towards subversion strategies ("heresy") (Bourdieu, 1993). However, for Bourdieu the struggle does not mean that less powerful agents are attempting to destroy the field and invalidate or terminate the game. The fact that all field agents are engaged in the game presupposes a doxa (another key term of Bourdieu's work)–universal unspoken acknowledgement and acceptance of a fact of life, in this case the existence of the field, of the game, and stakes and interests that are worth struggling for (Bourdieu, 1993). All agents who

participate in the struggle "help to reproduce the game by helping to produce belief in the value of the stakes" (p. 74).

Bourdieu's discipline-specific competition can be exemplified on a case from earth sciences, where it has recently resulted in changes in communicative practices. Every year, earth scientists from around the globe meet at the biggest disciplinary event: the American Geophysical Union (AGU) Meeting in San Francisco. From my conversations with earth scientists, I know that when submitting abstracts to the AGU, they are not worried about being accepted–everyone who has the budget for it will go. However, as the numbers of attending scientists have swelled to well over 20,000 in recent years, most attendees will be allocated "only" a poster presentation. Being selected to give a talk is much more prestigious, but much less likely, especially for junior discourse community members (oral presenters are frequently selected based on their renown in the field). In order to enable more researchers to communicate their findings, two additional alternate formats were introduced in 2016: a panel discussion (a five-minute presentation followed by a question and answer session) and a lightning talk (a short oral presentation of a poster in a forum). The alternate formats may thus change the hierarchy of disciplinary oral genres.

Bourdieu himself saw his notions of habitus, capital, and field not so much as theoretical for the sake of theorising, but rather as "a set of thinking tools to be visible through [empirical] results they yield" (Wacquant, 1989, p. 50). He clearly looked down on freely-floating theories that perpetuated vague metadiscourse and could never be a substitute for empirical research (Wacquant, 1989). Indeed, Bourdieu's thinking tools have been the theoretical foundations of a plethora of empirical studies, both qualitative and quantitative, in the contexts of language and linguistics, education, cultural studies,

welfare systems, online video gaming, water management, sociology of culture, etc., some of which can be found in various edited volumes (e.g., Albright et al., 2018; Grenfell, 2011; Grenfell & Lebaron, 2014; Hilgers & Mangez, 2014). The volume edited by Albright et al. also aimed to "attend to what one can and cannot do with using Bourdieu's field analysis" (p. ix), and provided space to researchers who complemented/developed Bourdieu's concepts with critical discourse analysis, post-Foucauldian governmentality studies, or LCT, the latter contributed by Maton. Maton's chapter starts with a summary of Bourdieu's pursued aim in social science: to achieve a new gaze onto the world (metanoia), which will allow moving past (breaking/rupturing from) immediate sensual perceptions of objects, which provide only illusionary vision, towards seeing underlying relational principles. "Bourdieu emphasised *a relational mode of thinking* that conceives phenomena as realisations of generative principles that are relationally defined" (p. 251, original emphasis). Maton characterised Bourdieu's relational as having a horizontal and a vertical dimension. The horizontal relations of concepts prevent understanding practices based solely either on actors' dispositions or their social contexts, which Maton interprets as being illustrated by Bourdieu's well known formula from *Distinction* (1984, p. 101): "[(habitus) (capital)] + field = practice", meaning that practices in fields are determined by actors' habitus (dispositions) and the amount of capital they have, which translates into their positions in a field. The verticality of relations is given by the fact that an agent's habitus, capital, position in a field, and practices are only one option of many other possible realisations thereof. One possible realisation's underlying characteristics are derived from relations to other possible realisations.

Maton's horizontal and vertical characteristics of Bourdieu's concepts are intuitively appealing, but their use does not accord well with my, admittedly limited, reading of Bourdieu's extensive oeuvre. For example, the famous formula as well as mentions of horizontality and verticality appear in *Distinction* in the chapter *The social space and its transformations*. Bourdieu does construct "a three-dimensional space" constituted by "volume of capital, composition of capital, and change in these two properties over time" (p. 114). The change ("reconversion", p. 125) in "the volume of the type of capital already dominant in the asset structure" (p. 132) enables "vertical movements, upwards or downwards, in the same vertical sector, that is, in the same field" (p. 131); the change of "one type of capital into another or of one sub-type into another sub-type" enables "transverse movements, from one field to another" (p. 131), either horizontally or between levels (higher or lower in another field). Horizontal and vertical thus represent axes of types and amounts of capital in social space. I cautiously speculate that Maton's horizontal and vertical characteristics better prepare the reader for the use of the spatial dimensions in LCT, which are described later in Maton's chapter.

Maton then states that "Bourdieu's concepts do not fully embody the relational thinking he called for" (p. 252). To support his assertion, Maton notes that, for example, while habitus is defined as a structured structure as well as a structuring structure, it is not clear how to conceptualise the structure in terms of its underlying principles as a possible structure among other structures (the systems of settings of underlying principles that differentiate the one from many possible realisations). Without conceptualising these structures, similarities/differences/changes over time or bases of match/mismatch (e.g., in habitus) are difficult to systematically illustrate. The same

issues apparently apply to capital, field, position. Bourdieu, Maton says, "began to reveal" (p. 253) the underlying properties through dichotomous types[25], but the binaries, while pointing to what should be relationally conceptualised, do not afford conceptualisation of, for instance, degrees of difference or general properties.

Curiously, what Maton sees as not possible with Bourdieu's thinking tools is reminiscent of Bernstein's critique of Bourdieu and Passeron (as well as of reproduction theories in general) regarding education and pedagogical communication. Bernstein is not mentioned in Maton's chapter, though it appears likely that some cues may have been taken from Bernstein (e.g., 2003b), who, for example, pointed out about habitus that it

> "is essentially a cultural grammar specialized by class position and fields of practice. It is by no means clear what are the rules of these class-specialized grammars and fields of practice, nor is it clear how the specialized grammars are constructed and relayed in the process of transmission. But these are not the special objects of Bourdieu's project" (pp. 2-3).

Further, Bernstein (2003b) noted that a theory of cultural reproduction "has to be able, in the same theory, to translate micro into macro, and macro into micro, with the same concepts" (p. 147); that "theories of cultural reproduction ought to have within them very strong rules which enable the theorist or the researcher to say, 'This is the same,' 'This is a variation,' 'This is a change.'" (p. 147); more specifically about pedagogic communication, "there should be a theory of communication from which one can derive the distorted and the so-called undistorted" (p. 148). Without being able to generate "specific principles of description of the agencies of their concern" (p. 148),

[25] The argument of dichotomous types and the need for conceptual extension is reprised in the case of Bernstein (Section 3.3.1).

which, from Bernstein's perspective, Bourdieu and Passeron's concepts (or concepts from reproduction theories) could not do, "the concepts are diagnostic… their concepts specify what is to be described, they call for a description, but are unable to provide principles for that description" (p. 149).

Maton (2018) then proposes to build on or "complement" (p. 259) Bourdieu's framework with LCT rather than be in competition with it. He summarises a study that drew on the dimension of Specialization to describe how the principles underlying habituses of Chinese students at an Australian university clashed with dispositions (and their underlying structures) valued by their teachers, resulting in negative experience due to continued use of mismatched rules of achievement. Based on the study, three relational attributes of LCT are postulated. Firstly, LCT is able to conceptualise the underlying structures/principles of practices, positions, and dispositions, etc. Secondly, the topological property of LCT enables each possible instance of practice/position/ disposition to be characterised in relation to all other possible realisations of practice/ position/disposition. All possible realisations can be charted across the Cartesian plane as a scattergraph to reveal dominant types, range of diversity, or changes among them (e.g., over time). Third, the same LCT concepts can be enacted to study varied phenomena in diverse social fields.

Maton's (2018) acknowledgement of some inspiration for LCT by Bourdieu's thinking tools is one example of the proliferation of research projects that have drawn on Bourdieu, together serving as a testament to the popularity, influence, and insights offered by field theory (as mentioned above). This influence has also inevitably been accompanied by misinterpretations and misapplications (Grenfell, 2010) as well as adaptations and intended extensions (Grenfell, 2018). Interestingly, those for whom

Bourdieu failed/avoided/ignored/sidestepped/overlooked certain issues often resort to eight recurring strategies (Grenfell, 2010). Maton's (2018) account bears partial marks of strategy number two: Bourdieu is interpreted in a certain way, "pertinent" omissions are identified, and then a solution is proposed, frequently without contrasting the author's and Bourdieu's philosophical bases or cultural contexts. By itself, that does not mean that Maton's work should be discarded, as finding a gap in previous research is one of the most common raisons d'être for new studies. Moreover, Grenfell (2018, same volume as Maton (2018)) encourages continued work on Bourdieu's project as long as Bourdieu is clearly interpreted for what he did and why and what he did not do and why. As if in response, Maton (2018) writes that despite the power of Bourdieu's ideas, we also need to recognise the limits of Bourdieu's concepts and "augment" (p. 250) them so that Bourdieu's intention of relational thinking can be implemented. While that is in the spirit of Grenfell's call, one should be aware that "Bourdieu's ideas are themselves now part of an academic struggle, which inherently carries the *interests* of those expressing them" (Grenfell, 2010, p. 14, original emphasis).

3.2.2 Bernstein's pedagogic device and knowledge structures

The second significant influence on LCT was the work of Basil Bernstein. Bernstein was a British sociologist of education interested in the social and cultural explanations of inequality in educational achievement and in the nature of pedagogic discourse. His early research was focused on why working-class children fared worse in language-based school subjects (regardless of their native ability) than middle-class children. Bernstein (2003a) formulated the notions of public and formal language use by social classes, according to which working-class children were exposed to and used

mostly/virtually only the restricted code (highly context-dependent language, descriptive, with socially condensed meanings relying on shared knowledge, and consisting of shorter and simpler sentences with a narrower vocabulary range), while middle-class children were exposed to and used both the restricted and elaborated code (context independent, more analytical and abstract rather than purely descriptive, with explicit meanings, more grammatically complex, with richer vocabulary). This dual exposure and use prepared middle-class students better for schooling, where elaborated code was not only preferred but also expected. Another area that Bernstein (2000) was interested in was acquisition of knowledge, and he formulated the concepts of the recognition rule (being able to distinguish between contexts, and identify legitimate texts or practices for particular contexts) and the realisation rule (being able to produce appropriate texts or practices for particular contexts). Perhaps the best-known concept is that of the pedagogic device (Bernstein, 2000), that is, how pedagogic knowledge is established (recontextualised from other discourses), regulated, and legitimated. While the pedagogic device is concerned with knowledge, the focus is on the transformation of knowledge into pedagogic communication, and thus the pedagogic device represents "a theory of pedagogy, not knowledge" (Moore, 2013, p. 154), although Bernstein's work on knowledge structures had significant implications for the theory of knowledge (Moore, 2013).

Of Bernstein's theoretical framework, his conceptualisations of pedagogic codes (e.g., Bernstein, 1971, 2000), which refer to fields of recontextualisation and reproduction, and knowledge structures (Bernstein, 1999), which refer to (intellectual) fields of production, were especially influential for the development of LCT. Let us explore both concepts in turn.

Bernstein's (2000) pedagogic codes reflect the relationship between classification and framing. Classification is a term for the translation of power relations. Dominant power creates boundaries between specialised categories, in our case academic (sub-)disciplines. Within the boundaries, (sub-)disciplines develop their own discourses, identities, voices, and rules–since classification is either strong or weak (C+ / C–). Weakly classified/insulated (sub-)disciplines have less specialised discourses, identities, voices, and rules. The established boundaries then give rise to relations between and among (sub-)disciplines.

What is unique about a (sub-)discipline does not arise from what (sub-)disciplines develop within their space inside the boundaries, but rather from the relations between and among (sub-)disciplinary boundaries: Classification "refers to the degree of boundary maintenance between contents" (Bernstein, 1971, p. 49). Therefore, a (sub-)discipline is independent only to the extent to which it can delineate a boundary separating it from other (sub-)disciplines, a boundary that defines its social space and social division of labour. Breach of the boundary may endanger the identity of the (sub-)discipline, because the identity is the space separating individual (sub-)disciplines. Indeed, Bernstein (1971) referred to educational identity, in the sense of loyalty to being socialised into an academic subject, using Durkheim's term *sacred,* "the otherness of educational knowledge" (p. 56). An example of how an even perceived boundary breach may be viewed as dangerous for a (sub-)discipline's identity comes from Maton and Doran (2017a). In their paper on one of the LCT concepts, the authors explore English discourse from a sociological standpoint, and for that purpose define and work with a plethora of terms from linguistics. Yet, at the beginning of the paper, the authors make effort to clarify that their study does not attempt to formulate an LCT model of English

discourse, so that their paper is not perceived "by linguists as a hostile incursion onto their terrain and by sociologists as a failure to remain sufficiently sociological" (p. 53).

Framing represents the degree of control over the different forms of legitimate communication in pedagogic relations, for example, between parents and their children, or teachers and pupils; that is, "framing is about *who* controls *what*" (Bernstein, 2000, p. 12, original emphases). Framing is the internal control over the message that is transmitted and acquired in pedagogic contexts in terms of content selection, its sequencing, pacing, criteria, and "control over the social base which makes this transmission possible" (p. 13). Comparing classification and framing, Bernstein (2000) posits that classification limits what is legitimate discourse or voice, whereas framing controls how legitimate discourse is realised/communicated. Framing is either strong or weak (F+ / F–). Strong framing means that the transmitter (e.g., a teacher) tightly controls the selection of what is communicated, in what sequence it is communicated, at what pace it is communicated, the criteria and the social base allowing the communication. As a result of such framing, pedagogic practice is visible and explicit. Weak framing means that the acquirer (e.g., pupils) has more control over the communication; and consequently pedagogic practice is likely invisible. The strength of framing can vary with respect to individual elements; that is, framing can be strong over what is communicated and in what sequence, but weak over pacing.

Bernstein (1999) pointed out that previous work focused on elaborating pedagogic discourses, especially contrasting specialist and everyday knowledge, but the internal principles of the discourses received little attention and were taken for granted. To fill the gap, in his 1999 essay on knowledge structures, he developed two previously proposed broad categories of discourse: horizontal and vertical. Horizontal discourse

represents everyday and common-sense knowledge, which is "likely to be oral, local, context dependent and specific, tacit, multi-layered, and contradictory across but not within contexts" (p. 159). The knowledge transmitted in horizontal discourse is segmentally organised and represents a collection of strategies that can assist us in various everyday scenarios. The context dependence of such knowledge means that the strategies acquired for one context may be irrelevant for another (e.g., knowing how to tie our shoelaces does not help us choose the best shoes). At the same time, a context does not necessarily require only one correct strategy.

Each community has a reservoir of strategies, out of which each individual living in the community acquires a personal repertoire, comprising a common core from the community reservoir and specific strategies determined by his/her life contexts and situations. An individual's acquisition of horizontal knowledge is a "major cultural relay" (p. 160), accomplished tacitly through modelling or explicitly by showing or instructing in face-to-face interactions in the family, peer group or local community.

Vertical discourse represents academic knowledge or, in other words, "specialized symbolic structures of explicit knowledge" (Bernstein, 1999, p. 161). We can discern two forms of such knowledge: In the sciences, it is "a coherent, explicit, and systematically principled structure, hierarchically organised" (p. 159), while in the humanities and the social sciences, the knowledge is horizontal and "takes the form of a series of specialised languages with specialised modes of interrogation and specialised criteria for the production and circulation of texts" (p. 159). The knowledge transmitted via vertical discourse integrates meanings and is acquired through recontextualisation.

How do the two forms of vertical discourse develop? Hierarchical knowledge primarily integrates propositions in order to create more general propositions or more

abstract theories. This knowledge structure has been depicted by Bernstein (1999) as forming a triangle with a broadening base as it integrates and subsumes previous knowledge. Any disagreements between proponents of opposing propositions are preferably resolved by experimental procedures, and when not possible, propositions are either refuted or integrated into more general propositions. Horizontal knowledge, in comparison, consists of a series of specialised languages (e.g., the various -isms) with frequently contradictory criteria for what counts as legitimate knowledge, legitimate focus of inquiry, and legitimate contribution. Each language is represented by a group of its proponents whose symbolic capital is closely bound with the defence and propagation of the language. Opposing horizontal languages cannot be reconciled experimentally or by integration, so horizontal knowledge develops by introducing a new language, new legitimate questions, and new proponents.

Within vertical discourse, Bernstein (1999) distinguished horizontal knowledge structures with strong and weak grammars (i.e., knowledge structures with more or less explicit arrangement of technical concepts). Knowledge structures with strong grammars (e.g., economics, linguistics, and mathematics) are able to empirically describe, define, or model the phenomena they study comparatively precisely, though these knowledge structures (excluding mathematics) "often achieve their power by rigorous restrictions on the empirical phenomena they address" (p. 164). Knowledge structures with weak grammars (e.g., sociology, social anthropology, and cultural studies) are much less able to explicitly conceptualise the phenomena they focus on. In such knowledge structures, practitioners need to acquire a "gaze", either through explicit instruction or tacitly in social interactions with those who already possess it in order to discern what is regarded as legitimate knowledge.

As a closing note, one might add that Bernstein's concepts of knowledge discourses and structures are sometimes described as "more suggestive than explanatory" , "locked into an early (lexical) metaphorical stage of discussion" (Maton & Muller, 2007, p. 27). Nevertheless, it is only fair to point out that Bernstein (2000) conceded the limitation of dichotomous types in his conceptions, and by insisting that they are not "ideal types" (p. 123) acknowledged the existence of many forms his concepts could represent.

Bernstein's work was followed by Maton, who proposed to expand on it in several ways. For example, LCT aims to be applicable beyond pedagogic contexts. It also focuses on issues of struggle over status and legitimate practices. Then, while Bernstein's concepts apply to whole fields, LCT's ambition is for its concepts to apply to objects of study "at any level, from subject areas to individual pedagogic or textual practices" (Martin et al., 2020, p. 21). Also, Bernstein's strong/weak code modalities (C± / F±) offer binary types and four combinations of strengths. Bernstein (2000) himself noted of his concepts that "the organising principles underlying the early opposing dichotomous forms were limited in their generating power" owing to a very weak grammar, although "the powers of this grammar increased somewhat, by their replacement by the more general concepts of classification and framing" (p. 124). LCT seeks to further avoid binary types by placing its various code modalities on axes of Cartesian planes, thus arguing to encompass potentially infinite positions of relative strengths (i.e., topological characterisation).[26] In addition, Cartesian planes can be thought of as scattergrams, making visible the diversity of existing codes, what codes predominate, as well as changes over time. Moreover, Bernstein's theorised knowledge

[26] Taking advantage of these infinite positions is left to the users of the LCT framework. One way of doing it is proposed in Chapter 4 of this thesis.

structures, to which LCT proposes complementary knower structures, since "for every knowledge structure, there is also a knower structure, so to focus solely on knowledge structures is to see only one dimension of the fields" (Maton, 2014a, p. 72). LCT further proposes an expanded range of codes available aiming to describe multiple facets of practices and discourses. Bernstein's pedagogic codes of classification and framing as well as knowledge structures are subsumed into specialization codes, his elaborated and restricted codes into semantics, and codes describing autonomy, density, and temporality are added. On the following pages, I will define each of these LCT codes in turn.

3.3 Maton's Legitimation Code Theory (LCT)

When talking about the role of theory in his work, Bourdieu remarked "let me say outright and very forcefully that I never "theorize", if by that we mean engage in the kind of conceptual gobbledygook (*laius*) that is good for textbooks… I never set out to "do theory" or to "construct a theory" *per se*" (Wacquant, 1989, p. 50, original emphases). Differentiating thus between a theoretical theory—"often vacuous meta-discourse around concepts treated as intellectual totems" (p. 50)—and a scientific theory, Bourdieu described his work's theoretical concepts, which he preferred to call "thinking tools" (p. 50), as having been shaped by and inextricably linked to empirical research into practical problems and visible through the results they helped unravel. "Bourdieu's *theory of practice* begins in practice and ends in practice—the first empirical, the second scientific" (Grenfell, 2018, p. 278, original emphasis). He then "never felt the urge to retrace the genealogy of the concepts I have coined or reactivated" (Wacquant, 1989, p. 51). In essence, by distancing himself from theoretical

theories and arguing for theoretical concepts necessarily arising from engagement with practical problems, Bourdieu affirms Kant's well-known argument that "thoughts without content are empty, intuitions without concepts are blind" (Kant, 2007, p. 93). [27]

Drawing explicitly on Kant's famous argument, Maton (2016a) laments the separation, in empirical research into education and into society, between, on the one hand, theoretical concepts that cannot be or are challenging to apply to data, and, on the other hand, descriptive studies lacking theoretical underpinning such that their findings cannot be applied beyond specific contexts of the studies. To help researchers avoid these issues and other so-called false dichotomies[28], he puts forward Legitimation Code Theory as a solution. This is a bold move, whose high ambition is well illustrated when Maton (2005b) quotes (after acknowledgements) Basil Bernstein (from personal communication between the two): "The impossible is the only thing worth thinking about".

As what kind of theory is LCT presented? First, noticeably echoing the name of Bourdieu's theory of practice, LCT is described as "a practical theory of practice" (Maton, 2016a, p. 3), given that LCT concepts are to be "enacted in empirical studies to engage in genuine dialogue with data" (p. 3). Second, LCT is described as an explanatory framework or conceptual toolkit (again perceptibly echoing Bourdieu's use of the term thinking tools). In describing LCT as an explanatory framework, Maton (2014a, 2016a) draws directly on Archer (1995), who posits that an explanatory

---

[27] In the second part of his *Critique of pure reason* (section on Logic in general), Kant posits that all human knowledge originates from the faculty of receiving representations of objects (receptivity for impressions) and from our ability to think/conceptualise objects from the representations (spontaneity of concepts). "Neither concepts without an intuition in some way corresponding to them, nor intuition without concepts, can yield knowledge" (Kant, 2007, p. 92).

[28] As examples of false dichotomies, Maton (2016a) mentions quantitative vs. qualitative methodologies, humanism vs. science, or behaviour vs. meaning, which are, Maton declares, often presented as "jointly exhaustive and mutually exclusive" (p.1).

framework (or an explanatory methodology) in itself "neither explains nor purports to explain anything" (p. 6). Instead, for Archer, an explanatory framework is embedded as a link in the chain between social ontologies (represented, for example, by John Locke's Theory of Individualism), though not necessarily exclusively tied to any one of them, and practical social theories (originating from substantive studies). The main function of this intermediary link is for an explanatory framework to regulate "any form of conflationary theorizing at the practical level" (p. 6), meaning one-dimensional theorising, which may occur when researchers are forced to take sides in social theory dichotomies such as individual vs. society or voluntarism vs. determinism, then reduce the disfavoured side's properties/powers/tendencies of social reality to a byproduct of the favoured side (i.e., conflate the two), instead of acknowledging that both views of social reality may be interconnected and exert autonomous influence. Maton (2014a, 2016a) proposes to develop Archer's three-link chain by differentiating meta-theories (-isms) offered by social ontologies, theories offered by explanatory frameworks, and substantive theories afforded by substantive studies. LCT is presented as an explanatory framework, and while it is engaged dialogically with meta-theories (e.g., critical realism), and possesses its own ontological and epistemological assumptions (e.g., "relational analysis, generative theorizing, and a non-empiricist exploration of the organizing principles of practices" (Maton, 2014a, p. 16)), it does not aspire to be a paradigm (meta-theory). The framework is also said not to be identical with the outcome of concrete empirical research; it "invites use to generate explanations" (Maton, 2016a, p. 5) or conjectures regarding a studied phenomenon, but the framework should not be reduced to a collection of outcomes of substantive studies or one specific substantive account/theoretical model of knowledge/practice/education. Third,

discernibly using Bourdieu's and Bernstein's language, LCT is described as "a relational theory" (Maton, 2016a, p. 235) in the sense that its framework embodies a gaze (i.e., the dispositions behind modes of thinking, acting, and being, also referred to as taste/eye/sense/feel for something) and suggests to offer insights not limited to individualised concepts, but expressing relations between/among concepts in LCT dimensions. Last, Maton (2014a) states that LCT concepts neither "arise *ex nihilo*" nor "herald a revolutionary 'break'"(p. 129, original emphases). Instead, they arise from and cumulatively develop inherited concepts (mainly Bernstein's), and should be understood as a work in progress, since they are continually (re-)developed as research studies using them speak back to LCT.

This last point needs to be put into perspective. Especially after reading Maton's (2014a) book-length elucidation of the framework's origin and description of the dimensions of Specialization and Semantics, one might form an impression that LCT is somewhat unique in innovating on Bernstein's work. However, as Singh (2015) pointed out, Bernstein's ideas and concepts inspired theory-building and much continued work on the question of inner properties of knowledge across linguistics, philosophy, anthropology, and sociology. A systematic literature review by Maton, if it were present, would have "ensured that the claims about knowledge-building and theory innovation were more nuanced" (p. 488).

To elucidate the rest of LCT's name, we can refer to Bernstein's concepts. At the centre of LCT's name is the notion of a code, which can be generally[29] defined as a "regulative principle, tacitly acquired, which selects and integrates: a) relevant

[29] Bernstein also defines a specific code for writing in a given context, and changes the second half of the definition to a regulative principle, tacitly acquired, which selects and integrates: "a) orientations to meanings b) textual productions c) specialized interactional practices" (Bernstein, 2003b, p. 12).

meanings b) forms of realizations c) evoking contexts" (Bernstein, 2003b, p. 11). The term *legitimation* refers to the nature of communication sanctioned in a particular context, because Bernstein (2003b) noted that the definition of a code implied appropriate/inappropriate or legitimate/illegitimate meanings, realisations, and contexts, thus it "presupposes a hierarchy in forms of communication and their demarcation and criteria" (pp. 11-12). This hierarchy is implicit in LCT as well, although it is not a sub-type of Bernstein's code theory. Maton (2016b) states that legitimation codes "conceptualize organizing principles of practices, dispositions and contexts" (p. 240). There are several types of legitimation codes, each encapsulated in one of LCT's currently five dimensions–Specialization, Semantics, Autonomy, Density, and Temporality. Maton (2016a) describes an LCT dimension as composed of a set of concepts that capture the internal organising principles of practices and relations in various contexts (e.g., academic disciplines, professions, and cultures). In other words, the set of concepts in each dimension represents a type of legitimation code: specialization codes, semantic codes, autonomy codes, density codes, and temporality codes. The ambition behind each dimension is to reveal a different facet of the Legitimation Device, which represents the mechanism of how the organising principles are constituted, maintained, changed, and transmitted. Struggles over the Legitimation Device reflect actors' struggles for dominance in a field of practice, for example, struggles over legitimate ways of doing and communicating in a (sub-)discipline.

Given that LCT's sets of concepts within each dimension generate four modalities depending on the varying strengths of the concepts (e.g., within the dimension of Specialization, there are knowledge, knower, elite, and relativist codes), LCT has introduced a plethora of additional descriptors (with internal sub-divisions),

especially when compared to other already established frameworks such as that of Bourdieu, who worked principally only with habitus/field. The word *principally* is important, since, on the one hand, further concepts can be discerned in Bourdieu's work —capital, reflexivity, social class, hysteresis, autonomy, doxa, interest (also named by Bourdieu as illusio or libido), symbolic violence, conatus—albeit, on the other hand, as Grenfell (2014) cautions, the concepts should be seen as integrated rather than discreet and ultimately of common epistemology. If we accept Bernstein's (1999) criterion that "the justification for yet another language can only be whether, on the one hand, its use enables a more productive, a more general perspective, and on the other, whether it leads to new research possibilities and interpretations" (p. 158), the rationale for LCT dimensions and concepts should be sketched below in addition to describing the concepts themselves. As a side note, the first three dimensions have been theorised in most detail, with Specialization and Semantics by far the most frequently enacted dimensions in empirical studies. Specialization is also the dimension underpinning this study, which is why more space (regarding its stated genealogy and concepts) is devoted to it compared to the other dimensions. Comments on LCT are also concentrated around this dimension, but could be extended to the whole framework.

3.3.1 Specialization

When describing the nature of empirical research, Bernstein (2003b) remarked that "a paper is often not terminal but a beginning, an opening to an enlarged problematic and an initial development of the language of its articulation and research" (p. 6). Correspondingly, his own papers Bernstein described as "an on-going attempt" (p. 8), inviting thus others to continue working on problems he devoted himself to. This

invitation was accepted by Maton (e.g., 2005b, 2014a, 2016a)—and many others not mentioned by Maton, as noted above—who proposed a framework that aims to enable analysis of concepts from both contexts of knowledge production (Bernstein's (1999) knowledge structures) and knowledge recontextualisation and reproduction (e.g., Bernstein's (1971) classification and framing and educational knowledge codes) in order to explore to what extent practices are a reflection of social power/social hierarchies and to what extent they mirror ontological and/or epistemological principles. The declared aim is to offer resolution to the problem of knowledge blindness (i.e., undertheorising of the properties, powers, and tendencies of knowledge itself as opposed to focusing on *whose* knowledge) and more broadly to the knowledge paradox[30] (i.e., modern society is supposed to be centred around the notion of knowledge, while what knowledge itself is remains unclear) (Maton, 2014a). The aspiration in Maton's project to conceptualise properties of knowledge structures that could be applied beyond curriculum and pedagogy to all social fields of practice has been received positively by some for whom Maton's work "significantly developed … the potential within Bernstein's thinking" that "organization of knowledge is also an organisation of identities located within different positions of a social field" (Moor, 2013, p. 149). Others (e.g., Singh, 2015) have cautioned against LCT's narrow interpretation of knowledge, ways of knowing, and knowledge growth.

One avenue of continuing work on Bernstein's concepts that Maton (2014a) took was to propose knower structures to complement Bernstein's knowledge structures. According to Maton (2014a), the idea of knowledge-knower structures can be approached, for example, through the works of authors who weighed in on the debate

[30] Also pointed out by Young (2008), as I noted in Section 3.1.

about the culture of the sciences vs. the culture of the humanities in the 1950s and 1960s as well as through the lens of Bernstein's code theory's focus on organising principles of intellectual fields. In the two cultures debate, the humanities were said to have fragmented into strongly-bounded fields after the decline of the Classics, which used to serve as a unifying core. The fragmented fields then competed with each other for influence, but there was no common means of adjudicating between competing claims. This description bore likeness to Bernstein's horizontal knowledge structures. Regarding the Classics as a unifying core, the common element did not lie in principles or procedures, but in the dispositions (attributes, behaviours, worldview, taste, etc.) that education in the Classics was supposed to cultivate (one could see the dispositions as Bourdieu's habitus). At the top of an imaginary "triangle" was an ideal knower (a cultured English gentleman) who pursued broad general education for the sake of knowing, and who inevitably followed a particular social profile (white, male, upper social class, private school, Oxford/Cambridge educated). Knowers aspiring to reach this ideal or along the way to this ideal were integrated in various positions below the apex. The humanities could thus be said to comprise hierarchical knower structures. Analogously to Bernstein's strong and weak grammars for horizontal knowledge structures, hierarchical knower structures could also be divided into those with strong grammars (explicitly stated social or biological criteria of legitimate knowers) and those with weaker grammars (based, for example, on enjoying a particular educational trajectory).

In comparison, the sciences were characterised as being united around shared standards, assumptions, and approaches. While the sciences also comprised multiple specialisms, each producing new claims, these claims could be integrated within a

"triangle" with more general propositions towards the apex and more specific phenomena below towards the base. The culture of the sciences thus resembled Bernstein's hierarchical knowledge structures. Knowers in the sciences were said to comprise individuals with various habituses, with social and/or biological origins little relevant when the yardstick of achievement was more meritocratic, based on knowledge of specialist principles and procedures. The sciences could thus be said to resemble horizontal knower structures.

If, then, "the location of the 'hierarchical' in a social field of practice is the basis of its recontextualising principles" (Maton, 2014a, p. 71)—recontextualising rules derive from distributive rules, which "specialise forms of knowledge, forms of consciousness and forms of practice to social groups" (Bernstein, 2000, p. 28); that is, "the distributive rules mark and distribute who may transmit what to whom and under what conditions, and they attempt to set the outer limits of legitimate discourse" (Bernstein, 2000, p. 31)—then complementing knowledge structures with knower structures means that differences between social fields can be conceptualised not based on whether hierarchical structures are present or absent, but rather based on from what kind of hierarchies principles that legitimate achievement and positions emanate. Additionally, discourses and actors within social fields of practice are positioned not only within knowledge structures and knower structures, but "also establish different forms of relations to these two structures… epistemic relations to knowledge structures and social relations to knower structures" (Maton, 2014a, p. 72).

Going back to the debate about the cultural differences between the sciences and the humanities, Maton (2014a) noted that the shared identity in scientific culture was said to emanate from allegiance to knowledge that could be arrived at by adherence to

specialised procedures and principles. In comparison, in the humanities, identity and achievement were described as depending relatively less on following specialised procedures and principles that could provide knowledge of objects of study, but more weight was put on personal dispositions of knowers. In terms of relations to knowledge structures (epistemic relations) and to knower structures (social relations), the sciences classified and framed (to use Bernstein's terms) epistemic relations strongly, while in the humanities, social relations were classified and framed strongly. These characteristics of the sciences and humanities and their relations to knowledge and knower structures then bring us to the dimension of Specialization.

While I attempted above to preface and complement my summary of Maton with my reading of Bernstein to create a sense of cohesion where I thought Maton's account needed it, one should admit that the characteristics of the sciences and the humanities mentioned above are necessarily generalised. Existence of a neat dividing line is doubtful, though it effectively serves a simplified narrative. In both the sciences and the humanities, there are bound to be degrees of hierarchical and horizontal inner organisation of both knowledge structures and knower structures. Especially knower structures in the sciences are unlikely purely horizontal in the light of Bernstein's (2000) view of academic disciplines as singulars, each with its "rules of entry, examinations, licenses to practice, distribution of rewards and punishments" (p. 52). Further evidence somewhat complicating Maton's reporting that knowledge-making in science is (mostly/ fairly) blind to personal merits or creed is, for example, Dietz's (1977) commentary on how plate tectonics became the new revolutionary paradigm in modern geology: "There have been, and still are, some critics of plate tectonics, but they are few and have been virtually overwhelmed by wide general acceptance by the opinion makers or "invisible

college" within our scientific discipline" (pp. 1-2, original emphasis). Of course, more precise characteristics of individual disciplines need to be determined by empirical research. For now, Maton's rationalisation can be considered as a working model—with all theories naturally tending to produce models, as Bernstein (2000) noted—but between Maton's model and reality there necessarily exists a "discursive gap" as studied phenomena always generate a greater range of possibilities, making theories underdetermined (Bernstein, 2000, pp. 125-126).

Another avenue of continuing work on Bernstein's concepts described by Maton (2014a) was integrating classification/framing and collection/integrated codes into LCT. Bernstein (1971) described collection codes as marked by strong classification (content differentiation along subject lines) and framing (increased teacher discretion and decreased pupil discretion) of educational knowledge, while integrated codes, where integration means "*subordination* of previously insulated subjects *or* courses to some *relational* idea, which blurs the boundaries between the subjects" (p. 53, original emphases), possess weaker classification and framing. Both types of codes can have sub-types, differing in the relative strengths of classification and framing. The generally strong classification and framing of the collection codes is "redescribed" by Maton (2014a, p. 75) as stronger epistemic relations, while the generally weaker classification and framing of the integrated codes is redescribed as weaker epistemic relations. The strengths of social relations vary from one empirical case to another, Maton (2014a) notes, yet, as a rough generalisation he adds that "it is likely (and studies using code theory suggest)" (p. 76) that knower dispositions are less important under collection codes and more important under integrated codes. Unfortunately, Maton provides no citation for the reader to verify this assertion. On the one hand, a modicum of indirect

evidence for the claim can be taken from Bernstein (1971), who mentioned, for example in the European context, the engrained discipline among students to work unquestionably within received frames of the collection code as well as evaluation based on received educational knowledge (theoretically affording meritocratically-derived hierarchy of achievement within the education system, although meritocracy was disputed by Bourdieu, Bernstein, and NSOE sociologists of education—Section 3.1). On the other hand, Bernstein, as mentioned above, did not focus much on knower dispositions in his work, save for mentioning some content differentiation based on higher/lower achievement of students, so the absence in Maton (2014a) of any reference leaves the reader no other option but to either accept or doubt Maton's claim. Nevertheless, the above-redescribed characteristics of collection and integration codes again bring us to the dimension of Specialisation.

Specialization is described by Maton (2014a) as reflecting the fact that "practices and beliefs are about or oriented towards something and by someone" (p. 29). Specialization thus comprises two concepts: epistemic relations (ER) between practices and their objects of study or focus, and social relations (SR) between practices and their actors or subjects. In academic disciplines, which are concerned with knowledge claims, epistemic relations capture what can be legitimately studied and how it can be legitimately studied; social relations capture who can be a legitimate knower. Both epistemic relations and social relations[31] can vary independently from stronger (symbolised by a qualifier sign +) to weaker (symbolised by a qualifier sign –), which

[31] LCT concepts have their own written conventions. For example, Maton (2016b) stresses in the LCT glossary that concepts consisting of two words must always include both words. Thus, we should write epistemic relations and social relations, not epistemic and social relations.

mirrors how strongly the boundaries are controlled. The different strengths then reflect four main code modalities, named by Maton as (pp. 30-31):

- *Knowledge code* (ER+, SR–): The bases of legitimate practices and achievement depend on whether actors possess specialised knowledge of objects of study and of procedures, while the actors' personal characteristics are de-emphasised.
- *Knower code* (ER–, SR+): While specialised knowledge is downplayed, personal characteristics of actors are emphasised as the yardstick of achievement. These characteristics can be considered as born (e.g., natural talent), cultivated (e.g., artistic taste), or social (e.g., social class).
- *Elite code* (ER+, SR+): Legitimacy is based on having both specialised knowledge of objects of study and of procedures, and on being the right knower.
- *Relativist code* (ER–, SR–): Anything goes; that is, the bases of legitimate practices and achievement depend neither on possession of specialised knowledge, nor on personal dispositions of actors.

The four modalities of Specialization are shown in Figure 3.1 below.

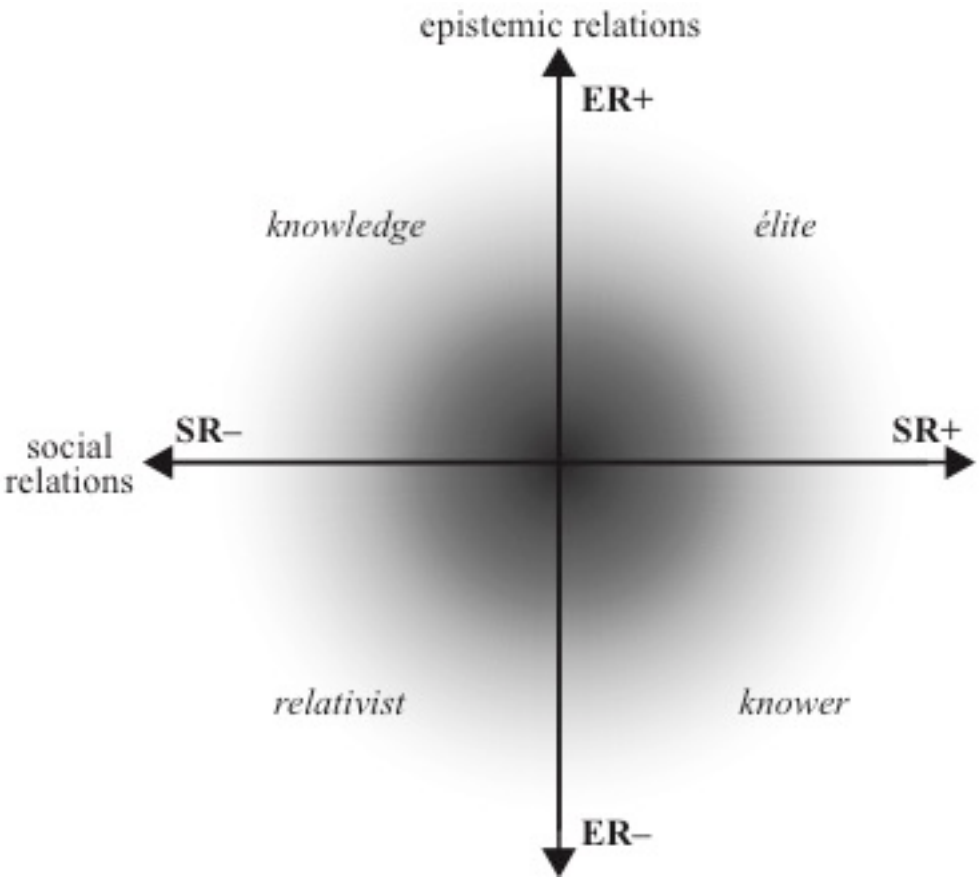


Figure 3.1 The specialization plane (Maton, 2014a, p. 30).

Based on Maton (2014a), one can further elucidate that in a discipline with a predominating knowledge code, legitimation is accomplished through referring to specialised knowledge. Legitimate contributions provide insights into legitimate objects of study through legitimate theoretical and methodological approaches, all of which may differ from other disciplines' legitimate objects of study and approaches to their study. Personal attributes of practitioners are considered unimportant, and everyone is equally positioned to make legitimate contributions in the field as long as they stay within the bounds of legitimate objects of study/approaches. In contrast, a discipline with a predominating knower code strongly emphasises legitimate knowers, who may assert insights into a variety of objects of study. The choice of topics and methods is at

the knowers' discretion, and potential conflicts in knowledge claims cannot be settled by epistemological means. Knowledge claims are underpinned by knowers' personal attributes, such as social origins or acquired gazes. Anyone else not possessing the same attributes is unable to make new knowledge claims or challenge existing ones.

Another level of conceptual development of Specialization is represented by the 4-K model (Maton, 2014a), which is a proposal of a fine-grained differentiation of both epistemic relations and social relations. This development is concerned with an endeavour to answer questions of how differences within fields that share the same Specialization code can be conceptualised, especially how such differences may potentially impact cumulative building of knowledge in social fields of practice.

One issue can be traced from Bernstein's formulation of strong and weak grammars. As a reminder, the term *grammar* has been explained as a "conceptual syntax" for empirical descriptions or modelling of empirical relations (Bernstein, 1999, p. 164); "ordering principles connecting the various levels of activity" (Bernstein, 2000, p. 185); or "historically developed methodologies that enable the synthesis of knowledge at increasingly higher levels of abstraction and generality through conceptual subsumption" (Moore, 2013, p. 86), all of which suggests that grammatically (i.e., strong grammar) could be applicable to hierarchical knowledge structures, not only with respect to horizontal knowledge structures as used in Bernstein's (1999) paper. But if, on the one hand, "grammaticality determines the capacity of a theory or a language to progress through worldly corroboration" (Muller, 2007, p. 71), and, on the other hand, some fields with strong grammars "often achieve their power by rigorous restrictions on the empirical phenomena they address" (Bernstein, 1999, p. 164), how is, Maton (2014a) asked, the reliability of empirical descriptions and modelling impacted, or what

explanatory power is achieved at such a cost? Is it sufficient to define epistemic relations to knowledge structures only with respect to precise descriptions (strong grammar)? Should the issue of what counts as a legitimate description or model of empirical phenomena be also considered? In addition, in the case of disciplines where the legitimate achievement is defined more in relation to hierarchies of knower structures—"weak grammar segmentation tends to be associated with *knower* claims" (Moore, 2013, p. 146, original emphasis)—how can the principles that generate legitimate gazes of knowers be conceptualised?

The questions above led Maton (2014a) to distinguish within epistemic relations *ontic relations* (OR) between (disciplinary) knowledge and (disciplinary) objects of study, and *discursive relations* (DR) between (disciplinary) practice or knowledge and other practices or knowledges. In other words, the strength of ontic relations represents how strongly a discipline delineates its legitimate objects of study, while the strength of discursive relations represents how strongly a discipline controls the legitimate methods and procedures for investigating its objects of study. The strengths of these relations can vary, and the relations can be strengthened or weakened, generating four insights of the epistemic plane (Maton, 2014a, p. 176):

- *Situational insight* (OR+, DR–) bounds strongly what objects of study are legitimate, but weakly how these objects can be studied. Disciplines with this insight are specialised by the problems they investigate, though these problems may be investigated using various methods and procedures.
- *Doctrinal insight* (OR–, DR+) does not delimit the objects of study, but controls legitimate approaches. Doctrinal insight disciplines are

specialised by their approaches, which can be applied to varied phenomena.

- *Purist insight* (OR+, DR+) asserts that only legitimate objects of study can be studied, and it must be done only by specific means. Studying other objects is not valued, and methods of study must not be applied to inappropriate objects. Disciplinary practitioners are thus limited in their choice of what to study an/or how to study it, and their professional identity is built on their knowledge of legitimate objects of study and appropriate use of legitimate methods.
- *Knower or no insight* (OR–, DR–) means that practices or disciplines weakly control what objects of study or methods are legitimate. This type of weaker epistemic relations may signal either a knower code (ER–, SR–) or a relativist code (ER–, SR–).

The typology of insights is shown in Figure 3.2 below.

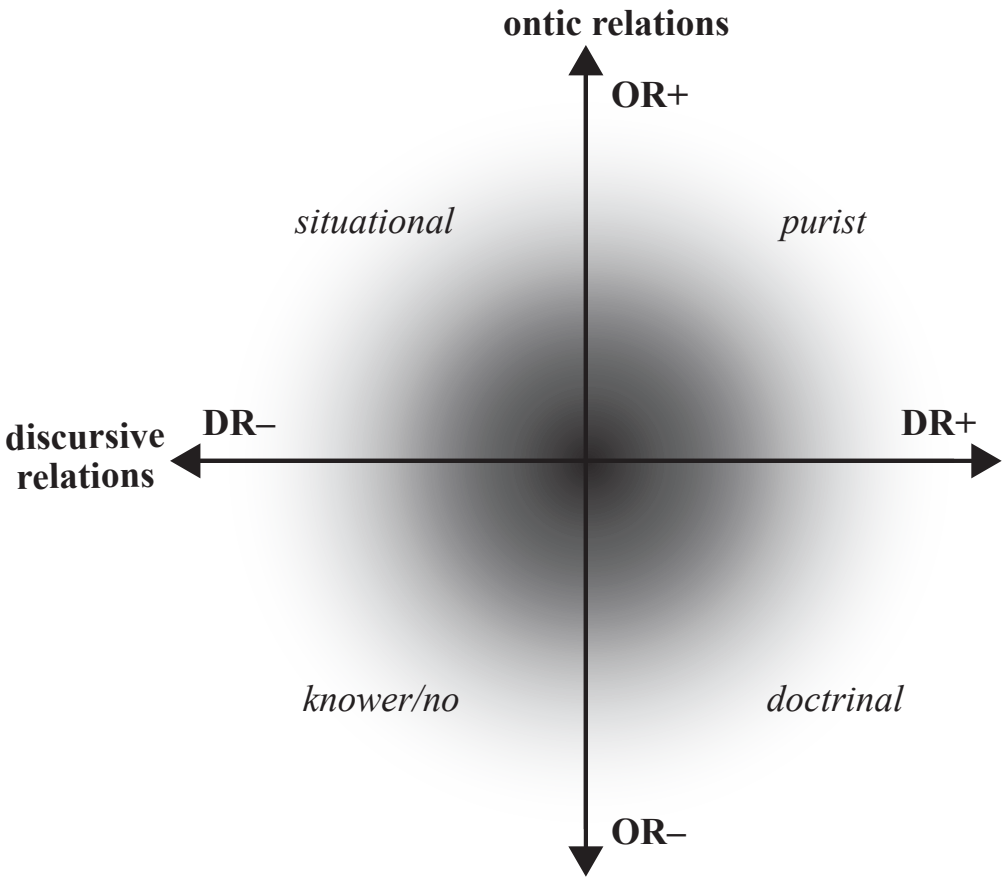


Figure 3.2 The epistemic plane - insights (Maton, 2014a, p. 177).

All four insights may appear in the same field (e.g., Wolff, 2017; 2020), but their predominance may be based on a particular problem investigated or the preferred ways of working either by individuals or procedures put in place by companies. For example, investigating the relationship between physics, mathematics, and logic, Wolff shows on several case studies of mechatronics engineers working on a variety of problems that to be successful problem solvers, the engineers need to recognise which insights (ways of thinking and acting) are required at which step of a problem-solving situation. She also finds that individuals whose natural preferences for certain insights clash with institutionally-imposed insights (e.g., larger companies may highly regulate which approaches are legitimate–a doctrinal insight) do not thrive in such environments.

In the model, Maton also distinguished social relations based on what kinds of knowers they represent and how knowers know. Hence, he differentiated between *subjective relations* (SubR)–how strongly knowledge claims bound and control legitimate kinds of knowers–and *interactional relations* (IR)–how strongly knowledge claims bound and control legitimate ways of knowing through interactions with significant others. The strengths of these relations can vary, and the relations can be strengthened or weakened, generating four gazes of the social plane (Maton, 2014a, pp. 185-187):

- *Social gazes* (SubR+, IR–) are possessed by knowers belonging to a specific social category (class, gender, etc.), regardless of interactions with others.
- *Cultivated gazes* (SubR–, IR+) are possessed by knowers who have interacted with significant others, for example, by having been continuously exposed to good models, or who have extensively participated in the field/discipline, or have been apprenticed under a recognised practitioner.
- *Born gazes* (SubR+, IR+) are possessed by knowers who both belong to the appropriate social category and who have been exposed to significant others.
- *Trained/blank gazes* (SubR–, IR–) are possessed by knowers with specialised knowledge, and may thus be part of either a knowledge code (ER+, SR–, a *trained gaze*), or a relativist code (ER–, SR–, a blank gaze). Unlike the first three gazes, which are characterised by stronger

social relations (SR+), the *trained/blank gazes* are characterised by weaker social relations (SR–).

The typology of gazes is shown in Figure 3.3 below.

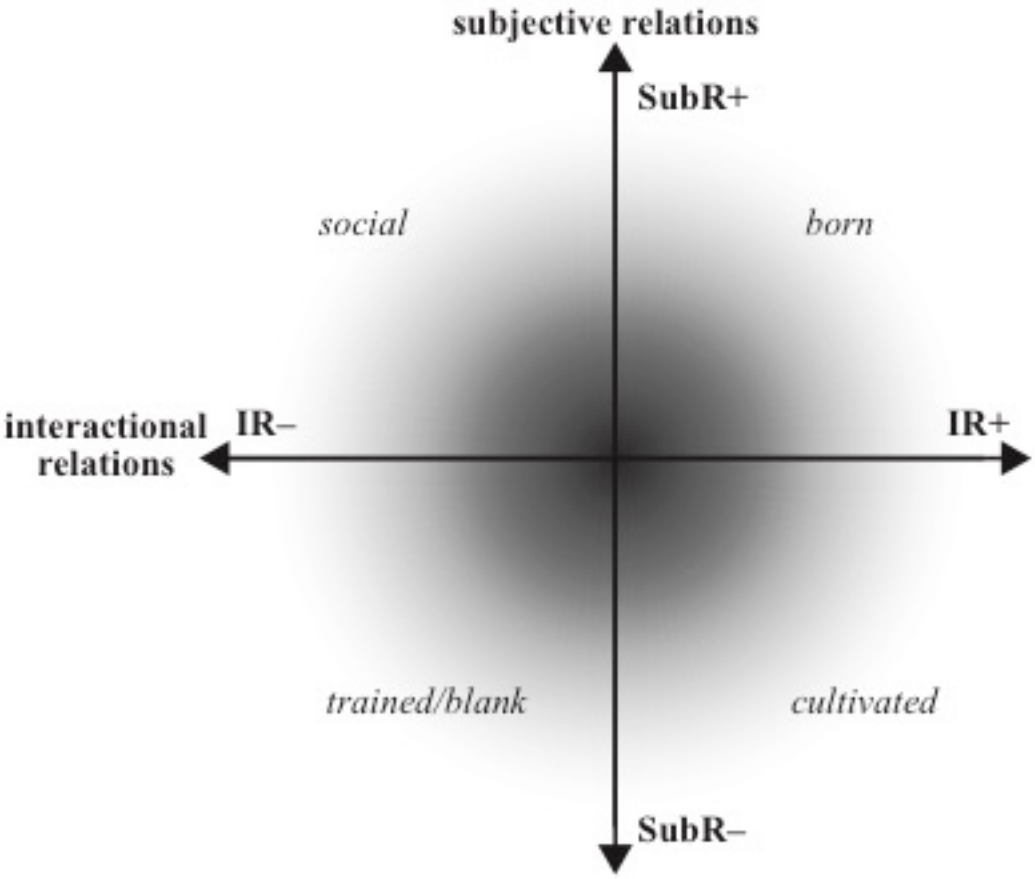


Figure 3.3 The social plane - gazes (Maton, 2014a, p. 186).

The four gazes do not characterise any discipline in absolute terms; that is, as a discipline evolves, the gazes can shift (and clash). Maton (2014a) exemplifies this shift on the case of British cultural studies. From the late 1950s, the discipline was characterised by a cultivated gaze (SubR–, IR+), given its focus on fostering students' critical discrimination of great works of culture, and on inculcating new disciplinary members with appreciation of famous works of cultural studies. The social background

of disciplinary members was insignificant, since the cultivated gaze could be taught to and learned by almost anyone, and the discipline was open to new members from the traditionally minority groups (e.g., the working class). In the mid-1970s, the arrival of various standpoint theories based on race, gender, class, ethnicity, or sexuality marked the beginning of an emphasis on one's social origin. Knowledge of, or critical discrimination of, previous works and educational background were de-emphasised, and instead, new approaches were theorised, privileging claims reflecting pre-existing gazes stemming from subjective experiences of members of the various social groups. This gaze based on primary experiences could not be taught or learned. The legitimation in the discipline thus shifted towards a social gaze (SubR+, IR–). In the 1990s, legitimate gazes in cultural studies again shifted, as the various post- (e.g., post-structuralism, post-modernism), critical, and deconstructive theories rose to prominence, reflecting stronger interactional relations (IR+). Legitimacy was not based on privileging primary experiences of social groups (SubR–), but on gazes "capable of recognising and realising ostensibly 'virtuous' or 'radical' stances" (p. 189).

The dimension of Specialization has, to date, been theorised in most detail of all LCT dimensions. The ambition has ostensibly been to cumulatively develop Bernstein's concepts to offer a more nuanced description of the principles underlying practices in social fields in order to resolve the knowledge paradox of modern societies (knowledge-based economies), described by Maton (2014a) as "knowledge is everything and nothing" (p. 1). Nevertheless, LCT's proliferation of types of knowledge insights and knowers' gazes has also been critiqued on the very same grounds offered by Maton as solutions. What if the core of the knowledge paradox, Singh (2015) asked, is "the need for certainty and control" (p. 489) rather than knowledge undertheorisation? If so, the

solution offered by LCT could result in "more ambiguity and uncertainty, which in turn leads to a flurry of knowledge production" (p. 489). Additionally, Singh suggested that continuation of Bernstein's project "may not be best served by categorising different participants as knowers with different gazes" (p. 492), especially when LCT, in constructing its concepts by criticising various standpoint theories, has not engaged critically with the scholarship in the criticised fields, unlike Bernstein; and has based its conceptual development on an earlier version of code theory, ignoring later development in Bernstein's work. The last point echoed Tyler (2014), who, while critiquing the theoretical underpinning of the dimension of Semantics (described below), wrote that LCT had drawn on the earlier version of concepts abandoned by Bernstein in the 1960s but picked up later by SFL. Instead, Singh (2015) proposed, that since Bernstein himself viewed contesting theories around shared objects of study as social division of labour among various middle-class factions vying for and possessing differing power/influence in intellectual fields, neo-Bernsteinian cumulative knowledge-building could as well focus on analysing the social division of labour, the strength of classification around the various factions, and the flow of ideas within and between the factions' boundaries. Singh's proposition is then for LCT to align itself more closely with Bernstein's oeuvre. Similarly, Tyler (2014) recommended that instead of rejecting the complete theoretical underpinning of LCT, the framework should be developed into "a more explicit, precise and discursively-informed set of analytical tools grounded in the trajectory of Bernstein's project" (p. 4).

3.3.2 Semantics

Young (2020) noted that Bernstein was the first to argue that debates about education needed to focus also on knowledge (and not only on the effect of education on students). Bernstein (2000) described his own work as having been primarily concerned with "how relations within are constituted" (p. 188), as what was missing from theories of cultural reproduction in education was an "internal analysis of the structure of discourse itself" (p. 4). This focus on the internal structure of educational discourse is also a starting point for Maton's (2014a) description of the background to the dimension of Semantics. According to the literature reviewed by Maton, education suffers from the problem of segmentalism, where knowledge is acquired by students as a series of isolated ideas and skills rather than cumulatively built on, trapping students in those contexts as they are unable to transfer what they learned into other settings. To begin exploring segmental and cumulative learning, Maton again draws on Bernstein's (1999) horizontal and vertical discourses and hierarchical and horizontal knowledge structures.

As mentioned in Section 3.2.2, horizontal discourses represent everyday (commonsense) knowledge acquired in discreet contexts, meaning that the strategies acquired for one context may be irrelevant for another. At the same time, a context does not necessarily require only one correct strategy. Vertical discourses represent academic or professional knowledge, which is organised into specialised symbolic structures. Vertical discourses may be hierarchical, developing by integration and increasing abstraction, or they may be horizontal, developing by lateral accumulation of new specialised languages segmentally separated. Knowledge structures also possess degrees of grammaticality, through which they can formulate more or less ambiguous referents,

and all theories possess languages of description (Bernstein's (2000) internal (L1) and external (L2) languages, which, respectively, describe internal constituents and their relationships and make the internal constituents visible by translating them into empirical referents). Maton (2014a) posited that Bernstein's conceptual distinction between vertical and horizontal discourses highlighted differences in context dependence. He further stated that context dependence was applicable in education contexts, since school knowledge could also be classified into hierarchical or horizontal, depending on whether lessons/courses/curricula presented new knowledge as integrated with/building on previous units/years or as segmented. What needed to be developed, Maton argued, was a continuum of differences arising from dichotomous types as well as the underlying organising principles that gave discourse/structure types their characteristic features.

According to Martin et al. (2020), the dimension of Semantics (not the same as semantics in discourse) "construes social fields of practice as semantic structures whose organizing principles are conceptualized as semantic codes" (p. 22). The codes reflect the strengths of two concepts, which capture the underlying organising principles: semantic gravity (SG) and semantic density (SD). Semantic gravity highlights the extent to which meanings relate to or depend on their contexts: The stronger the contextual dependence (i.e., the more localised the concept), the stronger the semantic gravity; the weaker the semantic gravity (i.e., the more generalised the concept), the less tied the concept is to its context of occurrence.

Semantic density reflects the extent of meaning condensation (i.e., complexity of practices). The stronger semantic density is, the more meanings are condensed in practices (the more complex practices are); the weaker the semantic density, the fewer

meanings are condensed. However, Maton (2013) cautions that the strength of semantic density of a practice is not inherent (to the practice by itself), but is always conditional upon the semantic structure within which it occurs. For example, the word *gold* in everyday contexts denotes a shiny, yellow metal used in jewellery or electronics, but in chemistry, *gold* is embedded in a periodical table alongside other elements, has its atomic weight and number of electrons, and conducts heat, etc.

Both semantic gravity and semantic density are always present in practices (Shay, 2013). While the concepts' relative strengths may vary from stronger (+) to weaker (–) independently (Maton, 2016a), in some cases, the strength of one concept significantly correlates with the strength of the other (Shay, 2013). For example, everyday knowledge learned through performing specific tasks in daily life (Bernstein's (1999) horizontal discourse) is embedded in context (SG+) and largely not based on theory or complex concepts (SD–); in contrast, academic knowledge (Bernstein's (1999) vertical discourse) is often independent of a specific context (SG–), which corresponds with its conceptual condensation, abstractness, and generalisability (SD+). Thus, the varying relative strengths of the semantic concepts reflect four main code modalities (Martin et al., 2020, p. 22):

- *Rhizomatic code* (SG–, SD+): The bases of achievement and of legitimacy consist of practices which are relatively context independent, and their meanings are condensed and complex.
- *Prosaic code* (SG+, SD–): Legitimate practices are relatively context dependent, and their meanings are simpler.
- *Rarefied code* (SG–, SD–): Legitimacy is underpinned by practices which are relatively context independent, and contain fewer complex meanings.

- *Worldly code* (SG+, SD+): The basis of achievement stems from practices which are relatively context dependent, and complex, condensing multiple meanings.

The four modalities of Semantics are shown in Figure 3.4 below.

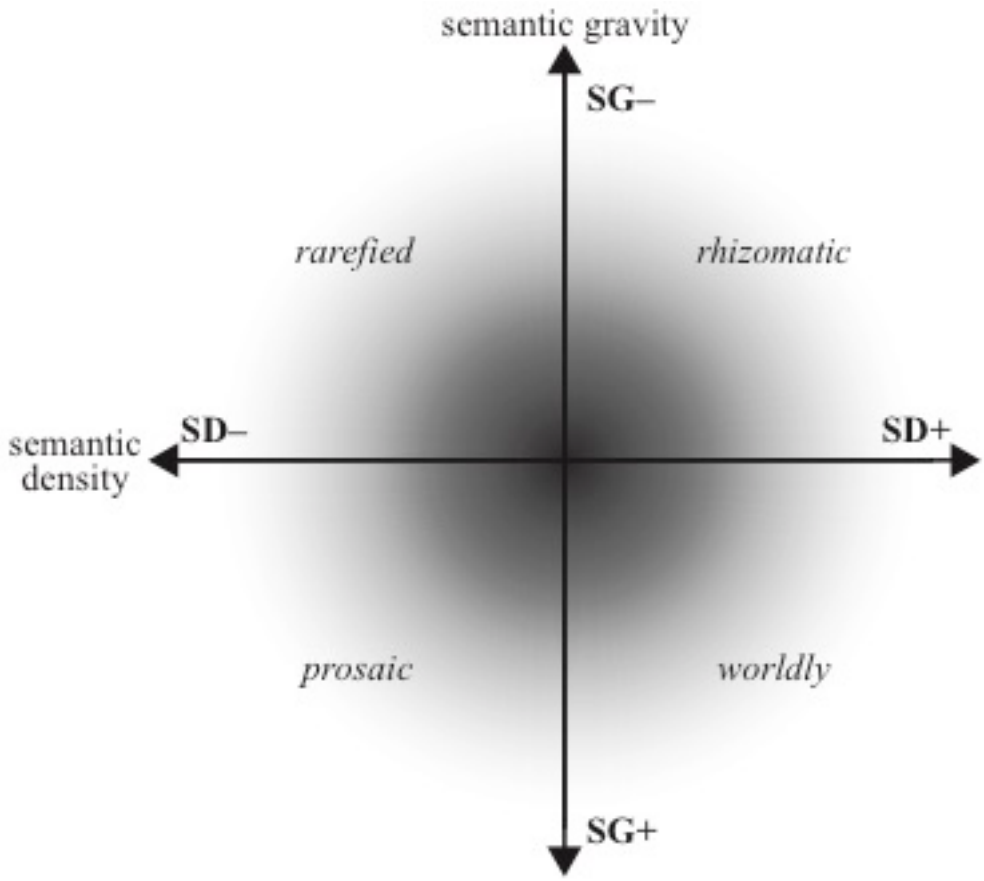


Figure 3.4 The semantic plane (Maton, 2016a, p. 16).

3.3.3 Autonomy

The question of autonomy went to the heart of Bernstein's pedagogic device. As a reminder, the concept of the pedagogic device was formulated in order to try to account for the selection of knowledge from available discourses that could be pedagogised, that is, recontextualised in educational contexts (Bernstein, 2000). The pedagogic device could be regarded as "the regulator of the autonomy" (autonomy as independence of the influence of other agents) of the education system, so that the

degrees of the relative autonomy could vary between a high level of autonomy, where education is controlled by educators based on intrinsic principles as an end in itself (e.g., traditional education), while education with a low level of autonomy is based on extrinsic principles as a means to an end (e.g., vocational training) (Moore, 2013, p. 156).

Further tracing of autonomy in Bernstein's (2000) work leads to the concepts of singulars and regions. Singulars (e.g., physics, chemistry, history, economics) are intellectual fields whose identities are inward facing, formed by delineating for themselves strong boundaries (strong classification and framing) around unique names, discourses, hierarchies, and practices. Regions (e.g., medicine, architecture, engineering) are recontextualised singulars, that is, "the interface" (p. 52) between intellectual fields and fields of external practice. Increasing regionalisation of knowledge, especially in higher education, weakens the autonomy of singulars (their discursive and political bases), and indicates the degree of external influence (e.g., by the markets that they serve). Internally oriented identity (introjected) is shifting to externally oriented identity (projected).

Autonomy was initially described by Maton (2005a) and later substantially revised (e.g., Maton & Howard, 2018, 2020). Maton and Howard (2018) state that this dimension is based on the premise that any set of practices comprises constituents (e.g., actors, ideas, artefacts, institutions, machine elements, body movements, sounds, etc.) that are related together in particular ways (e.g., explicit or implicit procedures, customs, tacit orthodoxies, mechanisms, formal rules, etc.). Autonomy then comprises two codes which explore the strengths of external boundaries (degrees of insulation) of constituents belonging to different practices (positional autonomy, PA) and the strengths

of boundaries around how the constituents from one practice are related to constituents from other practices (relational autonomy, RA). Autonomy thus develops Bernstein's classification and framing (Section 3.2.2). Both PA and RA can vary in strengths, with stronger autonomies (+) denoting stronger insulation and higher specificity of constituents (PA+) and their relationships to a field of practice (RA+), and with weaker autonomies (–) signifying that constituents (PA–) and their relations (RA–) may be shared with or drawn from other fields or contexts. The varying strengths thus generate four autonomy code modalities (Maton & Howard, 2018, p. 7), shown in Figure 3.5 below:

- *Sovereign code* (PA+, RA +): Valued ideas, dispositions, contexts, etc. originate from within a field or practice (strongly insulated constituents), and work according to its specific modus operandi–"internal constituents for internal purposes" (p. 7).
- *Exotic code* (PA–, RA–): Legitimacy is derived from constituents and ways of working derived from or shared with other practices or fields–"external constituents for external purposes" (p. 7).
- *Introjected code* (PA–, RA+): The bases of achievement are constituents shared with or originating from other contexts or practices but put to work according to internal ways of working–"external constituents turned to internal purposes" (p. 7).
- *Projected code* (PA+, RA–): Legitimacy emanates from constituents within the field but oriented towards a modus operandi from another field or context–"internal constituents turned to external purposes" (p. 7).

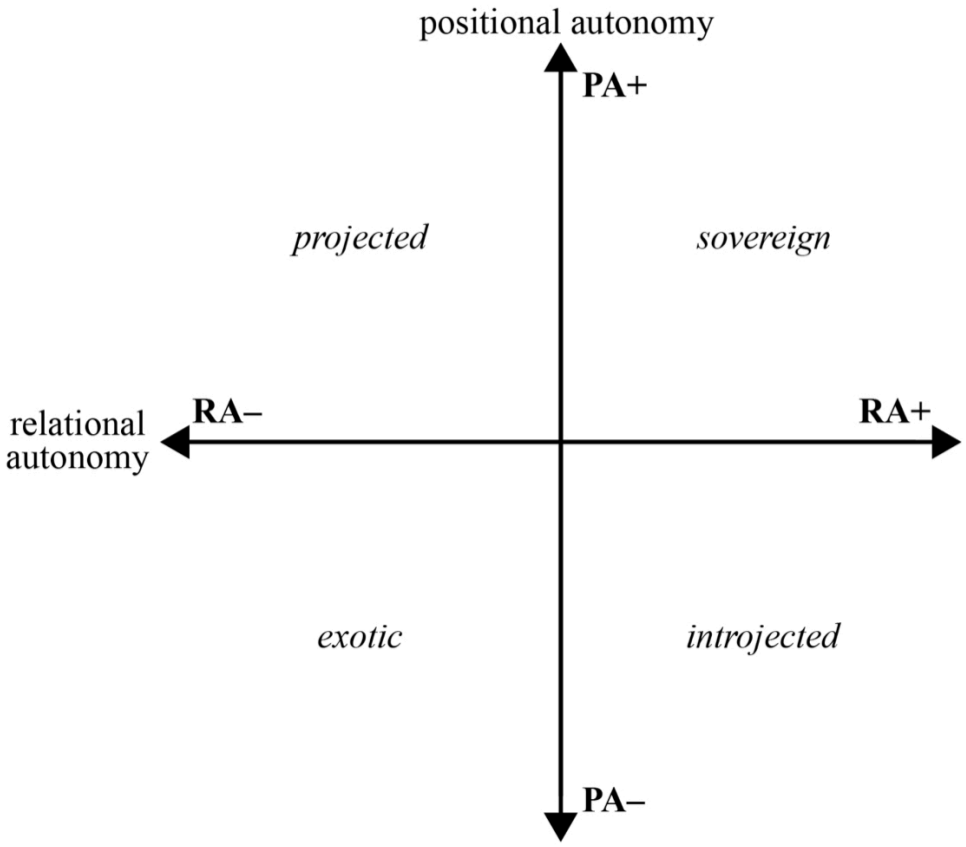


Figure 3.5 The autonomy plane (Maton & Howard, 2018, p. 6).

Maton (2005a) exemplifies the code modalities in his research on English higher education. He states that high-status English universities in the 1960s were strongly insulated from external influences and offered education based on various humanist ideals (PA+, RA+). In contrast, lower-status English colleges in the same period were influenced by external factors and offered vocationally-oriented education (PA–, RA–).[32] The other two modalities are exemplified as institutions strongly influenced by government policies or commercial interests but offering academically-oriented

[32] Vocationally/professionally oriented programmes tend to have a weaker relational autonomy (RA–) (Shay, 2016).

education (PA–, RA+), and institutions governed by academics but offering education meeting professional/vocational needs (PA+, RA–).[33]

3.3.4 Temporality

Maton has announced on the LCT website (www.legitimationcodetheory.com) that the original conceptualisation of this dimension (Maton, 2005b) has been completely revised, and its redefined form is under advanced testing. Briefly then, Temporality consists of the concepts of age and orientation (Maton, 2005b), reformulated in Maton (2016b) as temporal position (TP) and temporal orientation (TO). Positions in a social field may be more established (archaeo-) or newer (neo-), and their characteristics may emanate from traditional practices (retrospective) or may orientate towards newer practices (prospective), with the terms *retrospective* and *prospective* being based on Bernstein's (e.g., 2000, Chapter 4) pedagogic identities, which he distinguished based on "the temporal orientation of the resources used for their social construction" (p. xii). The 2005 version of Temporality is schematised in Figure 3.6 below.

		Orientation	
		Backward ($+F^t$)	Forward ($-F^t$)
Age	Older ($+C^t$)	archaeo-retrospective	archaeo-prospective
	Younger ($-C^t$)	neo-retrospective	neo-prospective

Figure 3.6 The modalities of Temporality (Maton, 2005b, p. 94).

[33] This last modality has been exemplified by Shay (2016) on the case of medical education in South Africa, where medical schools have reshaped the curriculum to graduate more general than specialist physicians in response to the call of the health professional body.

Maton (2005b) states that the idea of a higher-status university was associated with age (older was better). Old institutes acquired their wisdom and maturity, and built their traditions. Their current practices were inherited from their respected past ($+C^t$, $+F^t$). Indeed, the purpose of a traditional university education at higher-status institutes was to resist the commercialised and industrialised way of life in modern society. On the other hand, lower-status colleges were described as relatively new (modern), without tradition, and with practices oriented towards the future ($–C^t$, $–F^t$).

As Maton (2016a) reminds us, using Bourdieu's expression, research may uncover the specific "rules of the game" (p. 13), but the rules are not impervious to change. He explains that, in fact, while one code may dominate at a specific time, it may not be universally adhered to or unchallenged. More than one code can coexist, struggling for dominance. If competing codes exist within a social field, we can then describe degrees of code clash and code match. As a result of the struggle for dominance or a change in practices, foci or emphases, dominant codes may change over time. Such changes can either radically alter the rules of the game (code shifts), with far-reaching consequences for previously successful actors in the field, or they can change to a lesser extent (code drifts), such as when codes move within the same quadrant of the Cartesian plane.

3.3.5 Density

I am including the dimension of Density only for the sake of completeness, since this dimension has not been reformulated since its original conception in Maton (2005b), and to the best of my knowledge, has not been used widely. Moreover, Density

is not mentioned on the LCT website (www.legitimationcodetheory.com), so its current status appears unclear.

As stated by Maton (2005b), Density describes the "differentiation among positions *within* a field" (p. 88, original emphasis). The differentiation is given by the interaction between the concentration of student population or canonical texts, captured by the concept of material density (MaD), and the diversity of opinions and beliefs, captured by the concept of moral density (MoD). The basic premise is that a higher material density will likely result in changes to and diversification of belief systems. Both material density and moral density may vary independently from stronger (+) to weaker (–). The different strengths then reflect four main code modalities, as shown in Figure 3.7 below.

		Moral density	
		MoD+	MoD-
Material density	MaD+	large population, heterogeneous beliefs	large population, homogeneous beliefs
	MaD-	small population, heterogeneous beliefs	small population, homogeneous beliefs

Figure 3.7 The modalities of Density (Maton, 2005b, p. 90).

Maton (2005a) states that judgements about the quality of education frequently include attributes of size, quantity, and scale. For example, education provided by higher-status universities has been viewed as an intimate interaction between a small student population and their teachers, both sharing the same beliefs and values (MaD–, MoD–). In contrast, lower-status colleges have been characterised by offering education

in diverse disciplines to a large population of students from various backgrounds and with various beliefs (MaD+, MoD+).

3.4 Collaboration of LCT and SFL

To complement the account of LCT dimensions, a short section should perhaps be devoted to mutual influence between those who work within the LCT framework and some of those who primarily work within the Sydney School of Systemic Functional Linguistics, given that this collaboration has spawned both books (Christie & Martin, 2007; Christie & Maton, 2011; Martin et al., 2020; Maton et al., 2021) and articles (e.g., Macnaught et al., 2013; Martin, 2013; Matruglio et al., 2013).

If we view LCT as a neo-Bernsteinian framework (Section 3.3.1), the current LCT-SFL collaboration is a continuation of conversations between Basil Bernstein on one side and Michael Halliday and Ruqaiya Hasan on the other side, which started in the 1960s (Maton & Howard, 2020). One reason why some functional linguists turned to knowledge theories of Bernstein was their feeling that while SFL was "good at seeing how meaning is realised in texts", it was "very much weaker at understanding how meaning is distributed in society" (Christie et al., 2007, p. 239). The collaboration unfolded in several phases (Maton & Howard, 2020). In the first phase, the dialogue focused on coding orientations and their social distribution to meaning. The second phase in the 1990s explored pedagogic discourse and genre-based literacy. The third phase in the early 2000s involved the notion of knowledge structures in intellectual fields and the register category of field. In phase four in the late 2000s, LCT's Specialization was drawn on to investigate, for example, field and individuation. The

fifth phase in the 2010s centred on the dimension of Semantics and multiple concepts from SFL's metafunctions and strata.

The last two collaboration phases have been marked by, as Maton and Howard (2020) put it, complementary LCT-SFL analyses of the same data. This complementarity is illustrated by the authors on the example of Semantics: LCT can trace increasing abstraction of knowledge structures in terms of semantic gravity, while SFL can contribute by identifying that abstraction in text is achieved by using, for example, grammatical metaphor or periodicity. LCT and SFL are also said to influence each other's theory development. For example, Martin's use in classroom discourse analyses of the concepts of technicality and grammatical metaphor led Maton to address knowledge complexity by proposing the concept of semantic density, whose use in joint analyses in turn encouraged Martin to propose the concept of mass to address how SFL could conceptualise meaning complexity while encompassing field, mode, and tenor.

According to Maton and Howard (2020), the most common scenario of how studies in education combine LCT and SFL is by enacting LCT concepts to investigate knowledge practices, while SFL concepts explore the linguistic features through which the knowledge practices are realised. The same approach has been taken in part of the methodology of this thesis.

3.5 Application of LCT in research into organising principles of academic fields of practice

Maton (2016a) states that all five dimensions are "simultaneous: They explore not different practices but rather different organising principles that may underlie the same practices" (p. 18). The simultaneity means, he reminds us, that the same data may

be analysed from different perspectives in order to gain complementary insights. At the same time, individual dimensions or even individual concepts from each dimension may be researched alone.

Previous research using LCT has made use of all these options. For example, all five dimensions are used by Arbee et al. (2014) to investigate legitimate participation and achievement in the academic field of marketing. They describe marketing as characterised by a knower code, but the discipline's high positional autonomy (insulation from business contexts) makes it difficult for students to acquire knower dispositions for successful participation in praxis. McNamara (2010) also draws on all the dimensions to explore the structure of the academic field of nursing in Ireland. He finds that the discipline lacks its specific contents and has low positional autonomy (i.e., the discipline is only weakly insulated from external influences).

One dimension, Specialization, is used by Carvalho and Dong (2009) to show that what constitutes legitimate knowledge in four sub-disciplines of the same field of design markedly diverges: Engineering design is characterised by a knowledge code; fashion design emphasises a knower code; architectural designers need to possess both specialist knowledge and be the right kind of knowers; and legitimacy claims in the comparatively young sub-discipline of digital media design have seemingly not coalesced into a unified code. Luckett (2010) refers to epistemic relations and social relations to analyse the recontextualisation of a higher education curriculum in South Africa. She advises against a complete and undivided Africanisation of the curriculum, which would merely substitute one knowledge code for another, and advocates balancing knower codes with knowledge-code disciplines to facilitate competitive participation in a global economy.

Only one concept from Specialization, social relations, is employed by Cheung (2015) to investigate how a postgraduate student of applied linguistics expresses her as well as other authors' voices. The finding that the student's gaze shifts according to her rhetorical purpose indicates that budding academic writers can adapt their voices to employ legitimate forms of argumentation. Georgiou's (2016) study draws on one part of Semantics, semantic gravity, to analyse how first-year undergraduates in physics understand disciplinary concepts. The study shows that satisfactory achievement is not simply a matter of providing correct answers, but that students also need to display an appropriate balance of abstract concepts and concrete examples. Interestingly, those students who favour more abstract concepts tend to be less successful. In a differently focused study, both semantic gravity and semantic density were utilised by Clarence (2016) to explore teaching and learning in conversation with faculty from various sub-disciplines of political science. In the context of two academic development workshops, the author engaged faculty in conversation about the nature of knowledge in political science, the disciplinary and sub-disciplinary curricula, and teaching approaches. Such use of the semantics concepts was reported to have engendered lively productive debate about how semantic waves (given by strengthening and weakening of semantic density and semantic gravity) could help undergraduate students better grasp disciplinary content, make connections between specialist concepts across modules and years of study, and what students should be learning, when, and why.

The example studies mentioned above are methodologically qualitative (interviews of disciplinary insiders or students, discussions, questionnaires, analyses of curriculum documents and assessments) as have been the vast majority of LCT studies to date. This methodological preponderance thus leaves unfulfilled the stated

advantages of topological characterisation of practices (i.e., the possibility of capturing potentially infinite degrees of strengths on conceptual axes) in the Cartesian plane, or of capturing the dominant codes in fields where multiple codes coexist. To realise these advantages, a more quantitative methodology is needed.

3.6 EAP studies that have drawn on LCT

Some EAP studies have brought in LCT (or concepts from Bernstein later subsumed within Maton's LCT) only to explain their findings, while other studies have used LCT as their driving theoretical framework. The former studies could be represented by Cao and Hu's (2014) explanation of disciplinary influences on citation practices in terms of knowledge-knower structures (Section 2.4.2), with the same attribution invoked in Hu and Cao (2015) (Section 2.4.3) to explain differences in the disciplinary use of interactional metadiscourse (especially hedges and boosters). One clear benefit of drawing on a complementary perspective from the sociology of knowledge to interpret the studies' results is enhanced explanatory power. On the other hand, though, the studies used the sociological concepts only in a general way, by referring to general characteristics of (hierarchical/horizontal) knowledge structures and (hierarchical/horizontal) knower structures, and general propensity of knowledge-making practices in the natural sciences, social sciences, and the humanities to be underpinned by these structures, as described by Bernstein (1999) and Maton (e.g., 2007, 2014a). To illustrate, for example, Hu and Cao (2015, p. 20) state that "disciplines operating with a knower code depend more on the distinct individual characteristics of those constructing disciplinary knowledge" and so "claims to knowledge tend to be legitimated by appealing to knowers' personal voice, expertise, experience, and

authority". Based on this general description they conclude that "the higher incidence of boosters in the applied linguistics and education RAs…. seems to be a function of the knower code at work" (p. 20). Similarly, Cao and Hu (2014) conclude that "the stronger knower orientations of applied linguistics and education provide a plausible explanation for the more frequent use of integral citations" (p. 28).

These conclusions suffer from a theoretical and a methodological issue. First, the two studies refer to Bernstein's knowledge and Maton's knower *structures*, from which they jump to *codes*. For example, Cao and Hu (2014) note that "for our practical purpose, hierarchical knowledge-horizontal knower structures, which dominate the natural sciences, can be seen as constituting the two ends of a continuum of modes of specialization characterised, respectively, as the knowledge and knower codes (Maton, 2007)" (p. 28). Hu and Cao (2015) claim that knowledge code and knower code are "referred to as knowledge-knower structures by Maton" (p. 20). These formulations, I believe, show a misunderstanding of the relationship between *structures* and *codes*. The cited passage from Maton (2007), in fact, states that

> If we understand the discursive practices of intellectual fields as *knowledge-knower structures* that specialise actors and discourses in different ways, then the principles underlying these practices can be addressed in terms of their legitimation codes of specialization. This notion is based on the simple idea that actors and discourses are not only positioned in both a structure of knowledge and in a structure of knowers, but also establish different forms of relations to these two structures. One can thereby analytically distinguish between an *epistemic relation* (ER) to the knowledge structure and a *social relation* (SR) to the knower structure (p. 93).

Structures then do not equate *codes*. Put simply, the path from *structures* to *codes* leads via *relations* (ER and SR), because it is the combinations of strengths of ER/SR that determine Specialization *codes*, where *codes* describe the organising principles underlying knowledge-knower *structures*. Moreover, if we decide to interpret results with reference to *codes*, we can progress beyond limited number of *structures:* "the concepts of 'knowledge structures' and 'knower structures' can be left behind when one reaches specialization codes" (Martin et al, 2020, p. 21).

Second, I find it hard to agree with the studies' interpretations of the frequencies of boosters and integral citations as reflecting a knowledge or knower code of the studied disciplines. Cao and Hu (2014) and Hu and Cao (2015) seem to be placing academic fields into code quadrants based on a single indicator (i.e., mostly the attributed strength of SR), which is a shortcut that Maton and Howard (2016) warn against (born out of their research experience). Maton (2014a) clearly states that a position of a discipline in a particular Specialization dimension quadrant is given by the relative strengths of the constitutive relations–epistemic relations and social relations. In other words, each constitutive relation needs to be operationalised separately, so that its strength can be represented on the epistemic relations axis or the social relations axis. "Together, these two locations generate the position in the plane and thus the code. Therefore, to enact the concepts one should begin not with the codes but with the two relations whose relative strengths generate the codes" (Maton & Howard, 2016, p. 57). If, for example, we think of evidentiality usage in terms of social relations, as Hu and Cao (2015) seem to be doing, the absence of any operationalisation of epistemic relations means that their explanation of a knower/knowledge code at work is more of a speculation. As a side note, against Hu and Cao's explanation also somewhat argues the

finding of Hyland and Jiang (2016) that usage of boosters increased between 1965 and 2015 in electrical engineering RAs but decreased in applied linguistics. Through the optics of LCT, with respect to Hyland and Jiang's finding, we could perhaps be talking about strengthening and weakening of social relations (if, for the sake of argument, we admit that evidentiality can reveal something about the strength of social relations), but we certainly cannot position any discipline within a Specialization quadrant, or even speculate, in a similar vein to Hu and Cao, that higher incidence of boosting in electrical engineering in 2015 compared to 1965 means the field is moving towards or has moved into the knower quadrant.

Other types of EAP studies have been driven by LCT. Monbec (2018) conceptually looked at the issue of transfer of linguistic skills that students can learn in EGAP (English for General Academic Purposes) courses to various disciplinary contexts. Drawing on Specialization, she notes that transfer could be thought of as cumulative learning. Then strong epistemic relations (ER+) reflected in, for example, clearly and explicitly articulated curriculum, are likely to create more favourable conditions for such learning than strong social relations (SR+), where bases of achievement are implicit. However, she notes that EAP is a field underpinned by a knower code, with weak epistemic relations (e.g., debates whether EAP is actually an academic field; whether the focus should be general or specific EAP; there are concerns that many EAP textbooks and courses lack theoretical and research bases; knowledge about language can be confined to learning rules) and strong social relations (e.g., EAP has been described as intuitive; the native speaker is still in some parts of the world considered the best candidate for EAP positions). Thus, EAP's organising principles may hinder creating curricula that would assist transfer. Increasing the chances of

transfer thus implies constructing curricula underpinned by explicit functional knowledge of language (e.g., informed by SFL or genre approaches). Drawing on Semantics (semantic gravity), Monbec notes that cumulative learning for transfer is better achieved when knowledge is less dependent on (or tied to) its context. Applied to a curriculum, she recommends that knowledge/skills should first be presented at an abstract level (SG–), then practised (e.g., writing task, peer review, tutor's feedback) within the context of an EAP course (SG+), and then related to disciplinary contexts (e.g., analyses of disciplinary texts) so that students do not associate the knowledge only with the EAP course (SG–).

The issue of transfer is also at the core of a case study by Tann and Scott (2021), who drew on the dimension of Autonomy to investigate integration of EAP literacy skills and business ethics knowledge in an EAP course Management Communication at an Australian university. The course was originally conceptualised as a traditional grammar class (from LCT's Autonomy perspective, students were kept within one code, in which literacy concepts were used purely for literacy purposes), and received lower feedback rating due to the clash between students' expectations and school's perspective, with students not being able to transfer the literacy skills learned because they did not see the relevance of the course. The course was later substantially redesigned regarding pedagogy and delivery, while the content and readings remained almost identical. The redesigned course created more favourable conditions for learning how to apply language skills to a disciplinary context. To illustrate how, the authors analysed a recorded lecture from one of the redesigned course lessons, and showed that at different stages of the lesson, the lecturer moved from one autonomy code to another (autonomy tours; Maton & Howard (2018)). For example, when analysing a text in

order to teach students how to provide readers with an outline of the main issues in the initial paragraph, the lecturer started by eliciting what business keywords best summarise the content of individual sentences/lines (exotic code: business concepts for business purposes), then transitioned to learning about how the keywords in fact represented the main points of the text (introjected code: business concepts for literacy purposes), and finally connected the main points to individual parts of the text to reveal the texts' (genre) structure (sovereign code: literacy concepts for literacy purposes). The lecturer's progression from discussing a text's content to textual structure and development thus enabled students to integrate knowledge about literacy with their subject-area knowledge.

In the studies like those of Monbec and Tann and Scott, while the nature of knowledge-building may remain hidden from the students (specialization, semantics, and autonomy codes are not revealed), students can still benefit from the theoretical underpinning of the courses. But students can equally benefit when the nature of knowledge is made explicit to them. In another case study, this time of a pedagogical intervention, Kirk (2017) successfully prepared anthropology students to write a 2000-word reflective statement for a master's programme module. He drew on the concept of semantic gravity to scaffold the students' understanding of the form of knowledge present in a higher-scoring work (i.e., he revealed the basis of high achievement). By establishing and enacting three levels of context-dependence of meanings (more concrete experiences, more generalised patterns, and more abstract or theoretical content), he was able to trace with the students the sequencing of information (alternating between more context specific and less context dependent), and visualise it as the peaks and troughs of information waves. Explicitly revealing the form of

knowledge to the students then provided the students with the tools to revise their drafts, structure future assignments, and understand disciplinary expectations.

The last three studies summarised above explicitly relate themselves to EAP. If we draw on EAP's definition from Section 2.1.1, we can also include here studies concerned with researching students' work (to the best of my knowledge using most commonly the dimension of Semantics) in specific secondary and tertiary education subjects. One such example is the study by Szenes et al. (2015), who analysed high-achieving writing tasks (a critical reflection essay from social work and a reflective journal from business studies) aimed at eliciting from university students examples of critical thinking. Szenes et al. did not only yet again show the presence of semantic waves (as did Kirk, 2017), but they also found disciplinary differences in the semantic profiles–both waving and staging (ordering). For example, in the social work essay, the stage of drawing lessons from reflection appeared in a slightly different place, and exhibited milder weaving between context-dependent and independent information, while in business the same stage was marked by highly abstract and decontextualised writing.

As is apparent from this short overview, apart from Cao and Hu (2014) and Hu and Cao (2015), a large number of studies (including those summarised above) have focused on how EAP can draw on LCT to improve teaching and learning (e.g., by analysing successful high-achieving writing tasks, by structuring curricula, lessons, pedagogical interventions). The focus has been mostly–and perhaps naturally so–realised via the dimension of Semantics, with the dimension of Specialization used less commonly. Moreover, LCT concepts have been operationalised qualitatively. The predominant focus on teaching and learning leaves little explored whether LCT could

contribute to another large EAP research area, that of investigating disciplinary differences. The predominant methodological focus again leaves untapped LCT's potential for topological characterisation of practices. My study aims to partially fill the existing gaps.

3.7 Specific research questions

In this thesis, I draw on LCT and seek to investigate some of the organising principles underlying knowledge creation in marine chemistry and seismology. Specifically, the aim is to reveal any differences in knowledge-making practices by drawing on the dimension of Specialisation, that is, to investigate how these two social fields of practice establish their claims to being special vis-a-vis each other. I theorise two new realisations of the dimension of Specialisation: Visibility of knowledge and knowers, and Authoritativeness of knowledge and knowers. Visibility should capture how the sub-disciplines front/foreground their specialism, and thus draw clear, visible boundaries around themselves. Authoritativeness probes the link between degrees of conviction and acceptable communication of sub-disciplinary body of knowledge.

The ideas of Visibility and Authoritativeness are obvious candidates for examination, and they have been investigated in EAP studies, as is evident from the review of the selected EAP research areas in Section 2.4. Nevertheless, the reviewed EAP studies focused only on, what could from LCT perspective be considered, aspects of either epistemic relations or social relations. In this work, the insights are expanded by investigating whether the sub-disciplines' specialism is underpinned by emphasising the strength of epistemic relations to knowledge structures, social relations to knower structures, both, or neither.

This work then engages with LCT's Specialization in order to answer the overarching question introduced in Section 1.4, and repeated here for the reader's convenience.

> What are some of the legitimate knowledge-making practices and bases of achievement in the sub-disciplines of marine chemistry and seismology, as evident from the usage of selected linguistic features in published research articles, and how do these practices and bases of achievement differ?

To answer the overarching question, I investigate the sub-disciplinary usages of four selected linguistic features (i.e., Themes, citations, hedges and boosters, and authorial self-mention) in the light of what the usages reveal about legitimate relations to the sub-disciplinary objects of study and sub-disciplinary actors. Specifically, I research the following sub-questions:

1) How visible is disciplinary knowledge in sub-disciplinary research writing, as reflected by the use of topical Themes?
2) How visible are knowers in sub-disciplinary knowledge construction, as reflected in the use of integral and non-integral citations?
3) What is the level of authoritativeness (epistemic certainty) with which propositional content is expressed, as reflected in the use of hedging and boosting?
4) What level of knower authority is legitimate in sub-disciplinary research articles, as reflected in the use of exclusive personal pronouns?

The term *use* in the first question refers to the frequency of occurrence of disciplinarily more or less specific Themes; in the second question, *use* means textual integration of citations and the concomitant visibility of cited sources; in the third

question, *use* refers to the presence of expressions toning down or up the conviction with which propositions are presented; and in the last question, it refers to the frequencies of occurrence of personal pronouns in various authorial roles.

3.8 Summary

In this chapter, I have provided the background to as well as some critical appraisal of the theoretical framework, Legitimation Code Theory (LCT), underpinning this study. The framework aspires to provide the tools for describing the tacit organising principles of practices and for capturing how such practices change through time or in various parts of a text. LCT draws on the notion of struggle over control of the rules of the game in social fields of practice (Bourdieu, 1993) and on the concepts of hierarchical and horizontal knowledge structures (Bernstein, 1999), as well as the notions of classification and framing (Bernstein, 2000). LCT proposes five dimensions that seek to capture the internal organising principles underlying practices and relations in various contexts. Each dimension is characterised by four principal modalities in a Cartesian plane with infinite possibilities for positioning. Thus, the dimensions are said to not only reveal the similarities and differences in languages of legitimation of academic (sub-)disciplines, but also to uncover the bases of achievement. So far, LCT has been predominantly operationalised in most studies qualitatively, leaving little explored LCT's stated potential for topological characterisation. In addition, EAP studies have drawn on LCT mostly to improve teaching and learning, little utilising LCT to suggest insights into the organising principles of academic fields. The methodological and focal gaps thus constitute rationales for this study.

Chapter IV: Methodology

In this chapter, I detail the overall research design and its rationalisation, corpus collection, operationalisation of LCT codes, and data coding and analysis. The quantitative research design enabled me to investigate, in a principled manner, some of the organising principles encapsulated in the dimension of Specialization to uncover how disciplinary insiders in two earth science sub-disciplines legitimate their contribution through selected linguistic features.

4.1 Research design

The methodology unfolded in two steps. First, I constructed a corpus of RAs, one sub-corpus for each of the two earth science sub-disciplines (seismology and marine chemistry); next, I selected four linguistic features that reflected the dimension of Specialization (see Section 4.3 for more information on operationalising the dimensions), and created coding schemes that would allow for quantifying the organising principles and bases of achievement theorised by LCT.

4.2 Collecting corpus data

4.2.1 Constructing the sub-corpora

To answer my research questions, I constructed a specialist corpus comprising two sub-corpora of RAs reporting on empirical studies in seismology and marine chemistry published in a recent five-year period (2012-2017). The corpus construction was originally intended to comprise three successive steps (Alise & Teddlie, 2010): First, choosing two sub-disciplines; second, selecting sub-disciplinary journals; third, sampling RAs from the journals. However, identifying a sufficient number of specialist

sub-disciplinary journals proved challenging, and this step was in the end omitted (for details, see Section 4.2.1.2). In the remaining two steps of corpus construction, multilevel mixed-methods sampling (Teddlie & Yu, 2007) was used: following Alise and Teddlie (2010), in the selection of target academic disciplines, purposive sampling technique was applied; in the RA selection, stratified random sampling was employed. The sampling of RAs was limited to experimental research articles rather than other types that exist (there are at least also theoretical and review articles (Swales, 2004)), since experimental articles predominate in the selected sub-disciplines of earth sciences, which is not always the case in all (sub-)disciplines: for example, in astrophysics, "the subject matter does not lend itself to experimentation" (Tarone et al., 1998, p. 115), and logical argument papers are frequent.

4.2.1.1 Selecting sub-disciplines of earth sciences

The mother discipline, (i.e., earth sciences) and its sub-disciplines were selected through purposive sampling. This sampling technique is non-probability/non-random, and involves selecting units (e.g., people, settings, and institutions) based on specific questions a study is seeking to answer (Teddlie & Yu, 2007). My choice of earth sciences was motivated not only by my immediate teaching context (Section 1.1), but also by the fact that the discipline comprises an intriguing amalgam of sub-disciplines (i.e., specialisations), whose nature is likely different from some related disciplines (e.g., physics, chemistry, or biology), and yet their larger organising principles underlying knowledge building remain unexplored. At the same time, to the best of my knowledge, linguistic research into earth sciences has been rare, particularly with respect to research into one of its chief genres for knowledge creation and

dissemination–the RA, and consists of only several studies conducted by a handful of researchers (Section 2.4). Since earth sciences subsume these numerous sub-disciplines, I decided on two sub-disciplines that could hint at the range of nuances that co-exist within the field. One of the sub-disciplines, marine chemistry (sometimes termed chemical oceanography), belongs to the area of hydrosphere studies, and can be generally described as field- and laboratory-based, traditionally micro focused, and aimed at reconstructing the past and understanding the present. The other sub-discipline, seismology, belongs to the area of solid earth studies (i.e., studies of earth's solid surface and interior), and can be generally described as office-, laboratory-, and field-based, and predictive. I shall describe the two sub-disciplines below in slightly more detail.

According to Aki and Richards (2002), seismology is concerned with our planet's mechanical vibrations, which may be generated naturally by earthquakes or volcanic eruptions as well as artificially by explosions caused by human activity. These vibrations are recorded as waves or wiggles by seismograms, which appeared in the 1880s. Seismology is unique in that its study of seismic waves enables earth scientists to explore the otherwise unattainable depths of the Earth's interior with resolution better than other geophysics sub-fields. That is because seismic waves, while passing through the planet's interior, are less susceptible to deformation than other observed physical phenomena (e.g., heat, gravity, or electromagnetism). The higher accuracy and resolution can then reveal information about the planet's current internal state.

Seismologists do not focus only on the present-day conditions of the Earth's crust and crustal boundaries, but try to link existing fault motions to previous earthquakes in order to understand the evolution of fault systems (Chen et al., 2012).

This knowledge is key for understanding earthquake physics and for estimating seismic hazard (Chen et al., 2012). Essentially, the holy grail of risk assessment in seismology is the ability to forecast where earthquakes will strike and how strong they will be, just like meteorologists forecast tomorrow's weather (Cemen & Yilmaz, 2017). The pressure on seismologists to contribute to seismic hazard mitigation is given by the social, political, and economic consequences of devastating earthquakes (Aki & Richards, 2002).

In addition to fundamental earthquakes physics research and seismic hazard assessment, seismology has been applied to oil, gas, and mineral prospecting, structural engineering to aid in the design of earthquake-resistant buildings, and even in the detection and monitoring of nuclear explosions (Aki & Richards, 2002). Depending on the specific focus, seismology can be observational, computational, analytical, etc., with seismologists working in the field, on laboratory experiments, and on numerical simulations (Chen et al., 2012).

Methodologically, seismology is underpinned by numerical modelling. The models, which relate data recorded by seismograms to conditions in the Earth's interior via parameters, are largely mathematical, with rather simple physics contained in equations about motion (Aki & Richards, 2002). Yet, since the models are essentially simulations, seismological studies ensure that inferences and predictions they generate by modelling progress through a re-iterative cycle of "model formulation verification, simulation-based predictions, validation against observations, and data assimilation to improve the model and reinitiat[ion of] the cycle at a higher level where the model is deficient" (Chen et al., 2012, p. 4). The cycle requires fusion of high-resolution datasets, theories of how mechanical vibrations propagate through the Earth, and models of

seismic source mechanisms as well as of the Earth's interior, all processed by large-capacity computers (Aki & Richards, 2002).

The second sub-discipline is marine chemistry. Marine chemists focus on tracing the journey of chemical elements from their original sources, through the ocean reservoir, to sediment deposition, and on understanding the elements' concentration and behaviour (Chester & Jickells, 2012). Another focus, especially related to sediment records, which are viewed as archives of continental and oceanic history, is to reconstruct past environmental conditions (Futterer, 2006). Marine chemists thus need to understand chemical reactions that take place on time scales ranging from seconds to millions of years (Emerson & Hedges, 2008) to answer questions such as why the sea is salty and how carbon and nitrogen cycle in the environment (Chester & Jickells, 2012).

Despite the size of the oceans, the chemical processes in them are ultimately influenced by tiny microorganisms of less than 1mm in diameter (Chester & Jickells, 2012). "The oceanic chemical system is driven by a physical-chemical-biological trinity" (p. 395), especially by particulate–dissolved matter interactions, against the backdrop of a microcosm of particles scattered everywhere in the water (Chester & Jickells, 2012).

The development of the sub-discipline has been briefly summarised by Emerson and Hedges (2008). During the past several decades, marine chemistry has evolved considerably. From a field focused on discovering what chemical elements and compounds are in seawater and how the chemical constituents interact (i.e., pure marine chemistry), the discipline has morphed into an interdisciplinary field seeking to uncover mechanisms behind the distribution of chemical compounds in the seas and oceans, and the rates at which the distribution is occurring. So while some important research still

comes from pure marine chemistry, the field is at present largely oriented to studying the impact of chemical processes on whole oceans and the global climate. The evolution of marine chemistry has been greatly facilitated by developments in analytical techniques and instruments that enable sophisticated sampling and detection of even extremely small concentrations of chemical elements. Despite the technological progress, marine chemistry is still very much field based, involving the use of vessels (and diving) in sample collection, because long-term remote measurement by instruments installed in the field is still impeded by lower instrument stability and accuracy.

Based on the above-mentioned research histories, foci, and methods, the sub-disciplines of seismology and marine chemistry can be expected to differ from each other in terms of their organising structures and bases of legitimate contribution.

4.2.1.2 Selecting academic journals

As the second step of building my corpus, I originally intended to select sub-disciplinary journals through quota sampling and purposive sampling. Quota sampling was to ensure equal representation (Gravetter & Forzano, 2012), and given the scope of this study, four to five journals per sub-discipline were deemed sufficient. In purposive outlier sampling, I intended to choose from the most prestigious journals in the sub-fields in order to collect instances of model expert communication (Alise & Teddlie, 2010). To facilitate my selection, I referred to Thomson Reuters 2016 edition of the *Journal Citation Reports*, to select the journals with the highest five-year impact factors–rough indications of the relative positions of the journals in their respective fields.

After shortlisting the target journals based on their impact factors as well as information about their aims and scopes, and consulting specialist informants selected through convenience volunteer sampling (Teddlie & Yu, 2007), I realised that individual earth science journals tended to publish on relatively broad ranges of topics, with earth scientists also preferring to publish in more generalist journals to increase the potential reach of their articles. I thus decided to source directly for sub-disciplinary RAs on the Web of Science (n.d.). The advantage of such sourcing was that the RAs came from multiple journals, unrestricted by quota sampling.

4.2.1.3 Selecting RAs

Published RAs are important because they have achieved a legitimate status in the eyes of the specialist community (Hyland, 2004). My search on the Web of Science (n.d.) started by deciding the fixed filters: years of publication (2012-2017), language (English), and publication type (RA). For each of the sub-disciplines, I consulted disciplinary informants (all Assistant or Associate Professors at the Earth Observatory of Singapore or the Asian School of the Environment at my university) for keywords that would characterise seismology and marine chemistry. The recommended keywords for seismology were *earthquake source* and *velocity structure*; for marine chemistry, they were *cycling* and *trace metals*. Next, applying the fixed filters, I searched for combinations of the sub-disciplinary names and the keywords (Table 4.1).

Table 4.1

Combinations of disciplinary names and keywords in the Web of Science

Seismolog* + earthquake source + velocity structure	Marine chemistry + cycl* + trace metal
Seismolog* + earthquake source	Marine chemistry + cycl*
Seismolog* + velocity structure	Marine chemistry + trace metal

The combinations that yielded the largest numbers of results (*seismolog** + *earthquake source* and *marine chemistry* + *cycl**) were further refined by the Web of Science Categories in order to exclude irrelevant categories (e.g., *engineering electrical electronic* or *meteorology atmospheric sciences* for seismology, and *paleontology* or *fisheries* for marine chemistry). At the same time, the search results were still kept sufficiently broad (for seismology the categories comprised *geochemistry geophysics*, *mechanics*, *geosciences multidisciplinary*, *physics multidisciplinary*, *engineering geological*, *geology*, *physics mathematical*, and *geography physical*; for marine chemistry the categories were *geosciences multidisciplinary*, *geochemistry geophysics*, *chemistry inorganic nuclear*, *oceanography*, *chemistry multidisciplinary*, *chemistry organic*, and *chemistry analytical*). Next, I carefully perused all the RA abstracts yielded by the refined search, and identified a pool of 279 RAs that could be classified as empirical studies in seismology (called observational seismology, as opposed to theoretical and methodological papers, called theoretical seismology) and 149 RAs that could be classified as empirical studies in marine chemistry. Obviously, the 279 and 149 RAs did not represent the total number of RAs published in seismology and marine chemistry between 2012 and 2017. Nevertheless, the RAs were easily identifiable

through an automated search as belonging to the sub-disciplines. Next, to each pool, I applied a random sampling approach, in which any RA had the same chance of being selected, using random numbers generated by a computer programme (Teddlie & Yu, 2007). The generated lists of random RAs (titles and abstracts) were then vetted by two informants (Assistant or Associate Professors at the Earth Observatory of Singapore or the Asian School of the Environment at my university) per sub-discipline, and only RAs agreed on by both informants as being empirical and from their sub-discipline were further considered. The resulting corpus thus comprised 60 seismology RAs from 29 journals and 60 marine chemistry RAs from 20 journals. Most of the journals have a broad scope, beyond seismology and marine chemistry, with three journals (Earth and Planetary Science Letters, Geophysical Research Letters, and Journal of Geophysical Research series) being sources of RAs in both sub-corpora.[34] The resulting corpus structure confirms my informants' mention that earth scientists prefer to publish in journals with broader scopes to reach larger audiences. In fact, my informants' mention corresponds to Hyland and Jiang's (2019) assertion of the trend in academic writing towards reaching audiences beyond specialised communities.

As the last step, texts of the discussion and conclusion sections (for the rationale of why only these sections, see Section 4.3.2) of the randomly selected and vetted RAs were extracted, and following Hu and Cao (2015), figures, tables, and footnotes/ endnotes were removed. Additionally, titles, section and sub-section headings, paragraph numbers, equations, acknowledgements, data and resources (acknowledging uses of various commercial software, catalogues and tools post conclusions), and appendices were removed.

[34] The complete name list of the journals together with the numbers of corpus RAs in them can be found in Appendix I.

4.3 Relating the concepts of LCT's Specialization to selected linguistic features

To translate between internal concepts (classification and framing) and empirical referents, Bernstein (2000) uses the term "external language of description" (p. 132), which allows the internal concepts to be "activated as a reading device" (p. 133). Building on Bernstein, Maton (2016b) further elaborates, pointing out that LCT distinguishes three types of translation devices: external languages of description for translating between theory or concepts and data within a particular study; external languages of enactment for translating between theory or concepts and practices; and mediating languages for translating between theory or concepts and data not limited to a specific study but "embracing all empirical forms of a phenomenon" (p. 241). In the present study, the translation devices used were external languages of description.

One obvious way of relating concepts to empirical data is via linguistic realisations present in a text. Such a translation, or operationalisation in common parlance, is in line with the conception of language mirroring thought. As Vergaro (2015) points out:

> Though conceptualization does not lend itself to direct observation, it can be studied indirectly via language, due to the close relationship between linguistic and conceptual structure, i.e., linguistic meaning is equated with conceptualization, and linguistic structures are viewed as conventionalized conceptual units. (p. 3)

While thought is not identical with language, thought is mainly represented linguistically, since no other representational system proposed is universally accepted (Beaugrande & Dressler, 1981), and at the same time offers the possibility of recruiting

a finite system to produce an infinite number of expressions, with each expression being definitely interpretable in thought and sound (Berwick & Chomsky, 2016, p. 1). Therefore, I believe it is appropriate to represent the Specialization codes of marine chemistry and seismology through language expressions.

4.3.1 Specialization involved in disciplinary practices

Before describing how I investigated Specialization codes of the two sub-disciplines, I should perhaps briefly attend to the general notion of knowledge specialization in disciplines.

One of the functions of language is to serve as an instrument for describing our experience with the outer and inner world (Butt et al., 2012). But to organise this experience systematically in a scientific theory means that we have to develop a dedicated semiotic system (Halliday & Martin, 1996). As Halliday and Martin (1996) state, early scholars took this semiotic evolution very seriously. They tried to develop language that would not only precisely categorise existing knowledge in technical taxonomies, but also a language that would become the tool for scientific enquiry and the means of generating further knowledge. The evolution brought several lexical and grammatical innovations. Perhaps the most significant ones involved nominalizations (verbs were turned into adjectives and nouns) and extended nominal groups (clauses and prepositional phrases were embedded in them). Thus, the evolution of science spurred the evolution of scientific language: “The birth of science, from the union of technology with mathematics, is realised semiotically by the birth of grammatical metaphor from the union of nominalization with recursive modification of the nominal group” (Halliday & Martin, 1996, p. 15).

According to Halliday and Martin, once verbs became nominalised the words became technical. Phenomena were construed as if they were things: They were transformed and encapsulated into abstract objects.[35] When this language shift happened on a massive scale, it reconstructed the nature of our whole experience of the world (Halliday & Martin, 1996). Thus, when from the 17th century science became increasingly mathematised, it distanced its knowledge and language from Bernstein's horizontal discourse representing everyday language and common-sense understanding (Maton, 2007). This distancing and complexity of technical terms corresponded to the complexity of highly abstract conceptual structures, bringing with it the notion that learning science necessitated learning the language of science (Halliday, 2007). Learning a specialised language and body of knowledge then became the basis of specialization in science (Maton, 2007). Indeed, the status and prestige of aspiring academic fields is to an extent dependent on inaccessibility to the outsiders of the knowledge they produce (Hashem, 2007).

Explicating how knowledge is constructed, Martin (2007) gives an example from a geology textbook, showing how a description of a specialised activity sequence (how volcanoes are formed) introduces the disciplinary taxonomy, in which the technical words (lava, magma, volcano) function as participants (in an SFL sense; in SFL, participants are usually nominal groups (Thompson, 2013)). Martin (2007) considers geology to be an example of a hierarchical knowledge discourse. Translated into Maton's Specialization, such a discourse can be inferred as representing a knowledge code (ER+/SR–), which emphasises specialised knowledge of principles and procedures, and downplays the role of disciplinary actors (Maton, 2014a). A similar

[35] In fact, generalised or abstract concepts would be difficult to relay without nominalizations (Hinkel, 2004).

conclusion that geology represents a knowledge code can be inferred after reading Dressen and Swales (2000), who posit that in the part-genre of field accounts, experienced geologists prefer to let the phenomena under study tell their own story. The two last-mentioned studies are among the very few extant studies in applied linguistics investigating how earth scientists represent disciplinary knowledge, but as far as I am aware, no study has looked at earth sciences through the lens of LCT. In this study, I investigate knowledge-making practices in seismology and marine chemistry by operationalising the LCT dimension of Specialization in two different ways: one reflecting the Authoritativeness of knowledge and knowers; the other revealing the Visibility of disciplinary knowledge and knowers. I opted for double operationalisation of the same dimension of knowledge practices to allow for broader intra- and inter-disciplinary comparisons of aspects of legitimate practices.

4.3.2 The first step: Identifying knowledge claims

I argue that discussion and conclusion sections (where separate from discussion[36]) are RA part-genres suitable for investigating aspects of sub-disciplinary differences in what knowledge and knower practices are legitimate. Firstly, part-genres themselves can be considered sufficiently representative samples. As Swales (2004) notes while briefly reviewing studies of the historical development of the RA, individual sections of the research article (mainly introductions, methods, results, discussions, and conclusions) have effectively evolved into part-genres with distinct stylistic and rhetorical features. Secondly, discussions foreground research results, with the work of others serving as a means of support, comparison, or contradiction (Swales,

[36] In their paper on the changing structure of the empirical RA, Lin and Evans (2012) find that separate conclusions appear in almost 75% of the published articles in their corpus of 24 disciplines + 15 fields.

2004), include claims about how the reported research advances the field (Lin & Evans, 2012), all with the aim of persuading readers of the importance of the study by focusing on the implications of the results (Blackwell & Martin, 2011). In conclusions, new knowledge claims are summarised, and implications drawn. Discussion and conclusion sections are thus part-genres where readers can expect a high concentration of knowledge claims. Additionally, as Hu and Cao (2015) note, these RA sections likely constitute fertile ground for studying features of interactional metadiscourse (Hyland, 2005b), some of whose elements (hedging, boosting, and self-mention) are operationalised in this thesis as a means of revealing knowledge and knower practices in marine chemistry and seismology.

However, agreeing on how to define knowledge claims has proved elusive (Dahl, 2013), with differing conceptualisations put forward (e.g., Bloor & Bloor, 1993; Dahl, 2008; Myers, 1992). The absence of clear definitions poses a considerable challenge to the linguistic study of knowledge claims, making the identification of knowledge claims in RAs rather difficult. The absence of definitional consensus is further compounded by the fact that exactly what counts as a knowledge claim requires expert-level familiarity with the disciplines in question. As an applied linguist, I do not possess this kind of disciplinary background in seismology or marine chemistry, and indeed my initial attempts to identify knowledge claims in a reliable manner proved extremely challenging. As a result, I made the decision in this work to equate knowledge claims with propositions (statements or questions), as defined by SFL, at the level of ranking clauses, excluding modalizers. A detailed explanation is provided in the paragraphs below.

According to Halliday and Matthiessen (2014), clause as exchange fulfils two basic functions: giving and demanding. An equally fundamental distinction is what is being exchanged, that is, whether goods and services or information. When information is at the centre of exchange, clauses are either in the forms of statements (to give information) or questions (to demand information). Such clauses are propositions, which can be "affirmed or denied, and also doubted, contradicted, insisted on, accepted with reservation, qualified, tempered, regretted, and so on" (p. 138). Propositions then exclude exchanges of goods and services, which are expressed through offers and commands. Offers are not found in research articles, while commands appear as directives, but infrequently (e.g., Hyland, 2005). Directives were excluded from my data analysis.

The unit of my coding identifying a proposition was a ranking clause (i.e., the highest level in the scale of rank in the grammar of English, above phrase/group, word, and morpheme; Halliday and Matthiessen, 2014, p. 9). It is at the rank scale level of the clause that language allows us to communicate ideas and their connections in more detail and with more nuance compared to the ranks of group or word (Bloor & Bloor, 2013). In contrast, rankshifted clauses (i.e., clauses that have been downgraded, and function at the level of a phrase/group), as Halliday and Matthiessen (2014) state, are presented as presumed rather than being open for negotiation. Rankshifted clauses which are non-finite (non-finite clauses do not include the Finite element[37], that is, a marker of either the present/past/future time, such as auxiliary *be*/*have*/*do*, or a marker of modality–likely/unlikely–, such as modal *could*/*might*/*should*) "are even further

[37] Thompson (2013) states that the Finite can be identified by adding a tag question to a declarative clause. Thus, for example, in the two ranking clauses from Seis 18 *Therefore, they should be interpreted as a compressional tip line of the E–W fault, as attested by the focal mechanism of the largest shock.*, the Finite in the first clause is *should*, while the second clause in non-finite.

removed from the status of negotiability than finite ones" (p. 172). They are not anchored in the here and now, and hence are not challengeable. Additionally, non-finite clauses are mostly without grammatical Subjects, which are presupposed/implicit, but may elude identification. For example, in *We found a significant effect of temperature on the partitioning of Mg and Ba into otoliths for juvenile rockfishes, providing the information needed to interpret otolith elemental ratios in a coastal upwelling region.* (MChem 4), is it the authors *(we)* who provide information, or is it the *significant effect of temperature*? According to Halliday and Matthiessen (2014), "the question is really irrelevant; it is precisely the function of the non-finite to make it unnecessary to decide" (p. 469). Thus, non-finite ranking clauses were excluded.

How was a clause defined? Following Halliday and Matthiessen (2014), in this thesis, the term clause refers to a grammatical unit, whereas the term sentence refers to orthography (p. 8). A clause consists of one or more groups, where a verbal group (realising Process) is obligatory (Butt et al., 2012). Thus, the following orthographic sentence from my data includes two clauses/propositions (verbal groups underlined):

1) *The increased abundance of active suspension feeders (relative to deposit feeders) <u>is</u> also likely <u>to have stimulated</u> microbial denitrification and anammox, through an increase in the provision of [NO2–N] and [NO3–N]*
2) *as water <u>is</u> actively <u>moved</u> into the sediment during feeding (Howe et al. 2004; Dang et al. 2010).* (MChem 1)

Clauses containing mathematical notations (e.g., =, </>) instead of verbal groups were excluded from analysis.

When clauses were of the same type (i.e., dependent or independent), shared the grammatical Subject, were joined by a coordinating conjunction (i.e., *and*, *but*, *or*), and the subsequent clause(s) in such a nexus had their Subjects and/or parts of their verbal groups (e.g., the modal Finites and/or the auxiliary *be*) ellipted, the ellipted

elements were then presupposed (Butt, et al., 2012; Halliday & Matthiessen, 2014). For example, in

1) ... *we regard the HF, LP and VLP to have systematically different source positions*

2) *and wish to examine their significance in the context of the location procedures and associated uncertainties.* (Seis 54)

the Subject *we* is ellipted from but presupposed in the second clause. Where the whole verbal group was ellipted, the text was not considered a separate clause/ proposition (underlined in the following example):

> *The HF earthquakes may result from either brittle failure of higher strength lava flow materials (Heap et al., 2015), or more likely the rapid failure of critically stressed rock in the hydrothermal conduit.* (Seis 54)

Some orthographic sentences in my data start with what I shall term a modalizer–a name derived from Halliday and Matthiessen (2014), who distinguish modality types into modalization of propositions (i.e., statements or questions) and modulation of proposals (i.e., commands or offers). A modalizer is a clause, often a projecting or commenting clause, which explicitly signals the source (subjective or objective) of an epistemic attitude to the subsequent proposition. For example, a modalizing clause often includes a verb/expression limiting a writer's commitment, like a hedge (e.g., the subjective *We hypothesize that…* or the objective *Our results suggest that…*), or a verb/expression communicating a writer's conviction, like a booster (e.g., *We find that…*, *Our results demonstrate that…*). In my data, modalizing clauses, which in Systemic Functional Grammar concern probability and usuality, also encompass epistemically neutral verbs/expressions/attitude markers (e.g., *This means that …*, *It is important that…, It is surprising that…, We agree that…, We prefer that…*). A modalizer "frames" (Thompson, 2013, p. 247) the proposition it modalizes, and since it does not include disciplinary content in its own right, it was not considered as a proposition/knowledge claim. As Thompson (2013) puts it, modal meanings expressed in separate clauses stand outside the propositional meanings of the clauses that are modified. Further, also drawing on Halliday and Matthiessen (2014), I broadly characterised modalizing clauses as metaphorical realisations of probability or usuality, that is, I considered them

interpersonal grammatical metaphors. As interpersonal grammatical metaphors, the clauses function at the same time as projecting clauses as well as modal Adjuncts (i.e., *we believe –> possibly, it is likely –> probably).* The broad characterisation applies since "it is not always possible to say exactly what is and what is not a metaphorical representation of a modality" (p.689), with language providing an infinite number of realisations of expressing opinions, from subjective clauses (*we think/speculate/believe/suggest, etc.)*, objective clauses (*it is likely/probable/ suggested, etc.*), to prepositional phrases (*in my opinion, in all probability*), and modal Adjuncts (Halliday & Matthiessen, 2014).

Admittedly, though, at times, a modalizer may not be distinguished from a non-modalizer unambiguously. For the sake of clarity and systematisation, in this work, a modalizer was defined as a clause with a *lexical verb* preceding a *that clause* (e.g., *we suggest that*), or with *be* preceding a *that clause* (e.g., *an alternative explanation is that*), or with *be + adjective* preceding a *that clause* (e.g., *it is clear that*), or a combination of these (e.g., *it is not unreasonable to suggest that)*. Some modalizing clauses (underlined in the following example from MChem 21) may also be followed by clauses in which *that* has been omitted (e.g., *Extreme isotopic variation [[preserved in the Gaoyuzhuang Formation]] suggests strata were deposited in close proximity to a regional marine chemocline…* . (The double brackets [[ ]] mark a rankshifted/embedded clause.) This definition of a modalizer includes both explicit subjective (e.g., *we suggest that*) and explicit objective (e.g., *it is clear that*) modalization (Thompson, 2013), and encompasses the majority of cases of clauses that introduce disciplinary content in my corpus.

Excluding modalizers allowed analysis of disciplinary knowledge claims per se. For example, in the orthographic sentence below, the modalizing statements (underlined) were not part of the ranking clause analysis.

Although we may conclude

- *that macro-faunal bioturbation and bioirrigation activities had a stimulatory effect on microbial denitrification,*

it remains unclear

- *why we did not detect any change in archaeal or bacterial nitrifier (amoA gene) abundance in either sand or sandy mud sediments.* (MChem 1)

Had the sentences been parsed following Halliday and Matthiessen (2014), the disciplinary content would have been rankshifted, leaving only the modalizers in place of propositions.

An exception to the above rule of excluding modalizers from being propositions were cases of *be* + *adjective* (underlined) preceding non-finite clauses, such as:

> *In sites [[where carbonate saturation at the site of deposition is high]], it is probably reasonable [[to attribute SNW changes to surface conditions (as in Barker and Elderfield, 2002; Marshall et al., 2013; Weinkauf et al., 2013; Osborne et al., 2016, etc.)]].* (MChem 3)
>
> *In between these end-member scenarios, it may be difficult [[to untangle competing effects (Bijma et al., 2002)*, (MChem 3)

The clause complexes were analysed in this way because if the *be* + *adjective* structures had been excluded, the following non-finite clauses would not have been analysed either. Thus, the complete clauses would have been excluded from my data, even though it is clear that they express propositions relating to disciplinary content.

The above-described identification of propositions as knowledge claims suffers from certain limitations. While basing my analysis only on ranking clauses and separating of modalizers allowed me to identify disciplinary knowledge claims introduced by verbs (e.g., *we suggest that*…), an obvious limitation that comes to mind is exclusion of knowledge claims introduced by stance nouns (e.g., *the fact that, the hypothesis that*). *Stance noun* + *that* structures, as a type of *Noun Complement* structure, occur not infrequently across disciplines (Jiang & Hyland, 2015). Nevertheless, excluding *stance nouns* + *that* knowledge claims in this study avoided several potential problems:

1. Complex embedding of propositions/knowledge claims (in the example below, the triple brackets [[[ ]]] mark a rankshifted/embedded clause complex, the double brackets [[ ]] mark a rankshifted/embedded clause, and the double vertical lines || mark clause boundaries) may lead to a potentially more subjective (i.e., less replicable) analysis of what counts as a knowledge claim or how many claims are present in, for example,

   Macroseismic analysis of original reports [[coming from the epicentral zone]] provides arguments [[[to favour the hypothesis [[that the main shock is not

located near the epicentre of the 2005 Vallorcine earthquake]] || and thus that the 2005 and 1905 earthquakes did not occur along the same structure]]]. (Seis 46);

2. Complex relationships between knowledge claims within a single clause: For example, in *This is consistent with findings [[that microbial remineralization processes produce mainly low-molecular weight DOM <3kDa (Burdige and Gardner, 1998)*]] (MChem 8), where *this is consistent with* is a claim about the validity of an immediately preceding claim, while *finding* introduces a new knowledge claim; and
3. Conflict in the Authoritativeness (epistemic commitment) of verbs (Section 4.3.5) and stance nouns, for example, between *supported* and *hypothesis* in *The model comparison supported the hypothesis [[that the coupling of nitrification to denitrification would be higher at the Oligo site than the Poly site]]*. (MChem 30).

The coding sheet used to identify ranking clauses forms Appendix II.

In the next four sections, I describe how I used LCT's dimension of Specialization to investigate two areas of legitimate knowledge-building practices in marine chemistry and seismology which I conceptualise and call Visibility of knowledge and knowers and Authoritativeness of knowledge and knowers. To translate between linguistic data and epistemic relations and social relations of Specialization, I drew on topical Themes and citations for Visibility and on hedging and boosting and personal pronouns (authorial self-mention) for Authoritativeness.

4.3.3 Visibility of knowledge: Topical Themes

Scientific disciplines are said to specialise by *what* and *how* knowledge is expressed, with the expression involving meanings and language distinct from everyday

discourse (Maton, 2014a). Capturing the distinct language and meanings through a principled classification can then reveal the form that disciplinary knowledge assumes.

In this thesis, I propose that one such classification of disciplinary knowledge is by its visibility (prominence). Visibility of disciplinary knowledge is taken to represent the strength of epistemic relations. If epistemic relations reflect focus on specialised objects of study and procedures for studying such objects, then the more prominent these specialised objects and procedures are in disciplinary writing, the stronger epistemic relations are. Visibility in clauses can be represented by the types of topical Themes. That is, in (sub-)disciplines with strong epistemic relations, we can expect (sub-)disciplinarily-specific topical Themes to dominate.

Before defining topical Themes for the purpose of this thesis, I shall briefly explain the concept of a Theme. In Systemic Functional Linguistics, Halliday and Matthiessen (2014) define the Theme of a clause as the first group or phrase that has some experiential function (i.e., in transitivity analysis, it represents a participant, a circumstance, or a process), and that at the same time does not extend past this first element. This definition of the Theme does not remain uncontested, with alternative definitions including the Subject (for a brief overview, see Forey and Sampson, 2017). In the case of a clause complex, there are two possible thematic domains: One encapsulates the entire clause complex, while the other domain is at the level of each individual clause, whether the clause is independent or not.[38]

Themes can be broadly divided into three types. The Theme is either textual (connection of a message to the previous text), interpersonal (interactions between

[38] While the clause nexus domain would be more suitable for analysing textual progression, the clausal domain is what I have selected in this work, because the clausal domain allows me to expand the number of Themes realising the Visibility of knowledge and knower structures in seismology and marine chemistry.

interlocutors or the attitude/commitment they are taking), or topical (experiential element) (Halliday & Matthiessen, 2014). The textual Theme is expressed by a conjunction, conjunctive Adjunct, or continuative. The interpersonal Theme is either a vocative, modal Adjunct, or Finite (Thompson, 2013). The topical Theme is expressed as a Subject (unmarked Theme), Predicator, Complement, or a circumstantial Adjunct (marked Theme) (Bloor & Bloor, 2013). The topical Theme is obligatory, with every clause containing only one topical Theme; the other two types of Theme are optional (Halliday & Matthiessen, 2014). However, when more than one type of Theme occurs, with the topical Theme being always the last in the sequence, a clause is said to have a multiple Theme (Thompson, 2013).

In this thesis, I follow Halliday and Matthiessen (2014) and define the topical Theme as the first element that plays a role in transitivity. In other words, the topical Theme "ends with the first constituent that is either participant, circumstance or process" (p. 105). In the examples below, I provide examples from MChem 31 of the three constituents (underlined).

- *The surface DOC concentration around Antarctica (poleward of 50◦S) is in the range of 50 µmol/kg* (participant)
- *In the model, peak HgT concentrations occur between 1000 m and 1500 m.* (circumstance)
- *... see Hansell et al., 2009, Figure 1* (process – As mentioned in Section 4.3.2, commands are not propositions, and were thus excluded.)

Other types of Themes (textual and interpersonal) do not reflect Visibility of disciplinary knowledge well. Textual Themes comprise largely conjunctive Adjuncts, while interpersonal Themes comprise mostly modal Adjuncts (Table 4.2 below).

Some other linguists investigating Themes have preferred to focus on the grammatical Subject. For example, for MacDonald (1992), the Subject best expresses agency that reflects the rhetorical situation: the syntactic sense of the Subject (who does what or what a sentence is about) reveals the semantic sense (knowledge-making practices in a discipline through variously classified nouns). However, since what occupies the first position in a sentence is significant (e.g., Gosden, 1993) and draws the reader's attention–in other words, the first experiential[39] clause element is in the position of high visibility–even non-Subject Themes (marked Themes) can reveal disciplinary knowledge building.

As mentioned above, the unit of coding in this work is the ranking clause. In some studies (e.g., Ebrahimi & Khedri, 2011; Ghadessy, 1999; North, 2005), the T-unit (Hunt's (1965) Minimal Terminable Unit consisting of one main clause and all subordinate clauses attached thereto) was selected in order to analyse thematic progression. Thematic progression analysis was not a focus here, plus a T-unit would not have allowed me to consider Themes in independent ranking clauses. Also, I focused only on the topical Theme, since it is the experiential element that captures and represents the disciplinary content.

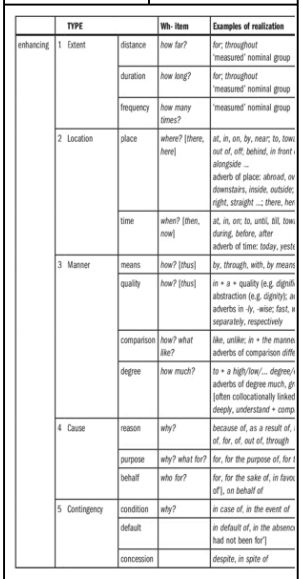
To identify topical Themes, I first created a rubric with multiple examples pooled from the literature as a guide in distinguishing topical from textual and interpersonal Themes.

[39] In Systemic Functional Grammar (SFG), a clause simultaneously fulfils three metafunctions: the ideational metafunction represents the world around us and our experience in the world (it is divided into experiential meanings, which describe our experience(s), and logical meanings, which describe relationships between/among them); the interpersonal metafunction describes the personal and social relationships enacted by clauses; and the textual metafunction describes how texts are coherently constructed/sequenced/organised (Halliday & Matthiessen, 2014).

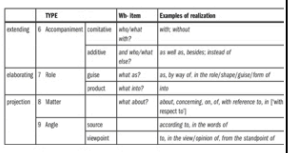
Table 4.2

A rubric with types of Themes and their example realisations (based on Bloor & Bloor, 2013; Halliday & Matthiessen, 2014; Thompson, 2013)

Textual Theme		**Interpersonal Theme**		**Topical Theme**	
Connection of a message to the previous text		Interactions between interlocutors or the positions they are taking		Experiences in the world (people/things, events, circumstances)	
Continuatives Note: Inherently thematic	umm, yes, right, well, no, oh, now…	Vocatives Note: Characteristically thematic	Henry!, Sir! My God! …	Subject Complement	usually nominal group or prepositional phrase
Conjunctions (structural Themes) Note: Inherently thematic since they always appear at the beginning, and locate the clause in a specific logico-semantic link to another clause.	and, or, but, so, yet, then, for, either, neither, nor, when, while, before, after, until, because, whether, although, since, that, to, by, with, supposing (that), assuming (that), given that, provided (that), so that to, as to, in spite of the fact that, well, anyway …	Modal Adjuncts Note: Characteristically thematic	probably, possibly, certainly, perhaps, maybe, usually, always, sometimes, seldom, often, generally, in general, broadly speaking, on the whole, in principle, regularly, for the most part, of course, surely, obviously, clearly, in my opinion, personally, no doubt, evidently, presumably, apparently, tentatively, provisionally, frankly, surprisingly, understandably, as expected, by chance, happily, surely, unfortunately, tragically, honestly, oddly, undeniably…	Predicator	verbal group
				Circumstantial Adjunct Note: Circumstantial Adjuncts can be predicated Themes; conjunctive Adjuncts cannot.	
Conjunctive (discourse) Adjuncts Note: Characteristically thematic as they frequently appear in thematic positions but may occur elsewhere too.	actually, also, alternatively, as a matter of fact, as a result, as far as X is concerned, at least, before that, besides, briefly, conversely, despite that, finally, for example, for instance, for this reason, however, in addition, in any case, in conclusion, in fact, instead, in other words, in the same way, in that case, in this respect, later on, leaving that aside, likewise, meanwhile, moreover, nevertheless, next, on the other hand, or rather, otherwise, soon, that is, therefore, thus, to be precise, to sum up, under the circumstances, with this in mind …	Finite elements Note: Characteristically thematic	am, is, are, was, were, do, does, did, have, has, had, shall, will, can, could, may, might, should, must, would…		
Wh-relatives (½ text., ½ top.) Note: Characteristically thematic	which, what, who, whose, where… + that	Wh-question words (½ int., ½ top.) Note: Characteristically thematic	who, what, why, where, when, how		



	TYPE		Wh- item	Examples of realization
enhancing	1 Extent	distance	*how far?*	*for; throughout* 'measured' nominal group
		duration	*how long?*	*for; throughout* 'measured' nominal group
		frequency	*how many times?*	'measured' nominal group
	2 Location	place	*where? [there, here]*	*at, in, on, by, near; to, towa* *out of, off; behind, in front* *alongside …* adverb of place: *abroad, ov* *downstairs, inside, outside;* *right, straight …; there, her*
		time	*when? [then, now]*	*at, in, on; to, until, till, towa* *during, before, after* adverb of time: *today, yest*
	3 Manner	means	*how? [thus]*	*by, through, with, by means*
		quality	*how? [thus]*	*in + a + quality (e.g. dignit* *abstraction (e.g. dignity); a* adverbs in *-ly, -wise, fast, w* *separately, respectively*
		comparison	*how? what like?*	*like, unlike; in + the manne* adverbs of comparison *diffe*
		degree	*how much?*	*to + a high/low/… degree/* adverbs of degree *much, gr* [often collocationally linked *deeply, understand + comp*
	4 Cause	reason	*why?*	*because of, as a result of, .* *of, for, of, out of, through*
		purpose	*why? what for?*	*for, for the purpose of, for t*
		behalf	*who for?*	*for, for the sake of, in favo* *of'], on behalf of*
	5 Contingency	condition	*why?*	*in case of, in the event of*
		default		*in default of, in the absenc* *had not been for']
		concession		*despite, in spite of*



	TYPE		Wh- item	Examples of realization
extending	6 Accompaniment	comitative	*who/what with?*	*with; without*
		additive	*and who/what else?*	*as well as, besides; instead of*
elaborating	7 Role	guise	*what as?*	*as, by way of, in the role/shape/guise/form of*
		product	*what into?*	*into*
projection	8 Matter		*what about?*	*about, concerning, on, of, with reference to, in ['with respect to']*
	9 Angle	source		*according to, in the words of*
		viewpoint		*to, in the view/opinion of, from the standpoint of*

Next, when identifying topical Themes, I also had to account for the Mood system. While most clauses in my corpora are declarative, there is the occasional interrogative and imperative. There are also special Thematic structures, namely Thematic equatives, predicated Themes, and Thematised comments (Table 4.3).

Table 4.3

Theme constituents in the three Mood systems (based on Bloor & Bloor, 2013) and special Thematic structures (based on Thompson, 2013)

Mood	Theme
Declarative	Participant/Subject (unmarked) *A more modest change in seawater Mg implies smaller corrections for Neogene foraminiferal Mg/Ca records.* (MChem 59) Circumstance/Adjunct (marked) *At mid-term, the consistent high productivity (low values of the %Fp index, Fig. 9) from MIS 12 to MIS 9 indicates high nutrient supply,…* (MChem 58)
Interrogative (yes/no)	Finite (interpersonal Theme) + Subject (topical Theme) *Can we consider the Mw 5.2 Lorca earthquake as the characteristic earthquake for this area?* (Seis 10)
Interrogative (wh-)	Question word (wh-word) *How can coccoliths in shallow sediments become enriched in heavy Zn instead?* (MChem 20)
Imperative	Process/Predicator *(see for example the 2001 left-lateral Skyros event; Benetatos et al., 2002, Ganas et al., 2005; the 2013 right-lateral Lemnos event, Ganas et al., 2014b)* (Seis 2)
Special Thematic structures	

Thematic equative	Theme = Rheme *[[What is more ambiguous]] is [[whether the exhumation of the lower striated part of the scarp and the more recent white strip at its base are due to seismic activity]].* (Seis 46)
Predicated Theme	A cleft sentence (*it +be + nominal/adverbial group*) that highlights a constituent, for example, to be contrasted with another constituent. *However, it is the presence of Mn–Fe oxides on sediments throughout the subterranean estuary that resulted in increased partitioning between the adsorbed and dissolved phase.* (MChem 49) *It is here that the Tonga-Samoa earthquake appeared to rupture across the trench axis, ...* (Seis 12)
Thematised comment	A comment on, evaluation of, or attitude towards a proposition *It is interesting that the aftershocks are concentrated along the edges of the presumptive rupture plane.* (Seis 51) *It is unfortunate that the compilation lacks d34SSPS data for sites in the sulfate concentration range of 5–8.5 mM.* (MChem 41)

The thematic analysis in this thesis was also impacted by the division of clauses in my corpus into modalizers and propositions (Section 4.3.2). To briefly reiterate, a modalizer is a clause, often a projecting or commenting clause, which signals an epistemic (probability, usuality, typicality, obviousness) or emotional attitude to the subsequent proposition, modifying/negotiating the degree of certainty about its validity (e.g., *we believe that*; *it is not unreasonable to suggest that; an alternative explanation is that*). A modalizer is not considered a proposition, since it may not include disciplinary content (a disciplinary proposition) in its own right. Thus-defined modalizers are interpersonal grammatical metaphors which function at the same time as projecting clauses as well as mood Adjuncts (i.e., *we believe* –> possibly, *it is not*

unreasonable to suggest that –> likely). Modalizers defined in this way have parallels to special thematic structures (Table 4.3 above) and issues in thematic analysis (e.g., Themes in reported clauses or in interpersonal grammatical metaphors) mentioned in Thompson (2013).

Adopting the interpersonal perspective's focus on function rather than the experiential perspective's focus on wording (Thomson, 2013), I here considered modalizers as realisations of interpersonal rather than topical Themes, and thus excluded them from my thematic analysis of knowledge visibility. First, the interpersonal perspective on modalizing clauses echoes an alternative approach to thematic analysis in which interpersonal and reporting sentences are understood as Orienting Themes (as in, e.g., North, 2005) or Contextual Frames, whose function is to usher in the true starting point of a clause (Thompson, 2013). Second, as Thompson (2013) points out, while modalizing clauses are the main clauses structurally, they are not the main clauses semantically. Modalizers can be analysed for transitivity[40] (mostly as mental or relational processes) because their modal meanings are metaphorically experientialized, but in their common semantic interpretation (e.g., *we suggest* (metaphorical) –> possibly (common interpretation)) they are not included in transitivity analysis. Therefore, the first group or phrase that has some function in the experiential structure of the clause lies within the proposition that modalizers introduce. Lastly, this exclusion mirrors the exclusion of modalizers from my Authoritativeness analysis of knowledge propositions (Section 4.3.5). I should also add that some Thematic equatives which precede a *that* clause (e.g., *one possible mechanism is that*) were also excluded, even though they are not mentioned above as interpersonal

[40] The kinds of verbs/processes that capture and construe our experience of the world

grammatical metaphors (which they could also be (*one possible mechanism is that* –> possibly). Nevertheless, such Thematic equatives introduce (sub-)disciplinary propositions, and hence are modalizers as well.

Other exclusions from my thematic analysis, with theoretical justifications based on Halliday and Matthiessen (2014), are the following:

- *which*/*who* in grammatical Subjects in non-defining relative clauses: They represent partly textual links to independent clauses (inherently textual Themes), and partly refer to experiential clausal elements (inherently topical Themes). The inherently topical thematicity of *who*/*which* does not occupy the topical Theme slot completely, leaving space for another experiential element before the Finite. Thus, given the focus of the thematic analysis on the Visibility of knowledge, clauses where *who*/*which* were the only Subjects followed by the Finite/Predicator were excluded; where *who*/*which* were followed by another participant or circumstance, that experiential element was taken as the topical Theme.
- Imperatives: They are proposals (i.e., they function as demands in exchanges of goods and services) and not propositions.
- *wh*- question words in *wh*- interrogatives: They are partly interpersonal and partly topical Themes.
- Embedded (rankshifted) clauses: Unless they are grammatical Subjects, they contribute negligibly to the thematic progression of texts.
- Non-finite clauses without Subjects: They are without topical Themes.

- Anaphoric elliptical clauses: Topical Themes are identical to, albeit absent and presupposed from, preceding clauses. Excluding ellipted Subjects avoids duplicity.

All identified topical Themes in the corpus were then categorised based on the strengths of epistemic relations into four groups along the epistemic relations axis, and assigned numerical values ranging from 1 (ER++) to –1 (ER– –). The highest value (ER++) denotes disciplinarily specific Themes; (ER+) represents general academic Themes (i.e., not specific to seismology or marine chemistry); (ER–) comprises non-academic expressions and references to people; and the lowest value (ER– –) is occupied by expressions that are either grammatical placeholders, or that refer to concepts retrievable elsewhere in the text (see Table 4.4 below).

Table 4.4

The coding table for Visibility of knowledge. Topical Themes are an external language of description for epistemic relations. The examples are from MChem 39 and Seis 39.

ER + + (1)	**ER + (0.5)**	**ER – (–0.5)**	**ER – – (–1)**
Disciplinarily specific Themes that include specialist expressions identifiable as belonging to the field (including discipline-specific locations)	Disciplinarily non-specific Themes comprising general/core academic expressions	People (self-mention, references, research community), non-academic expressions, and non-disciplinary places and time(s)	Referential Themes a) *It, there* dummy subjects b) *It, they* etc. non-personal references c) *Both, some, all, each*, etc. d) *That, those, this, these* (demonstratives)
The surface field of Alk is distinct from the actual alkalinity observations (first and last rows of Fig. 1).* (MChem 39) *The reduction of focal depths is more significant in the western segment…* (Seis 39)	*This feature has been termed the "Great Calcite Belt" (Balch et al., 2011) …* (MChem 39) *… and these various estimates agree closely as to strike and slip angles.* (Seis 39)	*Chen (2002) reported calcium concentrations along 150°W in the North Pacific from 8 to 55°N.* (MChem 39) *In the southwest, Aegean subduction accommodates that motion.* (Seis 39)	*There is some bias in the collection of data, …* (MChem 39) *This is a consistent feature regardless of the quality of the earthquake location set…* (Seis 39)
Excluded from analysis: 1. Modalizers 2. *which/who* as the sole Subjects in non-defining relative clauses 3. Imperatives 4. *wh*-question words (*who, what, why, where, when, how),* even in reported clauses 5. Embedded (rank-shifted) clauses, unless they are Subjects 6. Non-finite clauses without Subjects 7. Elliptical clauses			

While the last category, ER– –, is reducible to a relatively short list of easily identifiable pronouns/determiners, the other three categories are voluminous, with a potential for subjective classification. Therefore, to ensure principled and repeatable categorisation of identified topical Themes, as a general rule, expressions were categorised as ER++ if they appeared in a list of terms compiled from two earth science

dictionaries; Themes were categorised as ER+ based on a new Academic Vocabulary List (Gardner & Davies, 2013); ER–, non-academic expressions, were all others. However, there are still outstanding issues regarding dealing with technical vocabulary, to which I turn next.

Following Nation (2013), in this study, I defined technical vocabulary as words closely associated with the subject matter of a particular discipline. This close association means the words are frequent within or unique to that discipline, and as a result some technical vocabulary is likely unfamiliar to disciplinary outsiders. Yet, identifying technical words is far from straightforward. A strict dichotomy (technical vs. non-technical) may not always exist (Ha & Hyland, 2017), and the line between disciplinary and everyday meanings is in some cases permeable. Some specialist vocabulary can come from all vocabulary frequency bands (high, mid, and low), which can lead to the similarity or correspondence of technical meanings and the basic meanings the terms possess in everyday use (Nation, 2013)– The reader may remember Maton's (2013) example with the word *gold*, which was mentioned in Section 3.3. Thus, if a technical word is also a high-frequency word, it will be perceived as technical only if it is used in a technical sense (Chung & Nation, 2004), and if we are familiar with the taxonomy that the term is part of (Nation, 2013).

The permeability between technical/non-technical usage has an implication for this study: It raises a question of how technical words, especially those which are at the same time everyday expressions, were dealt with. I adopted the criterion of Chung and Nation (2004) that if the word appeared in a technical dictionary, it was a technical word. This criterion also agrees with Maton and Doran (2017a), who argue that while some words occur both as technical and everyday, if such words appear in disciplinary

texts, their meanings involve relationships with other meanings and taxonomies, locating them within complex meaning constellations that expand their epistemic meanings beyond meanings in everyday use. Thus, the words are technical, distinct from their everyday usage. Applied to earth sciences, for example, *fault*, *earthquake,* and *magma* appear in everyday usage, where they possess the same meaning as they do in disciplinary texts, but in everyday usage the words do not entail the differentiation and resonance to other meanings (to use Maton and Doran's (2017a) notions): *magma* can be *basaltic* or *rhyolitic*, implying varying contents of silica, implying varying degree of viscosity, implying varying eruption styles, implying varying geographical distribution, etc. At the same time, earth sciences include a relatively higher number of words whose meanings markedly diverge from the same words' everyday usage. For example, *doctor* is a cooling sea breeze in the tropics; a *bomb* is a non-volcanic rock which is still viscous when ejected from an erupting volcano; a *bed* is the smallest division of a stratified rock series, marked by a well-defined divisional plane from its neighbours above and below. These technical definitions then support Maton and Doran's assertion regarding embedding in complex constellations of meanings.

I adopted one of the approaches to identifying technical vocabulary mentioned by Chung and Nation (2004)–that of using a technical dictionary[41]. Other approaches exist as well. Using human raters and a scale, while accurate, is impractical for such an amount of text as in this study's sub-disciplinary corpora. Using computer software to extract technical terms from corpora is practical and relatively reliable but not accurate at identifying multi-word terms (Chung, 2003), a major disadvantage and limitation. Therefore, I believe that adopting the dictionary approach was the best option.

[41] When working with LCT dimensions, Maton and Doran (2017a) also refer to the option of using specialist dictionaries to distinguish technical words.

According to Chung and Nation (2004), it offers an average success rate of around 80%, and depending on the dictionary used, it can correctly identify 98% of technical terms. To ensure a higher accuracy rate, two dictionaries (i.e., the second edition of *McGraw-Hill dictionary of earth science* and the fourth edition of Oxford's *A dictionary of geology and earth sciences*) were used in the coding of topical Themes. To the best of my knowledge, separate specialist dictionaries of the various earth science sub-disciplines do not exist, so the dictionaries mentioned above are general earth science/ geology dictionaries, comprising terms from multiple sub-disciplines. Only the main entries, the headwords in boldface, were counted as technical entries. Each headword entry was counted separately, regardless of whether an entry was a single-word or multi-word unit; whether it appeared in the two dictionaries in various spelling forms (e.g., *bahada–bajada*); whether it was listed first as an acronym/abbreviation and later in its full form (e.g., *DDA value–depth-duration-area value*); whether two or more entries belonged to the same word family (e.g., *laterite–lateritic–laterization*); whether two entries were synonymous (e.g., *Antarctic vortex–polar vortex*); or whether an entry consisted of a number-letter/word-number/symbol-word combination (e.g., ^{14}C, *Bessel ellipsoid of 1841*, ΔT *method*). After being extracted, the headwords were visually scanned, and several types of items removed: headwords consisting of single letters (e.g., *B* or *G*); headwords that were only prefixes (e.g., *mega-*, *anti-*, or *meta-*), because they could potentially be included in many non-technical expressions; and standalone persons' names (e.g., *Lyell, Charles [1797-1875]*)[42]. Duplicate entries (i.e., identically spelt entries appearing in both dictionaries) were removed in order to create a list of

[42] While Maton and Doran (2017a) state that persons' names are also technical terms, my corpora consist of published experimental RAs, not disciplinary textbooks, which renders personal names related to the historical development of the field arguably less relevant.

unique technical terms. Combining the terms from both dictionaries avoided the bias inherent in relying on technical terms included in a single dictionary (Chung & Nation, 2004; Nation, 2013), and at the same time provided a sufficient number of technical terms (more than 16,000).

Then, following Hu and Cao (2015), part of the data was coded by another rater. Specifically, 5% of the total RAs, or 526 instances of topical Themes, were coded by a second rater, an Associate Professor in Applied Linguistics, who was trained on using the coding scheme. Inter-coder agreement was assessed with Cohen's kappa (measure of agreement reached .880), and any disagreements were discussed. The remainder of the data was coded by me, but where doubt persisted, individual Themes were again classified in consultation with the second rater.

4.3.4 Visibility of knowers: Citation practices

The relative strength of epistemic relations and social relations in RA introductions from the natural sciences, social sciences, and humanities was investigated by Hood (2011). She examined how disciplinary knowledge was projected, namely, sources that projected (e.g., researchers, research participants, and published work) and the content that was projected (e.g., analytical procedures, observations, and participants' accounts). Some of the projecting sources were citations (researcher/ year of publication), whose visibility (i.e., whether they are directly named in the text or represented by superscript numbers) was interpreted as indicating the strength of social relations.

Similarly to Hood's (2011) conception, in this study I used visibility of cited sources (Visibility of disciplinary knowers) as an external language of description for

LCT's social relations. This kind of investigation of citation–to study knowledge construction–falls traditionally within the sociological focus of citation analysis (White, 2004). I was guided by the question of how visible researchers were in discussion and conclusion sections from seismology and marine chemistry RAs. Research into citation practices across academic fields shows that overall, authors are less prominent, with names and years of publication in parentheses or represented by superscripts, but that their prominence tends to be higher in the social sciences and humanities compared to the natural sciences and engineering (Hyland, 2002a). While the general conclusion about lower prominence of authors in disciplinary writing may hold for earth sciences as well, whether any sub-disciplinary differences exist is unclear. Neither is it clear how homogeneous individual sub-fields are in citation practices, or whether authors' prominence increases in RA sections comprising higher concentrations of knowledge claims. Finally, the extant research does not make clear what impact author prominence has on the form of sub-disciplinary knowledge (i.e., Specialization in the Visibility of knowledge and knowers).

Cited sources can range from highly visible to obscured. To classify citations based on visibility, a visibility cline was created by combining and modifying three existing citation typologies. The basic distinction between integral and non-integral citations originates from Swales (1990) (see Section 2.4.2). I collapsed his four sub-types of integral citations (Subject, passive agent, part of a noun phrase, and adjunct structure) into two: grammatical Subject head noun positions and other positions. The reason is that while occupying the Subject position makes source authors the most visible (Hyland, 1999), the specific individual visibility of the remaining three sub-types is debatable. The Subject position in my classification means that the author's name is

the head noun (e.g., *Swales classifies*…), rather than a pre- or post-modifier of a head noun (e.g., *Swales's classification distinguishes*…), in which case authorial visibility would be reduced, and the Subject head noun would in fact refer to a concept or a procedure (Hood, 2011; Hyland, 1999). Hyland's (1999) three-point distinction (grammatical Subject, non-Subject, and noun phrase) also fails to clarify whether non-Subject positions are more visible than noun phrases. Thus, to reiterate, I decided to divide integral citations into Subject head noun positions and other positions only.

As for non-integral citation sub-types, Thompson's (2002) typology allows for a finer-grade distinction of both parenthetical and super-scripted citations as mentioned in Section 2.4.2, but again, the visibility differences of the individual sub-types are debatable. In addition, I deemed all instances of super-scripted citations found in my corpus to be sufficiently represented if collapsed into a single category (e.g., I disregarded Hood's (2010) distinction of the size of superscripts). That is because source authors referred to by super-scripted citations are obviously invisible to the reader on the page where they are referenced, and to differentiate their hypothetical degrees of visibility thus makes little practical sense. The last non-integral citation sub-type I considered is a coalescence of Swales' (1990) contrasting citation and Thompson's (2002) reference, and I categorised it as a parenthetical citation. The resulting hybrid classification is summarised in Table 4.5 below.

Table 4.5

Types of integral and non-integral citations modified and combined from Hyland (1999), Swales (1990), and Thompson (2002). The examples are from MChem 21; the parenthetical clause citation (underlined) is from MChem 2, and the superscripted citation is from Seis 48.

Integral	Non-integral
a) Subject head noun *Berelson et al. (2009) investigated the contents of CAS from carbonate and tufas in modern Walker Lake, western Nevada ...* b) Other positions *These concentrations are higher ..., but similar to values retrieved by Chu et al. (2007) in a low-resolution sampling of the Jixian Group.* *Marine sulfate concentrations inferred by the rate-dependent model of Kah et al. (2004), however, assumes a homogeneous sulfate reservoir.* *According to Evans and Mitchell (2011), the early Mesoproterozoic witnessed the breakup of the core of the Columbia (or Nuna)*	a) Parenthetical citation *Substantial loss (up to 80%) of CAS can occur during diagenesis (Gill et al., 2008),...* *they were sufficient to assess the specified validation criteria (guided by Rykiel (1996))* Contrastive citation/Reference *The presence of euxinic conditions, however, requires consideration of two distinct active sulfur reservoirs (cf. Thompson and Kah, 2012).* b) Super-scripted citation *A probable example of the strongest triggered seismicity is the series of earthquakes with M ~ 7.0 at the Gazli gas field in Central Asia [2].*

Citation visibility as a linguistic translation of the relative strength of Maton's Specialization codes is in agreement with Maton's (2010) assertion that knower visibility, in particular voice, is a characteristic of stronger social relations, hence knower-code disciplines. The citation visibility can be classified as follows:

Integral citations are more visible than non-integral ones (e.g., Hyland, 1999; Thompson, 2002). Integral citations hence represent stronger social relations (SR++ and SR+). The increased visibility of integral citations is partly due to the fact that in the natural sciences, the emphasis is given to ideas rather than to authors (Dowdey, 1992), so using an author's name is a rhetorical choice implying the names and the referenced RAs are important (Paul, 2000).

Out of the two integral citation sub-types, authors appearing as Subject head nouns are more prominent (hence I have designated them as SR++) than authors in other positions (Flottum, Dahl, & Kinn, 2006) (designated SR+).

Between the non-integral sub-types, I argue that citations in parenthetical clauses are more visible (SR–) than other parenthetical citations and super-scripted citations (SR– –), because the authors still play a role in the grammar of the parenthetical clause.

Table 4.6

The coding table for Visibility of knowers

SR ++ (1)	**SR + (0.5)**	**SR – (–0.5)**	**SR – – (–1)**
Integral citations		Non-integral citations	
Subject head (noun)	Other positions	In a parenthetical clause	Parenthetical/ Superscripted

All instances of citations in my corpus were classified based on the coding table. Interestingly, almost all the RAs in the corpus format in-text citations as author + year of publication; only two articles in my seismology sub-corpus (Seis 13 and Seis 48–not

from the same journal) use superscript numbers (Table 4.7 below). Seis 13 has no references in the discussion or conclusion sections; Seis 48 features six clauses with superscripted references, which were coded as SR– –.

Table 4.7

In-text reference formats of the corpus articles

Seismology		Marine chemistry	
No. of RAs	Citation format	No. of RAs	Citation format
58	Author(s) and year of publication	60	Author(s) and year of publication
2	Superscript numbers		

The counting of citations included name(s) of author(s) and collective references followed by a year of publication (e.g., *ANZECC (2000) guidelines*), the words like *the authors/the researchers*, as well as pronouns referring to other authors (*s/he, her, his, they, their, who, whose, others*). In agreement with Hu and Wang (2014) and Hyland (1999), references to laws/methods/parameters named after researches (e.g., *Redfield ratio, Rayleigh fractionation, Liebig's law, Green's function, Poisson's ratio*) were excluded. The rare references to websites (e.g., in MChem 33 or Seis 52) were also excluded. Each instance was counted (i.e., regardless whether each instance referred to the same or different proposition since I am not a subject specialist). Multiple citations in parentheses (e.g., in MChem 2: *...but are enhanced by curvature and sloping cross-channel bathymetry [Blanckaert and Graf, 2001; Shao et al., 2003]*) or connected with conjunctions such as *and*/*or* (e.g., in Seis 54: *This is highlighted in Dumufier and Rivera*

[1997] and Vavrycuk [2001]) were counted as one. They were counted separately in cases such as MChem 7: ... *attributed to either increased productivity (Meyer et al., 2011) or to intensified water-column stratification (Song et al., 2013),* because each citation clearly represented a new/different finding. Elided grammatical Subjects were not counted, for example (elided Subject supplied and underlined), *Yin et al. (1995a,b) reported gradients in vertical mixing and vertical nitrate flux in the Fraser River plume over a springs-neaps cycle and <Yin et al. (1995a,b)> attributed greater mixing during spring tides to greater bottom stirring, ...* (MChem 2).

Following the procedures in Hu and Cao (2015), part of the data was coded by another rater. Specifically, 4.2% of the total RAs, or 150 instances of referencing, were coded by a second rater, an Associate Professor in Applied Linguistics, who was trained on using the coding scheme. Inter-coder agreement was assessed with Cohen's kappa (measure of agreement reached 1.000). The remainder of the data was coded by me, but where doubt persisted, individual instances of citations were again classified in consultation with the second rater.

4.3.5 Authoritativeness of knowledge: Epistemic certainty of propositions

Apart from operationalising the LCT dimension of Specialization in terms of capturing the Visibility of sub-disciplinary knowledge and knowers, in this thesis, I also capture the forms of sub-disciplinary Specialization in terms of how authoritative knowledge claims and knowers (authors) are. Authoritativeness refers to the degree of epistemic certainty with which knowledge claims are put forward (more or less certain/ more or less tentative), and to the degree of assertiveness with which authors announce their role in knowledge construction and their contribution to the field. Generally, the

more assuredly a claim is presented, the stronger the epistemic relations; the more important an author's role, the stronger the social relations. To explain how I coded Authoritativeness of knowledge via hedging and boosting, I start broadly by first addressing the issue of arguability of propositions. I then describe the kinds of expressions I focused on in my corpus, the coding rubric, problematic cases, and cases that were excluded.

In an exchange of information, whenever propositions are asserted, they are up for agreement or disagreement (Butt et al., 2012). What is asserted is fundamentally embodied by the Mood element (Halliday and Matthiessen, 2014). In other words, in arguments about asserted propositions, the stakes rest with the relationship between the grammatical Subject and the Finite (Butt et al., 2012). The Subject assumes the position of modal responsibility, and the Finite element makes a proposition arguable by giving "it a point of reference in the here and now" (Halliday and Matthiessen, 2014, p. 144) in two ways: by reference to time (through tense) and by reference to the writer's judgement of the proposition as being likely/unlikely (through modality). Therefore, propositions are not arguable if they are missing either the Subject or the Finite, neither of which is recoverable from previous clauses (Butt et al., 2012).[43] Additionally, for a proposition to be arguable, one feature of finiteness is necessary: polarity (positive/ negative), that is, whether something is/is not a proposition (Halliday & Matthiessen, 2014, p. 144).

Propositions can be encoded as asserting either a definite yes or no, or they can be modified by degrees of probability, typicality, usuality, and obviousness (known as

[43] Recovering the Subject can be problematic. To remind the reader of an example from Section 4.3.2, in *We found a significant effect of temperature on the partitioning of Mg and Ba into otoliths for juvenile rockfishes, providing the information [[needed to interpret otolith elemental ratios in a coastal upwelling region]]*. (MChem 4), do *we* provide the information, or does *the significant effect* provide the information? As Halliday and Matthiessen (2004) state, "the question is really irrelevant; it is precisely the function of the non-finite to make it unnecessary to decide" (p. 404).

modality), so that they are somewhere between being definitely positive and definitely negative. Linguistically, this adjustment is done within the Mood Block either via a modal Finite (e.g., *could*, *might*, *would*), an adverbial group or a prepositional phrase functioning as a modal Adjunct (e.g., *occasionally*, *for the most part*, *possibly*, *obviously*), or via an interpersonal grammatical metaphor (Butt et al., 2012)–a modalizer in my corpora (Section 4.3.2 and below). Other linguistic devices beyond the Mood Block that modify propositions in my corpora include lexical verbs (i.e., the Predicator in Systemic Functional Grammar) (e.g., *suggest*, *show*, *demonstrate*), epistemic adjectives that modify the verb *to be* (i.e., Complements in Systemic Functional Grammar) (e.g., *be consistent with*, *be likely*, *be probable*), conditional conjunctions (e.g., *if*), and nouns (i.e., Complements in Systemic Functional Grammar) that follow the verb *to have* (e.g., *have the potential to*). All these expressions that modify propositions centre on verbs (note that a verb is a precondition for the existence of a proposition).

For the sake of clarity and simplicity, I refer to all the above-listed linguistic devices that modify propositions by degrees of probability, typicality, usuality, and obviousness as hedges or boosters. If a finite ranking clause is not modified by a hedge or booster, it is a bare proposition.

Hedges and boosters modify a propositional stance by either weakening or strengthening, respectively, its epistemic certainty, which in this work is called Authoritativeness of propositions/knowledge claims. To decide which specific expressions modified the strength of propositions, I drew on previous research, and compiled a list (Appendix III), though certainly not exhaustive, of hedges and boosters identified by Cao (2014), Farrokhi and Emami (2008), Hyland (1996b, 1998), Peacock

(2006), Swales and Feak (2012), and Varttala (2001). This list served a guiding purpose, with all expressions in my corpus examined in their contexts.

Apart from hedges and boosters, the strength of propositions can be modified by modalizers (i.e., clauses that frame propositions as interpersonal grammatical metaphors; Section 4.3.2). A modalizer clause may impact the Authoritativeness of the subsequent clause (a booster = +1; a neutral verb/attitude marker = 0; a hedge = –1), which itself can include modality (implicit subjective or implicit objective, Thompson, 2013), for example, in its modal Finite. One modalizing clause can modify more than one proposition, meaning that the strength of a modalizer applies to all the following clauses in the same clause complex. Thus, in *It is likely that the few events [[located on the continent]] are poorly located and actually originated in the Gulf of California.* (Seis 50), *It is likely* modifies the proposition *that the few events [[located on the continent]] are poorly located* and the proposition *and actually originated in the Gulf of California.*

All propositions in my corpus were classified following Table 4.8 below.

Table 4.8

The coding table for Authoritativeness of knowledge with corpus examples

ER ++ (1)	ER + (0.5)	ER – (–0.5)	ER – – (–1)
A bare proposition *This frequent lack of correlation of hydrographic parameters contrasts* bare verb *with some previous B/Ca core-top calibration studies (Foster, 2008; Yu et al., 2013, 2007)…* (MChem 13)	**A bare proposition with a modalizing hedge** *This implies* modalizing hedge *that these regional B/Ca-pH correlations are* bare verb *an artifact of the regional correlation of many carbonate system parameters and environmental parameters with one*	**A hedged proposition with a modalizing booster** *… and they concluded* modalizing booster *that V might* hedge *function in at least two disparate sites of action.* (MChem 38)	**A hedged proposition** *Hence the relation between B/Ca and* $B(OH)_4^-$ */DIC in our study and in published records is likely* hedge *also related to calcification rates.* (MChem 13)

<table>
<tr>
<td>A boosted proposition

We find booster no evidence [[that N. incompta or G. bulloides B/Ca values are affected by post-depositional dissolution]], …
(MChem 13)</td>
<td>another (Fig. S2).
(MChem 13)

Previous studies of planktic foraminiferal and inorganic calcite have proposed modalizing hedge that B/Ca in calcite is controlled bare verb by one of several related carbonate system variables.
(MChem 13)</td>
<td>And Ito et al. (2016) showed modalizing booster that this pattern may hedge have been caused by the combined effects of natural climate variability and the deposition of polluted dust over the North Pacific Ocean.
(MChem 10)</td>
<td></td>
</tr>
<tr>
<td>A bare proposition with a modalizing booster

Foster et al. (2013) found modalizing booster that small benthic foraminifera became bare verb more heavily calcified during the PETM, …
(MChem 3)</td>
<td>A boosted proposition with a modalizing hedge

…which indicates modalizing hedge that interannual variation must booster be considered when using otolith chemistry to evaluate movement in this system.
(MChem 4)</td>
<td rowspan="2">A boosted hedged proposition (rare, only 9 cases in MChem and 8 in Seis)

… while further multi-species culture experiments would hedge no doubt booster be beneficial.
(MChem 3)

The d³⁴S shifts [[initiated with deposition of the pyrite layer (Fig. 6c)]] may hedge certainly booster reflect potential diagenetic HS overprint [[that raised originally more negative d³⁴S values]].
(MChem 7)

Indeed booster, simple one-dimensional modeling of particle dynamics can hedge generate a similar shallow lithogenic metal maximum…
(MChem 6)

The high velocity in the upper crust [[observed in the southern part of the plateau]] and … clearly booster suggest hedge oceanic nature of the crust beneath the thrust zone (Gokarn et al., 2008).

(Seis 15)</td>
<td rowspan="2">A hedged proposition with a modalizing hedge

… so it is possible modalizing hedge that the relationship of these variables to B/Ca could hedge reflect their correlation with $\Omega_{calcite}$ more than a primary influence on B/Ca.
(MChem 13)</td>
</tr>
<tr>
<td>A boosted proposition with a modalizing booster

PERMANOVA analysis demonstrated modalizing booster that waters associated with weak upwelling show booster distinct Mg/Ca, Sr/Ca, and Ba/Ca ratios relative to strong upwelling and relaxation conditions.
(MChem 4)</td>
<td>A hedged boosted proposition

… future studies should booster perhaps hedge pay more attention [[to identifying the significance of any existing natural feedbacks]].
(MChem 23)

A hedged bare proposition

Apparently hedge this material is depleted bare verb in 13C relative to the bulk POC pool.
(MChem 56)</td>
</tr>
</table>

The range of options and numerical values in the table is limited to four, in line with all other coding classifications in this work, and bearing in mind that "modality is expressed in a wide range of ways which makes a fine-grained typology problematic" (Hyland, 2005, p. 177).

While the compiled list of hedges and boosters, together with the coding table above, ensured principled classification, the linguistic data in my corpora presented other issues that had to be addressed. I now turn to describing problematic cases and cases that were excluded.

One problematic issue was distinguishing at times how many propositions following a modalizer were in fact modified by that modalizer. For example, in MChem 1,

Our findings suggest that

1. *nitrogen transformation in shelf sea sediments is dependent on whether specific microbial transcripts are influenced by differences in the composition of the bioturbating macrofauna, environmental context (here, nutrient enrichment and sediment type), and recent history of anthropogenic disturbance (here, frequency of bottom fishing),*
2. *although these effects are not necessarily interactive*
3. *and their relative importance is context dependent (Wohlgemuth et al. 2017).*

the sentence may be read as *Our findings suggest* modalizing only the first proposition or the first plus the second. The modalizer cannot modify the third proposition (*and their relative importance is context dependent*), because the clause is referenced to another study and clearly not part of *our findings*. However, whether the reference relates only to this last proposition or even to the preceding one (*these effects are not*

necessarily interactive), in which case the modalizer would modify only the first proposition, is unclear, and would require me to refer to the source cited. As this cannot be conceivably done for all such instances in my corpus, specifying that in such cases a modalizer applies both to the first and second propositions in the clause complex seems reasonable. While such a decision may be disputed, I have made it in the light of my limited knowledge of seismology and marine chemistry, and applied it consistently in the coding of all data.

However, where a modalizer was already rankshifted because it modified (as a dependent relative clause) a nominal group of a preceding ranking clause, it remained rankshifted (see the examples from MChem 4 and MChem 60 below, with the rankshifted modalizers *finding* and *suggest* underlined).

> *These results agree with previous studies [[finding [[that it takes days to weeks for otolith elemental ratios to stabilize || and reach equilibrium with ambient water concentrations (Elsdon & Gillanders 2005, Miller 2011)]]]].* (MChem 4)

> *Furthermore, the U export rates in this study are in contrast to research [[that suggest [[that STEs are U sinks]]]] (Charette and Sholkovitz, 2006; Duncan and Shaw, 2003; Moore et al., 2011; O'Connor et al., 2015; Santos et al., 2011; Windom and Niencheski, 2003) (Table 1).* (MChem 60)

Another difficulty in some clauses was to decide whether it was indeed the finite verb that was modified by a modal Adjunct (e.g., *probably*, *mainly*, *clearly*). To reiterate, propositions are centred on verbal groups (Section 4.3.2), and when the verbal groups include the Finite elements, propositions are arguable (this Section above). Hence, finite verbal groups are the constituents I focused on, and I considered only

Adjuncts that modified finite verbal groups. For example, consider the word *mainly* (underlined) in the following extract: … *these integrated high-quality datasets point out that thrust-related fold growth in the offshore sector of the thrust belt, in the position [[affected by the recent earthquakes]], occurred mainly in Pliocene times* (Seis 42). Exactly which clause constituent does it modify? To differentiate whether *mainly* modifies *occurred* or *in Pliocene times*, I devised what I call the passive voice and predicated Theme tests, which allowed me to grammatically restructure clauses, revealing whether the modified clause constituent was indeed a finite verb or not. I restructured the clause as *these integrated high-quality datasets point out it was mainly in Pliocene times that thrust-related fold growth in the offshore sector of the thrust belt, in the position [[affected by the recent earthquakes]] occurred* to reveal that *mainly* is part of the Adjunct of time, and does not modify the verb *occurred*.

Another type of modification may be realised by modal Comment Adjuncts (e.g., *frankly*, *fortunately*, *interestingly*, *of course*, *regrettably*), which, however, express more of an affective rather than epistemic attitude (Bloor & Bloor, 2013). For example, in *Remarkably, the d13C SPE-DOC values in the shallow sediments are generally more ^{13}C depleted than the seawater values (Fig.2).* (MChem 8), the underlined Comment Adjunct was not considered as modalizing the proposition, but rather as being neutral (i.e., not affecting the epistemic strength of the proposition) as it describes the writer's emotional attitude (i.e., one kind of Hyland's (2005a) attitude markers).

Non-finite clauses, typically infinitive and participle clauses (underlined in the examples below), are not arguable, and hence were excluded from analysis.

Here, two separate chemostat pre-experiments were conducted in order to determine the appropriate dilution rate (30% per day). (MChem 15)

> *In a warming ocean, the surface heating, glacier melt, and increased precipitation at high latitudes can increase the ocean stratification, leading to the weakened mixing of O_2-rich surface waters into the thermocline*. (MChem 10)

However, when the Finite, usually modal, and the Subject were possible to recover from the previous clause in the same clause complex (underlined examples below), a non-finite clause was still coded for Authoritativeness, with the recovered Finite representing its epistemic certainty.

> *Active estuarine management [[focused on the improvement of bottom water dissolved oxygen for ecological health reasons]] must* booster *concurrently reduce anthropogenic nitrogen loads and* recovered booster *maintain high levels of microbial denitrification for net water quality improvement*. (MChem 2)

> *This process may* hedge *not be limited to the sediment surface, since Fe(II) may* hedge *be oxidized and* recovered hedge *precipitated as Fe(III)*… (MChem 8)

Questions and indirect questions (typically structures with *whether* or *whether or not*) were excluded, even though they included Subjects and Finites, because questions do not put forward any knowledge claim in an authoritative way. Another excluded category comprised stance nouns (e.g., Jiang & Hyland, 2015) and what Hyland (1996a) refers to as strategic hedges, described as references to inadequate/ limited knowledge, deficiencies in the model/method/theory, and experimental weaknesses (e.g., *viewed in this way*; *under these conditions*; *based on our model*). With only a few examples given by Hyland, I found strategic hedges to be an insufficiently defined category, likely prone to inconsistent identification during coding. I also

disregarded any impact of expressions such as *little* (underlined in the example below) on the epistemic certainty of propositions.

> *These findings corroborate the results of Fear et al. (2005), who observed little change in relative N_2 production despite highly variable denitrification rates across a range of salinity conditions in the Neuse River Estuary, North Carolina.* (MChem 30)

On the one hand, the word *little* could function as a hedge (i.e., as compared to *no change*), but it could also describe exactly what was observed (i.e., *little change*). In addition, formulations describing observations of *major/some/little changes* could be viewed as evaluative, serving to highlight/attract attention to or de-emphasise results rather than commenting on epistemic certainty. This area of academic writing is certainly more complex, and beyond the scope of this study, in which I use less fine-grained coding schemes. Here, a coarser classification suffices to investigate knowledge-building practices in seismology and marine chemistry throughout the lens of LCT. Studies focusing exclusively on the individual linguistic features could certainly develop much more detailed distinctions and descriptors.

Finally, following the procedures in Hu and Cao (2015), part of the data was coded by another rater. Specifically, 4.2% of the total RAs or 150 instances of referencing. Specifically, 3.3% of the total RAs, or 402 instances of hedging and boosting, were coded by a second rater, an Associate Professor in Applied Linguistics, who was trained on using the coding scheme. Inter-coder agreement was assessed with Cohen's kappa (measure of agreement reached .939). The remainder of the data was

coded by me, but where doubt persisted, individual instances were again classified in consultation with the second rater.

4.3.6 Authoritativeness of knowers: Use of exclusive personal pronouns

As mentioned in Section 2.4.4, the use of exclusive first-person pronouns in earth sciences has, to the best of my knowledge, been investigated only in the part-genre of the field account; yet, in earth sciences, exclusive personal pronouns are used in other parts of RAs. In an exploratory teaching project on how often personal pronouns are used, my students and I found a vast frequency range–from anywhere between fewer than 10 to more than 100 instances in RAs of comparable length. Such a range may not be entirely surprising. First, significant intra-disciplinary differences in the raw count of *I* in RAs have been found before (e.g., McGrath, 2016). Second, based on my conversations with senior faculty at the Earth Observatory of Singapore and the Asian School of the Environment and my experience in editing RAs written by expert disciplinary members, some writers clearly prefer to mention themselves explicitly in the text more frequently. For example, for one of the manuscripts I was editing, my instruction from the main author was to change the passive voice structures (e.g., *was analysed by using*) employed by the co-authors to active voice clauses with *we analysed*. The high frequency of exclusive personal pronouns in some earth science RAs does not necessarily mean that writing in the discipline is becoming "author saturated" (Geertz, 1988, p. 9). Rather, these examples show the possible variations within a discipline, with individual authors' usages "not [necessarily] affect[ing] their article's acceptance by the disciplinary community as a 'family member'" (McGrath,

2016, p. 96). From the viewpoint of a legitimate disciplinary contribution described by Maton (2016a), the variation may well signal that while one code could be dominant, it may not be universally accepted, and it may be challenged; or more codes may co-exist without one being dominant, but they may be competing to establish their dominance.

Given the variation in frequency and in preferential sentence constructions, it may be tempting to conclude that in earth sciences no specialization code is dominant, meaning that knowledge claims can be legitimated by a focus on either epistemic relations (specialised knowledge of procedures and processes) or social relations (who you are is important). Such a scenario would suggest that earth sciences as a discipline may be characterised by a code clash and a struggle for identity. Nevertheless, such a conclusion may not be warranted without investigation, and to avoid this error, we need to look more closely at how expert writers actually use first-person pronouns.

While some studies have analysed authorial roles expressed by personal and possessive pronouns in grammatical Subject and other positions (e.g., Duenas, 2007; Hyland, 2001; Kuo, 1999), others limited their focus to personal pronouns *I/we* (e.g., Martinez, 2005; McGrath, 2016; Starfield & Ravelli, 2006). In this work, from ranking or rankshifted clauses, I considered only the grammatical Subject pronouns *I* (single-authored articles in my corpora are MChem 14, Seis 13, Seis 22, and Seis 50) and *we* (both inclusive and exclusive), as well as words like *author(s)* in Subject positions if a text's author(s) referred to himself/herself/themselves. The focus on *I/we* Subjects stemmed from my interest in a very direct link between authorial agency and what happened in a clause. In terms of Systemic Functional Linguistics and its clause as exchange function, the Subject within the Mood Block[44] is "the one on which the

[44] The Mood Block comprises the Subject and the Finite, and sometimes a Mood Adjunct (Halliday & Matthiessen, 2014, pp. 175-176).

validity of the proposition is made to rest"—it is"responsible for the functioning of the clause as an interactive event" (Halliday & Matthiessen, 2014, p. 146). Also, as Hood (2011) points out, when researchers/authors do not function as the Subject head nouns but merely as pre-modifiers, clauses are about the head nouns rather than the authors. Thus, for example, the clause *Our results are consistent with constraints on magma body size from previous geophysical studies…* (Seis 14) is about the results and less directly about the researchers. Thus, possessive pronouns (e.g., *our* in the above example) are weaker representations of authorial voice then the first-person pronouns *I/ we* (Duenas, 2007). In addition, when personal pronouns are taken as instruments of promoting one's research methods, results, or contributions (Harwood, 2005c), *I/we* more directly highlight what authors want to be credited for.

To map exclusive first-person pronouns on Maton's (2014a) social relations, I selected the functional pronoun taxonomy of Tang and John (1999), mentioned in Section 2.4.4. Of their six originally defined authorial identities, only four were relevant in my research, because *'I' as the recounter of the research process* does not appear in the discussion or conclusion sections, and *'I' as the guide*, representing inclusive pronoun uses, was merged with *'I' as the representative* under a conventionally-named category *Inclusive We*. I mapped the remaining four identities to Maton's framework as follows:

We as originators

This is a high-authority role representing very strong social relations (SR++). The authors present new knowledge claims as results, findings, observations (announcing results), or conclusions. (This role is often marked by verbs such as *claim*, *suggest*, *argue* (McGrath, 2016), *find*, *identify*, *observe*, *conclude* (Cao,

2014).) They thus stake out their contributions to the field. Having the right and the ability to originate new ideas (Tang & John, 1999) to bring the field forward can be interpreted as having the insider credibility or legitimacy to participate in the disciplinary debate. Such credibility can be demonstrated by, for instance, being able to pinpoint issues that are currently relevant, or by being able to identify and fill a gap in disciplinary knowledge with something new to say (Hyland, 2004). By identifying the new contributions they make, writers position themselves as disciplinary scholars (Kamler & Thomson, 2006). In this role, writers refer to both their knowledge of disciplinary issues and their own contributions, stemming from immersion in disciplinary works. The knowledge cultivated through this immersion is then incorporated into their own work (Martin, 2016).

We as opinion-holders

Overall, this writer role is characterised by strong social relations (SR+). The role entails several facets. First, the authors comment either on previously published research or work of others in the field. (This role is often marked by mental cognitive verbs such as *think*, *believe*, or *assume* (McGrath, 2016; Starfield & Ravelling, 2006).) The ability to comment on disciplinary issues and to take a stance shows familiarity with disciplinary literature, which, as Hyland (2004) points out, is a marker of membership in the disciplinary community. Knowledge of disciplinary works is built based on immersion in and interaction with a disciplinary canon (Martin, 2016). Second, the authors comment on their research being reported on in the text. Both kinds of comments may include agreeing, disagreeing, pointing out strengths and limitations, interpretation,

difficulties, plausibility, expectations, preference, assumptions, recommendations, implications, or elaboration. Third, comments on methodology/procedures or interpretive frameworks belong here, as they either provide justification/backing to strengthen the authors' viewpoints, or reveal more information so that the reader can follow and evaluate the authors' train of thought. Here, personalising the claim through first-person pronouns may promote authors' competence as researchers (Harwood, 2005a) to persuade readers that the choices they made were disciplinarily appropriate so that readers "know how data became results in order to trust the author's claims" (Smagorinsky, 2008, p. 394). In this role, writers need to display specialised knowledge of suitable procedures of data collection and analysis vis-à-vis a situation or setting. An example of this third facet is the following sentence from Seis 4:

If all events clustered near $\zeta = 0^{\circ}$*, then we would likely invoke a single-processed model of shear faulting (with no opening or closing); if all events clustered near* $\zeta = 90^{\circ}$*, then we might invoke a single-process model of tensional (or compressional) cracks.*

Textual We

This role is less authoritative, characterised by weaker social relations (SR–). The authors are the organisers of their text, referring the reader through the use of a verb to information in a specific part of their research article, be it a section, a figure, or a table. The authors may also announce how their article is structured, and what topics will be dealt with where.

Inclusive We

This identity projects weak authoritativeness and thus weak social relations (SR– –). Through the same personal pronouns, the authors refer to themselves as well as to their readers. The pronouns thus stand for members of the disciplinary community.

The table below presents a summary of the classification together with examples from my corpus.

Table 4.9

The coding table for Authoritativeness of knowers

We as originators	*We* as opinion holders	Textual *We*	Inclusive *We*
SR++ (1)	SR+ (0.5)	SR– (-0.5)	SR– – (-1)
• Experimental or modelling results, findings, and observations • Conclusions • Knowledge claims based on results/ findings/ observations (inferences/ interpretations/ hypotheses/ propositions)	• Strengths/ Limitations • Comparisons (of results) with previous studies • Elaborating an argument • Assumptions/ Attributions • Expectations/If scenarios • Hypotheses (to be tested) • Opinions/Beliefs • Recommendations • Implications • Procedures, materials, and methods • Specifying interpretation frameworks • Research focus/ activities/approach	• Paper's goal and purpose • Organising and structuring the text • Drawing attention to parts of the text • Defining terms in the text • Often together with expressions such as *here*, *in this paper*, *in this study*	• Authors + readers • Disciplinary members • People in general

• *We use porosity/ velocity relations to show that the porosities needed to explain the observed velocities are larger than expected at 3–5 km depth.* (Seis 14) • *We find a compact 30 s long rupture with 5.0±0.15 m of peak slip. (Seis 57)* • *We introduced a detailed fault plane solution for the 19 October 2012 earthquake…* (Seis 44) • *Considering this range, we conclude that the classical model remains a possibility.* (Seis 4)	• *We regard the mechanical failure modes outlined by Heap et al. (2015) as more appropriate for VT and LP type activity found by Nishi et al. (1996) and Sherburn et al. (1998), …* (Seis 54) • *Although we prefer this explanation, we cannot rule out the possibility that above 10 s period the distribution of noise sources is heterogeneous, …* (Seis 11) • *We would expect more reliable solutions in a setting with larger magnitudes (say, up to M_W 4).* (Seis 4) • *We leave to further investigations the determination of a refined source model from the joint inversion of static and dynamic (waveforms) and surface displacements.* (Seis 52)	• *In this study, we examined recent seismotectonic structures and source parameters of earthquakes using the RMT method to have a better understanding on the deformation processes that control central Anatolia.* (Seis 21) • *We discuss site effects in Section S2, Supporting Information.* (Seis 4) • *We note there are sub-seismic arrivals visible in the spectrograms at group arrivals <0.5 km s−1.* (Seis 11) • *We refer to (1) as the classical model, following Dufumier & Rivera (1997), and (2) as the crack-plus-double-couple (CDC) model, following Minson et al. (2007).* (Seis 4)	• *It is widely believed that large earthquakes occurred in the eastern and western segments… However, if we exclude the 1605 event from large earthquakes from the seismogenic zone, it is not always the case.* (Seis 35) • *More paleosesimic data… are needed to investigate this hypothesis and to answer some arising questions: can we consider the Mw 5.2 Lorca earthquake as the characteristic earthquake for this area? … Can we expect peak slip at the inter segment zone as it has been suggested…?* (Seis 10) • *In order to study the dependency of the duration, we usually consider scaled duration, which is duration divided by moment…* (Seis 16)

At this point a question may arise: Why has Tang and John's (1999) taxonomy been selected as the base for my authorial roles taxonomy, when other functional

taxonomies classifying first-person pronouns exist (e.g., Harwood, 2005a; Hyland, 2002c; Sheldon, 2009; Starfield & Ravelli, 2006)? Firstly, Tang and John's taxonomy describes core categories of author identities that are found in my data. Also, the taxonomy is collapsible to a four-point scale, is adaptable to my research design, and maps on the concept of Authoritativeness of disciplinary knowers. Secondly, both Harwood (2005a) and Hyland (2002c) had different motivations to create their taxonomies: they aimed to compare writings of undergraduate students and writings of experts in published RAs, so the category of *expressing self-benefit* is neither applicable to RAs nor to the LCT-guided notion of Authoritativeness. Thirdly, Harwood's and Hyland's categories of *first person in acknowledgements/'I' as a disciplinary servant* are irrelevant, since acknowledgements were not part of the corpus collected for this study. Lastly, the remaining categories in Harwood's and Hyland's frameworks can be encapsulated within one of Tang and John's: *Methodological 'I'/ explaining a procedure* matches *'I' as the recounter of the research process*; *'I' to state a purpose/stating a purpose* and *'I' to define terms* match *'I' as the architect of the essay*; *'I' to state results and claims/stating results or claims* matches *'I' as the originator*; and *'I' to elaborate an argument/elaborating an argument* can be placed in between *'I' as the opinion-holder* and *'I' as the originator*.

Two other functional taxonomies which could have been selected should be mentioned: those of Starfield and Ravelli (2006) and Sheldon (2009). Starfield and Ravelli's taxonomy, which draws on Hyland's, Harwood's, and Tang and John's, was not selected for three reasons. First, Starfield and Ravelli combine *'I' as the guide* and *'I' as the architect*, which I preferred to keep separate due to their inclusive and exclusive personal pronoun usage. Second, *methodological 'I'*, *'I' as opinion holder*,

and *'I' as originator – claim maker* are already included in Tang and John's taxonomy. Third, a new category put forward by Starfield and Ravelli is *reflexive 'I'*, which is a narrative voice of an individual reflecting on his/her engagement in the research process. Yet, while *reflexive 'I'* is suggested to be distinctive, Starfield and Ravelli also admit that more research is necessary to "establish it firmly as a separate type" (p. 235). Sheldon's taxonomy, which draws on Tang and John's and on Starfield and Ravelli's, was not selected mostly for similar reasons.

Regarding coding, I searched the corpora for all instances of first-person pronouns (and words like *authors,* where they referred to the writers of the text; where *authors* referred to other researchers, the word was included in the category of citations–SR Visibility). Elided Subject pronouns were excluded, for example in (underlined)

... if we are to provide ecosystem-relevant evidence [[to underpin decisions [[that aim to secure the protection of natural capital (Pittman and Armitage, 2016)]]]], and ensure the sustainable management of coastal and shelf sea ecosystem services (Voss et al. 2013) (MChem 1)

or in

We also show that the incorporation of Ba/Ca into otoliths is reduced below 9°C and recommend caution [[in interpreting chemical signatures in wild fish [[that may experience cold temperatures]]]]. (MChem 4)

Then, following Hu and Cao (2015), part of the data was coded by another rater. Specifically, 10.8% of the total RAs, or 236 instances of first-person pronouns, were coded by a second rater, an Associate Professor in Applied Linguistics, who was trained on using the coding scheme. While this step was repeated in the coding of the other

three linguistic features, here it was important also because the semantic distinction between authorial roles can be fuzzy for analysts, although exploitable by writers (Harwood, 2005b). Pronouns may perform multiple functions at the same time, and may thus at times not be easily placed under one category (McGrath, 2016; Vladimirou, 2007). Therefore, I referred to both context/co-text and the verb semantics following the pronouns (Sheldon, 2009). If two interpretations were still possible, I followed McGrath (2016, p. 89) and "err[ed] on the side of the more authoritative category". That was true especially when differentiating between inclusive and exclusive *we* (Harwood, 2005b) even with the help of context (Vladimirou, 2007). Inter-coder agreement was assessed with Cohen's kappa (measure of agreement reached .728). The remainder of the data was coded by me, but where doubt persisted, individual instances were again classified in consultation with the second rater.

4.4 Post-coding analysis

The four-point coding scales for topical Themes, citations, hedging and boosting, and first-person pronouns are designed to take advantage of topological differences provided for by representing epistemic relations and social relations on axes and hence in the Specialization Cartesian plane. Since the degrees of strengths of relations to knowledge structures and knower structures can be infinite, capturing these degrees opens endless possibilities for visualising dominant codes as well as variations between/among and within sub-disciplines, either synchronically or diachronically. To make these infinite possibilities easy to calculate and graphically represent, I contained the strengths of relations within an interval from -1 (the weakest possible) to +1 (the strongest possible). I was then able to calculate a value somewhere within this interval

for each of the four linguistic features for each individual sub-disciplinary RA by first tallying the possible coding values from -1 to +1, and second by dividing the tallied total by the number of instances of the linguistic feature. For example, in RA MChem 3, there are 125 topical Themes: 88 are ER++ (+1); 18 are ER+ (+0.5); 9 are ER– (-0.5); and 10 are ER– – (-1). When the four values (i.e., +1, +0.5, -0.5, and -1) are added together for topical Themes in the RA, the sum aggregate is 83.5, which, divided by the number of topical Theme decisions (125), equals 0.668 (i.e., moderately strong relations to the knowledge structure) on the ER Visibility axis running from -1 to +1. In the same RA, there are 38 citations: 1 is SR++ (+1); 3 are SR+ (+0.5); 2 are SR– (-0.5); and 32 are SR– – (-1). When the values +1, +0.5, -0.5, and -1 are added together for citations in the RA, the sum aggregate is -30.5, which, divided by the number of citation decisions (38), equals -0.80263 (i.e., very weak relations to the knower structure) on the SR Visibility axis. The two final values for MChem 3 then place the RA higher up and to the left in the knowledge code quadrant.

For each relation of a sub-discipline (i.e., for ER for Visibility for marine chemistry, SR for Visibility for marine chemistry, ER for Visibility for seismology, SR for Visibility for seismology, etc.), I then calculated summary descriptive statistics: central tendency (through the measures of mean and median), as well as dispersion (through the measure of range from provided minimum and maximum values). While dispersion could also be expressed by calculating standard deviations (SD), which express how typical individual values differ from the mean values, I opted for the simple measure of range because of large differences in the dataset–large differences in the usage of the four linguistic features in individual corpus RAs. With these large differences, the probability of a typical individual value deviating from the mean by the

calculated SD would be lower. Overall, these summary descriptive statistics are useful measures given my overarching research question, where I aim to reveal the dominant Specialization code, as well as possible simultaneous existence of other minority Specialization codes. These statistics also sufficiently answer my specific research questions (how visible or authoritative knowledge/knowers are).

To identify sub-disciplinary differences in the realisations of the strongest possible relations to knowledge and knower structures, qualitative analyses were based on sub-samples (for the sake of manageability) of ER+1 topical Themes and all SR+1 citations. After inspection in context, the sampled topical Themes were divided into four categories: objects of study/occurring phenomenon; research process/method/approach/ instrument; research product; and geographical or study location/time period. These categories reflect the relative strengths (unquantified) of relations to legitimate problem situations (ontic relations) and to legitimate methods (discursive relations) used by the sub-disciplines. Investigation of SR+1 citations revealed influential individuals, institutions, and countries where the researchers reside, as well as age of publications that were prominently cited. The age of publications provided an insight into how current cited knowledge is in the sub-disciplines. Importantly, matching the type of published knowledge (e.g., methods, quantitative results, site-specific findings) with age of publications hinted at what knowledge prominently cited individuals are most frequently credited for.

4.5 Summary

To answer my research questions, above I have detailed ways of operationalising the LCT dimension of Specialization, both quantitatively and qualitatively. While the

operationalisation is not the only possible translation between underlying organising principles and language features (i.e., data), I believe I have selected features that reveal the range of possible bases of legitimate achievement in marine chemistry and seismology.

Chapter V: Results

In this chapter, I present the quantitative and qualitative findings. I begin with findings for the four individual linguistic devices as operationalisations of epistemic relations and social relations. These findings answer the four specific research questions posed in Section 3.6:

1) How visible is disciplinary knowledge in sub-disciplinary research writing, as reflected by the use of topical Themes?
2) How visible are knowers in sub-disciplinary knowledge construction, as reflected in the use of integral and non-integral citations?
3) What is the level of authoritativeness (epistemic certainty) with which propositional content is expressed, as reflected in the use of hedging and boosting?
4) What level of knower authority is legitimate in sub-disciplinary research articles, as reflected in the use of exclusive personal pronouns?

I then move onto comparing marine chemistry and seismology in terms of the organising principles of Specialization in the Visibility of knowledge and knowers and Authoritativeness of knowledge and knowers.

5.1 Visibility of knowledge and knowers

To answer the first two specific research questions, I profiled disciplinary writing with respect to percentages of various topical Themes and types of citations (references to other studies). As a brief reminder of the methodology, in Chapter IV I posit that if the Visibility of disciplinary knowledge reflects the relative strength of a field's epistemic relations (i.e., the field's prominence in its focus on specialised objects

of study and procedures), then the relative strength will be greater when more disciplinary topical Themes are used, and weaker when topical Themes are mostly represented by general academic, non-academic, or referential constituents. Likewise, if the Visibility of disciplinary knowers reflects the relative strength of social relations (i.e., the field's prominence in its focus on who originates a knowledge claim), then the relative strength will be greater when knowers (cited authors) appear in grammatically more prominent positions as Subjects, and weaker when cited authors appear in other positions or in parentheses.

5.1.1 Visibility in marine chemistry

Most knowledge claims in the marine chemistry sub-corpus, by a vast margin, start with disciplinarily specific topical Themes, followed by general academic, non-academic, and referential Themes, respectively, in descending proportions (Table 5.1; note that the median values can be taken to represent a typical discussion (and conclusion) section of a marine chemistry RA).

Table 5.1

The breakdown of descriptive statistics for Visibility of disciplinary knowledge in marine chemistry.

Themes	MEAN	MEDIAN	MIN	MAX	RANGE
Disciplinary **ER + + (1)**	71.5%	71.5%	47.9%	85.4%	37.5%
General/cor academic **ER + (0.5)**	12.6%	11.8%	3.4%	35.4%	32%
Non-academic **ER – (–0.5)**	8.2%	8%	0.9%	25.6%	24.7%
Referential **ER – – (–1)**	7.7%	7.3%	0%	17.5%	17.5%

While the fact that disciplinary Themes are the most frequent is not surprising on its own, interestingly, disciplinary Theme usage also shows the largest range from among the four studied Theme types. Nevertheless, examining the range in more detail reveals substantial uniformity: only six RAs in the sub-corpus employ ER+ + Themes in fewer than 60% of claims, while in 27 RAs disciplinary Themes begin at least 75% of claims. The remaining 27 RAs employ ER+ + Themes in the range between 60% and 75%.

General academic Themes appear much less frequently, yet uniformity is also obvious. The most common ranges are 10% to 20% of claims (33 RAs) and under 10% (22 RAs). Only five RAs contain more than 20% of general academic Themes.

Non-academic Themes are comparatively fewer, with 41 RAs employing them in under 10% of claims. Only three RAs feature non-academic Themes in more than 15% of claims. The remaining 16 RAs employ this type of topical Theme in the range between 10% and 15%.

A similar distribution applies to referential Themes (e.g., *it*, *both*, *there*, *this*), which are the least frequent. In 42 RAs, they appear at the beginning of fewer than 10% of claims (not at all in 2 RAs), and in 3 RAs they range from 16% to 17.5%. The remaining 15 RAs employ referential Themes in anywhere between 10% and 14% of claims.

Altogether, the findings reveal a convergence in writing, more pronounced with less visible Themes. While under half of marine chemistry RAs contain claims beginning with disciplinary Themes in 75% of cases, more than half of RAs feature 10% to 20% of general academic Themes, and almost three quarters use non-disciplinary and referential Themes in fewer than 10% of claims.

In the marine chemistry sub-corpus, knowers do not feature prominently in claims. The absolute majority of references are parenthetical, again by a vast margin, with the other three knower Visibility categories being rather marginal on average (Table 5.2; note that the median values can be taken to represent a typical discussion (and conclusion) section of a marine chemistry RA).

Table 5.2

The breakdown of descriptive statistics for Visibility of disciplinary knowers in marine chemistry.

Citation position	MEAN	MEDIAN	MIN	MAX	RANGE
Subject head noun **SR + + (1)**	8.9%	5.7%	0%	42.9%	42.9%
Other (non Subject) **SR + (0.5)**	8.5%	5.5%	0%	45.2%	45.2%
In a parenthetical clause **SR – (–0.5)**	1.4%	0%	0%	8.7%	8.7%
Parenthetical **SR – – (–1)**	81.3%	85.5%	11.9%	100%	88.1%

The most frequent category of parenthetical citations also covers the largest range from 11.9% to 100%. The minimum number is an outlier, though: the next lower-frequency SR – – RA contains more than three times as many instances (37.9%), and only five other RAs contain fewer that 60% of parenthetical citations (40.6%, 43.5%, 48.3%, 52%, and 59%). In comparison, the maximum number is not really an outlier, with 5 RAs containing exclusively parenthetical citations, and 16 other RAs featuring citations that are more than 90% of the parenthetical type. Indeed, at least 75% of citations are parenthetical in 46 RAs, showing a predominant tendency of marine

chemists to de-emphasise disciplinary knowers, when measured by the saliency of the grammatical position of citations. Only seven RAs employ parenthetical citations in the range of 60% to under 75%.

The next most common citation positions are SR + + and SR +, appearing with similar mean and median frequencies. Both also cover an almost identical range. Neither type, of course, appears in the 5 RAs with exclusively parenthetical citations mentioned above, with 0% representation in further 8 RAs for SR + + positions, and further 11 RAs for SR+ positions (only 2 of these 19 RAs are the same). If citations in Subject and non-Subject positions appear at all, most constitute fewer than 10% of the total citations (26 RAs for both SR + + and SR+ positions). The remaining percentile distributions are also almost identical: 13 RAs and 11 RAs comprise 10% to 20% of SR+ + and SR+ citations, respectively, and 8 RAs and 7 RAs contain more than 20% of SR+ + and SR+ citations, respectively.

The least common citation type is SR – (in a parenthetical clause). It does not appear at all in 40 RAs. In the remaining 20 RAs of the sub-corpus, SR– citations constitute fewer than 10% of the total citations in an article.

Taken together, the findings reveal uniformity in low knower Visibility. More than three quarters of RAs de-emphasise (SR– –) the marine chemistry knower in at least 75% of cases; in comparison, knowers are more prominent (more than 20% of citations are SR+ +) in only about a tenth of marine chemistry RAs.

The percent representations of the various types of topical Themes and citations can be converted to positions (from -1 to +1) on epistemic relations and social relations Visibility axes for individual RAs (Section 4.4). The positions of all topical Themes and citations in the marine chemistry corpus are tabulated (Table 5.3; note that the median

values can be taken to represent a typical discussion (and conclusion) section of a marine chemistry RA) and graphically represented on the Visibility axes below in Figure 5.1.

Table 5.3

The descriptive statistics for positions on ER and SR Visibility axes in marine chemistry.

	MEAN	MEDIAN	MIN	MAX	RANGE
Positions on ER Vis. axis based on Theme types	0.659	0.681	0.341	0.883	0.542
Positions on SR Vis. axis based on grammatical positions of citations	-0.689	-0.780	-1	0.536	1.536

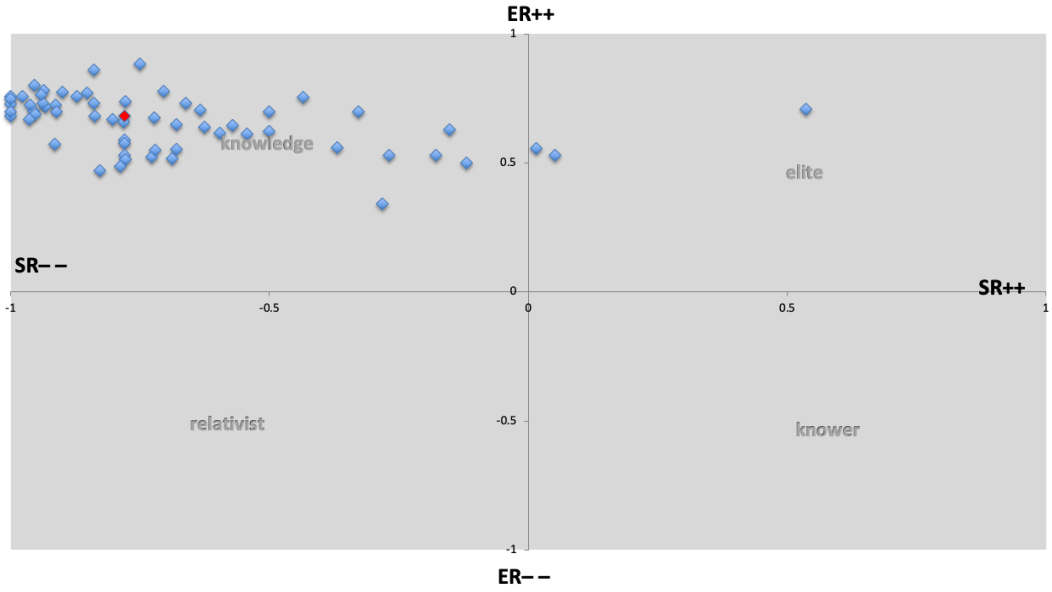


Figure 5.1 The positions of all marine chemistry RAs in the Visibility plane, as well as the median position (in red), clearly show the predominant knowledge orientation in Specialization.

If, then, in Specialization all "practices are about or oriented towards something and by someone" (Maton, 2016a, p. 12), the Visibility of knowledge and knowers in marine chemistry puts the discipline squarely in the knowledge code. In other words, legitimate writing about knowledge claims in marine chemistry emphasises sub-disciplinary topical Themes, signalling focus on specialised objects of study and procedures for studying such objects, while de-emphasising the role of actors in disciplinary claims. Additionally, more than 75% of the RAs are positioned between ER+0.5 and ER+1 and SR-0.5 and SR-1, signalling not only a very strong knowledge code orientation of writing practice, but also a substantial uniformity of practice, with only 13 RAs not in the sub-quadrant. Of the 13 RAs, 3 RAs are located within the elite quadrant, all above ER+0.5, showing that while not common, marine chemistry knowers are at times afforded a more visible status in originating sub-disciplinary claims.

5.1.2 Visibility in seismology

In the seismology sub-corpus, the majority of knowledge claims start with disciplinarily specific topical Themes, followed by similar mean values of general academic, non-academic, and referential Themes (Table 5.4; note that the median values can be taken to represent a typical discussion (and conclusion) section of a seismology RA).

Table 5.4

The breakdown of descriptive statistics for Visibility of disciplinary knowledge in seismology.

Themes	MEAN	MEDIAN	MIN	MAX	RANGE
Disciplinary **ER + + (1)**	72%	71.1%	41.2%	92.9%	51.7%
General/cor academic **ER + (0.5)**	9.1%	7.7%	0%	29.4%	29.4%
Non-academic **ER – (–0.5)**	10.7%	11%	0%	27.1%	27.1%
Referential **ER – – (–1)**	8.2%	7.1%	0%	23.7%	23.7%

In seismology, similarly to marine chemistry, disciplinary Theme usage shows the largest range, but here it is even larger. Yet again, a more detailed examination of the large range reveals that only five RAs in the sub-corpus employ disciplinary Themes in fewer than 60% of claims; 22 RAs employ disciplinary Themes in at least 75% of claims, and the remaining 33 RAs employ ER+ + Themes in the range between 60% and 75%.

General academic Themes appear much less frequently, in fact, not at all in five RAs. In further 35 RAs, general academic Themes start fewer than 10% of claims. That means that two thirds of the seismology RAs employ this type of Theme very little. The infrequent usage is further confirmed by only five RAs containing at least 20% of general academic Themes. The remaining 15 RAs fit in the range of 10% to under 20% of claims.

Similarly to general academic Themes, referential Themes are also less frequent. Again, two thirds of the RAs employ them in fewer than 10% of claims (with 0% in

three RAs). In the remaining 20 RAs, claims start with referential Themes in 10% to 20% of cases (18 RAs), and more than 20% of cases (2 RAs).

Non-academic Themes are more frequent, with half of the RAs employing them in the range between 10% and under 20%. Four RAs feature non-academic Themes in more than 20% of claims. The remaining 26 RAs employ this type of topical Theme in under 10% of claims (4 RAs not at all).

Altogether, the findings reveal that the majority of the seismology RAs employ claims which begin with disciplinary Themes in 60% to 75% of cases, non-academic Themes in the range of 10% to 20%, and general academic and referential Themes in under 10% of cases.

Regarding knower visibility in claims, while seismologists do not overall feature in knowledge making prominently, the median values are noticeably different compared to marine chemistry, showing that knowers in this sub-discipline are more prominent than knowers in marine chemistry, when measured by the syntactic positions of citations. The majority of references are still parenthetical, though by a smaller margin, as median values of SR+ + and SR+ citations are nearly two to three times larger, respectively, compared to marine chemistry (Table 5.5; note that the median values can be taken to represent a typical discussion (and conclusion) section of a seismology RA).

Table 5.5

The breakdown of descriptive statistics for Visibility of disciplinary knowers in seismology.

Citation position	MEAN	MEDIAN	MIN	MAX	RANGE
Subject hea noun **SR + + (1)**	13%	9.8%	0%	62.5%	62.5%
Other (non Subject) **SR + (0.5)**	18%	13.4%	0%	100%	100%
In a parenthetica clause **SR – (–0.5)**	1.6%	0%	0%	15.4%	15.4%
Parenthetical **SR – – (–1)**	64%	69.2%	0%	100%	100%

Two seismology RA discussion sections, Seis 13 and Seis 53, do not include any citations. They can thus be considered outliers, and will not be included in the citation breakdown that follows.

The most frequent category of parenthetical citations spans the maximum range of 100%. In fact, in 10 RAs, parenthetical citations are the only type (100%). This low knower Visibility in one sixth of the RAs is partially counterbalanced by the fact that parenthetical citations do not appear at all (0%) in three RAs, and in further 16 RAs, their representation is below 60%. In 14 RAs, parenthetical citations make up 60% to under 75% of cases, and in 14 RAs at least 75% of citations are parenthetical. Counting the 10 RAs mentioned earlier, only 24 seismology RAs feature citations which are parenthetical in at least 75% of cases–a drop from the 46 RAs observed for the same category in marine chemistry.

The next most common citation types, though distribution-wise similar, are SR + and SR + +, respectively. Both cover large ranges. SR+ citations make up 100% in two

RAs, but these RAs are outliers, as the closest percent representations are 60%, 50% (2 RAs), 45%, 42%, 40% (2 RAs), 38%, and 33% (2 RAs). SR+ + citations never comprise more than 63% of all cases per article, with the next frequent representations reaching 55%, 50%, 41%, 40%, and 33% (2 RAs). Six and eight RAs contain more than 20% (but under 33%) of SR+ and SR+ + citations, respectively. Most SR+ and SR + + citations constitute the 10% to 20% frequency range (16 RAs for both). SR+ citations do not appear at all in 7 RAs, and SR+ + in 12 RAs, in addition to the 10 RAs with exclusively parenthetical citations mentioned above, making the under 10% range the least common for SR+ + (5 RAs) and SR+ (7 RAs).

The least common citation type is SR – (in a parenthetical clause). It does not appear at all in 45 RAs. In eight RAs of the seismology sub-corpus, SR– citations constitute fewer than 10% of the total, and in the remaining five 10% to 20%.

Taken together, the findings reveal higher knower prominence in seismology than in marine chemistry, given largely by the fact that parenthetical citations predominate (i.e., they constitute at least 75% of the citations in RAs) in only slightly more than one third of the sub-corpus (compared to more than two thirds in marine chemistry). Other notable contributing factors to higher knower Visibility in seismology are that SR+ citations constitute more than 30% of citations in 12 RAs and SR+ + more than 30% in 7 RAs, and that the under 10% range is much less common in seismology than in marine chemistry.

The percentages of the various types of topical Themes and citations in seismology can be converted to positions (from -1 to +1) on epistemic relations and social relations Visibility axes. The positions are tabulated (Table 5.6; note that the median values can be taken to represent a typical discussion (and conclusion) section of

a seismology RA) and graphically represented on the Visibility axes below in Figure 5.2.

Table 5.6

The descriptive statistics for positions on ER and SR Visibility axes in seismology.

	MEAN	MEDIAN	MIN	MAX	RANGE
Positions on ER Vis. axis based on Theme types	0.630	0.614	0.324	0.911	0.587
Positions on SR Vis. axis based on grammatical positions of citations	-0.428	-0.5	-1	0.7	1.7

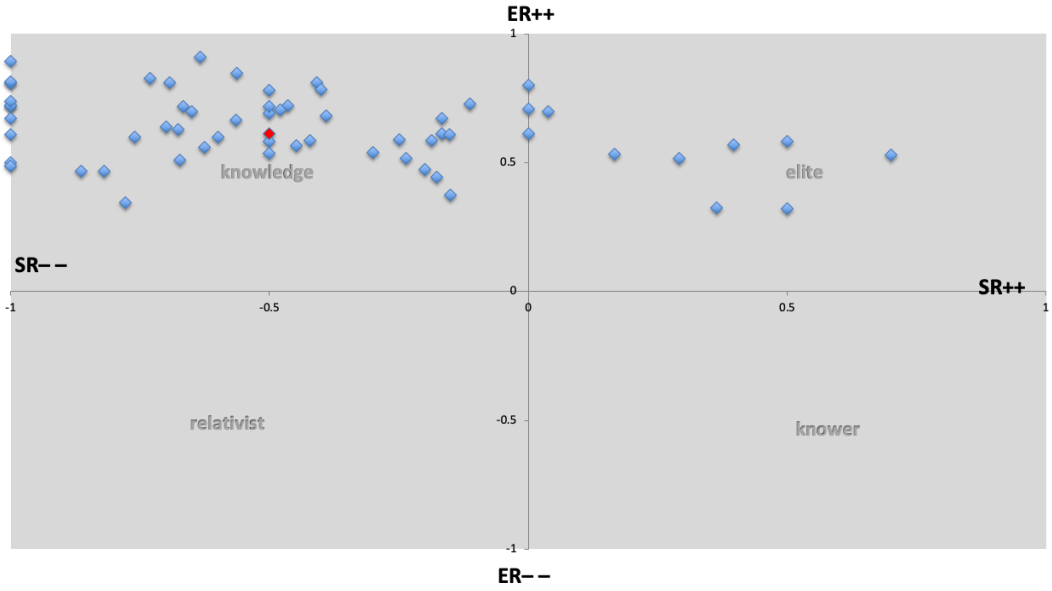


Figure 5.2 The positions of all seismology RAs in the Visibility plane, as well as the median position (in red), clearly show the predominant knowledge orientation in Specialization as well as increased knower Visibility.

The median values place seismology in the centre of the knowledge code quadrant, showing a stronger emphasis on specialised objects of study and procedures for studying such objects, while emphasising less the role of seismologists in claim making. Nevertheless, the spread of the 60 RAs is noticeable, and more than half of them (32 RAs) are not positioned within the smaller sub-quadrant (between ER+0.5 and ER+1 and SR-0.5 and SR-1) as was the case with most marine chemistry RAs. The spread is attributable to increased knower Visibility, although eight seismology RAs are below +0.5 on the ER axis as compared to only three marine chemistry RAs. Moreover, eight seismology RAs are positioned in the elite quadrant, again as compared to three marine chemistry RAs. The spread signals that legitimate depiction of knowledge and especially knower Visibility in seismology is less uniform, and at the same time that norms of what counts as acceptable disciplinary writing are more flexible.

5.2 Authoritativeness of knowledge and knowers

To answer the third and fourth specific research questions (i.e., about the level of epistemic certainty with which content is expressed and the level of knower authority), I profiled my corpus with respect to proportions of bare, boosted, and hedged claims and also with respect to the usage of first-person pronouns (i.e., which authorial role the pronouns fulfil: *we* as originators, *we* as opinion holders, textual *we*, or inclusive *we*). As a brief reminded of the methodology, in Chapter IV I posit that if Authoritativeness of disciplinary knowledge reflects the relative strength of a field's epistemic relations, then the relative strength will be greater when claims are more authoritative (bare or boosted), and weaker when disciplinary content is presented more tentatively. Likewise, if Authoritativeness of disciplinary knowers reflects the relative strength of social

relations, then the relative strength will be greater when knowers (authors) assume more authoritative roles as originators of claims or opinion holders, and weaker when they appear in the text-organising role or inclusively with other disciplinary members.

5.2.1 Authoritativeness in marine chemistry

If we equate knowledge claims with propositions (Section 4.3.2), about two thirds of marine chemistry claims are presented either as bare, boosted, or bare/boosted + modalized (Section 4.3.5). Claims of the lowest epistemic certainty (ER – –) are the next most frequent, while medium certainty and medium uncertainty claims are the least common, respectively (Table 5.7; note that the median values can be taken to represent a typical discussion (and conclusion) section of a marine chemistry RA).

Table 5.7

The breakdown of descriptive statistics for Authoritativeness of disciplinary knowledge in marine chemistry.

Certainty	MEAN	MEDIAN	MIN	MAX	RANGE
High **ER + + (1)**	61%	62%	39.6%	75.7%	36.1%
Medium high **ER + (0.5)**	7.5%	7%	0%	18.4%	18.4%
Medium low **ER – (–0.5)**	1.1%	0.8%	0%	5.9%	5.9%
Low **ER – – (–1)**	30.4%	29.4%	13%	56.4%	43.3%

Within the ER+ + category, between the minimum and maximum values, most RAs (27) employ the highest level of certainty in the range of >60% to 70% of their claims. The next most frequent high-certainty interval is >50% to 60% of claims in 17 RAs. The remaining 16 RAs in the sub-corpus can be divided equally between those

where the ER+ + claims make up not more than 50% of the claims and those where high-certainty claims are between >70% and the maximum of 75.7%.

Lowest-certainty claims (ER– –) are about half as frequent as the previous category. In comparison, they cover a larger range, whose examination reveals that minimum and maximum values are outliers: ER– – claims constitute under 20% of propositions in only two RAs, and more than 40% in five RAs. The majority of the sub-corpus RAs employ this level of certainty in more than 20% to 30% (30 RAs) and in the range of more than 30% to 40% (23 RAs).

Bare or boosted propositions with modalizing hedges or hedged bare/boosted propositions (ER+) are significantly less common. They represent under 10% (quite commonly under 5%) of claims in 44 RAs. The remaining 16 RAs employ them in the range of 10% to 18.4%.

ER– claims are infrequent. One type of ER– claims, as the reader may remember–boosted hedged propositions–appeared in the whole sub-corpus only nine times (Section 4.3.5). ER– claims do not appear at all in 21 RAs, and in the remaining RAs these claims do not make up more than 5.9% of propositions (in fact, often not more than 1%).

Altogether, a typical marine chemistry article is composed of propositions which are highly certain in more than 60% to 70% of cases, tentative in 20% to 30% of cases, with the remaining claims being of the medium-high certainty type.

Regarding knower Authoritativeness in claims by means of the first-person pronoun *we* (and *I* in the case of the single-authored RA MChem 14), marine chemists assume in the majority of the RAs authoritative roles of SR + + and SR+ frequently, while the organising textual role (SR–) is about half that frequent. Marine chemistry

authors-knowers use the inclusive *we*, through which they assume a communal identity (SR– –), the least frequently (Table 5.8; note that the median values can be taken to represent a typical discussion (and conclusion) section of a marine chemistry RA).

Table 5.8

The breakdown of descriptive statistics for Authoritativeness of disciplinary knowers in marine chemistry.

Self-mention role	MEAN	MEDIAN	MIN	MAX	RANGE
Originators **SR + + (1)**	36.5%	35.9%	0%	100%	100%
Opinion holders **SR + (0.5)**	33.7%	30.7%	0%	100%	100%
Textual **SR – (–0.5)**	12.3%	6.5%	0%	100%	100%
Inclusive **SR – – (–1)**	5.9%	0%	0%	100%	100%

Before describing the proportions of individual self-mention roles in more detail, I should address the obvious question suggested by an even cursory look at the table: Given that all types of *we* constitute "all or nothing" in at least some RAs (i.e., there are RAs in which each type of *we* is the only type of self-mention represented, while other RAs contain no self-mention), how many individual instances of self-mention are actually there in each article? Personal pronouns usage in research articles has been variously described in the EAP literature (Chapter II) as a challenge to objectivity (Starfield, 2015), marked (Harwood, 2005a), highly controversial (Hyland, 2001), and thus likely to be scrutinised (Starfield, 2015). Can we then suppose they may be used on the whole sparingly?

Table 5.9

Usage frequency of first-person pronouns in discussions (and conclusions) of marine chemistry RAs.

Number of pronouns	0	1	2 to 10	11 to 20	21 to 30	31 and above
Number of RAs	7	6	27	13	3	4

The raw counts in the table above show that writers of marine chemistry RAs certainly do not shy away from self-mention (even in single-authored MChem 14, both instances of *I* are of the SR+ + type). The differences in the raw count cannot be simply explained by article length. For example, the RAs with zero counts include both long, medium length, and short texts, and RAs with more than 20 instances include long and medium-to-shorter texts. Instead, the differences very strongly suggest that *we* usage is a stylistic feature, as well as *potentially* a philosophical one (Starfield, 2015, emphasis mine). Another interesting finding is that the number of RAs with no personal pronouns is counter-balanced by RAs with at least 21 instances. Even within the most common usage category of 2 to 10 pronouns per article, 15 of the 27 RAs employ personal pronouns at least 6 times. Then, Hyland and Jiang's (2016) suggestion that a growing usage of self-mention in research writing can also reflect the growing tendency in the sciences to highlight individual contributions and interpretations may be borne out.

One more observation regarding raw frequencies should be mentioned: the cases of single usage. If a personal pronoun is used only once per RA, what degree of knower Authoritativeness does it represent? Do marine chemistry authors make the single instance count, and represent themselves as highly Authoritative? The answer is that all roles are represented, and no tendency can be discerned. In the six RAs with a single

personal pronoun, there are two instances of SR + + and SR –, and one instance of SR+ and SR– –.

Returning to the frequency of the four types of self-mention, and excluding the RAs with no and one personal pronoun, the most common type, SR+ +, makes up at least 75% in 7 RAs (100% in 3 RAs), and between 50% but under 75% in 13 RAs. The finding that the SR+ + authorial role comprises at least 50% of personal pronouns in one third of the sub-corpus RAs shows a trend for marine chemists to assert their contribution in knowledge creation. SR + + pronouns comprise between 25% and under 50% in 15 RAs, and under 25% in 12 RAs (0% in 4 RAs).

The next most common personal pronoun type, with almost identical mean representation, is SR +. It makes up at least 75% in 6 RAs (100% in 1 RA), and between 50% but under 75% in 15 RAs, which are essentially the same proportions as SR+ + pronouns. Also similarly, SR+ pronouns comprise at least 50% of cases in one third of the sub-corpus RAs, again showing a higher degree of Authoritativeness of the marine chemist knower. SR + pronouns comprise between 25% and under 50% in 15 RAs, and under 25% in 11 RAs (0% in 5 RAs).

SR– personal pronouns are less common, reflected by their representing 75% of the cases in only one RA. The next closest usage range of 25% and under 50% applies to 7 RAs (5x 25% + 28.6% + 30%), meaning that when this type of pronoun appears, it makes up mostly under 25% of authorial self-mention (39 RAs, out of which 0% in 16 RAs).

The least common personal pronoun type is SR – –. It appears in merely 13 RAs: once making up 100% and once 50% of the cases, and under 25% in the remaining 11 RAs (largely under 10%, in fact).

Taken together, when marine chemists use personal pronouns to highlight their authorial role, SR+ + and SR+ types are roughly equally represented. SR– and SR– – types are either absent from most RAs or make up under 25% of self mention.

The percentages of statements at differing levels of epistemic certainty and the types of authorial roles in marine chemistry can be converted to positions (from -1 to +1) on epistemic relations and social relations Authoritativeness axes. The positions are tabulated (Table 5.10; note that the median values can be taken to represent a typical discussion (and conclusion) section of a marine chemistry RA) and graphically represented below in Figure 5.3.

Table 5.10

The descriptive statistics for positions on ER and SR Authoritativeness axes in marine chemistry.

	MEAN	MEDIAN	MIN	MAX	RANGE
Positions on ER Vis. axis based on Theme types	0.338	0.344	-0.148	0.667	0.814
Positions on SR Auth. axis based on grammatical positions of citations	0.414	0.530	-1	1	2

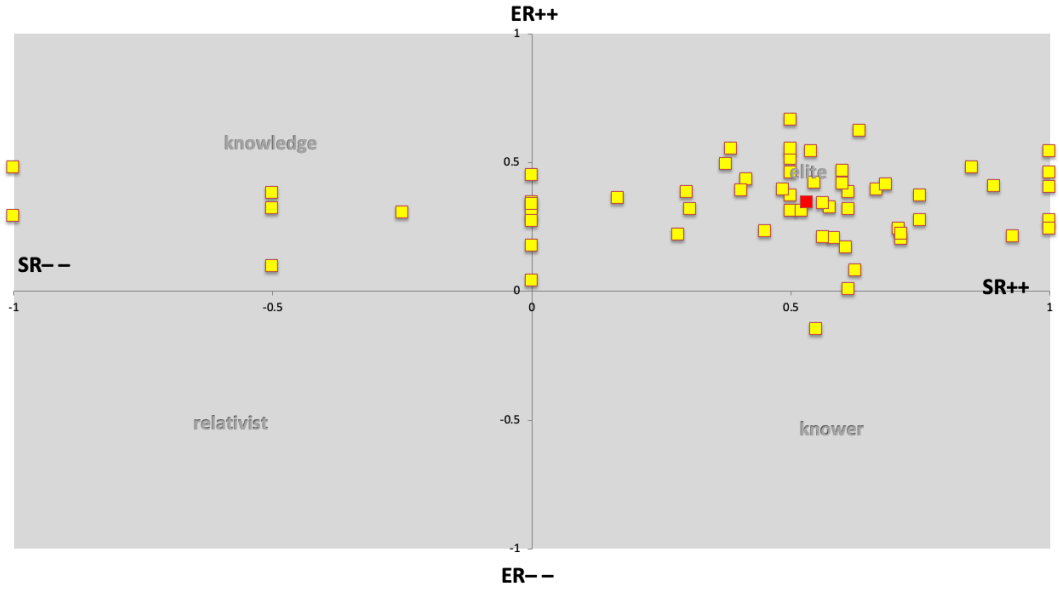


Figure 5.3 The positions of all marine chemistry RAs in the Authoritativeness plane, as well as the median position (in red), clearly show the predominant elite code orientation in Specialization as well as quite high knower Authoritativeness.

The median values position marine chemistry towards the centre of the elite code quadrant, showing that most disciplinary knowledge claims are presented with a medium-high degree of certainty, and marine chemistry knowers assert their personal role in knowledge creation. While a spread is noticeable (even if we disregard the seven RAs which do not feature any first person pronouns, and which are placed only on the epistemic relations axis), it is less pronounced in epistemic relations, where virtually all disciplinary claims are presented between slightly above 0 (neutral) and 0.5 (medium certainty). Regarding the strength of social relations, an interesting result is the fact that while some authors still de-emphasise the knower role in knowledge creation (the six RAs in the knowledge code quadrant), such authors are in the minority. More than half the sub-corpus articles (37) appear between +0.5 and +1 on the social relations axis, testifying to the knower importance, and authors of 10 RAs place the marine chemistry

knower between +0.17 and +0.5. A somewhat surprising finding is that one article is positioned in the knower code quadrant, making the discipline spread across three Specialization codes. This article is likely an outlier in the field, but at the same time shows that legitimate disciplinary writing in marine chemistry is to a degree flexible.

5.2.2 Authoritativeness in seismology

The median values of Authoritativeness of knowledge in the seismology sub-corpus are very close to the median values in marine chemistry: about two thirds of seismology claims are presented either as bare, boosted, or bare/boosted + modalized. The lowest epistemic certainty (ER – –) claims are about three times less common, but still the second most frequent. Medium certainty and medium uncertainty claims remain by far the least common, respectively. Even the minimal and maximal percentages are close, with the most notable differences in range for ER+ + and ER– – claims (Table 5.11; note that the median values can be taken to represent a typical discussion (and conclusion) section of a seismology RA).

Table 5.11

The breakdown of descriptive statistics for Authoritativeness of disciplinary knowledge in seismology.

Certainty	MEAN	MEDIAN	MIN	MAX	RANGE
High **ER + + (1)**	67%	65.9%	39.4%	94.7%	55.4%
Medium high **ER + (0.5)**	6%	5.8%	0%	16%	16%
Medium low **ER – (–0.5)**	1.2%	0%	0%	6.9%	6.9%
Low **ER – – (–1)**	25.8%	23.9%	3.6%	53.6%	50%

Within the ER+ + category, the extended range with a maximum of 94.7% (compared to 75.7% in marine chemistry) does not represent just a handful of outliers: in fact, 15 RAs, a quarter of the sub-corpus, employ the highest level of certainty in at least 75% of their claims, that is, beyond the maximum range in marine chemistry. Increased use of highly certain claims in seismology is also evident in the fact that the most frequent high-certainty interval, >60% to under 75%, includes 29 RAs, and that the interval of >50% to 60% of claims, which was the most frequent in marine chemistry, includes only 10 seismology RAs. In the remaining six RAs in the sub-corpus, ER+ + claims make up not more 50% of the total.

The lowest-certainty claims (ER– –) in seismology also cover a more extended range compared to marine chemistry. Unlike in marine chemistry, especially the minimum but also maximum values in seismology are less of outliers: these claims constitute under 20% of propositions in 15 RAs, mirroring the increased certainty of disciplinary claims, and more than 40% in 5 RAs. The majority of the sub-corpus RAs (29) employ this level of certainty in 20% to 30% of cases. In the remaining 11 RAs, the most common range is more than 30% to 40%.

Medium-certainty propositions (ER+) are significantly less common. They represent under 10% of claims in 48 RAs (0% in 11 RAs). The remaining 12 RAs employ them in the range of 10% to 16%.

ER– claims appear in only fewer than half of the sub-corpus articles. One type of ER– claims, boosted hedged propositions (Section 4.3.5), appears in the whole sub-corpus only eight times. ER– claims do not feature at all in 32 RAs, and in the remaining 28 RAs these claims do not make up more than 6.9% of the propositions (often 2% or 3%).

Altogether, a typical seismology article comprises propositions which are highly certain (ER+ +) in more than 60% and under 75% of cases, tentative (ER– –) in 20% to 30% of cases, with the remaining claims being of the medium-high (ER+) certainty type. ER– claims are seldom used.

Turning to knower Authoritativeness, authors in seismology describe their contributions to knowledge creation through SR+ and SR + + roles, in that order of frequency, while the textual role, SR–, is much less common. The least authoritative role, SR– –, is used sporadically (Table 5.12; note that the median values can be taken to represent a typical discussion (and conclusion) section of a seismology RA). The seismology sub-corpus includes three single-authored articles: Seis 13, Seis 22, and Seis 50. The first RA has no personal pronouns; the second includes the pronoun *I*; and the third one includes only one pronoun–*we* (underlined in the sentence below), perhaps to acknowledge the help received by the individuals listed in the article's acknowledgements in the research process: *"To verify [[whether this observation is not due to limitations in detection capability of the RESNES network]], we searched the Bulletin of the International Seismological Centre (ISC) for earthquakes [[occurring in our region of interest]]."*

Table 5.12

The breakdown of descriptive statistics for Authoritativeness of disciplinary knowers in seismology.

Self-mention role	MEAN	MEDIAN	MIN	MAX	RANGE
Originators **SR + + (1)**	28.5%	20%	0%	100%	100%
Opinion holders **SR + (0.5)**	35.8%	33.3%	0%	100%	100%
Textual **SR – (–0.5)**	9.7%	0%	0%	100%	100%
Inclusive **SR – – (–1)**	4.2%	0%	0%	66.7%	66.7%

The self-mention roles are once more used in an all-or-nothing fashion (i.e., there are RAs in which each type of *we* is the only type of self-mention represented, while other RAs contain no self-mention). Thus, before moving on to describing the proportions and percentages of the four self-mention roles, I will again provide an overview of the raw frequencies with which personal pronouns are used in the sub-corpus articles.

Table 5.13

Usage frequency of first-person pronouns in discussions (and conclusions) of seismology RAs.

Number of pronouns	0	1	2 to 10	11 to 20	21 to 30	31 and above
Number of RAs	13	6	32	5	3	1

The raw counts show that seismologists employ self-mention less than writers of marine chemistry RAs. Noticeable are especially the differences in the zero pronouns group (twice as common in seismology) and the more than 11 pronouns per RA

frequencies (half as common in seismology). Whether the disciplinary differences in the absolute numbers can be ascribed to article length (seismology RAs are on average substantially shorter) is beyond the scope of this work, as normalised counts per 1,000 or 10,000 words by themselves do not reveal anything about the form of knowledge or knower structures. Within seismology, article length is not a straightforward and satisfactory explanation either. Although the four RAs in the last two categories of the table (i.e., more than 21 pronouns per text) belong to the medium-length to long disciplinary texts, the five articles in the 11-20 range include several texts that are longer than the texts in the higher ranges, and the RAs with zero counts include both short and medium-length texts. More plausible explanations for the differences are stylistic preferences, perhaps a philosophical choice, as well as tendencies to make one's contribution known. These explanations are supported by analysis of Seis 59, one of the shortest texts in the corpus, whose 15 sentences contain 5 instances of *we*–a rate of one pronoun in every third sentence.

Regarding the six articles with only one pronoun, seismologists make the single instances count, so to speak: four of the articles include SR+ authorial roles, one SR+ +, and one the SR– role.

Returning to the frequency of the four types of self-mention, and excluding the RAs with no and one personal pronoun from the following description, the most common type, SR+, makes up at least 75% of personal pronouns in 7 RAs (100% in one RA), and between 50% but under 75% in 11 RAs. Thus, one third of the sub-corpus articles feature a medium-high level of authoritative pronouns in at least 50% of cases. These numbers are similar to the proportions of the most common type of authoritative personal pronouns in marine chemistry, SR+ +, but the difference in the degree of

Authoritativeness shows that seismologists represent themselves slightly less authoritatively. SR + pronouns comprise between 25% and under 50% in 13 seismology RAs, and under 25% in 10 RAs (0% in 6 RAs).

The SR+ + authorial role is the next most common, though the overall lower degree of knower Authoritativeness compared to the marine chemistry sub-corpus is evident here, too. While the SR+ + category makes up between 50% but under 75% of personal pronouns in 14 RAs, this proportion is counterbalanced by 14 RAs in which it comprises under 25% (0% in 5 RAs) of the cases. In further nine RAs, these pronouns appear in the range of between 25% and under 50%, which cannot be offset by only four RAs with at least 75% (100% in 1 RA) representation.

SR– and SR– – authorial roles feature minimally in the seismology corpus. Most (in 34 RAs for SR– and 36 RAs for SR– –) make up under 25% of the cases, though SR– roles are absent from 20 RAs and SR– – from 32 RAs. In the remaining ranges, both types appear almost accidentally: in the 25% to under 50% range, SR– pronouns are used in five RAs, and SR– – pronouns in three; in the 50% to under 75% range, both authorial roles appear twice (2x 50% for SR– and 50% + 65% for SR– –). None appears in the range above.

Taken together, seismologists represent their level of knower Authoritativeness most frequently as SR+ and SR + + in the 50% and above range, though in the case of SR+ + the Authoritativeness is offset by an equal number of RAs where the under 25% range predominates. The SR– and SR– – roles are scarcely used, and when they are, they constitute fewer than 25% of the total cases.

We can now convert the percentages of statements at differing levels of epistemic certainty and the types of authorial roles in seismology to positions (from -1

to +1) on epistemic relations and social relations Authoritativeness axes. The positions are tabulated (Table 5.14; note that the median values can be taken to represent a typical discussion (and conclusion) section of a seismology RA) and graphically represented below in Figure 5.4.

Table 5.14

The descriptive statistics for positions on ER and SR Authoritativeness axes in seismology.

	MEAN	MEDIAN	MIN	MAX	RANGE
Positions on ER Vis. axis based on Theme types	0.437	0.441	-0.1	0.895	0.995
Positions on SR Auth. axis based on grammatical positions of citations	0.373	0.449	-0.5	1	1.5

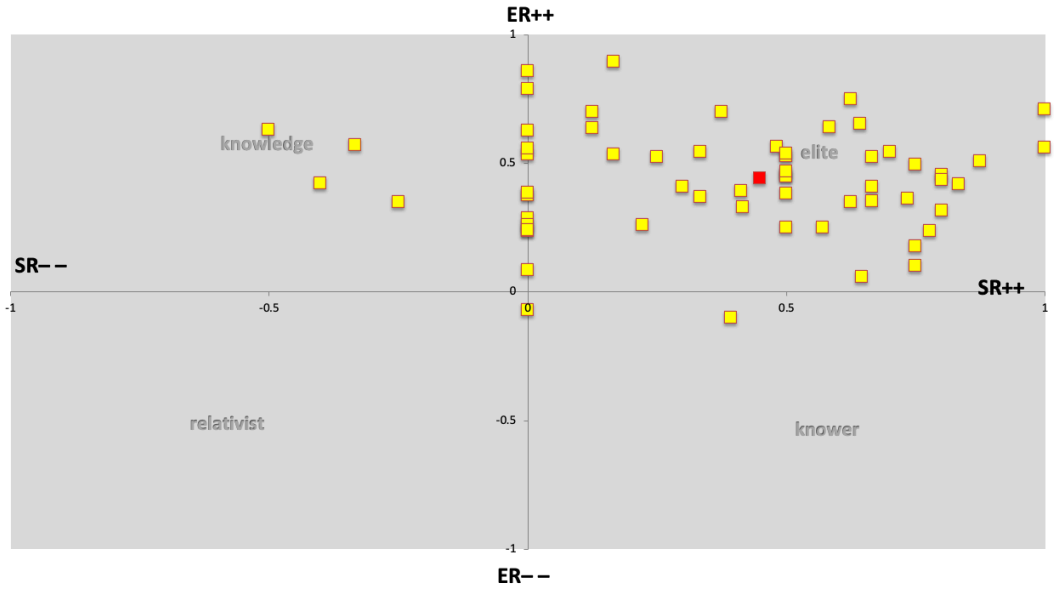


Figure 5.4 The positions of all seismology RAs in the Authoritativeness plane, as well as the median position (in red), clearly show the predominant elite code orientation in Specialization as well as higher levels of knowledge and knower Authoritativeness.

Similarly to marine chemistry, the median position in seismology is towards the centre of the elite code quadrant. Nevertheless, unlike marine chemistry, where a large number of texts cluster around the median position, the scatter of seismology texts along both ER and SR axes shows less uniformity (or more flexibility) in what is regarded as legitimate portrayal of knowledge and knower Authoritativeness. More seismology texts are above the +0.5 point on the ER axis, in line with the extended ER+ + range described above, showing that many knowledge claims are presented with a very high degree of epistemic certainty. Fewer seismology texts appear between +0.5 and +1 (29 vs. 37 in marine chemistry) on the social relations axis, but more (14 vs. 10) are placed between +0.17 and +0.5, again mirroring the above-mentioned overall lower knower Authoritativeness. The spread along the SR axis in seismology does not reach beyond -0.5, thus accounting for fewer texts compared to marine chemistry in the knowledge quadrant. Seismology also occupies the knower quadrant, a third Specialization code, though this article is likely an outlier in the field.

5.3 Summary of quantitative results

In this section, I have presented and qualitatively illustrated the quantitative results. The findings show that both sub-disciplines occupy the same quadrants in the Cartesian planes, and hence represent the same Specialization codes in both Visibility of knowledge and knowers and Authoritativeness of knowledge and knowers. Also, both sub-disciplines feature articles in which the writing styles and the portrayal of knowledge and knowers, as measured by the four linguistic features investigated, are vastly different; that is, in each sub-discipline, there are articles which appear on the extreme left or in the centre of the knowledge code quadrant and articles in the middle

or on the extreme right of the elite code. Existence of such differences in writing styles among individual articles of the same sub-field attests to existence of flexibility in the legitimate depiction of knowledge and knower in marine chemistry and seismology. On the other hand, in both types of Specialization, writing in marine chemistry displays clearer uniformity, with the tendency of marine chemistry texts to cluster around the median values distinguishing the sub-field from the more diffuse writing style of seismology. In the language of LCT, legitimate forms of knowledge, the rules of the game, can be interpreted slightly more loosely in seismology.

5.4 Qualitative results

The above-presented quantitative results have revealed that both marine chemistry and seismology occupy the same Specialization plane quadrants, although the distributions of individual corpora articles along the epistemic relations and social relations axes in Visibility and Authoritativeness present distinctive visual patterns. In the Visibility of knowledge and knowers, the distribution patterns of the sub-disciplines are more alike than in Authoritativeness, which then encourages further investigation, this time qualitative, into differences between these sub-disciplines of earth science. Of special interest are the differences between ER+1 (topical Themes) and SR+1 (integral citations that are grammatical Subjects) realisations. Topical Themes include specialist expressions identifiable as belonging to the field (including discipline-specific locations), and qualitative differences in *what* is fronted through thematisation could offer interesting epistemic relations insights into sub-disciplinary specialisation. In addition, through such a qualitative investigation, we can follow a much less trodden path of LCT research, and go one level deeper into the forms of sub-disciplinary

knowledge than what is afforded by the investigation of epistemic relations and social relations using the Specialization plane. Investigating highly visible citations can expose who the movers and shakers in the fields are–be they specific researchers or institutions or simply countries where the most visibly cited research originates, how old the cited knowledge is, and what kind of contributions knowers are prominently cited for. This kind of qualitative investigation of citations is, to the best of my knowledge, rare in EAP in general and non-existent in earth sciences.

5.4.1 Topical Themes in marine chemistry and seismology

Investigating ER+1 topical Themes qualitatively will necessarily be a preliminary exploration, since my corpus contains anywhere between 10 and 20 thousand ranking clauses. Given the scope of this work, I randomly selected a small sub-sample of 50 orthographic sentences per sub-discipline, yielding 60 ranking clauses in marine chemistry and 52 in seismology. The sub-sampled disciplinary topical Themes can be divided into roughly four categories based on the type of content that is fronted (topical Themes underlined):

a) Object of study/occurring phenomenon

Biological pump-driven changes in the prevalence and spatial distribution of marine anoxia can also help to explain an apparent contradiction between Phanerozoic trends in anoxia and hypothesized controls on the structure of animal ecosystems. (MChem 57);

The Mw 1/4 8.6 mainshock was followed two hours later by a Mw 1/4 8.2 aftershock, … (Seis 30);

b) Research process/method/approach/instrument

... the quantification of the particulate material release is challenging due to the difference in element concentrations between the fresh and saline water. (MChem 34);

Full moment tensor inversion of P waveforms at the frequencies [[used in this study (2–10 Hz)]] is a delicate procedure, ... (Seis 4);

c) Research product

Such isotopic values >20‰ heavier than the average values for seawater sulfate cannot be easily explained in the absence of an anoxic, stratified water column (Planavsky et al., 2012; Xie et al., 2012). (MChem 21);

Our moment tensor catalogue [[represented by the CDC model]] is presented in Figs 16(c) and (d). (Seis 4);

d) Geographical or study location/time period

Throughout the Phanerozoic there is a first-order inverse relationship between the carbon and sulfur isotope records, ... (MChem 21);

Between these segments, different deformations have been recognized southwestern from Lake Svan-Nur (areas "e, f" in Fig. 4). (Seis 23).

The last type of ER+1 Themes includes only five instances (four in marine chemistry and one in seismology), and given the overall small sub-sample, I have not considered this type further.

In both sub-fields, the most common disciplinary topical Theme type is object of study/occurring phenomenon. The small groups of sub-sampled Themes represent well the foci of the two sub-fields, as described in Section 4.2.1.1: marine chemistry is concerned with chemical elements and their circulation, concentration, distribution, reactions, and interactions in the seas and oceans, as well as how the elements are

influenced by biological and physical processes, as reflected in topical Themes such as *Ca and Mg*, *deep mixing*, *the transformation of Sr isotopic compositions of the particles*, *extreme enrichment of ^{34}SCAS in the Gaoyuzhuang formation*, *the shoaling of the aragonite saturation in the coastal zone*, *an intensification of the biological carbon pump in the productive coastal zone*, or *the decrease in oxygen solubility due to warming and the increase in oxygen demand from warming and weathering-induced heightened productivity*; seismology is concerned with earthquakes and their physics, seismic waves, fault motions, and natural hazard, as reflected in topical Themes such as *the Mw 1/4 8.1 1998 Balleny Islands earthquake*, *the generation mechanism of the near-fault ground motion pulses in the case of dip-slip faulting*, *R2 shears during coseismic rupturing*, *the aftershock distribution*, or *another cause of the subsidence along the surface ruptures*. Thus, beyond the obvious divergence in the studied phenomena, the most common topical Themes in the sub-disciplines do not qualitatively differ.

Themes concerned with research processes/methods can be considered together with Themes about research products, not only because research product Themes are significantly more frequent within the sub-sampled Themes, but because most research product Themes can also reveal, or at least hint at, legitimate ways in which objects of study are investigated. For example, from *our speciation modelling results* we can derive modelling; *86% of the Sr in the suspended particles* hints at calculations; and from *the distribution of relocated aftershocks along the profile* we can derive relocation procedures. These two Theme types combined are about twice more frequent in seismology. In addition, Themes in seismology provide us with information about approaches specific to that field (seismic wave inversion, aftershock relocation, backprojection, source–station geometry determination), whereas methods/products

mentioned in the marine chemistry sub-sampled Themes include only general (directly named or plausibly derived) procedures for both sub-fields (calculation, modelling, use and interpretation of instrumental and geological records).

Based on the qualitative analysis of the small sub-sample of disciplinary topical Themes, I propose two conclusions that clearly warrant further investigation. First, the higher specificity and frequency of mention of/allusion to research methods in seismology (despite a smaller Theme sub-sample than the sub-sample for marine chemistry) suggest that visibility of methods may play a more important role compared to marine chemistry when discussing research findings. Second, we can interpret these qualitative types of Themes, that is, Visibility of knowledge, following Maton's (2014a) 4-K model (Section 3.3). The 4-K model proposes a fine-grained differentiation of both epistemic relations (as four types of knowledge insights given by the relative strengths of ontic relations and discursive relations) and social relations (as four types of knower gazes given by the relative strengths of subjective relations and interactional relations), with types of insights relevant in this case. Themes concerning objects of study/ occurring phenomena can be interpreted as representing ontic relations (legitimate problem situations). Themes concerning methods/approaches can be interpreted as representing directly discursive relations (legitimate approaches), while Themes concerning research products as representing discursive relations indirectly. In both marine chemistry and seismology, the specificity of Themes concerning objects of study/occurring phenomena signals strong ontic relations in knowledge Visibility. The more general nature of directly mentioned (or plausibly derived from research products) methods/approaches in marine chemistry suggests weaker discursive relations in knowledge Visibility, while the combination of specific and general approaches in

seismology suggests more strongly bounded discursive relations in knowledge Visibility. Combining the strengths of ontic relations and discursive relations, we can propose marine chemistry and seismology sub-disciplinary insights (Figure 5.5).

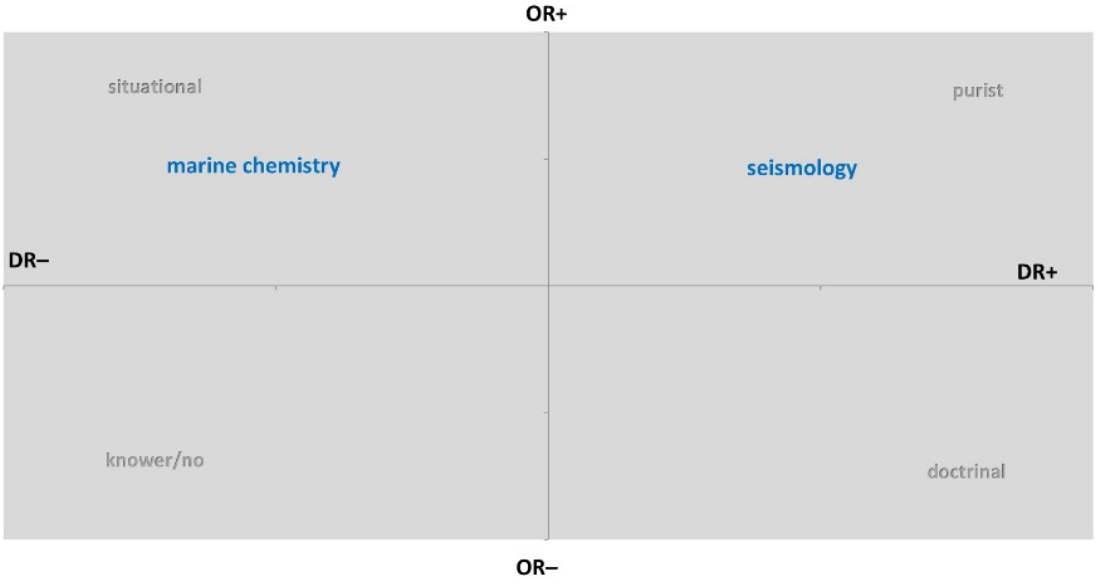


Figure 5.5 Proposed positions of marine chemistry and seismology in the epistemic Visibility plane.

5.4.2 Citations as grammatical Subjects: The movers and shakers in marine chemistry and seismology

5.4.2.1 Influential individuals, institutions, and countries

Moving on to exploring SR+1 citations, this time all the instances in my corpus, we can ask who the movers and shakers in marine chemistry and seismology are. To answer this question, we could count the number of instances in the corpus that an author et al. are referred to in a Subject position. Such an approach might not prove too insightful, as some discussion sections include multiple references to the same author et al., simply because that paper/research serves as a point of comparison (e.g., Zheng et

al. (2014a) in MChem 31), or the author et al.'s contribution to knowledge building in the field (e.g., a method, result, or comment) is divided into several successive clauses or sentences (e.g., Donner et al. (2015) in Seis 39). A better method of identifying influential individuals is to find whether any researcher (an individual author or first author) has been cited in multiple RAs in the corpus. In marine chemistry, there are four such researchers, while in seismology there are only two. Except for Matthew S. Fantle, an Associate Professor researching in geochemistry, who is cited in Subject positions for three different papers (two in MChem 28 and one in MChem 59), the other five knowers are cited for two papers, always one paper per RA (that is, for example, Professor Hiroo Kanamori is cited as the first author of his 1984 paper in Seis 9 and his 1974 paper in Seis 35). Alternatively, individuals could also be considered influential if they are cited in Subject positions in the same article for multiple papers they (first-)authored. Such a conception of influence expands the list of prominent authors by an additional 14 in marine chemistry and 6 in seismology, each author cited in connection with 2 papers. The final count of 18 influential marine chemistry authors out of 156 researchers in Subject positions and 8 seismology authors out of 112 reflects that only about 10% of disciplinary knowers are more influential[45]. Nevertheless, here caution against over-interpretation is advisable, since these 10% of authors are cited either in only two articles or for two different papers, and are thus only one SR+1 citation away from the remaining 90%.

More robust claims to influence and leadership in the fields arise if the conception of movers and shakers is expanded to institutions and countries where the

[45] The higher number of prominent authors in marine chemistry may initially seem somewhat counter intuitive, given that above I show seismology knowers to be on the whole more prominent. Nevertheless, this finding can be explained by the fact that marine chemistry discussion and conclusion sections are generally longer and include more references than in seismology.

prominently cited research originates. The numbers of such leaders remain small. In marine chemistry, there are four countries where most of the SR+1 cited authors hail from: the USA (70 authors), the UK (16 authors), Germany (15 authors), and France (12 authors). In the USA, the 70 authors work at many institutions, of which the most frequently appearing are the University of South Carolina (4x), Georgia Institute of Technology (4x), Northwestern University (3x), Columbia University (3x), Princeton University (3x), the State University of New York at Stony Brook (3x), the University of California in Santa Cruz (3x), the University of Tennessee (3x), and the Woods Hole Oceanographic Institution (3x). Among UK universities, only two are more prominent: the University of Southampton (5x) and Plymouth Marine Laboratory (3x). In Germany, most authors come either from Alfred Wegener Institute for Polar and Marine Research (5 appearances) or from Helmholtz Centre for Ocean Research in Kiel (4 appearances). All the French institutions appear one time each. Of note are also Hokkaido University and the University of British Colombia, from where four and three authors work, respectively, even though neither Japan nor Canada is among the top four countries.

In seismology, the leading countries where prominently cited disciplinary knowers work are the USA (27 authors), Japan (15 authors), and Turkey (9 authors)–all countries where earthquakes are frequent. Among the leading institutes are the US Geological Survey (7x), the Woods Hole Oceanographic Institution (2x), the University of California in San Diego (2x), Northwestern University (2x), the Massachusetts Institute of Technology (2x), Colombia University (2x), and the California Institute of Technology (2x). Among Japanese institutions are Tohoku University (2x), the University of Tokyo (2x), Kyoto University (2x), and Nagoya University (2x). In Turkey, two prominently cited researchers hail from Ankara University and Istanbul

Technical University. Of note are also institutes from countries outside the top three: again Germany's Alfred Wegener Institute for Polar and Marine Research (3x), Italy's Istituto Nazionale di Geofisica e Vulcanologia (5x), New Zealand's GNS Science (4x), and UK's Oxford (4x) and Cambridge (3x). Egypt's National Research Institute of Astronomy and Geophysics appears five times, although this institute is cited in the same article (Seis 43) about an earthquake in Egypt's Central Gulf of Suez, and is authored by Egyptian researchers. The results are summarised below in Table 5.15. Note, that several institutions (highlighted in red in the table) produce visibly-cited research in both sub-fields. Included in the last row of the table are also leading institutions outside of the leading countries.

Table 5.15

The leading countries and institutions from which prominently cited authors hail.

Marine chemistry		Seismology	
Leading countries	**Leading institutions**	**Leading countries**	**Leading institutions**
USA (70 authors)	University of South Carolina Georgia Institute of Technology Northwestern Universi Columbia University Princeton University State University of Ne York at Stony Brook University of Californi in Santa Cruz University of Ten nessee Woods Hole Oceanographic Institution	USA (27 authors)	US Geological Survey Woods Hole Oceanographic Institution University of Californi in San Diego Northwestern Universi Massachusetts Institute of Technology Colombia University California Institute of Technology
UK (16 authors)	University of Southampton Plymouth Marine Laboratory	Japan (15 authors)	Tohoku University University of Tokyo Kyoto University Nagoya University
Germany (15 authors)	Alfred Wegener Institu for Polar and Marine Research Helmholtz Centre for Ocean Research	Turkey (9 authors)	Ankara University Istanbul Technical University
France (12 authors)	No institution appears more than once.	–	–
Other (non-leading) countries	Hokkaido University University of British Colombia		Alfred Wegener Institu for Polar and Marine Research Istituto Nazionale di Geofisica e Vulcanolog GNS Science University of Oxford University of Cambridge

5.4.2.2 Pace of knowledge shift

Another interesting aspect of qualitative difference between the two sub-disciplines of earth science is the age of the works for which authors are cited in SR+1 positions (Table 5.16 below). In both marine chemistry and seismology, most of the SR+1 authors are recognised for research published either in the same decade that the corpus RAs were published (i.e., in the contemporary decade)[46] or in the decade immediately preceding (unsurprisingly, more citations refer to the preceding decade as my corpus does not cover the complete second decade of the 21st century). However, marine chemistry authors are cited primarily for more recent contributions, with the visibility of works dating prior to 2001 rapidly tapering off. In comparison, in seismology, the proportion of authors cited for contemporary contributions is noticeably smaller than in marine chemistry, while the proportion of authors cited for older papers is markedly higher. This finding begs the question of whether knowledge (or certain types of knowledge) becomes obsolete more rapidly in marine chemistry, which is suggested by the generally greater number of references in that sub-discipline compared to seismology. In that way, a larger amount of newly published knowledge in marine chemistry may replace older knowledge in the sub-disciplinary spotlight.

[46] My corpus includes RAs published only between 2012 and 2017.

Table 5.16

The percentages of the total number of RAs whose authors have been cited in SR+1 positions and of RAs by influential authors published in several decades show distinct sub-disciplinary trends.

	MChem		Seis	
	Total RAs	**Influential RAs**	**Total RAs**	**Influential RAs**
2011-2017	37.4%	35%	26.7%	25%
2001-2010	45.4%	46%	40%	12.5%
1991-2000	10.9%	19%	15%	37.5%
1981-1990	1.7%	0%	9.2%	12.5%
1971-1980	4%	0%	4.2%	12.5%
1970 and prior	0.6%	0%	5%	0%

The same general temporal trends in what sub-disciplinary knowledge is cited are revealed when we consider only the 18 and 8 influential disciplinary writers and their papers, 37 RAs in marine chemistry and 16 RAs in seismology. While marine chemistry traces the same sinusoid as above throughout decades, influential seismology knowers are cited more for older work. Even more intriguingly, seismological work done in the decade prior to the year 2000 is more frequently cited than contemporary contributions, with a quarter of the authors noted for papers published four to five decades ago.

To investigate the temporal trends further, we can try to identify the kinds of contributions that influential authors are cited for through time. In marine chemistry, influential authors publishing in the contemporary decade are cited almost exclusively

for two kinds of contributions. The first type is represented by procedures/methods/ approaches, such as

> *Zhang et al. [2015] include Hg sorption into DOC and DOC-associated OCRR in the Hg chemistry but do not conduct simulations to evaluate the depth-dependent Hg distribution in the oceans.* (MChem 31);
>
> or
>
> *Kriest and Oschlies [2011] identify a discretization problem [[where detritus flux is overestimated by popular transport schemes due to the variable level thickness of ocean models]].* (MChem 31).

Alternatively, contributions are site- or area-specific numerical results or descriptions of chemical concentrations or compositions, for example:

> *Moore et al. (2011) estimated that during each tidal cycle up to 70 104 mol DOC is exported by submarine groundwater discharge from the back barrier area of Spiekeroog (ca. 73 km²).* (MChem8);
>
> *Recently, Fantle (2015) suggested that the more open system dynamics at Site 1171 influenced recrystallization rates and that advection maintained local elemental and isotopic disequilibrium (with respect to Sr concentration and 87Sr/86Sr ratios) between the bulk solid and coexisting pore fluid.* (MChem 28);
>
> or
>
> *Lorenzoni et al. (2013) showed that while DOC concentrations remain relatively constant across the Cariaco Basin redoxcline, dissolved organic P (DOP) concentrations are elevated, exceeding waters immediately above and below the redoxcline by ~20–40 nM.* (MChem 55).

In the second half of the decade preceding the publication of corpus RAs, site- or area-specific findings continue to predominate. However, findings of influential authors published in the first half of that decade are of a more generalised nature, mostly concerned with behaviour of minerals or chemical elements under specific conditions, for example,

> *As Elzinga et al. (2006) explained, the surface defect density and types of surface sites available may differ significantly between aged calcite crystals equilibrated with solution and freshly cleaved calcite in solution at disequilibrium.* (MChem 20);
>
> *...while Poulton (2003) and Poulton et al. (2004) have showed that a significant proportion of the Fe(II) remains associated with the oxide surfaces for a considerable period of time during the reaction with sulfide, ...* (MChem 45);
>
> or
>
> *For example, Pedersen et al. (2005, 2006) found that ferrihydrite was transformed to lepidocrocite and goethite at low Fe(II) concentrations and into goethite and magnetite at high Fe(II) concentrations.* (MChem 45).

In even older publications, from the early 2000s or the 1990s, influential authors report still more general processes/trends for specific minerals, or area-specific findings involving the workings of large-scale dynamics, for example:

> *Van Cappellen and Qiu (1997) describe substantially lower reactivity of opal with age and depth in sediment.* (MChem 52);
>
> *Van Cappellen et al. (2002) show that the specific dissolution rates of core-top siliceous sediments can be up to three orders of magnitude lower than fresh surface ocean diatoms.* (MChem 52);

or

Beaufort et al. (2001) also document significant peaks at 30 ka in the productivity signal of the Equatorial Pacific Ocean. These authors assumed that these spectral peaks reflect the modulation phase of precession in the Equatorial Pacific Pool related to the Boreal Monsoon and ENSO-like dynamics. (MChem 58).

The trend in what influential authors are cited for through time supports the notion of more rapid obsolescence of reported knowledge in marine chemistry. The latest prominently cited authors' findings tend to be concrete and site specific, but they are probably almost completely replaced by new concrete and specific findings in the following decades. Older findings that retain their visible status in the sub-discipline are mostly more generalised and non-site specific, and thus better express timelessness.

The temporal distribution of what influential authors in seismology are cited for follows a different path. In fact, there is no discernible trend. Cited site-, area-, or event-specific findings are both contemporary and date back five decades, for example,

Schlindwein and Riedel (2010) found sounds of ongoing weak explosive activity at the rift flanks in 2001. (Seis 38);

Thakur (1992, 1998) reported existence of extensional faults beneath the Chamba syncline [[that cut the upper structural levels of the GHS]]. (Seis 60);

or

Kanamori & Cipar (1974) showed that a large and slow deformation occurred in the epicentral area of the major failure of the 1960 Mw 9.5 Chilean earthquake. (Seis 35).

One possible reason why site-, area- or event-specific findings are still reported in seismology decades after their original publication is the focus of the reported studies as well as the studies that cited them. Thakur (1992, 1998) is cited in a paper on seismicity in the Kangra-Chamba region, which was the area of a giant 1905 earthquake, and which is now experiencing intense seismicity. The potential for a new devastating earthquake may thus necessitate taking older findings into consideration. Similarly, Kanamori and Cypar's (1974) finding concerns the largest-ever instrumentally recorded earthquake. Its physical mechanism is referred to because the paper in which Kanamori and Cypar are cited focuses on a tsunami that in 1605 struck the western coast of Japan, with the tsunami having been disproportionately large compared to the low-intensity of reported shaking. Clearly, a repeat of such an event in the present day could place more people in harm's way, not to mention the material damage.

Another type of knowledge that influential authors are cited for concerns a specific tectonic or volcanic system behaviour–either proposals explaining the source processes of seismic signals (seismic signals accompany magma or plate movement), or long-term behaviour, such as:

> *Schlindwein (2012) proposed that these swarms are related to sites magmatic accretion [[that are distributed unevenly along the axis of ultraslow spreading ridges]] (Dick et al., 2003; Sauter and Cannat, 2010).* (Seis 38);
>
> *Ohminato et al. (1998) suggests that high Poisson ratios (~0.33) may be required to explain moment tensor solutions for tensile crack excitations [[inferred at that volcano]]. (Seis 54);*
>
> or

> *Platzman et al. (1994, 1998) carried out a paleomagnetic study in central Anatolia and they showed a CCW rotation, which presents over at least 10–12 Ma.* (Seis 21).

Also cited are proposals of physics-based models, general or area-specific, for example:

> *Jackson (1992) has proposed a different model, based on partitioning.* (Seis 39);
>
> or
>
> *Kanamori et al. [1984] proposed an equivalent force system appropriate for a volcanic eruption.* (Seis 9).

Overall, the cited findings reported by more influential authors in seismology better retain their currency and immediacy, since they are relevant to studies of recurring or potentially recurring natural hazards, or are concerned with generalised inner workings or physics of volcanic systems or tectonic plate movements. Older authors/findings may thus be prominently cited less so depending on their recency, but rather based on their relevance to a natural hazard being studied at a particular time (e.g., Kanamori et al. (1984) appears in an RA about a volcano in Ecuador that started an eruption cycle in 2010).

As a mini-test of whether the explanation for differences in time trends proposed based on the works of influential authors holds for non-influential authors, we can explore what types of knowledge are the oldest references in the corpus (i.e., those published before 1971). In marine chemistry, only one lead author falls in the category:

> *... by Deevey et al. (1963). In the concluding remarks, the authors state, "Meanwhile, we point to Green Lake as isotopically labeled 'queer' and wonder how many geologic records of meromictic lakes may be contributing to the confusion of geochemists."* (MChem 41).

The finding is site specific, but it is used to launch into a more general statement. As a side remark, interesting in this particular case is also the fact that Deevey et al. are acknowledged by means of a direct quotation, which is, based on my experience reading earth science research in the last several years, highly unusual. The choice to quote seems to confirm Hyland and Jiang's (2019) point that authors reserve "quotes for when they consider the original author's words to be the most vivid and effective way of presenting their case" (p. 72), because the authors of MChem 41 follow the quote by noting that "this perhaps jocular statement points to an additional important consideration when applying results from modern euxinic systems to sedimentary records of ancient oceans" (p. 52). Thus, although site specific, the oldest reference in the marine chemistry corpus in fact follows the trend of more generalised knowledge retaining a prominent citing status in the sub-discipline.

In seismology, there are six oldest SR+1 knowers, from 1944, 1946, 1951, 1952, 1962, and 1968. The earlier four references are site- or event-specific, for example:

> *Debrach (1946) and Rothé (1951) report damage to the submarine cables Brest–Casablanca and Brest–Dakar after the earthquake.* (Seis 41);
>
> *In the southern segment the previous researchers described normal fault displacements of 0.35–0.8 m (Treskov and Florensov 1952; Solonenko 1981) [[that could be either seismotectonic or seismogravitational displacements of the surface]].* (Seis 23).

These knowers are cited for investigations of either the largest-ever recorded submarine earthquake in the Northeast Atlantic basin (the 1944, 1946, 1951 references) or one of the strongest earthquakes in the Baikal rift zone. Studying again the underwater event in Seis 41 is seen as contributing to tsunami hazard assessment of the ocean basin, and

revisiting the seismotectonics of East Siberia in Seis 23 is presented as a continuation of interrupted work: The earthquake spurred investigations in the area, but attention was diverted several years later by a stronger event. The remaining older references acknowledge the knowers' contributions to descriptions of general physical processes:

> *Mogi (1962) showed an increase in rock heterogeneities results in higher b-values.* (Seis 24); and
>
> *Scholz (1968) showed the b-value varies inversely with stress level indicating that the seismic events grow to larger sizes if the stress level is high and that the b-value is higher in ductile rocks than in brittle rocks.* (Seis 24).

This mini-test has shown that older knowledge cited in the earliest references in the seismology corpus follows the pattern revealed by qualitatively evaluating knowledge contributions of influential authors: The currency of knowers' contributions cited depends little on the decade of publication, but more on their relevance to the studied natural hazards or the generality of the physical principles described.

If we continue investigating the differences in what knowledge contributions the remaining non-influential sub-disciplinary authors (i.e., those not cited in multiple corpus RAs or not cited for multiple papers) appearing in SR+1 positions have been acknowledged for, we find acknowledgement mostly for a method, approach, procedure, or sampling; a (numerical) finding, result, observation, or a more certain conclusion; a tentative finding or conclusion such as a suggestion, assumption, interpretation, inference, or proposition; textual information or a comment; and an overview of someone's research. Only one lead author, in seismology, is cited in connection with an evaluation of the cited work (the symbol << >> denotes an intervening clause):

This discrepancy may be explained if, << >>, Wallace et al. [2007] have overestimated the amount of deformation on the boundary between the Canterbury/Otago block and the Southern Alps and underestimated the strain in the north Canterbury plains. <<by employing a block model [[which assumes [[that most of the deformation occurs on the block boundaries]]]]>> (Seis 8).

I have further divided the contributions into what I term stand-alone knowledge claims and compound knowledge claims. A stand-alone knowledge claim is concerned with either a result, a method, a comment, etc., expressed in a single orthographic sentence, for example:

Tyrrell [1999] showed that phytoplankton control the deepwater N:P ratio through the competition between nonfixers and N fixers. (MChem 5)

or

Sodoudi et al. (2009) use the IASPIE91 earth model (Kennett & Enghdal 1991) (Seis 27).

A compound knowledge claim, which can be expressed in a single orthographic sentence or in several adjacent or subsequent sentences, comprises one of the following combinations of information:

- a (numerical) finding, result, observation, or a more certain conclusion + a tentative conclusion such as a suggestion, assumption, interpretation, inference, or proposition, for example:

  Brady and Walther (1989) report that, under alkaline pH conditions, dissolution rates are controlled by detachment of silicon from the negatively charged surface of silicate minerals. Brady and Walther (1989) argued that this effect can be attributed to the higher acidity of Si–OH groups << >>, and thus the

easier deprotonation of silanol groups, <<which results in weakening of the siloxane bonds>>. The authors state that hydrolysis of Al–O–Si bonds may be the rate-determining step, although this has only been observed under acidic pH conditions (pH < 3). (MChem 52);

- a method, approach, procedure, or sampling + a (numerical) finding, result, observation, or a more certain conclusion, for example:

  Likewise, Jin et al. (2006) found high CaCO3 export at 40° latitude using global nutrient and alkalinity observations and a biogeochemical model; (MChem 39);

- a method, approach, procedure, or sampling + a tentative conclusion such as a suggestion, assumption, interpretation, inference, or proposition, for example:

  Earlier Seeber et al. (2001) and Biswas (2005) had inferred such motion in this region based on structural and tectonic features. (Seis 5);

- a method, approach, procedure, or sampling + a (numerical) finding, result, observation, or a more certain conclusion + a tentative conclusion such as a suggestion, assumption, interpretation, inference, or proposition, for example:

  Using gravity data, Gettings and Griscom [1988] note that the structure near 3 km depth on the west flank is not sufficiently dense [[to be solid mafic rock]] and considered the case of either partial melt or felsic material [[decreasing the density]]. (Seis 14);

- other combinations (e.g., an overview of someone's research + a (numerical) finding, result, observation, or a more certain conclusion; an overview of someone's research + a method, approach, procedure, or sampling; or multiple textual information or comments), for example:

Ukawa and Ohtake [1987] analyzed a LFE beneath Izu-Ooshima volcano and concluded that a traction force [[induced by a magma flow]], << >>, is superior to a DC source. <<which can be treated as a single unidirectional force>> (Seis 6).

In marine chemistry, about 85% of stand-alone knowledge claims for which knowers are prominently cited pertain to findings and conclusions. Of these, the absolute majority (by far) are presented as results, observations, or epistemically certain conclusions. Tentative conclusions are more than four times less frequent. Many (about half) of the epistemically certain results and observations are site or area specific, or present numerical values, and are recent (published in the contemporary decade or only several years earlier), for example:

Nicholls and Trimmer (2009) found anammox activity accounted for <1-11% of N2 production across 8 temperate UK estuaries, ... (MChem 2);

For example, Woodson et al. (2013) measured higher Ba/CaSeawater in samples [[collected ~10 – 70 km offshore from locations with strong upwelling conditions]]. (MChem 4);

Gellatly and Lyons (2005) also reported stratigraphic shifts of up to 20‰ over intervals of less than 200 m from the Mesoproterozoic Belt Supergroup (~1.45 Ga), Montana, ... (MChem 21);

For instance, Charette and Sholkovitz (2006) found U removal rate of 0.4 mmol m2 d1 on the north-eastern coast of North America. (MChem 60).

Such a large number of recent numerical and site- or area-specific results mirrors the types of knowledge that influential authors are cited for in the same or recent decades, and supports the observation of "faster science" or faster turnover of findings (i.e., site-

or area-specific findings and knowledge claims are superseded by newer site- or area-specific findings or knowledge claims as site/area conditions change/develop or methods/technologies advance). The observation is further supported by the fact that in almost all the pre-2000 references, SR+1 knowers' findings/results are of a more general character, for example:

> *Thus, Price and Calvert (1973) demonstrated that, relative to anoxic sediments, oxic ones are enriched in iodine.* (MChem 37);
>
> *Aller (1980) indeed showed that burrows lead to significant decreases in porewater Si(OH)4 concentrations.* (MChem 42);
>
> *Weres et al. (1981) observed that the rate-limiting step in silica precipitation changed depending on the ambient concentration of adsorbed silicic acid. (MChem 52);*

The second most common type of stand-alone claims, though accounting for only about 7%, is textual information (i.e., what the cited author noted or stated, or what a text is about). About half are focused on phenomena, while the rest are purely text-informational, for example:

> *... however Whitehead et al. (2010) cautions that this trigger level may not be entirely appropriate for Tasmanian estuaries.* (MChem 2);
>
> *Ridame and Guieu [2002] noted that very small (i.e., only a few drops) rainfall events can contribute significant amounts of dust to the western Mediterranean.* (MChem 16);
>
> *Pichat et al. (2003) give little description of the non-carbonate phases present in their samples, although they mention clays and ferromanganese coatings, ...* (MChem 20).

Among compound knowledge claims, the most common types are methodology + a more certain conclusion, and a more certain conclusion + a tentative conclusion (suggested explanation or interpretation).[47] The combination of methodology/approach/ procedure with findings allows the author to either strengthen the finding, allow the reader to assess the finding, or explain differences in findings, for example:

> *Likewise, Jin et al. (2006) found high CaCO3 export at 40° latitude using global nutrient and alkalinity observations and a biogeochemical model;* (MChem 39);
>
> *Hain et al. (2014) demonstrated that negative discrepancies between observations of atmospheric D14C and a carbon cycle box model could be reconciled by imposing a deepening of the thermocline during HS1 and Younger Dryas Stadial (YDS) in the model.* (MChem 40);
>
> *Similarly, Bach et al. (2016) observed an increase in picoeukaryotes under increased pCO2, but no pre-industrial treatments were included.* (MChem 15).

A more certain conclusion + a tentative conclusion (suggested explanation or interpretation) compound claims are predominantly recent, most published one to four years prior to the article where they are cited (with the exception of one knowledge claim published in 1989, and one published 14 years earlier). In these claims, the cited knowers either explain their specific finding, or formulate the finding's implication, for example:

> *Scranton et al. (2014) noted a dichotomy between the dissolved inorganic N:P ratios [[measured in Cariaco Basin deep waters]] (approximately 16:1) relative to the rate of nutrient accumulation (11:1) and the overall ratio of total N:P in t he sinking particulate flux (approximately 5:1 to 12.5:1). These authors*

[47] The compound claim category that I term other combinations is as frequent as a more certain conclusion + a more tentative conclusion, but within the category, no combination appears more than once or twice.

hypothesized that the ratio difference was due to changes in the relative proportions of sinking particulate source material [[enriched in P relative to C and N]] and argued that both terrestrial and redoxcline sources were the likely culprits. (MChem 55);

Sugie and Yoshimura (2013) reported that the [Si]/SA and NSiP of P. pseudodelicatissima decreased with increasing pCO2 under various [Fe'] conditions suggesting that the effect of increasing CO2 levels is species-specific or differs between centric and pennate diatoms. (MChem 35).

In seismology, authors are also, unsurprisingly, cited in the vast majority of stand-alone knowledge claims for findings or conclusions, though these differ in their level of Authoritativeness (more or less epistemically certain). Compared to marine chemistry, slightly more than half of seismology findings and conclusions are introduced by tentative verbs (suggest, assume, interpret, propose, infer). What is mostly being tentatively advanced are knowledge claims about a behaviour of a particular tectonic or volcanic system or physical laws governing such systems, supporting my observation above that knowledge prominently cited in seismology is of a more general character, or is proposed to be generally relevant, and its currency then decays more slowly even decades after publication. For example:

As Gledhill et al. [2011] point out, this pattern of faulting seems to have occurred in at least two other recent earthquakes in the vicinity. (Seis 8);

Stein and Liu (2009) propose that the strain energy release from large intraplate earthquakes may dominate the local strain energy budget for many years following the main shock. (Seis 50);

Alternatively, MacLeod and Sherrod [1988] suggested that the presence of a long-lived magma body and repeated eruptions results in a thermal anomaly [[that extends from the magma body to the surface]]. (Seis 14);

As Molnar and Lyon-Caen [1973] suggest, if the P-axes of the earthquakes are more closely aligned than their slip vectors, then it is reasonable [[to suggest]] that the regional stress field – rather than the orientations of block boundaries – controls the distribution of deformation. (Seis 8).

The second most common type of stand-alone claims is textual information. It is slightly more frequent than in marine chemistry, but given the overall smaller number of SR+1 cases, the proportional representation of textual stand-alone claims in seismology is more significant (15%). Textual claims are also largely without time limitation on their relevance, even though such references are almost always very recent, for example,

Laperdin and Kachura (2010) noted the occurrence of similar phenomena after tens of years in the epicentral zone of the 1957, Ms =7.6 Muya earthquake. (Seis 23);

Nevertheless, Bukchin (2006) states that the combined inversion of Love and Rayleigh waves may help to overcome the problem due to their different amplitude ratios at different periods. (Seis 51).

Among compound knowledge claims, the most common types are methodology + a more certain conclusion and other combinations, although within the latter type, the most frequent combination is also a more certain conclusion with an overview, where the overview serves to introduce the conclusion (the example from Seis 6 above). The presence of methodology in compound claims again suggests that methodology is part

of an argument for or against a knowledge claim, allowing the author to strengthen its robustness and the reader to judge the claim's reliability, for example:

> *Based on GPS data, Márquez-Azua and DeMets (2003) suggest that the rate of extension in northern Mexico ranges from 1 to 3 mm=yr, at the 95% confidence level.* (Seis 45);
>
> *…where Nemati et al. (2011) computed a P-wave velocity for the upper mantle of 8.0 km s−1 and a Moho depth of 34 km using 458 travel times [[recorded in the IGUT network]],…* (Seis 27);

Alternatively, methodology serves as a point of comparison between findings, for example:

> *Castro et al. (2010) located a larger number of events near Cananea because they eliminated events [[that occurred only on the schedule [[provided by the mining company]]]] or events [[having wave forms typical of blasts]].* (Seis 50).

5.5 Summary of qualitative results

My investigation of the Visibility of knowledge (topical Themes) and knowers (integral citations that are grammatical Subjects) has revealed intriguing qualitative differences. While sub-disciplinary topical Themes concerned with objects of study strongly frame legitimate problems of marine chemistry and legitimate problems in seismology, topical Themes concerned with methods and research products frame legitimate approaches in marine chemistry less strongly than in seismology. The results can be interpreted as Specialization in the Visibility of epistemic relations following Maton's (2014a) 4-K model (Section 3.3), which proposes a fine-grained differentiation

of epistemic relations as four types of knowledge insights given by the relative strengths of ontic relations (between disciplinary knowledge and disciplinary objects of study) and discursive relations (between disciplinary practice or knowledge and other practices or knowledges). Based on the results, marine chemistry represents situational insight (i.e., the fronted Themes highlight a specialist nature of knowledge through sub-disciplinary objects of study, but a rather generalist nature of knowledge through methods and procedures for investigating these objects of study), while seismology represents purist insight (i.e., knowledge is presented as specialist through both objects of study and methods for their investigation).

Identifying lead authors as movers and shakers in the sub-fields based on citations of separate RAs has shown that very few individuals are cited marginally more than the rest. A clearer picture of influence emerges when we consider countries and institutions from where the lead authors hail.

Types of knowledge that knowers are prominently cited for reveal that marine chemistry findings which are site-or area-specific and numerical may become obsolete faster, while findings that are more general retain their relevance longer. In comparison, site- or area-specific and numerical findings in seismology retain their relevance throughout decades, as their currency is ensured by the sub-discipline's focus on recurring natural hazards. Seismological findings concerning general behaviour or physical laws remain current regardless of their time of publication.

In the following chapter, these findings will be compared to extant research in LCT and EAP/ESP to situate and discuss the findings of the study more broadly and to highlight the contribution of this study.

Chapter VI: Discussion

At the beginning of the thesis, I set out to explore a facet of disciplinarity of earth sciences, specifically some of the larger organising principles underlying knowledge building in the field. Two representative sub-disciplines were identified, and then investigated by combining and drawing on a framework from the sociology of knowledge (LCT) and linguistic research (EAP). In this chapter, the findings are discussed from the two perspectives, thus showing how this study aspires to further both research strands. The quantitative results regarding Themes, citation practices, hedging and boosting, and personal pronouns are discussed from LCT perspective in Section 6.1.1 as the Visibility and Authoritativeness of knowledge and knowers. The results of the sub-sample of topical Themes are discussed from LCT perspective in Section 6.1.2 as knowledge insights. The results of investigation of prominently cited authors are discussed from LCT perspective in Section 6.1.3 as a potential knower gaze. Age of knowledge and pace of knowledge change are discussed from LCT perspective in Section 6.1.4. The quantitative results regarding Themes, citation practices, hedging and boosting, and personal pronouns are discussed from EAP perspective in Section 6.3.

6.1 LCT perspective

6.1.1 Knowledge and knowers in marine chemistry and seismology

"For every knowledge structure there is also a knower structure… viewing the knowledge practices of intellectual fields as comprising knowledge-knower structures that specialize actors and discourses in different ways thereby provides a means of conceptualising differences among intellectual fields."

Maton (2014a, p. 72)

The overarching question posed at the start of this investigation into the underlying organising principles of earth sciences asked:

> What are some of the legitimate knowledge-making practices and bases of achievement in the sub-disciplines of marine chemistry and seismology, as evident from the usage of selected linguistic features in published research articles, and how do these practices and bases of achievement differ?

The theoretical starting point, as is evident from the opening quote, is Maton's (2016a) premise that all "practices are about or oriented towards something and by someone" (p. 12), that is, LCT's dimension of Specialization. Following on from this premise, Specialization is underpinned by the concepts of epistemic relations (between practices and objects of study) and social relations (between practices and actors of the practices), which can be stronger or weaker, thus generating four Specialization codes (knowledge, knower, elite, and relativist). In an attempt to uncover how specialised knowledge-making practices in marine chemistry and seismology are, this work proposes two new forms of such practices: Visibility of (sub-)disciplinary knowledge and knowers, and Authoritativeness of (sub-)disciplinary knowledge and knowers. By visualising the results in Cartesian plane plots, I have tried to map the sub-disciplinary topography of selected language forms. The results of investigating the Visibility of knowledge and knowers in the sub-disciplines show that both reside predominantly within the knowledge code, reflecting a stronger emphasis on specialised objects of study and procedures for studying such objects (stronger epistemic relations) through sub-disciplinary topical Themes (ER++), while emphasising less the role of knowers (authors) in claim making (weaker social relations) through dominant use of parenthetical citations (SR– –). Nevertheless, the sub-disciplinary knowledge code

profiles noticeably differ. In marine chemistry, more than three quarters of the corpus RAs are positioned within a smaller area of ER+0.5 and ER+1 and SR-0.5 and SR-1, signalling not only a very strong knowledge code orientation of writing practice, but also a substantial uniformity of practice. This convergence in writing is more obvious with less visible types of Themes. While under half of the marine chemistry RAs contain claims beginning with disciplinary Themes (ER++) in 75% of cases, more than half of the RAs feature 10% to 20% of general academic Themes (ER+), and almost three quarters use non-disciplinary (ER–) and referential (ER– –) Themes in fewer than 10% of claims.

In comparison to marine chemistry's concentration and general uniformity, seismology's profile is more diffuse. More than half of the RAs are not positioned within the smaller sub-quadrant (between ER+0.5 and ER+1 and SR-0.5 and SR-1), largely as a result of increased knower Visibility. The diffusion suggests that legitimate depiction of knowledge and especially knower Visibility in seismology is less uniform, pointing to a greater flexibility of norms of what counts as acceptable sub-disciplinary writing. The distribution of Themes shows that the majority of the seismology RAs employ claims which begin with disciplinary (ER++) Themes in 60% to 75% of cases, general academic (ER+) Themes in under 10% of cases, non-academic (ER–) Themes in the range of 10% to 20%, and referential (ER– –) Themes in under 10% of cases. Also, eight seismology RAs are below +0.5 on the ER axis as compared to only three marine chemistry RAs.

Both sub-disciplinary profiles also stretch into the elite quadrant, although only three RAs from marine chemistry versus eight RAs from seismology, showing that at times sub-disciplinary knowers are afforded a more visible status in originating claims.

Most commonly, though, Visibility of knowers (authors) is not prominent. The absolute majority of references in marine chemistry are parenthetical (SR– –), by a vast margin, with the other three knower Visibility categories (SR++, SR+, and SR–) being rather marginal on average. In seismology, the Visibility of knowers is higher. Parenthetical citations predominate (i.e., they constitute at least 75% of the citations in RAs) in only slightly more than one third of the sub-corpus (compared to more than two thirds in marine chemistry). Other notable contributing factors to higher knower Visibility in seismology are that SR+ citations constitute more than 30% of citations in 12 RAs and SR+ + more than 30% in 7 RAs, and that the under 10% range is much less common in seismology than in marine chemistry.

The results of investigating the Authoritativeness of knowledge and knowers in the sub-disciplines show legitimate bases of achievements shift to another Specialization plane quadrant. This time, Authoritativeness practices in marine chemistry and seismology reflect mostly the elite code (the median values place the sub-fields in the centre of the quadrant), which means that both epistemic relations to specialised objects of study and methods of inquiry as well as social relations to specialised knowers are stronger. In marine chemistry, knowledge claims are presented with a medium-high degree of certainty, with virtually all RAs positioned on the ER axis between slightly above 0 (neutral) and 0.5 (medium certainty). This uniformity of positions reflects the fact that a typical marine chemistry article is composed of propositions which are highly certain in more than 60% to 70% of cases, tentative in 20% to 30% of cases, with the remaining claims being of the medium-high certainty type. In seismology, the scatter along the ER axis is again noticeable, with comparatively more RAs positioned above ER+0.5, reflecting those texts contain

knowledge claims of a very high degree of certainty. Nevertheless, a typical seismology article is similar to a marine chemistry RA: it comprises propositions which are highly certain (ER+ +) in more than 60% and under 75% of cases, tentative (ER– –) in 20% to 30% of cases, with the remaining claims being of the medium-high (ER+) certainty type. ER– claims are seldom used. Interestingly, and somewhat surprisingly, owing to lower knowledge Authoritativeness, in one RA in both sub-disciplines, both fields also occupy the knower code quadrant, though the two texts are likely outliers.

Regarding the Authoritativeness of sub-disciplinary knowers, less concentration and more flexibility is evident is both fields, even though several RAs feature no first-person pronouns (7 in marine chemistry and 13 in seismology). In marine chemistry, in more than 60% of the RAs, authors assert their personal role in knowledge creation strongly, appearing between SR+0.5 and SR+1. This reflects that *We as originators* (SR++) and *We as opinion holders* (SR+) are the most common roles. In 16% of the RAs, first-person pronoun usage reflects overall knower Authoritativeness between SR+0.17 and SR+0.5. Only in 10% of the RAs are the knowers less Authoritative, thus shifting the articles into the knowledge code quadrant. In seismology, knowers are somewhat less authoritative, which results in a larger scatter along the social relations axis in the elite code quadrant. Only under 50% of the RAs fall between in the higher Authoritativeness spectrum of SR+0.5 and SR+1, but 23% are found in the lower to medium Authoritativeness range between SR+0.17 and SR+0.5. Fewer texts are placed in the knower code quadrant.

The above-summarised results (of the quantitative investigation) have revealed predominantly knowledge and elite codes as bases of achievement in marine chemistry and seismology. While superficially similar, the sub-disciplinary bases of achievement

are, in fact, distinct. The results reveal varying degrees of uniformity and flexibility within the dominant quadrants, as well as among quadrants (as indicated above, multiple Specialization quadrants are represented in both Visibility and Authoritativeness). The results thus affirm Maton's (2016a) premise that bases of achievement may not be universally adhered to. However, what this uniformity and flexibility reflect about the power dynamics in the sub-disciplines is hard to establish conclusively, since this research is synchronic, and the results are only quantitative. The results are a snapshot of writing practices within a short five-year window, representing a still image of the times–for as time periods are characterised and can be discerned by subjects of intellectual discussion and methods of approaching them (Bourdieu, 1971), periods may also be characterised by the prevailing use of linguistic features; the results may be pointing to changing practices (code shifts, code drifts; Maton, 2016a); they may be hinting at the fact that some actors (i.e., those authors whose RAs are outliers or closer to the margins of spread) do not recognise or understand predominating practices (Maton, 2014a); they could be suggesting an ongoing struggle for dominance (if so, then more obvious in seismology) between actors who control the power relations, and wish to conserve their social capital, versus those actors who want to subvert the extant rules and change what counts as legitimate contribution in order to increase their social capital in the sub-fields (Bourdieu, 1993); or they could be reflecting all of the above.

The results lend some support to Martin et al.'s (2020) contention that Bernstein's (1999) typology of knowledge structures (horizontal vs. vertical discourse and vertical discourse divided into horizontal and hierarchical) is a useful concept to start thinking with, but less useful analytically. Drawing on Bernstein, marine chemistry and seismology represent vertical discourse (i.e., explicit knowledge organised in

specialised symbolic structures), simply by virtue of being academic fields. Within vertical discourse, the strong epistemic relations in both Visibility and Authoritativeness point to hierarchical knowledge structures (Fig. 1.6 in Martin et al., 2020, p. 20) (integrating knowledge of different phenomena at lower levels to form generalised explanations), although at this point Bernstein's typology is not easily anchored to the theorisation of and data for Visibility and Authoritativeness. At the same time, with Bernstein's concepts, moving beyond a binary division is challenging, and topologies of sub-disciplinary distinctions are hard to visualise. Furthermore, knower structures are not featured in his typology. Of course, knower structures can be and have been extrapolated based on Bernstein's concepts (e.g., in Maton 2014a)[48], but we are still left with binary types that are difficult to apply to data. This application difficulty then led Martin et al. (2020) to contend that Bernstein's typology "can be left behind when one reaches specialization codes" (pp. 20-21).

Further observations, both affordances and potential problems, regarding LCT's ambition to offer enhanced explanatory power can be made based on this study's engagement with the first level of Specialization (results stemming from engagement with the 4-K model are reported in the following section), raising as many conceptual questions as it provides answers.

First, the methodology employed can be repeated and reapplied, allowing the Visibility and Authoritativeness of marine chemistry to be compared to those of seismology or of other fields of practice. The same can apply to other forms of

[48] Hierarchical knower structures are based on a conception of the dispositions (character, sensibilities) of an ideal knower at the pinnacle, and knowers at lower levels based on their developing dispositions. Horizontal knower structures characterise fields where the basis of achievement is knowledge of scientific principles and procedures rather than personal attributes.

discursive Specialization if suitable linguistic features can be identified to operationalise them.

Second, each specific realisation of Visibility or Authoritativeness, as represented in this study by each individual RA from marine chemistry or seismology, can be compared to all other possible realisations in the same sub-discipline or in other sub-disciplines. In this way, each possible realisation can reveal its degree of match or mismatch vis-à-vis the realisations dominating the sub-disciplinary practice. This comparison can then avoid, as Maton (2018) noted, statements similar to "agents feel like "fish out of water" because their habituses do not match the field" (p. 252, original emphases) without revealing the basis of the fit.

Third, conceptualising Visibility and Authoritativeness as comprising the strengths of epistemic relations and social relations suggests that a particular practice can be explored in terms of more than one indicator (knowledge and knower). This duality resembles Maton's (2018) summary of Bourdieu's stand that practices originate from both agents' dispositions and positions in the field. Maton (2018) sees these two factors as horizontality of conceptual relations in Bourdieu's work (my comment on this interpretation is in Section 3.2.1), thus "offering a corrective to accounts that explore [practices] only either [as] the attributes of agents or their social contexts" (p. 251). At the same time, potentially as a contradiction, there is a question of whether relational horizontality is impacted when practices are researched using single concepts (e.g., only social relations): "indeed, each concept can be enacted alone" and "the basis of choosing how many and which concepts to enact cannot be legislated in advance–it is a matter of possessing the right gaze" (Maton, 2016a, p. 18). This quotation can be

contrasted with Maton and Howard's (2016) critique of especially qualitative studies drawing on Bourdieu's framework:

> This is particularly the case for studies using individual concepts, such as 'habitus' or 'cultural capital', decontextualized from Bourdieu's wider framework, indicating perhaps that, when not thinking in terms of 'field', methods appropriate to thinking relationally do not come to mind. (p. 52, original emphases)

True, the quoted sentence was used in the context of Bourdieu's methodological pluralism vs. fully qualitative methodologies of most studies utilising Bourdieu's concepts, but my point about horizontal relationality still appears valid.

Four, if one realisation (an agent's habitus or position in a field) can be explored in relation to other possible realisations, which Maton (2018) sees as vertical relationality, it is unclear how linguistic realisations of Visibility and Authoritativeness can be understood vertically. Verticality has, to my knowledge, been defined by Maton (2014a) as increasing abstraction and subsumption in knowledge structures or increasing proximity to an ideal knower at the apex of an imaginary triangle of knower structures. Bourdieu (1984) referred to verticality in the sense of hierarchical positions in social fields. Should, then, verticality of Visibility and Authoritativeness be viewed as proximity of individual realisations to a predominant sub-disciplinary realisation, itself represented by a mathematical mean or median, either of which is, in turn, contingent on variations in usage? Or should verticality be understood with respect to linguistic practices of influential individuals? Or does verticality even matter here?

Five, to quote Maton (2018) again, relationally-understood practice means conceptualising vertically "by revealing the relational X of, say, a habitus in order to

analyse "horizontally" by relating that X to the X, W, Y or Z of capital, position, practice, etc." (p. 254). Ergo, understanding the verticality of, say, Visibility, which is not clear as I have just pointed out above, can then help relate it horizontally to the verticality of Authoritativeness to move closer to a more complete perspective on discursive Specialization in sub-disciplinary writing. But if Visibility and Authoritativeness as conceptualised in this study are only one possible conceptualisation of other possible conceptualisations of Visibility and Authoritativeness (i.e., other linguistic or non-linguistic features might exist that could also be argued to represent the two forms of Specialization), then another level of horizontality within horizontality exists, with the occupancy of Specialization codes and/or spatial distribution within the Specialization plane diverging from my results.

The existence of this horizontality within horizontality mirrors the possibility highlighted in Singh's (2015) question mentioned in Section 3.3.1: What if the knowledge paradox cannot be resolved by the LCT framework, because LCT modelling increases ambiguity and uncertainty due to proliferation of types of knowledge and knowers? Of course, Singh asked the question because she suggested that the paradox for present-day knowledge society may not be as Maton (2014a) (or others) defined it, but it could lie in the need for certainty and control.

Singh's question points to another potential issue. If the conceptualisations of Visibility and Authoritativeness adopted in this study can be related horizontally to other possible conceptualisations of Visibility and Authoritativeness, essentially generating segmentally ordered conceptualisations, what should we make of Maton's (2014a) statement that an effect of myopic modelling of knowledge is a "seductive illusion" that "if studied, knowledge must be endlessly typologized" (p. 8)? Here,

Maton refers to models that offer "segmental typologies and taxonomies of limited explanatory power" (p. 8). Perhaps LCT may run into the same problem of these other models.

6.1.2 Knowledge insights in marine chemistry and seismology

Additional aspects of legitimate knowledge-making practices have been revealed in the qualitative part of my study. The first part of this investigation was focused on ER+1 topical Themes, and drew analytically on the distinction between ontic epistemic relations (OR) and discursive epistemic relations (DR) (4-K model, Section 3.3). In both marine chemistry and seismology, of the four types of topical ER+1 Themes (1. object of study/occurring phenomenon, 2. research process/method/ approach/instrument, 3. research product, and 4. geographical or study location/time period), objects of study represent strong ontic relations (OR+), clearly separating the foci of the two sub-fields. Themes expressing research processes/methods and research products are more specialised (unique to the field) and frequent in seismology, suggesting strong discursive relations (DR+), while in marine chemistry, the same topic types tend to be more general (with procedures not unique to marine chemistry), suggesting weaker discursive relations (DR–). The relative strengths of ontic relations and discursive relations then reveal two knowledge insights: marine chemistry's situational insight, according to which fields are specialised by the problems they investigate, although not by the methods and procedures used to investigate such legitimate problems; and seismology's purist insight, according to which clearly defined legitimate objects of study can be investigated only through specific means.

The fact that sub-disciplines of one field occupy different specialization or insight quadrants is not unusual. For example, Carvalho and Dong (2009) investigated four sub-fields of design studies, and found that engineering design represented a knowledge code; fashion design was situated in the knower code quadrant; architecture design occupied the elite quadrant; and digital media design sat at the intersection of knowledge and knower codes. Neither does occupying different quadrants clash with my quantitative results showing that the sub-disciplines occupy the same quadrants at a more general Specialization level. Each set of codes within LCT offers the possibility of investigating the same practices from another perspective, thus potentially revealing additional rules of the game (to use Bourdieu's metaphor), reaching beyond binary classifications, and avoiding dichotomies.

6.1.3 Potential born knower gaze

Investigation of SR+1 citations has allowed me to consider influential individuals, institutions, and countries (i.e., the movers and shakers) in the sub-fields. While any claim to individuals' influence based on the number of SR+1 occurrences found in this study is somewhat tenuous (the small number of first authors identified are one or two citations away from everyone else), leadership of countries and institutions is more robust. The country that accounts for the highest number of prominently visible citations in both sub-fields is the USA, which is not surprising given the number of institutions in the country and USA's traditionally dominant position in academia. More interesting, though, is the fact that several institutions are likewise prominent in both sub-fields: USA's Northwestern University, Columbia University, and Woods Hole Oceanographic Institution, together with Germany's Alfred Wegener Institute for Polar

and Marine Research. This finding offers an avenue for future intriguing investigation into institutional affiliations of (not only) highly visible authors (SR+1) in marine chemistry and seismology. The future investigation could be theoretically anchored in the Weberian concept of social closure, that is, "the process by which social collectivities seek to maximise rewards by restricting access to resources and opportunities to a limited circle of eligibles" (Parkin, 1979, p. 44). Given the competitive nature of academic publishing,[49] among the various applications of social closure, of special interest should be the competitive perspective, which has, for example, been suggested as an explanation for a revealed over-representation of publications originating from elite sociology departments in four journals between 1960 and 2010 (Perrucci et al., 2019). Elite departments enjoy a cumulative advantage, which is cyclical and self-perpetuating, including higher salaries and lower teaching loads, which allows the departments to attract faculty who can engage in more frequent and cutting-edge research, hire more research and support staff, recruit academically more successful students, and expose these students to latest ideas and novel methods and techniques. Researchers from such departments are then more likely to submit high quality research, publish in leading journals, and set trends.[50] Seen through the lens of LCT's 4-K model (Section 3.3), we can extrapolate that the over-representation of elite departments then points to a born knower gaze, constituted by strong subjective relations (SubR+), where knower legitimacy is bound to being a member of an elite institution (i.e., one is able to benefit from tangible material and human resources and

[49] The competitive nature is manifested in multiple ways, for example, the high rejection rates in some prestigious journals (Gould, 2014); the number of papers that remain uncited five years after publication (Hamilton, 1991); the existence of false academy–or numerous predatory journals (e.g., Eriksson & Helgesson, 2017); or the fact that scholarly reputation, prestige, and ranking are linked to publishing (Perrucci et al., 2019).

[50] This self-perpetuating system of rewards in science was famously described by Merton (1968) as the Matthew effect, both at individual and institutional levels. A similar self-perpetuating cycle was described by Hyland and Salager-Meyer (2008) (Section 1.4).

from the status the membership confers) as well as strong interactional relations (IR+), where knower legitimacy derives from interactions with significant others (i.e., exposure to and interaction with other high-impact individuals at the institutions, and exposure to new ideas). Whether prominent institutions identified in my study (listed above), or institutions from which all SR+1 authors in my corpus hail, or indeed institutions of all cited authors, reflect such born gaze cannot yet be concluded, but this possibility is certainly worth investigating, both for the sub-disciplines and for prominent earth science journals.

6.1.4 Age of knowledge and pace of knowledge change

Further investigation of SR+1 citations has shown that in both sub-fields prominently cited knowledge was published mostly either in the same decade (contemporary decade) as the corpus articles or in the immediately preceding decade. In marine chemistry, the knowledge claims are primarily contemporary to recent, while visibly featured knowledge from before 2001 rapidly tapers off. In contrast, in seismology, recent to contemporary contributions are noticeably fewer. At the same time, knowledge from before 2001 is cited prominently much more frequently. This difference suggests more rapid change (or faster obsolescence) of reported sub-disciplinary propositions in marine chemistry, linked to the types of prominently cited knowledge claims. In marine chemistry, faster obsolescence is characteristic of findings which are site-or area-specific and numerical, while findings that are more general retain their relevance longer. In contrast, in seismology, site- or area-specific and numerical findings retain their relevance for many decades, as their currency is ensured by the sub-discipline's focus on recurring natural hazards. Similarly, knowledge claims

concerning general behaviour or physical laws remain current regardless of their time of publication.

The trends in the age of references and the pace of knowledge change revealed have, what might be called, LCT potential: The findings could in the future be linked to the LCT dimension of Temporality. When this dimension was initially formulated in Maton's PhD (Maton, 2005b), it operated with ideas of age (with older institutional and disciplinary practices venerated as wise and having stood the test of time) and the rate of change (with some disciplines focused on canonical knowledge, and knowledge progressing by revisiting and slowly adding details to established truths). The initial formulation of the dimension was used in two studies known to me, one of the discipline of marketing at a South African university (Arbee et al. 2014) and the other of academic nursing in Ireland (McNamara, 2010). In both studies Temporality was investigated through interviews of practitioners or teaching staff and students. In addition, Arbee searched the disciplinary literature to trace the changes in the definition of marketing and in disciplinary trends, and also looked at course documents and assessment tasks. No quantitative operationalisation was presented. Temporality has since been undergoing revision (currently, it is under advanced testing). At the present, Temporality is known to comprise the concepts of temporal position (TP) and temporal orientation (TO), yielding four temporal codes: prospective (TP+/TO+), retrospective (TP–/TO–), restoration (TP+/TO–), and renovation (TP–/TO+) (Maton, 2016b). Barring any major change in the underlying concepts in the revision process, I speculate my findings could reveal the temporal facet of the rules of the game in writing about knowledge and knowers in marine chemistry and seismology. At the same time, the

findings constitute tentative steps towards operationalising Temporality with respect to academic writing.

6.1.5 Contribution to LCT

One contribution of this work to LCT is the proposed conceptualisation of two forms of knowledge in the dimension of Specialization (i.e., literally how social fields of practice establish their claims to being special vis-a-vis other fields of practice): Visibility of knowledge and knowers and Authoritativeness of knowledge and knowers. The two proposed forms are accompanied by translation devices so that the forms can be applied to data. Firstly, I have posited that if the Visibility of disciplinary knowledge reflects the relative strength of a field's epistemic relations (i.e., the field's prominence in its focus on specialised objects of study and procedures), then the relative strength will be greater when more disciplinary topical Themes are used, and weaker when topical Themes are mostly represented by general academic, non-academic, or referential constituents. Topical Themes, it will be remembered, are the first elements that play a role in transitivity, and express the aboutness of clauses. Likewise, I have posited that if the Visibility of disciplinary knowers reflects the relative strength of social relations (i.e., the field's prominence in its focus on who originates a knowledge claim), then the relative strength will be greater when knowers (cited authors) appear in grammatically more prominent positions as Subjects, and weaker when cited authors appear in other positions or parentheses. Secondly, to translate between Authoritativeness and my data, I have postulated that if the Authoritativeness of disciplinary knowledge reflects the relative strength of a field's epistemic relations, then the relative strength will be greater when claims are expressed with higher degrees of

epistemic certainty (bare or boosted), and weaker when disciplinary content is presented more tentatively. Finally, if the Authoritativeness of disciplinary knowers reflects the relative strength of social relations, then the relative strength will be greater when knowers (authors) assume more authoritative roles as originators of claims or opinion holders through first-person pronouns, and weaker when they appear in the text-organising role or inclusively with other disciplinary members.

These two new forms of Specialization illustrate that the dimension lends itself to investigating multiple facets of what makes social fields of practice special, although a potential drawback is horizontal segmentalisation of knowledge and knower types (Section 6.1.1). Visibility and Authoritativeness thus join a plethora of already-investigated forms of Specialization as diverse as, for example, English paper examiners' comments on what constitutes successful papers (Anson, 2017), profiling sub-disciplines of the field of design in order to create a suitable e-learning environment (Carvalho et al., 2015), types of EFL learners who benefit from blogging activities (Chen, 2015), the design of games used to teach students of dental technology at university (Vahed et al., 2016), or the practice of writing about abstract symbols as an apprenticeship in Freemason lodges (Poulet, 2016). Ultimately, by expanding the already enacted conceptulisations of Specialization, Visibility and Authoritativeness aim to contribute to uncovering additional underlying principles and forms of knowledge.

Forms of knowledge, as pointed out by social realists, have real-world consequences (e.g., Maton, 2014a; Howard & Maton, 2011). LCT's stated goal is to reveal these consequences especially in education. To date, LCT studies have found, for example, that educational success may be impacted by students' ability to appropriately select the level of abstractions and combine them with concrete examples (Georgiou,

2016); that knowledge building may be fragmented instead of cumulative, limiting educational achievement (Martin, 2013); that educational difficulties of foreign students at a university can be correlated with a code clash between on the one hand the students' educational dispositions (what constitutes valued or desired achievement) transferred from their home country and on the other hand the host institution's constructivist teaching practices whose underlying dispositions remain hidden (Chen, 2010); or that the shifting expectations of a school subject at primary, secondary, and GCSE levels can impact the subject's uptake/popularity (Lamont & Maton, 2008). Such real-world impact is implied by the findings of this study as well, since the results show some of the knowledge and knower forms present in successful sub-disciplinary writings (bearing in mind that publication in a journal is a sign of acceptance by peers, as mentioned several times before in this thesis). For example, Visibility (or fronting) of the right kind of knowledge (as represented by topical Theme types both in the quantitative and qualitative results) can impact the recognition of a genre (Lavid, 2000) and discipline (e.g., Gosden, 1993); Visibility of knowers (as reflected in the proportion of integral types and non-integral citations) has been linked to ontological-epistemological-methodological assumptions of academic fields (e.g., whether knowledge is external to us or influenced by our cognition; whether knowledge can be objectively described or needs to be experienced subjectively) (Hu, 2018), and thus can also be recognised as writing in a discipline; Authoritativeness of knowledge (as represented by the degree of epistemic certainty in hedging and boosting) can mark a contribution worthy of publication–as Swales and Feak (2012) remark, one should be "confidently uncertain" (p. 163); Authoritativeness expressed by first-person pronouns

can highlight personal contribution, which, Hyland and Jiang (2016) noted, can be aimed at securing credit and furthering academic careers.

Furthermore, if we accept the postulate of LCT's Specialization that all knowledge is about something and by someone, findings of this study show the varying extents to which forms of knowledge in marine chemistry and seismology are impacted by relations to objects/methods of study and relations to actors of research. Some of these varying extents are demonstrated quantitatively as well as qualitatively through the two proposed forms of Specialization. Especially the plotted quantitative results show a wide range of what counts as acceptable sub-disciplinary writing–the further apart individual sub-disciplinary RAs are in the Specialization plane, the better they demonstrate that, even in marine chemistry, which is found to be more uniform. For example, through the differing use of integral citations and first-person pronouns across the 60 RAs from the same sub-discipline, authors portray knowledge either as more objectively constructed, free from human agency, or as a process in which researchers play an active role. My results thus suggested that the positivism vs. social constructivism dichotomy, referenced briefly in Section 2.3.3, offers less potential for insight if the dichotomic positions are to be associated wholesale with specific academic fields or even with general divisions of academic fields (e.g., into soft vs. hard, pure vs. applied).

The position expressed at the end of the previous paragraph perhaps merits further explanation. In general, the stance that dichotomies are either less insightful or unhelpful is, of course, not new. The limits of dichotomic view of knowledge or knowledge production are inevitably mentioned by social realists (e.g., Maton, 2014a, 2016a; Van Krieken et al. 2014; Young, 2008) but also by Bourdieu (e.g., 1990b): "Of

all the oppositions that artificially divide social science, the most fundamental, and the most ruinous, is the one that is set up between subjectivism and objectivism" (p. 25). For social realists, knowledge is both socially constructed and also objective/with real effects (Section 3.1); for Bourdieu, both modes of knowledge "are equally indispensable to a science of the social world" (p. 25), and their gains should be preserved. In the specific case that is referred to in the paragraph above–differences or tendencies in disciplinary writing–acceding to a wide applicability of the dichotomy despite this study's results would, I suggest, imply one of two positions: one, a member of a (sub-)discipline exists, to use Bourdieu's (e.g., 1984) language, within a structured structure, "the objectivist mode" (Grenfell, 2011, p. 27)–this is how we write; two, a member of a (sub-)discipline socially constructs a structuring structure, "the subjectivity mode" (Grenfell, 2011, p. 27) in order to build a facade of impersonality about the phenomena in the physical world, or to capture, more subjectively, the human world irreducible to controlled variables. My position on (sub-)disciplinary writing is both informed by this study's findings and transplanted from Bourdieu's view of habitus: a (sub-discipline) is conterminously a structured structure and a structuring structure. The uniformity in my results hints at what is acquired by being exposed to the modus operandi in the (sub-)field, while the differences show that language usage is not wholly determined by it.

At the same time, inter- as well as intra-sub-disciplinary uniformity and differences found in marine chemistry and seismology raise questions about the reasons for "socially sanctioned relative degrees of orthodoxy and non-orthodoxy" (Grenfell, 2011, p. 49) related to legitimate ways of expressing disciplinary knowledge. These questions cannot be answered without at least a diachronic study (since fields are

subject to change) as well as ethnographic investigation. Absence of an ethnographic component, or even better Boudieusian methodology comprising research object construction, three-level field analysis, and participant objectivation (e.g., Grenfell, 2010) is certainly a limitation of this study, for analysing language usage in terms of individual differences themselves does not reveal the degree to which "such variation is deeply affected by personal background–educational, professional and social *habitus*– and the specific *field* sites of operation" (Grenfell, 2011, p. 227, original emphases).

In summary, then, dichotomies such as positivism versus social constructionism or soft versus hard disciplines are broad-brush divisions which may be more useful to initiate our thinking of the underlying ontological-epistemological-methodological assumptions of disciplines, or for the most general of comparisons, but they may be less useful if we try to apply them to data. Therefore, for nuanced categorisation underpinned by empirical data, rather than thinking in terms of typologies (types of two, three, four, etc. categories), we should think where possible in terms of typologies-topologies (infinite positions within typologies). With more studies designed to reveal the topological variation of (sub-)disciplinary knowledge-making practices, we could then be in a better position to replace binary perspectives (the stated aim of social realism).

The same recommendation and caution of limitation applies to LCT-inspired studies where broad coalitions of disciplines like natural sciences/humanities/social sciences (Hood, 2011) on the one hand and specific sub-disciplines (e.g., of applied linguistics, education, and psychology, (Hu & Cao, 2015)) on the other hand are described as representing knowledge or knower codes. This thesis shows that if we delve into more detail, we may very well find that (sub-)disciplines occupy a particular

code when one form of knowledge/knower realisation is investigated, but another code when another realisation is the focus, as is the case of marine chemistry and seismology occupying predominantly the knowledge quadrant in Visibility and the elite quadrant in Authoritativeness. Additionally, this thesis shows that (sub-)disciplines can occupy multiple codes simultaneously even when the same knowledge/knower realisation is examined.

The recommendation to think in terms of typologies-topologies where possible or feasible brings me to the biggest methodological contribution of this study to LCT: quantification of external languages of description (i.e., the translation devices), which enabled me to chart more exact positions of each RA within Specialization code quadrants, thus taking advantage of at least two advances LCT concepts are said to offer compared to Bernstein's concepts: topologies and within-subject differences (Martin et al., 2020). While the positions are still approximate due to the four-point coding scales used (see Section 7.3), to the best of my knowledge, such quantitative precision as in this work has been described only in Maton and Howard (2016), who recounted developing an instrument for a study of perceptions of six subject areas by university students. Their survey instrument comprised three questions about knowledge and dispositions necessary to succeed in each of the subjects, with four possible answers to each question (coded by numerical values 1-4, and then calculating mean values to arrive at positions on the epistemic relations axes and social relations axes). Two other studies quantified their results. First, epistemic relations and social relations were quantified in a survey of designers' perceptions of four design disciplines (Carvalho et al., 2015). The same 4-point Likert scale as above was used and means calculated, but the results were not used to pinpoint exact positions, but only to place individual design

disciplines within one of the four Specialization quadrants. Both studies are versions of the same instrument. Carvalho et al. (2015) used another survey instrument (calculating mean values from 0 to 3 depending on whether participants completed a sentence with 1 to 3 words associated either with epistemic relations or social relations), but again only to determine approximate positions in quadrants. Second, Chen (2015) surveyed students' attitudes to seven pedagogical aspects of an L2 blogging project. Each aspect was coded as being either knower or knowledge oriented, with students scoring them on a 6-point Likert scale from 0 (least beneficial) to 5 (most beneficial). The mean values showed how strongly students valued the knowledge- or knower-oriented aspects; no precise position was determined, and elite/relativist quadrants were not mentioned. Except for these three studies, all other studies employing LCT Specialization (or indeed other dimensions) that I am aware of are more qualitative, determining code modalities only approximately–indeed, Maton and Howard (2016) stated that studies utilising LCT had been "overwhelmingly" (p. 50) qualitative. In general then, LCT research has not yet taken fuller advantage of topologies. In addition, none of the studies was designed to reveal precise intra-disciplinary differences. Such a conclusion is not necessarily negative, as "one only needs as much theory as the problem-situation demands" (Maton, 2016a, p. 20). At the same time, determining approximate positions predominantly qualitatively and representing disciplines as a whole does have caveats we need to be aware of.

Another methodological innovation this study proposes is the plethora of new external languages of description: two sets for ER/SR[51] and one set for OR/DR, and potentially SubR/IR, and TP/TO. Especially the ER/SR pairs are quantifiable and

[51] One of the SR concepts was inspired by Hood's (2011) examination of projecting sources (Section 4.3.4).

robustly anchored in SFL theory or previous EAP research. All the external languages of description can be applied to published RAs from any academic (sub-)discipline to tease out some of its internal organising principles.

Last but not least, by focusing on earth science, at the level of sub-disciplines, this work expands the range of academic disciplines whose underlying principles LCT-inspired studies investigated, for example, mechatronics engineering (Wolff, 2017; 2020), engineering design (Carvalho & Dong, 2009), nursing (McNamara, 2010), music (Lamont & Maton, 2008), marketing (Albee et al., 2014), humanities (psychology, English language and literature, historical studies, film and media) (Luckett & Hunma, 2014), or graphic design (Giloi & Belluigi, 2017). This work also expands the cooperation between LCT and EAP, which is an area where some studies already exist, for example, knowledge-knower structures drawn upon by Hu and Wang (2014) or Hu and Cao (2015); Autonomy employed by Tann and Scott (2020); or Semantics utilised by Kirk (2017).

6.2 EAP perspective

While the main thrust of the thesis' contribution is towards LCT theory and methodology, this work also draws on disciplinary debates and findings in EAP.[52] In fact, in section 2.3 I argued that the thesis is premised on the existence of discipline-specific ways of communicating. Owing to this link, the results can be linked to several issues debated in the EAP community beyond the already briefly mentioned disciplinary

[52] Within the context of broader English language teaching, my study belongs to English for Specific Academic Purposes (ESAP), following Jordan's (1997) division (Section 2.1.1).

knowledge construction (i.e., positivism vs. social constructionism) and disciplinary community dissection according to Becher (1989).

6.2.1 Discourse community

Three decades ago, Swales (1990) formulated six criteria by which a discourse community (not only academic) can be identified. The criteria are sufficiently broad yet delimiting to usefully demarcate communities. My research, both the process and the results, allows me to comment on characteristics 1, 2, and 4, whose summaries are repeated below for the reader's convenience.

1) Discourse community members share public goals. These goals do not simply equate studying the same object or subject, since it is goal sharing which is criterial.

Swales (1990) exemplifies his point by writing that "the fact that the shared object of study is, say, the Vatican, does not imply that students of the Vatican in history departments, the Kremlin, dioceses, birth control agencies and liberation theology seminaries form a discourse community" (p. 25). While we may agree, in principle, with Swales' comment that discourse community goals cannot be defined based purely on their objects of study, we perhaps need more nuance in the case of academic discourse communities, whose objects of study are the communities' most discernible public goals. In terms of Bernstein's (2000) classification, disciplines draw boundaries to create their social space and divide labour with other disciplines; it is the strength of boundary insulation that protects disciplinary identity, discourse, voice, and rules. Following Bernstein, the object of study would then be criterial. The research reported

in this thesis suggests that these boundaries can be more permeable (i.e., Swales' non-equation applies) when particular public goals are shared with (an)other (sub-)discipline(s). For example, in my corpus multiple seismology RAs potentially "trespass" into the territory of volcanology by studying magma motions. In those RAs, seismologists are not trying to appropriate another discourse community's object of study or become volcanologists, but rather they share volcanology's public goal of trying to predict eruptions. Compared to volcanology, seismology focuses on earthquakes (its typical object of study) that magma motions generate, and applies typical seismology tools, thus maintaining their separate discourse community. Both discourse communities can benefit from each other's perspective (move closer to fulfilling their goals), and at the same time remain distinct, having other own public goals.

2) Discourse community members communicate through agreed channels (e.g., meetings, journals, and newsletters).

These agreed channels should not be understood as exclusively narrow–as being directed at one narrowly-specialised discourse community. In fact, agreed channels can be shared within a larger discourse community. For example, many earth science journals publish on a broad range of topics, and earth scientists often publish in these "generalist journals", as some of my colleagues confirmed, to reach larger audiences. This tendency is also reflected in my corpus: the same journals are the sources of both marine chemistry and seismology RAs (altogether up to 20 percent of the corpus RAs). Existence of such broader or shared channels in earth sciences acknowledges the interconnectedness of sub-disciplinary objects of study, as also illustrated above with volcanology and seismology, facilitates sub-disciplinary collaboration, encourages take-

up of findings across sub-disciplines (i.e., a study's results from one sub-discipline are the starting point of a study from another sub-discipline), and ultimately assists the broader community pursue its greater public goal of understanding our planet's processes and evolution.

4) Discourse community members communicate via genres that represent what the community values and how it operates. The genres bespeak appropriate topics, format, as well as ways of arguing and negotiating.

The results reported in this thesis show that appropriate formats and ways of arguing represent guidelines which can be stricter or more flexible. Swales (1990) himself recognises this, noting that the norms that communities use may range on a continuum from stringent to developing. In my corpus, some of the RAs are still recognised as belonging to a sub-discipline despite their vastly different positions in the Cartesian plane. Indeed, the differences visualised in Cartesian planes support Prior's (1998) claim that disciplinary communities are constituted by negotiating one's membership through participating, thus reaching an approximate consensus. Additionally, my results show that appropriate topics include temporal consideration, given by knowledge currency and knowledge obsolescence.

6.2.2 Specific vs. general EAP

The findings of this study support the position that EAP teaching should be discipline specific. In fact, we should not stop there. In the case of disciplines that are broadly constituted, such as earth sciences, we should teach appropriate forms and ways of writing and arguing at the sub-disciplinary level, if these are known, if the composition of the student body makes it feasible (e.g., a sub-disciplinarily relatively

homogeneous class of students), and if our students need exactly that (i.e., we are teaching students who are learning to write for publication in their field). My position here agrees with Swales' (2019) viewpoint that the level of genre analysis engaged in should follow from the pedagogical aim. He gives examples of eight genre analysis levels to illustrate the different needs of a class where the aim is to grasp the difference between academic papers and academic presentations versus a class of post-docs learning to write papers in their sub-field. If then the conditions of information availability, class homogeneity, and students' needs are met, it may indeed be difficult to agree with Widdowson's (1983) position that a specific focus is too narrow, limits students' communicative competence, and disadvantages students by teaching prescribed linguistic behaviour. As I argued above, forms of knowledge at the sub-disciplinary level have real-world impact. Moreover, the range of flexibility in the sub-disciplinary writing revealed by this thesis could raise students' awareness that writing is not a formulaic strait jacket. The range of flexibility revealed also raises questions about Spack's (1988) suggestion that disciplinary writing should be taught by disciplinary specialists. Disciplinary specialists are highly unlikely equipped with the knowledge of linguistic and sociological theories which is needed to arrive at the results reported in Chapter V. Even if they were, the aforementioned range visualised in this thesis through Cartesian planes is highly likely beyond a single specialist's experience and/or awareness.

From a certain perspective, my findings may seem to offer support for the anti-specificity claim: visually comparing the Cartesian planes for marine chemistry and seismology either in the Visibility or Authoritativeness of knowledge and knowers reveals overlaps (e.g., in the regions around the median values). For most (though not

all) of the individual RAs in one sub-discipline, another similarly positioned RA from the other sub-discipline can be found. Hence, if writers stay within this overlapping territory, the use (frequency, types) of the investigated features may not seem specific to a sub-discipline. Nevertheless, sub-disciplinary profiles should be seen as more than a collection of isolated representative RAs. Here, we could again refer to the clear differences in, first, the possible ranges–within single quadrants, the spread to two (in the case of Visibility) and three (in the case of Authoritativeness) quadrants, the higher or lower numbers of RAs not in the predominant quadrant–and second, in the clustering–for example, in marine chemistry, in the case of the Visibility of knowledge and knowers, more than 75% of the RAs are positioned between ER+0.5 and ER+1 and SR-0.5 and SR-1. Ranges and clustering may be influenced not only by the studied phenomenon within the same sub-field or the results arrived at, but also by individual preferences. This is acknowledged by Hyland (2005a) in one of his studies, when he points out that crafting of a disciplinarily acceptable authorial persona "is an act of personal choice, and the influence of individual personality, confidence, experience, and ideological preference are clearly important"… since "we are not the instruments of our disciplines" (p. 191). At the same time, the clustering shows that conventions in some sub-disciplines certainly seem to be stronger than in other sub-disciplines. Other clear sub-disciplinary differences beyond ranges and clustering are qualitative, for example, in the ages of citations and the types of findings/contributions that authors are prominently cited for.

Altogether, my findings support the EAP specificity camp, although not all the camp's proponents/standpoints. For example, Hyland's (2017a) formulation that "essentially, my argument is that each discipline draws on different lexical,

grammatical, and rhetorical resources to create specialized knowledge" (p. 9) could, perhaps, be too strong. Rather, disciplines can draw on the same linguistic resources, but to differing extents, which Hyland himself shows in many of his studies. The results reported in this thesis show that differences exist even within sub-disciplines, where they are more or less pronounced. Thus some RAs are still recognised as belonging to a sub-discipline despite their vastly different positions in the Cartesian plane.

This last point then leads me to comment on how research into linguistic differences by specificity proponents is frequently conducted and/or how findings are reported. In particular, I want to mention the influential research of Ken Hyland, not only because his approach has spanned several decades and inspired many similar studies, but also because three of the linguistic features quantified in this study (citations, self-mention, and heading and boosting) are subjects of his numerous single- and co-authored stance and engagement publications (e.g., 1998, 1999, 2000, 2001, 2002c, 2003b, 2005a, 2008a, 2016, 2019). In his research, Hyland reports differing distributions of various linguistic features (e.g., in addition to the ones just mentioned, rhetorical questions, directives, attitude markers) through quantitative means and through interviews with disciplinary insiders, to demonstrate the socially-situated nature of disciplinary knowledge building (from the point of view of LCT's framework, Hyland looks at how linguistic features portray/build knowledge and how they express the knower). The quantitative distributions in his work are presented as average values per 1,000 or 10,000 words, or per whole article rather than individual sections, which arguably reveals only a partial picture. Taking a specific example, what are readers (who may be novice writers or their teachers) supposed to derive from knowing that the pronoun *we* (as the most common form of self-mention reported) appears per 10,000

words 24 times in biology, 25.4 times in applied linguistics, 23.3 times in electronic engineering, and 39.3 times in physics (Hyland, 2001), beyond the obvious that physics seems different and using *we* in these disciplines is acceptable? How exactly will such information help anyone write in their disciplines? This information would be more insightful if accompanied for each discipline by details about the range to visualise (not necessarily in Cartesian planes) what is possible, and about clustering to visualise what predominates. If, in my research, the reported results comprised only the mean and median values of Visibility and Authoritativeness, which are rather similar, the intra- and inter-sub-disciplinary differences would remain concealed. Obviously, even this partial picture can expose some general tendencies, and is a useful starting point of comparison. Nevertheless, for research into (sub-)disciplinary writing differences to wield more explanatory and pedagogical power, and in order to more helpfully illustrate to those who are learning to write in (sub-)disciplines that "what counts as convincing argument, appropriate tone, persuasive interaction, and so on, is managed for a particular audience" (Hyland, 2017a, p. 9), EAP now needs to expand on pioneering approaches (or circumscribed studies, as Swales (2019) puts it). I hope I have shown that one possible, though certainly not the only, path is for EAP research to draw on LCT.

6.3 Research into linguistic features

To tease out further snippets of characteristics of earth sciences, in this section, I compare findings from the research reviewed in Chapter II into Themes, citation practices, hedging and boosting, and self-mention with this study's results about the four linguistic features. This comparison comes with a large caveat: the linguistic

features were in this study used for the purpose of translating between LCT concepts and my corpus data. Hence, the specific research questions asked in Section 3.6 (e.g., What level of knower authority is legitimate in sub-disciplinary research articles, as reflected in the use of exclusive personal pronouns?) served the purpose of assisting me with capturing the relative strengths of epistemic relations and social relations in the discussion and conclusion sections. And so did the methodological coding of the linguistic features, which, in the case of topical Themes and hedging and boosting, were categorised into typologies that only partially correspond to existing typologies in EAP (or SFL). The caveat thus underscores that neither the questions nor the methodology was formed for purely EAP research.

6.3.1 Themes

The just-mentioned caveat of the methodological variation between this study and other studies is compounded by the methodological variation among studies of Themes. Some studies (e.g., Ghadessy, 1999; North, 2005) used the T-unit (i.e., one main clause and all subordinate clauses attached thereto, Section 4.3.3) while this study also included examination of Themes of independent subordinate ranking clauses. Other studies focused on Themes as represented only by the grammatical Subject (e.g., MacDonald, 1992; Montemayor-Borsinger, 2003), while Themes investigated in this study included Circumstances as well, although most of the Themes in this work are expressed by grammatical Subjects, as is generally the case in texts (Halliday & Matthiessen, 2014; Gosden (1993) quantified this tendency as well). Comparison of this study's findings about Theme usage with other studies of Themes is thus challenging, although some approximate comparison can be attempted.

One possibility is to see whether this study's findings bear out two characteristics of earth sciences posited by Kleinhans et al. (2010) in Section 1.2.1: first, earth sciences have been said to incorporate elements of both a descriptive historical science and a natural science associated with causality and law formation; second, earth sciences are said not to be reducible to physics and chemistry. To review the two characteristics, this study's findings need to be compared with findings about Theme usage in history, and also in natural sciences in general and in physics and chemistry in particular.

Only two studies I am aware of included classification of grammatical Subjects/ Themes for research writing in history (Ghadessy, 1999; MacDonald, 1992). Unfortunately, Ghadessy's focus on abstracts makes his findings incomparable with mine, and MacDonald's categories of Themes are incompatible with categories in my study. The only conclusion that can thus be arrived at is less than insightful: that most grammatical Subjects (or topical Themes) in MacDonald's and this study are, unsurprisingly, (sub-)disciplinarily specific, referring to findings, methods, and inferences. This conclusion does not allow me to verify the first characteristic of earth sciences based on my thematic analysis.

As for comparison of Theme usage in marine chemistry and seismology with natural sciences in general, this study's result is in line with North (2005), who found that science majors portrayed knowledge building as emanating from plain facts (in my case sub-disciplinarily specific Themes) rather than building knowledge as constructed/ negotiated by individuals (personal pronouns, integral citations). Regarding comparing this study's findings with findings about physics and chemistry in particular, some reference can be made vis-a-vis Gosden (1993), who found that discussions in

chemistry, physics, and biology RAs employed 63.7% of Real World grammatical Subjects (research objects, procedures, and nominalised mental processes), and 12.5% of grammatical Subjects were first-person *we* (called *Discourse participant*) and integral citations (called *Interactive Participant*), although these percentages were undifferentiated by discipline. In my study, in marine chemistry, research objects and procedures (included in my ER++ category of sub-disciplinarily specific Themes) are thematised in 71.5% of cases and in seismology in 72% of cases. *Participant* Themes (roughly equivalent to my ER– category) appear in 8.2% (marine chemistry) and 10.7% (seismology) of the corpus clauses. The higher ER++ values and the lower ER– values in the earth science sub-disciplines, together with the fact that ER– Themes encompass more than *Participant* Subjects (i.e., in addition to self-mention and integral citations, there are non-academic expressions, non-disciplinary places and times), thus likely lowering even further the proportion of *we* and integral citations Subjects, hint that marine chemistry and seismology are distinct from chemistry and physics, on which they are based, and that earth science cannot be simply reduced to these disciplines.

6.3.2 Citation practices

Keeping in mind the limitation of comparing this study's findings, which are valid for discussion/conclusion sections, to previous research valid for whole RAs, several characteristics transpire. Citation types results in marine chemistry and seismology testify to altogether lower visibility of referenced authors, especially in marine chemistry, where at least 75% of citations are parenthetical in more than 75% of the corpus RAs. In seismology, 75th percentile of parenthetical citations applies to 40% of the RAs, but SR– – citations are still the most frequent type overall. These results

agree with Hyland's (2002a) general finding that authors in natural sciences and engineering tend to be less prominent than their counterparts in the humanities and social sciences. Nevertheless, the predominance of parenthetical citations in marine chemistry and seismology does not fully agree with three other previously reported general findings. First, Hood's (2011) assertion that natural science authors are rendered invisible largely by superscript citations: In my corpus, only 2 out of almost 50 journals use superscripts. Second, that integral citations in the natural sciences are comparatively infrequent (Hyland, 1999; Okamura, 2008). The operant word is *comparatively.* True, the mean and median values (respectively, in parentheses) for integral citations in marine chemistry (SR++: 8.9% and 5.7%; SR+: 8.5% and 5.5%) and seismology (SR++: 13% and 9.8%; SR+: 18% and 13.4%) are *comparatively* low, but in some RAs, integral citations appear more frequently. In seismology, SR++ citations comprise anywhere between 33% and 63% of all citations in seven RAs, and SR+ citations make up 100% in two RAs, and between 33% and 60% in further 10 RAs. While these findings could be partially explained by reflecting only the more discursive sections of RAs, the fact that the citation style of the several seismology RAs is so different, points towards personal rather than (sub-) disciplinary modus operandi. Third, according to Hyland (1999), integrally cited natural science/engineering authors are more frequently in passive or adjunct grammatical structures (equating SR+ citations in this study), which somewhat reduces authorial visibility. I found practically the same values for SR+ and SR++ (grammatical Subjects) citations in marine chemistry, and only slightly higher SR+ numbers in seismology, suggesting that at least in these two sub-fields, if authors use citation types consciously (Swales (2019) doubts that many stylistic

decisions are deliberate), they simply consider only in-sentence and parenthetical types when deciding how prominent cited authors should be.

Another aspect of disciplinary difference in citation usage is the age of cited references. In the natural sciences and engineering, references tend to be more recent, with the average age of references reported for biology (theses and dissertations) being 9 years (Pecorari, 2006) and for engineering (journal RAs) 13 years (Young, 2014), while in the humanities and social sciences (theses and dissertations in education and linguistics), the average age ranged from 19 to 22 years (Pecorari, 2006). Similar, though more nuanced, age trends have been reported in a study of 14 disciplines by Halevi (2013). For example, references in a history journal (Journal of Medieval History) were newer (i.e., published in the last 15 years) in 44% of cases; in physics journals, 67% of references were recent; in biochemistry journals, references were newer in 71% of cases; and in earth and planetary sciences, 71% of citations were recent. In this thesis, investigation limited to SR+1 references (summarised above in Section 6.1.4) reveals a comparable tendency to favour articles published in the last 17 years, with an earth science average of 74.75%. This combined average, however, conceals a significant difference between marine chemistry (82.8%) and seismology (66.7%), with the numerical distance almost equivalent to Halevi's reported difference between arts and humanities (60%) and engineering (79%).

With this comparison, I wish to digress, and briefly comment on what could be regarded as a somewhat problematic practice. The hinting above at the fact that the sub-disciplinary difference between marine chemistry and seismology is equivalent to the difference between hard and soft disciplines is intentionally misleading. Suggesting or concluding that two, three or four randomly selected disciplines represent the natural/

social sciences/humanities, as is frequently the case in the published literature and as also appears multiple times in this work when published sources are cited, needs to be considered as a broad extrapolation and a convenient shortcut. From Halevi's (2013) reported numbers we can clearly see that the "united states of hard disciplines" is not a monolithic disciplinary bloc, but rather "an association" which accommodates significant variety: for example, in computer science 51% of references were reported as newer vs. 68% in environmental science vs. 83% in material science. Thus, results regarding (sub-)disciplinary differences should perhaps be extrapolated into conservative statements rather than sweeping generalisations.

One explanation for the difference in the average age of references *could* lie in the well-known division of disciplines into urban and rural (Becher & Trowler, 2001), based on the number of researchers engaged in investigating specific problems. The former disciplines (e.g., physics, chemistry) are described as highly competitive and fast moving, since a large number of individuals are pursuing a relatively small range of narrowly defined problems. Articles are published fast, with papers generally on the shorter side. In the latter fields (e.g., history, linguistics), the range of questions/ problems investigated is broader, and the ratio of problem/researchers is relatively low. Since fast-paced fields tend to publish articles more frequently, more recent references should thus be available. This line of reasoning does suggest a partial explanation for the two earth science sub-disciplines investigated here, if we acknowledge that within urban fields, to which marine chemistry and seismology could be counted, some (sub-)disciplines are faster paced than others. A quick survey of marine chemistry and seismology sub-disciplinary corpora shows that five journals in marine chemistry and five in seismology are sources for 32 and 29 RAs, respectively. Two of the marine

chemistry journals publish 24 issues a year, one 16 issues, one 12 issues, and one 10, for a total of 86 issues. In seismology, one journal publishes 24 issues annually, two journals 12 issues, and two journals 6 issues, for a total of 60 issues. One of the marine chemistry journals (24 issues per year) also states that submission to initial decision takes only 15 days, and accepted articles are published on average 186 days following submission. Seismology journals, if information is available, are somewhat slower: two journals state that initial decisions will be known 60 and 56 days after submission (publication in the latter journal is on average in 231 days).

The potential explanation above is complicated by a finding that reference age differences do not correlate with the speed of growth in academic fields. While I do not have data specifically for marine chemistry and seismology, Halevi (2013) investigated the growth rate for 14 disciplines in the number of published RAs between 1960 and 2011 against the temporal tendencies in referencing, showing that more published articles and hence a plethora of potential recent (defined as having been published between 1996 and 2011) references do not determine the proportion of old to new references. For example, the fastest growing field, medicine, has 71% recent references; for the second fastest-growing engineering the percentage climbs to 79%; physics has grown three times faster than earth and planetary sciences, yet the proportions of recent references are 67% and 71%, respectively. With Halevi's result in mind, a potential explanation could be that the age of references in marine chemistry and seismology is influenced by relevance to a given research focus (Section 5.4.2.2): creating a record of changes in conditions in a specific place vs. general description of natural processes vs. focus on recurring phenomena. Applying this explanation to other (sub-)disciplines could prove insightful.

The third aspect of disciplinary differences considered here is the number of citations. More numerous citations have been reported for the humanities and social sciences (up to four times in some disciplines) than for the natural sciences texts (Hyland, 1999). The finding of the present investigation of an average of 36.7 citations in marine chemistry and 14.5 in seismology is in general agreement, and so is their combination into an earth science average of 25.6 with Halevi's (2013) reported 17.1 for earth and planetary sciences. All these reported numbers need to be interpreted with caution, though. First, this earth science average (25.6 citations) yet again conceals a huge difference between marine chemistry and seismology, which might at least partially be explained by the two-to-three times longer marine chemistry discussions/ conclusions compared to seismology, as well as the fact that marine chemistry authors frequently reference (at least in SR+1 citations) recent location-specific numerical results. Second, the numbers of citations reported in this study are valid only for discussions and conclusions, and thus the actual average numbers per complete RA can be higher, perhaps close to the lower values reported for the social sciences/humanities. Third, the numbers reported in previously published studies may confirm or disconfirm any gleaned trends. For example, Hyland (1999) himself acknowledges that molecular biology, with a high number of citations per paper, does not follow the trend reported for the natural sciences. On the other hand, Hu and Wang's (2014) finding about general medicine (which is a related field to Hyland's molecular biology) conforms to the trend. Flottum et al.'s (2006) finding about medicine is in line with Hyland's high number of references. Halevi (2013) reports very different numbers for molecular biology (50.7) and medicine (37.1), and likewise high and low values for various natural sciences and social sciences/humanities. Granted, the discrepancies might be due to methodological

differences (corpus size, sampling procedure, etc.). The question then is whether we, as language analysts, should better keep to the discipline rather than generalise beyond it. Thus, for example, it certainly is worth considering the generalised explanation of the numerical difference by natural scientists' focus on well-defined and narrowly specific issues as opposed to the concern with broader and topically diverse problems in the social sciences and humanities (Hyland, 1999), but this proffered reason should be validated against results based on large samples from multiple individual disciplines. Hyland's explanation appears difficult to reconcile with Becher and Trowler's (2001) informants' mentions that biologists are socially expected to "read everything relevant" (p. 114), historians may see citations as a mark of solidarity, and mathematicians approach problems "with a freshness or vision unclouded by the knowledge of how others had tried to tackle them before" (p. 114).

Another explanation has been speculated about by Flottum et al. (2006): that a higher number of references can reflect more established/older and larger disciplines. The authors acknowledge that their data do not provide sufficient evidence to confirm this. Neither is the hypothesis supported by the history of the two sub-disciplines in my study. While first serious observational seismology endeavours are from 1857, theory of wave propagation started to be formulated already in the early 1800s (Shearer, 2009). In comparison, the father of chemical oceanography, Robert Boyle, published his tract on seawater saltiness in 1673, but the foundations of modern physical, chemical, biological, and geological oceanography date to the 1870s and 1880s (Pilson, 2013). Both sub-disciplines, in the modern sense, are thus roughly equally old.

The above paragraphs have shown that while some citation characteristics reported for natural sciences (either in general, or for "related" disciplines of physics

and history) agree with the results reported here, discrepancies also exist, and could be partially explained by the difference in the focus of my corpus compared to corpora used in previously published research. Another explanation worth considering lies in the premise expressed earlier in the discussion section: describing textual characteristics as reflecting the rough division of disciplines into hard/soft or natural/social/humanities was useful as a starting point. We now need to move beyond it.

6.3.3 Hedging and boosting

Through hedging and boosting, authors express their opinions, evaluations, and levels of certainty, thus positioning themselves as authorities on the (sub-)disciplinary objects of study as well as direction and development of their field. The research direction into hedging and boosting that is most relevant to the present study is the focus on the disciplinary differences in their usage. Two general findings of that research are that hedges are more frequent than boosters in all disciplines (e.g., Hyland, 1998; 2005a), and that boosting is less frequent in the natural sciences than in the humanities and social sciences (e.g., Peacock, 2006), with the humanities/social science disciplines employing almost three times more boosters (Hyland, 1998). While the results arrived at in this study do not allow me to comment on the frequency of hedging/boosting in the natural sciences versus the humanities and social sciences, they allow me to comment on the reported proportion of hedged and boosted statements. Curiously, my findings do not seem to support the dominance of hedges. Of course, such an assertion must be tentative, especially since my methodology combines bare and boosted statements into one ER++ category. Nevertheless, in both earth science sub-disciplines investigated here, ER++ statements are the most common types of claims, with three quarters of

marine chemistry RAs comprising at least 50% of ER++ claims. In seismology, the proportion is even higher. In comparison, ER– – statements (i.e., hedged statements), which are comparable to previous research, make up 20 to 40% of claims in most of the marine chemistry RAs (more so in the range of 20 to 30%). In seismology, half the RAs are also in the 20 to 30% range, but a quarter of the RAs is below 20%. Thus, hedged statements are two to two and a half times less frequent than bare/boosted statements.

To explain why this study's findings (more ER++ than ER– – claims) diverge from the tendencies reported in the previous research (more frequent hedges overall, less frequent boosting in the natural sciences), one obvious reason is that bare claims are not differentiated from boosted ones, and a potentially high proportion of bare claims could have influenced the results. For example, in seismology, where the proportion of ER++ claims is higher than in marine chemistry, bare claims may be used to describe earthquakes that occurred (time, place, magnitude, damage caused, tsunamis generated etc.), such as ... *these four events and the Amorgos event are the largest Aegean Sea earthquakes [[to occur during the last 100 years]]* (Seis 55), or to depict the conditions in the field, either known from previous studies or arrived at during the study in question, such as *East of here there is a transitional, weakening USL zone, evident in the overlapping possible USL and no USL regions closer to the coast* (Seis 58).

In addition, I speculate that another reason for my study's divergence is that marine chemists and seismologists may need to convince audiences of reliability and plausibility of findings that are often inherently temporally and/or spatially uncertain. In earth sciences in general, disciplinary theories and hypotheses are oftentimes underdetermined given that direct causes of processes shaping the earth can be difficult to establish (Kleinhans et al., 2010), especially if they happen on time scales longer than

human life (e.g., changes of ocean chemistry over millennia) or in the planet's deep interior (e.g., causes of earthquakes of different magnitudes and their occurrence prediction). This underdetermination of many studied phenomena, which is likely tacitly accepted by disciplinary audiences, may then compel marine chemists and seismologists to present the accumulated body of their sub-disciplinary knowledge as emanating from experimental and/or research activities and procedures, such as from seismometer-recorded data, observations, chemical analyses, mathematical models, or (very commonly in my corpus) calculations/equations. Basing findings on such activities and procedures could then explain the high frequency of certain expressions in the corpus–the various forms of the words *calculate*, *observe*, *model*, and *show.* Verb forms of these research words are less likely to be hedged, and are more likely presented as bare or boosted (in the case of *show*) propositions.

6.3.4 First-person pronouns

First-person pronouns are perhaps the clearest indication of authorial presence in a text. They have been much more commonly found in the humanities and social sciences rather than in the natural sciences (Hyland, 2001), as the former disciplinary members are said to wish to identify themselves with their arguments and claim credibility, while the latter scholars subscribe more to an impersonal "empiricist ideology" (Hyland, 2005a, p. 181) by downplaying their role. Also, use or non-use of first-person pronouns has been argued to reflect disciplinary differences in methodology and epistemology, with self-mention more common in fields more conscious of human agency behind knowledge construction versus fields where data should speak for itself (MacDonald, 1992). This general tendency is in line with Dressen's (2003, 2014b)

described paucity of self-mention in field geologists' writings (Section 2.4.4). Findings of this study, however, do not only sharply contrast with Dressen's, but they also suggest the conclusion that marine chemistry and seismology could belong to (sub-)disciplines with a higher incidence of overt authorial presence. With average values of 9.8 (marine chemistry) and 5.6 (seismology) first-person *we* pronouns per RA discussion/conclusion section, the sub-disciplines will likely match or supersede Hyland's (2001) most "self-mention-prone" disciplines of physics and biology, with 12.7 and 11.5 instances of *we*, respectively, per complete article (not speaking of the reported hard fields average of 8.7).[53] However, when the frequency of *we* pronouns in the two earth science sub-disciplines is compared to findings by Kuo (1999), a potentially different conclusion suggests itself (again bearing in mind my corpus comprises only discussions/conclusions). Kuo, who investigated some of the same disciplines as Hyland, found very high frequencies of self-mention in 12 complete articles from computer science (a total of 795 instances of *we*), 12 RAs from electronic engineering (325 instances), and 12 from physics (274 instances). Such frequencies rival the numbers reported by Hyland (2001, 2003b) for disciplines from the humanities and social sciences!

The fact that the comparisons of my results, Hyland's and Kuo's results suggest potentially opposing general tendencies (firstly, that marine chemistry and seismology authors seem/do not seem to explicitly place themselves into their texts more frequently than authors in many other hard fields; secondly, that there are/are not notable

[53] Confusingly, some of the average values that Hyland (2001) tabulates differ from the values reported in Hyland (2003b), despite the fact that the same corpus was used for both publications. For example, the numbers for physics in Table 1 in Hyland (2001) do not correspond to the numbers for physics in Table 6 in Hyland (2003b): Hyland lists 55.9 (after deducting self-citations) per 10,000 words in his earlier paper, but 58.5 (including abstracts) in the later one. The same, although with a smaller discrepancy, is true for hard and soft fields averages (11.9 and 33.6, respectively, in Hyland (2001) versus 12.1 and 34.2, respectively, in Hyland (2003b)). How Hyland calculated the averages in both papers, or to what extent rounding up/down occurred, is not immediately transparent.

differences in self-mention frequencies between soft and hard disciplines) certainly begs an explanation. The discrepancies could at first easily be attributed to the size, and hence representativeness, of individual corpora: Kuo's corpus comprises only 12 articles from a single journal per discipline; Hyland's corpus with 30 articles from 10 journals per discipline appears more robust, although even his corpus is sampled from fewer sources than my corpus of 60 articles per sub-discipline published in 20 (marine chemistry) and 29 (seismology) journals. Yet, the data sources and corpora sizes are sound in each individual study, because considerations of representativeness are balanced with practicality, manageability, and especially intended purpose. Kuo's study aimed to reveal the multifaceted nature of personal pronoun usage by examining the communicative intentions (either own role or relationship with readers) with which first-, second-, third-person, and indefinite pronouns are employed in three disciplines; one of Hyland's study's aim was to quantify authorial presence in eight disciplines, and thus the data size had to be kept manageable for a single researcher; my study's main aim was to attempt to uncover the rules of the game using LCT in the discursive RA sections. It is then these differences in aims that render straightforward numerical comparisons problematic, and any general tendencies that numerical comparisons between and among (sub-)disciplines may point towards need to be considered in the light of the studies' limitations.

Regardless of issues of direct comparability, compounded by the fact that average values can conceal vast individual differences, the frequencies found in this study of first-person pronouns in marine chemistry and seismology do not support Hyland's (2005a) interpretation of impersonal empiricist ideology. Firstly, raw frequencies do not bear it out. Only 17% of my corpus RAs do not feature any first-

person pronoun, as compared to 24% featuring at least 11 instances. These differences, which likely exist in all fields, show the possible variations within sub-disciplines. But even within the most common category (49% of the corpus RAs) of raw counts of personal pronouns in marine chemistry and seismology (2 to 10 instances), 27 RAs (22.5% of corpus RAs) feature at least six instances of *we*. Secondly, most self-mentions in my corpus reflect the SR++ (originator) and SR+ (opinion holder) roles, which are more authoritative (Cheung & Lau, 2020; Tang & John, 1999) but also higher risk roles (Harwood, 2005a; Martinez, 2005): that is, originated claims and expressed opinions supported by authors' personal investment through *we* pronouns are also points where the threat to face and potential attack by readers are greater. These results are not unusual in natural sciences or hard disciplines. For example, Martinez (2005) found that in biology, personal pronouns most frequently appeared in the discussion to restate results to claim responsibility for novel findings, and in results sections to state results, and restate methodology that led to the novel results. Harwood (2005c) presented examples of how physicists and computer scientists used *we* to promote the novelty or achievement of their work and dispute previously published findings. Even Hyland (2003b) stated that personal pronouns were used in connection with announcing results in about 20% of self-mention cases in biology, physics, and engineering. If, then, the usage of personal pronouns signifies, among other issues, a philosophical choice between positivistic objectivity and socially constructed reality, as Starfield (2015) posits, marine chemists and seismologists who use first-person pronouns (i.e., the majority in my corpus) may be taking the middle road: They acknowledge (or even promote) their roles in research and in originating claims, thus admitting to some subjectivity (notably so when expressing beliefs and opinions in the SR+ role), and not

pretending that objectivity is established by eliding human agency. Yet at the same time, by constructing sub-disciplinary knowledge often through research acts, as I have described above, they show allegiance to the ontological assumption that reality is external to humans rather than constructed, and the epistemological assumption that reality can be described objectively as being shaped by laws and causality.

The studies above (Harwood, 2005a; Hyland, 2003b; Kuo, 1999; Martinez, 2005) demonstrate that writers from the traditionally hard disciplines are willing to appear in higher risk roles. This willingness should be considered in conjunction with change over time. Significant increases in self-mention have been documented in varied fields (sociology, electrical engineering, and biology, though a decrease in applied linguistics) in the past 50 years, with the largest jump in the period from 1985 to 2015 (Hyland & Jiang, 2016). It is not unreasonable to propose that the growth in raw frequencies has been accompanied by increased willingness of authors to appear in higher risk roles. Accordingly, we could speculate that the prevalence of SR++ and SR+ roles in my corpus from 2012 to 2017 reflects this growing inclination. If indeed true for all so-called hard fields, then perhaps it is time to revise the traditional description of natural sciences as being characterised by authorial effacement. Put differently, we may need to rethink whether higher authorial presence is still one of the defining characteristics of disciplines influenced by the constructionist paradigm.

6.4 Summary

This chapter has described the main contributions of the present study to LCT. Specifically, this study expands the realisations of the dimension of Specialization by proposing two new forms taken by knowledge in academic sub-fields of earth science–

Visibility of knowledge and knowers and Authoritativeness of knowledge and knowers– together with translation devices so that the forms can be applied to data. The biggest methodological contribution of the present study to LCT is quantification of the proposed external languages of description (i.e., the translation devices), which allows more exact positions of each RA to be charted within Specialization code quadrants, thus taking full advantage of topological spaces in the Cartesian plane. Furthermore, in this chapter, where possible, the findings of this study have been related to research in EAP, especially to studies of Themes, citation practices, hedging and boosting, and first-person pronoun self-mention. Based on this study's results, caution is advised regarding subscribing to the strong versions of traditional dichotomies (e.g., positivism versus social constructionism or hard versus soft fields). In addition, in this chapter, the general versus specific EAP debate has been revisited and commented on, with the results of the present study supporting the latter position.

Chapter VII: Conclusion

7.1 Mapping disciplinarity in earth sciences

Through extensive work with both novice and expert earth scientists, in large part centred on manuscripts of experimental research papers, I learned about the field's variety of specialisations and their differing communication conventions. At the same time, by probing the characteristics of earth sciences when engaging with works of philosophers of science, I learned that earth sciences combine elements of historical and natural sciences, which results in an epistemology unlike that of traditional experimental sciences on which many earth science sub-disciplines are based, and brings concomitant issues of autonomy and (non-)reducibility of the field to disciplines such as physics and chemistry. This communication diversity and debates about the field's characteristics raise intriguing questions about the nature of disciplinarity of earth sciences, yet, the field has remained largely unexplored in EAP/ESP. Therefore, in this thesis, I set out to partially fill the gap by investigating some of the legitimate knowledge-making practices and their underlying larger organising principles in two earth science sub-disciplines, marine chemistry and seismology.

I combined linguistic inquiry into four selected features (topical Themes, citations, hedging and boosting, and first-person pronouns) in three traditional EAP/ESP research perspectives–corpus, genre, and ethnographic–with an approach from the sociology of knowledge, Legitimation Code Theory (LCT). LCT, the main theoretical framework of this study, allowed me to interpret textual practices concerned with frequencies, types, and usage of the selected features as reflections of sub-disciplinary knowledge and knower structures, and expose some attributes of legitimate disciplinary knower participation and legitimate ways of referring to disciplinary knowledge.

Specifically, after proposing two new forms of the dimension of Specialization– Visibility of knowledge and knowers and Authoritativeness of knowledge and knowers– and creating translation devices so that the conceptualisations could be applied to data, I was able to quantitatively reveal the degrees to which sub-disciplinary knowers were prominent and epistemically authoritative in the process of constructing the specialist body of knowledge in the discussion and conclusion sections of published experimental research articles, as well as the degree of prominent and authoritative portrayal of specialist knowledge. This quantitative methodology, which is rare in LCT research, allowed the study to take advantage of the potential for topological characterisation of practices.

The quantitative findings revealed that in Visibility, relations to knowledge structures are stronger than relations to knower structures, testifying to an emphasis on objects of study rather than on sub-disciplinary actors. In Authoritativeness, both relations to knowledge and knower structures were strong, evidencing the important role of both sub-disciplinary objects of study and sub-disciplinary actors in knowledge building. Also, while both sub-disciplines displayed the same general tendencies, knowledge building in marine chemistry was clearly more uniform across individual RAs, whereas in seismology legitimate forms of knowledge and of knower participation could be interpreted slightly more loosely.

Furthermore, my LCT-underpinned investigation shone some light on qualitative sub-disciplinary differences in knowledge structures, and hinted at a potentially additional knower structure. This examination of the Visibility of knowledge has suggested that objects of study strongly frame legitimate problems of marine chemistry and legitimate problems in seismology, but that methods and research products frame

legitimate approaches in marine chemistry less strongly than in seismology. Interpreted as Specialization in the Visibility of relations to knowledge structures, in marine chemistry what is studied is distinct to the sub-field but not how it is studied, while in seismology both what is studied and how it is studied are distinct to the sub-field. Attempts at identifying the movers and shakers in the sub-fields based on the number of citations in Subject positions in separate RAs did not yield a clear picture of leadership of individuals, because very few authors are cited marginally more than the rest. A more discernible trend of influence emerges regarding institutions and countries where prominently cited research originates, with several institutions being prominent in both marine chemistry and in seismology. This finding suggests a possible link, worth further investigation, between authors from these institutions and a specific underlying knower structure.

Types of knowledge that knowers are prominently cited for revealed that marine chemistry findings which are site-or area-specific and numerical may become obsolete faster, while findings that are more general retain their relevance longer. In comparison, site- or area-specific and numerical findings in seismology retained their relevance throughout decades, as their currency was ensured by the sub-discipline's focus on recurring natural hazards. Seismological findings concerning general behaviour or physical laws remained current regardless of their time of publication.

The combination of linguistic and sociological approaches in this study contributes to moving EAP/ESP away from a narrow focus on textual practices to larger questions of (sub-)disciplinary knowledge building, thus answering calls within our field for a more comprehensive approach to the study of the discourse of science in order to better understand disciplinarity and how language functions in the social

settings of intellectual fields. In addition, to enhance explanatory power (and pedagogical utility), investigations of textual practices need to be more fine-grained than at the level of disciplines, should move beyond the commonly used quantitative averages and pre-defined lists of features. Last but not least, comparison of my findings about the usage of the selected linguistic features with other EAP/ESP studies showed that the nature of disciplinarity in earth sciences does not fit the mould on either side of broad dichotomies such as positivism vs. social constructionism or hard vs. soft disciplines, underscoring the limits of such generalised divisions.

7.2 Pedagogical implications

Apart from theoretical and methodological contributions of this thesis to Legitimation Code Theory research and to EAP/ESP, the results of this study have obvious pedagogical implications. The findings will, of course, be most informative and directly relevant for junior scholars in marine chemistry and seismology. But the findings are generally relevant to junior scholars from other (sub-)disciplines, as they can raise the scholars' awareness of the existing possibilities and variability in the (sub-)disciplinary usage of topical structure of texts, of referencing, hedging and boosting, and authorial self-mention, thus illuminating the potential for emphasis, highlighting, and self-promotion. In addition, all junior scholars will most likely find interesting the aspect of temporal relevance and currency of knowledge claims.

The last point brings to the fore the fact that truly improving the writing of junior scholars will require incorporating the findings into the discussion of broader socio-cultural practices, not only from an EAP/ESP perspective. In this sense, junior scholars should be made aware of the forms that disciplinary knowledge and knowers

take in published articles. Even if students realise only one thing–that published disciplinary knowledge has a describable form, which, although not immutable, contributes to one's writing being recognisable as doing a discipline–they will make an important step towards becoming more effective academic communicators. For as Bhatia (2017) aptly noted, disciplinary communication significantly involves negotiation of meaning with members of the relevant discourse community, which in turn involves knowing [on the part of junior scholars] the accepted communicative practices in order to achieve intended effects, and that is why communication with disciplinary peers cannot be limited to mere knowledge of appropriate technical lexis, syntactic structures, and conventions of punctuation, all strung coherently together. Of course, most students eventually tacitly acquire the norms of disciplinary communication through immersion, though the process usually takes longer, and based on my teaching and consulting experience, is fraught with frustration and anxiety. Familiarising junior scholars with forms of knowledge explicitly, or revealing the rules of the game (Maton 2014a), can make the learning journey more straightforward and pleasant, and hopefully pique the students' curiosity, and encourage them to explore their (sub-)disciplinary texts and delve into them in search of more than just the specialist information conveyed.

Let me now sketch three general suggestions[54] for incorporating my findings into teaching, applicable to students from all research (sub-)disciplines. First, students can benefit from being made aware of the fact that academic (sub-)disciplines differ beyond the obvious features of investigated subjects, phenomena, utilised approaches and procedures, and technical lexis. To this end, students can be introduced to the

[54] I shall not repeat pedagogical suggestions that I am aware of, for example, offered in RAs on hedging (Hyland, 1996a) or boosting (Peacock, 2006)).

Specialization dimension of Legitimation Code Theory and the four Specialization codes: knowledge, knower, elite, and relativist. Before exploring their respective disciplinary texts, students can take a disciplinary survey (Figure 7.1 below) adapted from Maton and Howard (2016) to illustrate how LCT, just introduced to them, can be applied. The first item in the survey instrument addresses epistemic relations, while the remaining two explore social relations. Within the social relations items, the talent item focuses on subjective relations, and the taste and feel item on interactional relations. The survey instrument thus facilitates investigation of not only Specialization, but also of the knower dimension (the 4-K model, Section 3.3). The survey can lead to a follow-up discussion with eliciting from students more specific examples from their (sub-)disciplinary practices, drawing on students' understanding and experience, and stimulating their thinking of the inter-relationships of epistemic relations and social relations relevant to their academic fields. Here, teaching classes comprising students from various fields can better help raise students' awareness of (sub-)disciplinary differences. As the survey instrument is quantitative, and the Likert-scale options can be assigned numerical values, students can also be asked to administer the survey to the teaching staff in their departments in order to gain insights into the Specialization codes of the various subjects within their majors.

In your opinion, how important are these things for being good at [the subject]?

	Not at all	Not very	Quite	Very
Skills, techniques and specialist knowledge	☐	☐	☐	☐
Natural-born talent	☐	☐	☐	☐
Taste, judgement or a developed 'feel' for it	☐	☐	☐	☐

Figure 7.1 Scaled quantitative instrument (Maton & Howard, 2016, p. 58).

Second, students can explore (sub-)disciplinary texts for the same linguistic features investigated in this thesis. Of course, the investigation will be scaled down and simplified. For example, students can investigate individual RA sections for topical Themes and hedging and boosting, given the more extensive analysis required, while they can compile small corpora to explore citations and self-mention. Further analytical simplification may include orthographic sentences as units of analysis instead of ranking clauses; topical Themes of sentences will be considered only at their beginnings, irrespective of how many clauses each clause complex comprises; the list of general academic Themes (ER +0.5) can be shortened; and hedging and boosting expressions can also be limited to shorter lists. Even with such downscaling and simplification, misclassification will need to be accepted and tolerated, as the purpose of the textual exploration is for students to get a rough idea of how the linguistic features are employed and what they potentially reveal about the shape of knowledge in their (sub-)fields, and not to produce rigorous analyses. I have already analysed types of citations and self-mention with several groups of my earth science students, and I can attest to the fact that even analyses with some margin of error can highlight differences that can lead to fruitful discussions. The findings can again be linked back to LCT, both to individual relations (epistemic and social) and to the four Specialization codes. In addition, I believe that the findings themselves can also be of interest without relating them to LCT. For instance, while the predominance of technical vocabulary in thematic positions will be of little surprise, what students might find interesting is the large range, that is, the possible difference between RAs in which claims begin almost exclusively with disciplinary topical Themes and RAs where technical Themes are at the start of fewer than half the claims. The latter types of RAs are outliers in marine

chemistry and seismology, but the proportions in other earth science sub-fields remain unexplored. Similarly intriguing can be investigating whether topical Themes related to methods deal mostly with approaches exclusive to a (sub-)field. With self-mention, students may find informative the differences in the raw counts of first-person pronouns, as well the frequencies of the four roles (originator, opinion holder, textual, and inclusive).

Third, students can investigate other linguistic features, and then relate their findings back to LCT. One idea is to identify the most common reporting verbs following citations, and classify them after Hyland (2002a) into research/cognition/ discourse acts. Research acts reporting verbs often relate to quantitative findings (e.g., *show*, *observe*) and procedures (e.g., *analyse*, *calculate*), and could then represent stronger epistemic relations. In comparison, some discourse acts verbs (e.g., *say*, *argue*) and cognition acts verbs (e.g., *assume*, *believe*) could signal weaker epistemic relations due to their more subjective connotations. Another linguistic feature that students could investigate are formulations of statements of findings. They could identify statements such as *Smith et al. demonstrate that*/*we demonstrate that*/*our results demonstrate that*/ *the model demonstrates that*, and classify them based on the strength of social relations (in my examples, the first two formulations would represent stronger social relations and the last one the weakest).

The proposed pedagogical applications will offer students opportunities to investigate not only variable usage of language, but also how the variability corresponds to deeper-level differences among academic fields and sub-fields in what is considered appropriate communication, and what are the potential bounds of legitimate form of contribution. In this way, pedagogical use of this study's results and LCT can answer the

call of scholars such as McGrath (2016) that when teasing out the differences in knowledge construction in academic disciplines, we need to move from the coarse grain of disciplines to a finer grain of sub-disciplines.

7.3 Limitations

In this section, I will try to heed Bizzell's (1992) advice and "achieve a critical distance on our knowledge" (p. 138) by first considering the theoretical lenses at the core of this thesis, for such lenses provide insights only if used within appropriate focal distances.

LCT aims to solve what it sees as knowledge undertheorisation in the sociology of education/sociology of knowledge by focusing on *forms* that knowledge and practices take. These forms are generated by organising principles, which LCT posits underpin knowledge and practices (Maton & Chen, 2020). While a focus on form is certainly a possible starting point in solving knowledge undertheorisation, a question that needs some contemplation, specifically in education, is to what extent form can be separated from content and context. LCT research on disciplinary writing practices at university has shown a positive impact on students' achievement (e.g., Georgiou, 2016; Kirk, 2017), as noted in Sections 3.5 and 3.6. Yet, if legitimate forms of knowledge and practices are determined by their content as well as their context–and as I show in this thesis, personal stylistic variation seems also significant–then researching forms exclusively can lead to providing students with empty moulds. For some students, that will be enough some of the time; not for others.

Another issue one might consider is the shape that LCT's theorisation of forms of knowledge and practices takes. LCT's four legitimation codes (in this thesis, the

knowledge, knower, elite, and relativist Specialisation codes) per dimension reflect the varying strengths of two concepts in a Cartesian plane. Four codes, especially considering LCT's declared emphasis on topological fine-grading (i.e., infinite differentiation of degrees of strengths of concepts), offer markedly more differentiation than binaries (e.g., Bernstein's horizontal vs. hierarchical knowledge structures). Nevertheless, what if forms of practices are shaped by interactions of more than two concepts, leading to more than four code modalities? LCT proposes, at least via Specialization's 4-K model, another level of differentiation, but what if further differentiation is needed at the non-4-K level?

The acknowledgement that sets of practices can assume a variety of forms (Maton & Chen, 2020) means that research guided by LCT can unnecessarily proliferate forms of knowledge and knowers (Singh, 2015) through droves of operationalisations, even of the same practices. For example, what if the discourse Visibility and Authoritativeness of knowledge and knowers can be conceptualised in (an)other way(s) than in this study (i.e., if other linguistic or discursive features are identified that could also be argued to represent the Visibility and Authoritativeness of knowledge and knowers)? Adjudicating between multiple forms and deciding which potentially impact practices more, or which should be emphasised would then need principled decisions, which LCT cannot at the moment provide guidance on. Also, could one say that such research contributes to illuminating complexity of language use, or could revealing many possible forms of practices potentially mire research in minutiae, thus increasing ambiguity and uncertainty?

Several other limitations pertaining to the operationalisation of concepts and data processing in this thesis should also be mentioned. One of the first such limitations

that comes to mind is the exclusive focus on the discussion and conclusion sections as well as equating knowledge claims with propositions (ranking clauses), reasons for which were given in Section 4.3.2. These two interlinked issues potentially impact the results reported in this thesis to a small degree. The aim here of attempting to reveal some of the knowledge making practices is well served by focusing on knowledge claims, whose concentration is arguably highest in discussions and conclusions. Obviously, not all ranking clauses in the two sections can be considered knowledge claims. For example, some ranking clauses are more likely about the text, such as *We compared the amplitude of infrasound with the corresponding seismic moment in Figure 17* (Seis 9), while some are clearly about procedures, such as *...we sampled only a few high-TOC shale beds.* (MChem 7). Nevertheless, such ranking clauses are in the minority, are arguably much less common in discussions/conclusions than in the remaining RA sections, and are (in the case of clauses referring to procedures) in fact part of arguments leading towards pronouncements of knowledge claims. Thus, the example clause from MChem 7 is immediately followed by the following clause: *In summary, Os-isotopic chemostratigraphy defines several magmatic or meteoritic pulses [[that caused highly unstable seawater chemistry during the latest Changhsingian]]*. As for ranking clauses referring to the text, not all such clauses are clear-cut cases of non-claims, for example, *We next examine the process [[[that might generate the three coupled event types || based on their timing and location]]]* (Seis 54). Likewise, a clause from the same RA, *... our moment tensor inversion (Section 5.2.2) does not provide information about the elastic rock conditions at the source* is immediately followed by *Instead, these may be viewed as inversion parameters for discontinuities (e.g., faults)*. This blurred line between claims and non-claims, absence of a clear

consensus on what constitutes a knowledge claim, and lack of technical sub-disciplinary knowledge on my part all make, I believe, focus on ranking clauses an acceptable strategy.

Another limitation comes in the form of the four-point scales used to code topical Themes, citations, hedges and boosters, and first-person pronouns. The scales were adopted for reasons of practicality and to create a coding scheme which mapped onto LCT quadrants, not because the four options represent all possible types there are, even if that meant that expressions which could have been in separate classes were merged. Thus, for example, Authoritativeness of knowers expressed by first-person pronouns has one large category, *We as opinion holders*, which subsumes roles such as comments on research of others, limitations of own studies, recommendations for further research, or opinions/beliefs. These roles are arguably not risky or authoritative to the same degree. Similarly, if we take the following two clauses from Seis 27:

Although this hypothesis could potentially explain the formation of this bulge in the geomorphology, …

… footwall uplift may have warped the overlying sediments to form a topographic high.,

we see evidently varying levels of commitment (one clause less certain than the other), yet both were classified as ER– – (ER-1). Nevertheless, the categories were selected to best suit the data in the corpus, and at the same time to keep the coding manageable. Importantly, the scales had to be equidistant and numerically equal, despite the fact that some ER/SR realisations have the potential to be coded on finer scales than others. Selecting finer scales may be more suitable for studies of individual linguistic features, but even there compromises necessarily exist. As Hyland (2005a) noted about hedging

and boosting, "It should also be borne in mind that evaluation is expressed in a wide range of ways which makes a fine-grained typology problematic.", with these ways being more or less explicit, and more or less "opaque for the analyst" (p. 177), for example, because of textual silences that are significant only to disciplinary insiders.

Hyland's remark about hurdles faced by linguists-analysts coding texts written for readers from other disciplines leads then to a potentially problematic practice of relying on predefined lists of hedges and boosters in some applied linguistics publications. For example, two following words (and there are, undoubtedly, many more cases) are listed as hedges in all the lists known to me: *indicate* and *about*. However, in

> *A recent statistical analysis of historical earthquakes in New Zealand [Nicol et al., 2016] indicates about half of all large historical earthquakes (Mw > 7.0) ruptured faults [[that would not have been identified as active prior to the event]] (Seis 52)*

indicates is hardly a hedge since the sentence deals with quantitative information about how many events ruptured what faults. Indeed, if we refer to the referenced text, Nicol et al. write *"We find that about half of all historical earthquakes Mw ≥ 7.0 ruptured faults that, based on today's state of knowledge of active fault locations, would not have been identified as active prior to the event."* (p. 1299). The fact that the verb *indicate* is not always a hedge was confirmed when I came across it in a manuscript I was editing with its first author (an Assistant Professor in seismology). He confirmed that *indicate* can often (but not always) express certainty, and that a specific context clearly signals to disciplinary readers what is intended. Similarly, the adverb *about*, especially when

followed by numerical values, does not necessarily express tentativeness or restricted validity. For example, in

> *Zn on pyrolusite, corundum, and Al(OH)3 surfaces was about 0.1‰ heavier than in solution, and Zn adsorbed on goethite and birnessite surfaces was about 0.2‰ lighter than solution. (MChem 20)*

about expresses certainty, as communicated to me by two of my colleagues (Assistant and Associate Professors in the field) when I approached them to ask how certain the writers were. My colleagues' usage and understanding of *about* does not align, at least in this and similar cases, with Hyland's (1998) explanation that words like *about* or *approximately* are attribute hedges, which do not limit propositional certainty but "limit the scope of the accompanying statement… either restricting the temporal or qualitative range of the claim or its generalizability" (p. 362). An alternative explanation for the role the words indicate and about play may be that they express 'disciplinarily acceptable and expected inaccuracy', especially when followed by numbers, possibly because a precise numerical value is not required to support a writer's argument (Hyland, 1998).

The above coding issues are neither new nor unique to this study. For example, Lewin (2005), conducted a small-scale investigation in which she compared hedges which authors identified in their own texts with hedges identified by readers. She found that the readers considered more statements as hedged than were intended by the writers. Moreover, neither the readers nor the authors identified multiple expressions traditionally considered as hedges in linguistic literature: the readers identified only 50% of such hedges, while "the most striking contrast [was] between the identifications by authors and by linguists" (p. 171). In this work, I have attempted to overcome these

analytical hurdles by paying close attention to the context, by inter-coding portions of data with another coder, and by asking disciplinary insiders. Nevertheless, as not all the data can be consulted with disciplinary informants, there will inevitably remain cases which those for whom the analysed texts were written would understand differently.

Last but not least, within this thesis, ethnography proper (Section 1.1.2.3) was engaged with only marginally, and practically all the prolonged and continued engagement that I have had with text writers from various earth science specialisations over their texts happened as part of my regular work at the EOS. The biggest reasons for the omission were my focus on predominantly quantitative textual analyses and access to a sufficient number of faculty specialising in marine chemistry and seismology. Thus, my results lack complementary insider perspectives on the dynamic between texts and the milieu of their production.

7.4 Recommendations for further research

The results of this study have shown that both marine chemistry and seismology share some of the underlying principles of legitimate contribution, expressed by the same Specialization codes: the knower code in Visibility and the elite code in Authoritativeness. Importantly, however, the results also clearly show some of the within-code differences: The scatter plots in Chapter V reveal that while the median values of both fields are rather close, individual sub-corpora articles in marine chemistry are more tightly concentrated within the knower and elite quadrants, while seismology RAs evidence more diffuse knowledge and knower practices. Teasing out of even more substantial within-code differences could be attempted by applying in the future Maton's (2014a) 4-K model (Section 3.3), and identifying new linguistic features

that could function as external languages of description between modalities of epistemic relations–insights (situational, purist, doctrinal, and knower/no) and modalities of social relations–gazes (social, born, cultivated, and trained/blank).

The 4-K model could further be drawn upon when investigating knowledge insights (ER+1 topical Themes), whose qualitative investigation in this study was limited. The idea is worth pursuing further quantitatively. Also, the potential presence of the knower born gaze (via institutional affiliations of all cited authors), as suggested in Section 6.1.3, could be probed further.

Another area worth exploring is the potential diachronic Specialization code shift or code drift. Similarly to studies such as Hyland and Jiang's changes of stance (2016) or citation practices (2017) between 1965 and 2015, future research into knowledge and knower practices in marine chemistry and seismology could look into whether and how Visibility and Authoritativeness have changed in the past decades. Such research could provide insights into multiple questions, such as whether the higher concentration around median points in marine chemistry and greater diffusion in seismology have been key features of sub-disciplinary knowledge and knower representations for a long time; whether the concentration and diffusion are the results of processes that started several decades ago (i.e., whether we are seeing increasing or decreasing uniformity); whether marine chemistry and seismology were more similar in the past; whether one sub-discipline is reaching a point or starting a process that the other sub-discipline is likely to follow; whether one sub-discipline (if so, then perhaps seismology) may still be in the process of consolidating knowledge and knower representations in writing; etc.

Yet another area of future research could be focused on whether the results of SR+1 investigations (pace of knowledge change) correlated in any way with the ranking/standing of the journals in which the prominently cited authors originally published the research which was then cited in my corpus. Additionally, seismology citations could be investigated to see whether any correlation exists between the markedly higher number of older citations (compared to marine chemistry) and journal rankings.

LCT is said to offer the possibility of investigating the same practices from multiple perspectives, both through the lens of other dimensions or their individual constitutive concepts. Temporality has already been mentioned in Section 6.1.4, but another obvious candidate may be Semantics. For example, the wording of topical Themes in the sub-disciplines could be coded for their semantic density (degree of meaning condensation, Section 3.3) following the translation instrument of Maton and Doran (2017a), possibly after some adaptation. The same instrument could be used to code larger stretches of text, although due to the high demand of word-for-word coding on time, abstracts would be more suitable than whole discussions. The same focus on abstracts could be used for the second concept from Semantics, semantic gravity, to analyse the degree of meaning abstraction and draw a semantic curve or line (Section 3.3).

The above-mentioned recommendations for further research suggest how EAP/ ESP researchers could draw on the LCT toolkit when attempting to shine additional light onto the inner workings of academic fields. They also suggest an alternative or parallel path of inquiry into characteristics of (sub-)disciplinary writing. It is along this path, I hope, that more EAP/ESP researchers will choose to travel.

References

Aki, K., & Richards, P.G. (2002). *Quantitative Seismology* (2nd ed.). University Science Books.

Aktas, R.N., & Cortes, V. (2008). Shell nouns as cohesive devices in published and ESL student writing. *Journal of English for Academic Purposes, 7*(1), 3-14.

Alangari, M., Jaworska, S., & Laws, J. (2020). Who's afraid of phrasal verbs? The use of phrasal verbs in expert academic writing in the discipline of linguistics. *Journal of English for Academic Purposes, 43*, 1-13.

Albright, J., Hartman, D., & Widin, J. (Eds.). (2018). *Bourdieu's field theory and the social sciences*. Singapore: Palgrave Macmillan.

Alise, M.A., & Teddlie, C. (2010). A continuation of paradigm wars? Prevalence rates of methodological approaches across the social/behavioral sciences. *Journal of Mixed Methods Research, 4*(2), 103-126.

Allaby, M. (Ed.) (2013). *A dictionary of geology and earth sciences* (4th ed.). Oxford, UK: Oxford University Press.

Anson, D.W.J. (2017). Examining the examiners: The state of senior secondary English examinations in Australia. *Australian Journal of Language and Literacy, 40*(2), 135-146.

Anthony, L. (2011). Products, processes and practitioners: A critical look at the importance of specificity in ESP. *Taiwan ESP International Journal, 3*(2), 1-18.

Anthony, L. (2013). A critical look at software tools in corpus linguistics. *Linguistic Research, 30*(2), 141-161.

Arbee, A., Hugo, W., & Thomson, C. (2014). Epistemological access in marketing: A demonstration of the use of Legitimation Code Theory in higher education. *Journal of Education,* (59), 39-63.

Archer, M.S. (1995). *Realist social theory: The morphogenetic approach*. Cambridge; New York: Cambridge University Press.

Argamon, S., Chase, P., & Dodick, J. (2005). The languages of science: A corpus-based study of experimental and historical science articles. *Proceedings of the Annual Meeting of the Cognitive Science Society* (Vol. 27, No. 27).

Argamon, S., & Dodick, J. (2006). Corpus-based study of scientific methodology: Comparing the historical and experimental sciences. In J.G. Shanahan, Y. Qu & J. Wiebe (Eds.), *Computing attitude and affect in text: Theory and applications* (pp. 215-231). Springer: Dordrecht.

Argamon, S., Dodick, J., & Chase, P. (2008). Language use reflects scientific methodology: A corpus-based study of peer-reviewed journal articles. *Scientometrics, 75*(2), 203-238.

Atkinson, D. (1998). *Scientific discourse in sociohistorical context: The Philosophical Transactions of the Royal Society of London, 1675-1975*. Hoboken, NJ: Routledge.

Backlund, I. (1992). Theme in English telephone conversation. *Language Sciences, 14*(4), 545-564.

Banks, D. (2017). The extent to which the passive voice is used in the scientific journal article, 1985-2015. *Functional Linguistics, 4*(1), 1-17.

Basal, A., & Bada, E. (2012). Use of first person pronouns: A corpus based study of journal articles. *Energy Education Science and Technology Part B: Social and Educational Studies, 4*(3), 1777-1788.

Basturkmen, H. (2009). Commenting on results in published research articles and masters dissertations in Language Teaching. *Journal of English for Academic Purposes, 8*(4), 241-251.

Basturkmen, H. (2012). A genre-based investigation of discussion sections of research articles in dentistry and disciplinary variation. *Journal of English for Academic Purposes, 11*(2), 134-144.

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

Becher, T. (1989). *Academic tribes and territories*: *Intellectual enquiry and the cultures of disciplines*. Milton Keynes, Bristol: The Society for Research into Higher Education & Open University Press.

Becher, T., & Trowler, P. (2001). *Academic tribes and territories: Intellectual enquiry and the culture of disciplines* (2nd ed.). Society for Research into Higher Education & Open University Press.

Bernstein, B. (1971). On the classification and framing of educational knowledge. In M.F.D. Young (Ed.), *Knowledge and control: New directions for the sociology of education* (pp. 47-69). Collier-Macmillan.

Bernstein, B. (1999). Vertical and horizontal discourse: An essay. *British Journal of Sociology of Education, 20*(2), 157-173.

Bernstein, B. (2000). *Pedagogy, symbolic control and identity: Theory, research, critique* (Rev. ed.). Lanham, MD: Rowman & Littlefield Publishers.

Bernstein, B. (2003a). *Class, codes, and control. Basil Bernstein. Volume I, Theoretical studies towards a sociology of language.* London; New York: Routledge.

Bernstein, B. (2003b). *Class, codes, and control. Basil Bernstein. Volume IV, The structuring of pedagogic discourse*. London; New York: Routledge.

Berwick, R., & Chomsky, N. (2016). *Why only us : Language and evolution*. The MIT Press.

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

Bhatia, V.K. (2002). A generic view of academic discourse. In J. Flowerdew (Ed.), *Academic discourse* (pp. 21-39). London: Longman.

Bhatia, V.K. (2017). *Critical genre analysis: Investigating interdiscursive performance in professional practice*. London; New York: Routledge.

Bhatia, V.K., & Gotti, M. (2006). Introduction. In V.K. Bhatia & M. Gotti (Eds.), *Explorations in specialized genres* (pp. 9-17). New York: Peter Lang.

Biber, D. (2006). Stance in spoken and written university registers. *Journal of English for Academic Purposes, 5*(2), 97-116.

Biber, D. (2009). A corpus-driven approach to formulaic language in English: Multi-word patterns in speech and writing. *International Journal of Corpus Linguistics, 14*(3), 275-311.

Biber, D. (2015). Corpus-based and corpus-driven analyses of language variation and use. In B. Heine & H. Narrog (Eds.), *The Oxford handbook of linguistic analysis* (2nd ed.; pp. 159-191). New York: Oxford University Press.

Biber, D., Conrad, S., & Reppen, R. (1998). *Corpus linguistics: Investigating language structure and use*. Cambridge, UK: Cambridge University Press.

Bizzell, P. (1992). *Academic discourse and critical consciousness*. Pittsburgh, PA: University of Pittsburgh Press.

Blackwell, J. and Martin, J. (2011). *A scientific approach to scientific writing*. New York: Springer.

Blommaert, J. (2007). On scope and depth in linguistic ethnography. *Journal of Sociolinguistics, 11*(5), 682-688.

Bloor, M., & Bloor, T. (1986). Languages for Specific Purposes: Practice and Theory. CLCS Occasional Paper No.19, 1-35.

Bloor, M., & Bloor, T. (1993). How economists modify propositions. In: W. Henderson, T. Dudley-Evans (Eds.), & R. Backhouse (Eds.), *Economics and Language* (pp. 153–169). London; New York: Routledge.

Bloor, M., & Bloor, T. (2013). *The functional analysis of English: A Hallidayan approach* (3rd ed.). London; New York: Routledge.

Bourdieu, P. (1971). Systems of education and systems of thought. In M.F.D. Young (Ed.), *Knowledge and control: New directions for the sociology of education* (pp. 189-207). Collier-Macmillan.

Bourdieu, P. (1984). *Distinction: A social critique of the judgement of taste*. Cambridge, MA: Harvard University Press.

Bourdieu, P. (1986). The forms of capital. In J.G. Richardson (Ed.), *Handbook of theory and research for the sociology of education* (pp. 241-258). New York: Greenwood Press.

Bourdieu, P. (1990a). *In other words: Essays towards a reflexive sociology*. Cambridge, UK: Polity.

Bourdieu, P. (1990b). *The logic of practice*. Stanford, CA: Stanford University press.

Bourdieu, P. (1993). *Sociology in question*. Thousand Oaks, CA: Sage.

Bourdieu, P., & Passeron, J.C. (1977). *Reproduction in education, society and culture*. London; Beverly Hills: Sage Publications.

Bourdieu, P., & Passeron, J.C. (1979). *The inheritors: French students and their relation to culture*. Chicago; London: The University of Chicago Press.

Brett, P. (1994). A genre analysis of the results section of sociology articles. *English for Specific Purposes, 13*(1), 47-59.

Bucholtz, M., & Hall, K. (2005). Identity and interaction: A sociocultural linguistic approach. *Discourse Studies, 7*(4-5), 585-614.

Budin, W. (2002). Beyond books and journals: Electronic and web based resources to facilitate research. Retrieved from http://www.state.nj.us/health/ccr/documents/summer2002news.pdf

Butt, D., Fahey, R., Feez, S., & Spinks, S. (2012). *Using functional grammar: An explorer's guide* (3rd ed.). South Yarra, Australia: Palgrave Macmillan.

Canagarajah, A.S. (2002). *A geopolitics of academic writing*. Pittsburgh, PA: University of Pittsburgh Press.

Cao, F. (2014). *Metadiscourse in research articles: A comparative study across disciplines and research paradigms* [Unpublished doctoral dissertation]. Nanyang Technological University.

Cao, F., & Hu, G. (2014). Interactive metadiscourse in research articles: A comparative study of paradigmatic and disciplinary influences. *Journal of Pragmatics, 66*, 15-31.

Carvalho, L., & Dong, A. (2009). Legitimate design: A sociology of knowledge account of the field. *Design Studies, 30*(5), 483-502.

Carvalho, L., Dong, A., & Maton, K. (2015). Foregrounding knowledge in e-learning design: An illustration in a museum setting. *Australasian Journal of Educational Technology, 31*(3), 328-348.

Casanave, C.P. (2008). Learning participatory practices in graduate school: Some perspective-taking by a mainstream educator. In C. P. Casanave & X. Li (Eds.), *Learning the literacy practices of graduate school: Insiders' reflections on academic enculturation* (pp. 14-31). Ann Arbor, MI: The University of Michigan Press.

Cemen, I., & Yilmaz, Y. (Eds.). (2017). *Active global seismology: Neotectonics and earthquake potential of the eastern mediterranean region*. ProQuest Ebook Central.

Charles, M. (2006). Phraseological patterns in reporting clauses used in citation: A corpus-based study of theses in two disciplines. *English for Specific Purposes 25*(3), 310-331.

Charles, M. (2013). English for Academic Purposes. In B. Paltridge & S. Starfield (Eds.), *The handbook of English for Specific Purposes* (pp. 137-153). Malden, MA: Wiley-Blackwell.

Chen, P., Chiu, H., Cochran, E., Duan, B., Higher Education Press., Li, Y., Li, Y., Malin, P., Wu, R., Xie, X., & Yin, X. (2012). *Imaging, modeling and assimilation in seismology*. De Gruyter.

Chen, R. T. H. (2010). Knowledge and Knowers in Online Learning: Investigating the effects of online flexible learning on student sojourners [Unpublished doctoral thesis]. University of Wollongong.

Chen, R. T. H. (2015). L2 Blogging: Who Thrives and Who Does Not? Language Learning & Technology, 19(2), 177–196.

Cheng, A. (2015). Genre analysis as a pre-instructional, instructional, and teacher development framework. *Journal of English for Academic Purposes, 19*, 125-136.

Cheng, F.W, & Unsworth, L. (2016). Stance-taking as negotiating academic conflict in applied linguistics research article discussion sections. *Journal of English for Academic Purposes, 24*, 43-57.

Chester, R., & Jickells, T. (2012). *Marine geochemistry*. Hoboken, NJ : John Wiley & Sons.

Cheung, L.M.E. (2015). Legitimizing the knower's multiple voices in applied linguistics postgraduate written discourse. *TESOL International Journal, 10*(1), 62-76.

Cheung, Y. L., & Lau, S. W. L. (2020). Authorial voice in academic writing: A comparative study of journal articles in English Literature and Computer Science. *Ibérica, 39*(1), 215-242.

Christie, F., & Martin, J.R. (Eds.). (2007). *Language, knowledge and pedagogy: Functional linguistic and sociological perspectives*. New York: Continuum.

Christie, F., Martin, J.R., Maton, K., & Muller, J. (2007). Taking stock: Future directions in research in knowledge structure. In F. Christie, & J. R. Martin (Eds.). *Language, knowledge and pedagogy: Functional linguistic and sociological perspectives* (pp. 239-257). London; New York: Continuum.

Christie, F., & Maton, K. (Eds.). (2011). *Disciplinarity: Functional linguistic and sociological perspectives*. London: Continuum.

Christie, F., & Maton, K. (2011). Why disciplinarity? In F. Christie & K. Maton (Eds.), *Disciplinarity: Functional linguistic and sociological perspectives* (pp. 1-9). London: Continuum.

Chung, T.M. (2003). A corpus comparison approach for terminology extraction. *Terminology, 9*, 221-246.

Chung, T.M., & Nation, I.S.P. (2004). Identifying technical vocabulary. *System, 32*(2), 251-263.

Clarence, S. (2016). Exploring the nature of disciplinary teaching and learning using Legitimation Code Theory semantics. *Teaching in Higher Education, 21*(2), 123-137. https://doi.org/10.1080/13562517.2015.1115972

Cleland, C.E. (2001). Historical science, experimental science, and the scientific method. *Geology, 29*(11), 987-990.

Cleland, C.E. (2013). Common cause explanation and the search for a smoking gun. *Geological Society of America Special Papers, 502*, 1-9.

Cobb, T. (2013). *VocabProfilers*. Retrieved from http://www.lextutor.ca/vp/comp/

Coffin, C. (2009). Incorporating and evaluating voices in a film studies thesis. *Writing and Pedagogy, 1*(2), 163-193.

Cortes, V. (2013). The purpose of this study is to: Connecting lexical bundles and moves in research article introductions. *Journal of English for Academic Purposes, 12*(1), 33-43.

Cotos, E., Link, S., & Huffman, S. (2016). Studying disciplinary corpora to teach the craft of discussion. *Writing and Pedagogy, 8*(1), 33-64.

Cotos, E., Huffman, S., & Link, S. (2017). A move/step model for methods sections: Demonstrating Rigour and Credibility. *English for Specific Purposes, 46*, 90-106.

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

Coxhead, A., & Nation, I.S.P. (2001). The specialized vocabulary of English for academic purposes. In J. Flowerdew & M. Peacock (Eds.), *Research perspectives on English for academic purposes* (pp. 252-267). Cambridge, UK: Cambridge University Press.

Dahl, T. (2008). Contributing to the academic conversation: A study of new knowledge claims in economics and linguistics. *Journal of Pragmatics, 40*(7), 1184-1201.

Dahl, T. (2013). Knowledge claims. In C.A. Chapelle (Ed.), The encyclopedia of applied linguistics. Chichester, West Sussex, UK: Wiley-Blackwell.

De Beaugrande, R., & Dressler, W.U. (1981). *Introduction to text linguistics*. London: Longman.

Dietz, R.S. (1977). Plate tectonics: A revolution in geology and geophysics. *Tectonophysics, 38*, 1-6.

Ding, A., & Bruce, I. (2017). *The English for Academic Purposes practitioner: Operating on the edge of academia.* Cham, Switzerland: Palgrave Macmillan.

Dodick, J., Argamon, S., & Chase, P. (2009). Understanding scientific methodology in the historical and experimental sciences via language analysis. *Science & Education, 18*(8), 985-1004.

Dowdey, D. (1992). Citation and documentation curriculum. In M. Secor & D. Charney (Eds.), *Constructing rhetorical education* (pp. 330-351). Carbondale: Southern Illinois University Press.

Dressen, D. (2002). Textual silence in scientific research articles: Recontextualizations of the field account in geology. *Hermes: Journal of Language and Communication Studies, 28*, 81-107.

Dressen, D. (2003). Geologists' implicit persuasive strategies and the construction of evaluative evidence. *Journal of English for Academic Purposes, 2*(4), 273-290.

Dressen-Hammouda, D. (2008). From novice to disciplinary expert: Disciplinary identity and genre mastery. *English for Specific Purposes, 27*(2), 233-252.

Dressen-Hammouda, D. (2014a). Measuring the voice of disciplinarity in scientific writing: A longitudinal exploration of experienced writers in geology. *English for Specific Purposes, 34*, 14-25.

Dressen-Hammouda, D. (2014b). Place and space as shapers of disciplinary identity: The role of indexicality in the emergence of disciplinary writing expertise. In J. Bemford, F. Poppi, & D. Mazzi (Eds.), *Space, place and the discursive construction of identity* (pp. 71-106). Bern, Switzerland: Peter Lang.

Dressen, D., & Swales, J.M. (2000). "Geological setting/cadre geologique" in English and French petrology articles: Muted indications of explored places. In A. Trosberg (Ed.), *Analysing professional genres* (pp. 57-76). Amsterdam, te Netherlands: Benjamin.

Dueñas, P.M. (2007). 'I/we focus on…': A cross-cultural analysis of self-mentions in business management research articles. *Journal of English for Academic Purposes, 6*(2), 143-162.

Durkheim, E. (1995). *The elementary forms of religious life* (K.E. Fields, Trans.) New York: The Free Press. (Original work published 1912)

Ebrahimi, S.F., & Chan, S.H. (2015). Research article abstracts in applied linguistics and economics: Functional analysis of the grammatical subject. *Australian Journal of Linguistics 35*(4), 381-397.

Ebrahimi, S.F., Chan, S.H., & Ain, N.A. (2014). Discourse functions of grammatical subject in results and discussion sections of research article across four disciplines. *Journal of Writing Research, 6*(2), 125-140.

Ebrahimi, S.F., & Khedri, M. (2011). Thematicity in research article abstracts: A cross disciplinary study. *Educational Quest–An International Journal of Education and Applied Sciences, 2*(3), 287-292.

El-Issa, A. (2016). *Theme as an index of genre in discourse analysis: A functional approach*. AuthorHouse.

Emerson, S., & Hedges, J. (2008). *Chemical oceanography and the marine carbon cycle.* Cambridge, UK: Cambridge University Press.

Engber, C.A. (1995). The relationship of lexical proficiency to the quality of ESL compositions. *Journal of Second Language Writing, 4*(2), 139-155.

Eriksson, S. & Helgesson, G. (2017). The false academy: Predatory publishing in science and bioethics. *Medicine, Healthcare and Philosophy, 20*(2), 163-170.

Faigley, L., & Hansen, K. (1985). Learning to write in the social sciences. *College Composition and Communication, 36*(2), 140-149.

Farrokhi, F., & Emami, S. (2008). Hedges and boosters in academic writing: native vs. non-native research articles in applied linguistics and engineering. *Journal of English Language Pedagogy and Practice, 1*(2), 62-98.

Feak, C.B., & Swales, J.M. (2009). *Telling a research story: Writing a literature review*. Ann Arbor, MI: The University of Michigan Press.

Firestone, W.A. (1987). Meaning in method: The rhetoric of quantitative and qualitative research. *Educational Researcher, 16*(7), 16-21.

Flottum, K., Dahl, T., & Kinn, T. (2006). Academic voices: Across languages and disciplines. Amsterdam; Philadelphia: John Benjamins Publishing Company.

Flowerdew, J. (2002). Introduction: Approaches to the analysis of academic discourse in English. In J. Flowerdew (Ed.), *Academic discourse* (pp. 1-17). London: Longman.

Flowerdew, J. (2003). Signalling nouns in discourse. *English for Specific Purposes, 22*, 329-346.

Flowerdew, J. (2013). English for research publication purposes. In B. Paltridge & S. Starfield (Eds.), *The handbook of English for Specific Purposes* (pp. 301-321). Malden, MA: Wiley-Blackwell.

Flowerdew, J., & Peacock, M. (2001). Issues in EAP: A preliminary perspective. In J. Flowerdew & M. Peacock (Eds.), *Research perspectives on English for Academic Purposes* (pp. 8-24). Cambridge, UK: Cambridge University Press.

Flowerdew, L. (1998). Corpus linguistic techniques applied to textlinguistics. *System, 26*(4), 541-552.

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

Forey, G. & Sampson, N. (2017). Textual metafunction and theme: What's it' about? In T. Bartlett & G. O'Grady (Eds.), *The Routledge handbook of systemic functional linguistics* (pp. 131-145). London; New York: Routledge.

Frodeman, R. (1995). Geological reasoning: Geology as an interpretive and historical scince. *GSA Bulletin, 107*(8), 960-968.

Futterer, D.K. (2006). The solid phase of marine sediments. In H.D. Schultz & M. Zabel (Eds.), *Marine geochemistry* (2nd ed., pp. 1-25). Berlin, Germany: Springer-Verlag Berlin Heidelberg.

Gardner, S. (2008). Mapping ideational meaning in a corpus of student writing. In C. Jones & E. Ventola (Eds.), *New developments in the study of ideational meaning: From language to multimodality* (pp. 169-188). London: Equinox Publishing.

Gardner, D., & Davies, M. (2013). A new academic vocabulary list. *Applied Linguistics, 35*(3), 305-327.

Gee, J.P. (1990). *Social linguistics and literacies: Ideology in discourses*. Basingstoke, UK: The Falmer Press.

Gee, J.P. (2015). *Social linguistics and literacies: Ideology in discourses* (5th ed.). London; New York: Routledge.

Geertz, C. (1988). *Works and lives: The anthropologist as author*. Stanford, CA: Stanford University Press.

Georgiou, H. (2016). Putting physics knowledge in the hot seat: The semantics of student understandings of thermodynamics. In K. Maton, S. Hood, & S. Shay (Eds.), *Knowledge-building: Educational studies in Legitimation Code Theory* (pp. 176-192). New York: Routledge.

Ghadessy, M. (1999). Thematic organisation in academic article abstracts. *Estudios Ingleses de la Universidad Complutense, 7*, 141-161.

Giloi, S., & Belluigi, D.Z. (2017). Underlying knowledge-knower structures in graphic design: Contributing to establishing a cohesive language for use in graphic design education. *Art, Design & Communication in Higher Education, 16*(1), 7-22.

Gosden, H. (1993). Discourse functions of subject in scientific research articles. *Applied Linguistics, 14*(1), 56-75.

Gould, J. (2014, November 3). How to get published in high-impact journals: Big research and better writing. *Naturejobs*. Retrieved from http://blogs.nature.com/naturejobs/2014/11/03/how-to-get-published-in-high-impact-journals-big-research-and-better-writing

Gravetter, F.J., & Forzano, L.A.B. (2012). *Research methods for the behavioral sciences* (4th ed.). Belmont, CA: Wadsworth Cengage Learning.

Gray, B. (2010). On the use of demonstrative pronouns and determiners as cohesive devices: A focus on sentence-initial *this/these* in academic prose. *Journal of English for Academic Purposes, 9*(3), 167-183.

Gray, B., & Cortes, V. (2011). Perception vs. evidence: An analysis of *this* and *these* in academic prose. *English for Specific Purposes, 30*(1), 31-43.

Grenfell, M. (2010). Working with *habitus* and *field*: The logic of Bourdieu's practice. In E. Silva, & A. Warde (Eds.), *Cultural analysis and Bourdieu's legacy: Settling accounts and developing alternatives* (pp. 14-27). London; New York: Routledge.

Grenfell, M. (Ed.). (2011). *Bourdieu, language and linguistics*. London; New York: Continuum.

Grenfell, M. (Ed.). (2014). *Pierre Bourdieu: Key concepts* (2nd ed.). London; New York: Routledge.

Grenfell, M. (2018). Afterword: Reflecting in/on field theory in practice. In J. Albright, D. Hartman, & J. Widin (Eds.), *Bourdieu's field theory and the social sciences* (pp. 269-291). Singapore: Palgrave Macmillan.

Grenfell, M., & Lebaron, F. (Eds.). (2014). *Bourdieu and data analysis: Methodological principles and practice*. Bern, Switzerland: Peter Lang.

Ha, A.Y.H., & Hyland, K. (2017). What is technicality? A technicality analysis model for EAP vocabulary. *Journal of English for Academic Purposes, 28*, 35-49.

Hafner, C.A. (2018). Genre innovation and multimodal expression in scholarly communication: Video methods articles in experimental biology. *Ibérica*, *36*, 15-42.

Halevi, G. (2013). Citation characteristics in the arts & humanities. *Research Trends, 32*, 23-25.

Halliday, M.A.K. (2007). *The language of science*. Beijing: Peking University Press.

Halliday, M.A.K. (2016). *Aspects of language and learning*. Springer.

Halliday, M. A. K., & Martin, J. R. (1996) *Writing science: Literacy and discursive power*. London: Falmer Press.

Halliday, M.A.K., & Matthiessen, C.M.I.M. (2004). *An introduction to functional grammar* (3rd ed.). London: Hodder Education.

Halliday, M.A.K., & Matthiessen, C.M.I.M. (2014). *Halliday's introduction to functional grammar* (4th ed.). London: Routledge.

Halliday, M.A.K., McIntosh, A., & Strevens, P. (1965). *The linguistic sciences and language teaching*. London: Longman.

Hamilton, D.P. (1991). Research papers: Who's uncited now? *Science, 251*, 25.

Harwood, N. (2005a). 'I hoped to counteract the memory problem, but I made no impact whatsoever': Discussing methods in computing science using *I*. *English for Specific Purposes, 24*(3), 243-267.

Harwood, N. (2005b). 'We do not seem to have a theory…The theory I present here attempts to fill this gap': Inclusive and exclusive pronouns in academic writing. *Applied Linguistics, 26*(3), 343-375.

Harwood, N. (2005c). 'Nowhere has anyone attempted… In this article I aim to do just that': A corpus study of self-promotional *I* and *we* in academic writing across four disciplines. *Journal of Pragmatics, 37*(8), 1207-1231.

Harwood, N. (2006). (In)appropriate personal pronoun use in political science: A qualitative study and a proposed heuristic for future research. *Written Communication, 23*(4), 424-450.

Harwood, N. (2009). An interview-based study of the functions of citations in academic writing across two disciplines. *Journal of Pragmatics, 41*(3), 497-518.

Hashem, M. (2007). Becoming an independent field: Societal pressures, state and professions. *Higher Education, 54*(2), 181-205.

Hawking, S. (1993). *Black holes and baby universes and other essays*. New York: Bantam.

Hilgers, M., & Mangez, E. (Eds.). (2014). *Bourdieu's theory of social fields*. London: Routledge.

Hinkel, E. (2004). *Teaching academic ESL writing: Practical techniques in vocabulary and grammar*. Mahwah, NJ: Lawrence Erlbaum.

Ho, V., & Li, C. (2020). Achieving a better understanding of writer-reader interaction in academic discourse: The Metadiscourse-Stance-Participation framework. *Selected papers from the 29th International Symposium on English teaching, Taipei 2020*. (pp. 58-73).

Holmes, R. (1997). Genre analysis, and the social sciences: An investigation of the structure of research article discussion sections in three disciplines. *English for Specific Purposes, 16*(4), 321-337.

Hood, S. (2010). *Appraising research: Evaluation in academic writing*. Basingstoke, UK; New York: Palgrave Macmillan.

Hood, S. (2011). Writing discipline: Comparing inscriptions of knowledge and knowers in academic writing. In F. Christie & K. Maton (Eds.), *Disciplinarity: Functional linguistic and sociological perspectives* (pp. 106-128). New York: Continuum.

Hood, S. (2016). Ethnographies on the move, stories on the rise: Methods in the humanities. In K. Maton, S. Hood, & S. Shay (Eds.), *Knowledge-building: Educational studies in Legitimation Code Theory* (pp. 117-137). New York: Routledge.

Howard, S., & Maton, K. (2011). Theorising knowledge practices: A missing piece of the educational technology puzzle. *Research in Learning Technology, 19*(3), 191-206.

Hsu, W. (2011). The vocabulary thresholds of business textbooks and business research articles for EFL students. *English for Specific Purposes, 30*(4), 247-257.

Hu, G. (2018). Disciplinary knowledge making and academic discourse. In Y. Leung, J. Katchen, S. Hwang, & Y. Chen (Eds.). *Reconceptualizing English language teaching and learning in the 21st century* (pp. 553-573). Taipei, Republic of China: Crane Publishing.

Hu, G., & Cao, F. (2015). Disciplinary and paradigmatic influences on interactional metadiscourse in research articles. *English for Specific Purposes, 39*, 12-25.

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

Hu, M., & Nation, I.S.P. (2000). Unknown vocabulary density and reading comprehension. *Reading in a Foreign Language, 13*(1), 403–430.

Hunt, K. W. (1965). A synopsis of clause-to-sentence length factors. *The English Journal, 54*(4), 300-309.

Hutchinson, T., & Waters, A. (1987). *English for Specific Purposes: A learning-centred approach*. Cambridge, UK: Cambridge University Press.

Hyland, K. (1996a). Nurturing hedges in the ESP curriculum. System, *24*(4), 477-490.

Hyland, K. (1996b). Talking to the academy: Forms of hedging in science research articles. *Written Communication, 13*(2), 251-281.

Hyland, K. (1998). Boosting, hedging and the negotiation of academic knowledge. *Text 18*(3), 349- 382.

Hyland, K. (1999). Academic attribution: Citation and the construction of disciplinary knowledge. *Applied Linguistics, 20*(3), 341-367.

Hyland, K. (2000). Hedges, boosters, and lexical invisibility: Noticing modifiers in academic texts. *Language Awareness, 9*(4), 179-197.

Hyland, K. (2001). Humble servants of the discipline? Self-mention in research articles. *English for Specific Purposes, 20*(3), 207-226.

Hyland, K. (2002a). Activity and evaluation: Reporting practices in academic writing. In J. Flowerdew (Ed.), *Academic discourse* (pp. 115–130). London: Longman.

Hyland, K. (2002b). Specificity revisited: How far should we go? *English for Specific Purposes, 21*(4), 385-395.

Hyland, K. (2002c). Authority and invisibility: Authorial identity in academic writing. *Journal of Pragmatics, 34*(8), 1091-1112.

Hyland, K. (2002d). Options of identity in academic writing. *ELT Journal, 56*, 351-358.

Hyland, K. (2003a). Genre-based pedagogies: A social response to process. *Journal of Second language Writing, 12*(1), 17-29.

Hyland, K. (2003b). Self-citation and self-reference: Credibility and promotion in academic publication. *Journal of the American Society for Information Science and Technology, 54*(3), 251-259.

Hyland, K. (2004). *Disciplinary discourses: Social interactions in academic writing*. Ann Arbor, MI: The University of Michigan Press.

Hyland, K. (2005a). Stance and engagement: A model of interaction in academic discourse. *Discourse Studies, 7*(2), 173-192.

Hyland, K. (2005b). *Metadiscourse: Exploring interaction in writing*. London; New York: Continuum.

Hyland, K. (2006a). Disciplinary differences: Language variation in academic discourses. In K. Hyland & M. Bondi (Eds.) *Academic discourse across disciplines* (pp. 17-45). Bern, Switzerland: Peter Lang.

Hyland, K. (2006b). *English for Academic Purposes: An advanced resource book*. London; New York: Routledge.

Hyland, K. (2007). English for professional and academic purposes: Writing for scholarly publication. In D. Belcher (Ed.): *Teaching language purposefully: English for Specific Purposes in theory and practice* (pp. 1-27). New York: Cambridge University Press.

Hyland, K. (2008a). Persuasion, interaction and the construction of knowledge: Representing self and others in research writing. *International Journal of English Studies, 8*(2), 1-23.

Hyland, K. (2008b). As can be seen: Lexical bundles and disciplinary variation. *English for Specific Purposes, 27*(1), 4-21.

Hyland, K. (2009). *Academic discourse: English in a global context*. London: Continuum.

Hyland, K. (2011a). Disciplinary specificity: Discourse, context, and ESP. In D. Belcher, A.M. Johns, & B. Paltridge (Eds.), *New directions in English for specific purposes research* (pp. 6-24). Ann Arbor, MI: The University of Michigan Press.

Hyland, K. (2011b). Welcome to the machine: Thoughts on writing for scholarly publication. *Journal of Second Language Teaching Research, 1*(1), 58-68.

Hyland, K. (2012a). Corpora and academic discourse. In K. Hyland, C.M. Huat, & M. Handford (Eds.), *Corpus applications in applied linguistics* (pp. 30-46). London: Continuum.

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

Hyland, K. (2013a). Writing in the university: Education, knowledge and reputation. *Language Teaching, 46*(1), 53-70.

Hyland, K. (2013b). ESP and writing. In B. Paltridge & S. Starfield (Eds.), *The handbook of English for Specific Purposes* (pp. 95-113). Malden, MA: Wiley-Blackwell.

Hyland, K. (2014). Dialogue, community and persuasion in research writing. In L. Gil-Salom & C. Soler-Monreal (Eds.), *Dialogicity in written specialised genres* (pp. 1-20). Amsterdam, the Netherlands: John Benjamins.

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

Hyland, K. (2016a). Academic publishing and the myth of linguistic injustice. *Journal of Second Language Writing, 31*, 58-69.

Hyland, K. (2016b). Methods and methodologies in second language writing research. *System, 59*, 116-125.

Hyland, K. (2016c). General and specific EAP. In K. Hyland & P. Shaw (Eds.), *The Routledge handbook of English for Academic Purposes* (pp. 17-29). London; New York: Routledge.

Hyland, K. (2017a). English in the disciplines: Arguments for specificity. *ESP Today, 5*(1), 5-23.

Hyland, K. (2017b). Metadiscourse: What is it and where is it going? *Journal of Pragmatics, 113*, 16-29.

Hyland, K., & Jiang, F.K. (2016). Change of attitude? A diachronic study of stance. *Written Communication*, 1-24. doi: 10.1177/0741088316650399

Hyland, K., & Jiang, F.K. (2019). Points of reference: Changing patterns of academic citation. *Applied Linguistics*, *40*(1), 64-85.

Hyland, K., & Jiang, F.K. (2021). A bibliometric study of EAP research: Who is doing what, where and when? *Journal of English for Academic Purposes, 49*, Article 100929.

Hyland, K., & Salager-Meyer, F. (2008). Scientific writing. *Annual Review of Information Science and Technology*, *42*(1), 297-338.

Hyland, K., & Tse, P. (2005). Evaluative *that* constructions: Signalling stance in research abstracts. *Functions of Language, 12*(1), 39-63.

Hyland, K., & Tse, P. (2007). Is there an academic vocabulary? *TESOL Quarterly, 41*(2), 235-253.

Hymes, D. (1974). *Foundations in sociolinguistics: An ethnographic approach*. Philadelphia, PA: University of Pennsylvania Press.

Hyon, S. (1996). Genre in three traditions: Implications for ESL. *TESOL Quarterly, 30*(4), 693-722.

Jaworski, A., & Coupland, N. (2014). *The discourse reader* (3rd ed.). London: Routledge.

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

Jiang, F.K., & Hyland, K. (2018). Nouns and academic interactions: A neglected feature of metadiscourse. *Applied Linguistics, 39*(4), 508-531.

Jobling, R. (1973). A traditional theme in general sociology and its relevance for the study of universities. In R. Brown (Ed.), *Knowledge, education, and ultralight change* (pp. 203-218). London: Tavistock Publications.

Johns, A.M. (1988). The discourse communities dilemma: Identifying transferrable skills for the academic milieu. *English for Specific Purposes, 7*(1), 55-60.

Johns, A.M. (2013). The history of English for specific purposes research. In B. Paltridge & S. Starfield (Eds.), *The handbook of English for specific purposes* (pp. 5-30). Malden, MA: Wiley-Blackwell.

Johns, A.M., & Dudley-Evans, T. (1991). English for specific purposes: International in scope, specific in purpose. *TESOL Quarterly, 25*(2), 297-314.

Johnson, S., Linton, P., & Madigan, R. (1994). The role of internal standards in assessment of written discourse. *Discourse Processes, 18*(2), 231-245.

Jordan, R.R. (1997). *English for academic purposes: A guide and resource book for teachers*. Cambridge, UK: Cambridge University Press.

Journal Citation Reports (2016). *Sciences edition*. New York: Thompson ISI.

Kamler, B., & Thomson, P. (2006, April). *Doctoral writing: Pedagogies for work with literatures*. Paper presented at the AERA annual meeting, San Francisco.

Kanoksilapatham, B. (2012). Structure of research article introductions in three engineering subdisciplines. *IEEE Transactions on Professional Communication, 55*(4), 294-309.

Kanoksilapatham, B. (2015). Distinguishing textual features characterizing structural variation in research articles across three engineering sub-discipline corpora. *English for Specific Purposes, 37*, 74-86.

Kant, I. (2007). *Critique of pure reason* (2nd ed.). Palgrave Macmillan UK. https://doi.org/10.1007/978-1-137-10016-0

Kastens, K.A., Manduca, C.A., Cervato, C., Frodeman, R., Goodwin, C., Liben, L.S., … & Titus, S. (2009). How geoscientists think and learn. *Eos, Transactions, American Geophysical Union, 90*(31), 265-272.

Khedri, M., Heng, C.S., & Ebrahimi, S.F. (2013). An exploration of interactive metadiscourse markers in academic research article abstracts in two disciplines. *Discourse Studies, 15*(3), 1-13.

Kim, C., & Crosthwaite, P. (2019). Disciplinary differences in the use of evaluative *that*: Expression of stance via *that*-clauses in business and medicine. *Journal of English for Academic Purposes, 41*, 1-14.

Kirk, S. (2017). Waves of reflection: Seeing knowledge(s) in academic writing. In J. Kemp (Ed.), *EAP in a rapidly changing landscape: Issues, challenges and solutions* (pp. 109-118). Proceedings of the 2015 BALEAP conference. Reading: Garnet.

Kleinhans, G., Buskes, C.J.J., & de Regt, H.W. (2010). Philosophy of earth science. In F. Allhoff (Ed.), *Philosophies of the sciences* (pp. 213-236). Malden, MA: Wiley-Blackwell.

Kuo, C. H. (1999). The use of personal pronouns: Role relationships in scientific journal articles. *English for Specific Purposes, 18*(2), 121-138.

Kuteeva, M., & McGrath, L. (2015). The theoretical research article as a reflection of disciplinary practices: The case of pure mathematics. *Applied Linguistics, 36*(2), 215-235.

Kwan, B.S.C. (2009). Reading in preparation for writing a PhD thesis: Case studies of experiences. *Journal of English for Academic Purposes, 8*(3), 180-191.

Kwary, D.A. (2011). A hybrid method for determining technical vocabulary. *System*,

39(2), 175-185.

Labov, W. (1972). *Sociolinguistic patterns*. Philadelphia, PA: University of Pennsylvania Press.

Lamont, A., & Maton, K. (2008). Choosing music: Exploratory studies into the low uptake of music GCSE. *British Journal of Music Education, 25*(3), 267-282.

Lavid, J. (2000). Contextual constraints on thematization in written discourse: An empirical study. In P. Bonzon, M. Cavalcanti, & R. Nossum (Eds.), *Formal aspects of context* (pp. 37-47). Dordrecht: Kluwer Academic Publishers.

Le, T.N.P., & Harrington, M. (2015). Phraseology used to comment on results in the discussion section of applied linguistics quantitative research articles. *English for Specific Purposes, 39*, 45-61.

Lea, M.R., & Street, B. (1999). Writing as academic literacies: Understanding textual practices in higher education. In C.N. Candlin & K. Hyland (Eds.), *Writing: Texts, processes and practices* (pp. 62-81). London: Longman.

Lemke, J.L. (1995). Textual politics: Discourse and social dynamics. London: Taylor & Francis.

Leong, P.A. (2014). The passive voice in scientific writing. The current norm in science journals. *Journal of Science Communication, 13*(1), 1-16.

Lewin, B.A. (2005). Hedging: An exploratory study of authors' and readers' identification of "toning down" in scientific texts. *Journal of English for Academic Purposes, 4*(2), 163-178.

Licker, M.D. (Publisher) (2003). *McGraw-Hill dictionary of earth science* (2nd ed.). New York: McGraw-Hill.

Lillis, T. (2008). Ethnography as method, methodology, and "deep theorizing": Closing the gap between text and context in academic writing research. *Written Communication, 25*(3), 353-388.

Lillis, T., & Curry, M.J. (2006). Professional academic writing by multilingual scholars: Interactions with literacy brokers in the production of English-medium texts. *Written Communication, 23*(1), 3-35.

Lillis, T., & Curry, M.J. (2010). *Academic writing in a global context: The politics and practices of publishing in English*. London: Routledge.

Lillis, T., Hewings, A., Vladimirou, D., & Curry, M.J. (2010). The geolinguistics of English as an academic lingua franca: Citation practices across English-medium national and English-medium international journals. *International Journal of Applied Linguistics, 20*(1), 111-135.

Lillis, T., & Scott, M. (2007). Defining academic literacies research: Issues of epistemology, ideology and strategy. *Journal of Applied Linguistics, 4*(1), 5-32.

Lim, J.M.H. (2006). Method sections of management research articles: A pedagogically motivated qualitative study. *English for Specific Purposes, 25*(3), 282-309.

Lim, J.M.H. (2012). How do writers establish research niches? A genre-based investigation into management researchers' rhetorical steps and linguistic mechanisms. *Journal of English for Academic Purposes, 11*(3), 229-245.

Lin, L., & Evans, S. (2012). Structural patterns in empirical research articles: A cross-disciplinary study. *English for Specific Purposes, 31*(3), 150-160.

Liu, J., & Han, L. (2015). A corpus-based environmental academic word list building and its validity test. *English for Specific Purposes, 39*, 1-11.

Lores, R. (2004). On RA abstracts: From rhetorical structure to thematic organisation. *English for Specific Purposes, 23*(3), 280-302.

Love, A. (1993). Lexico-grammatical features of geology textbooks: Process and product revisited. *English for Specific Purposes, 12*(3), 197-218.

Lu, X., Casal, E., & Liu, Y. (2020). The rhetorical functions of syntactically complex sentences in social science research article introductions. *Journal of English for Academic Purposes, 44*, Article 100832.

Luckett, K. (2010). Knowledge claims and codes of legitimation: Implications for curriculum recontextualisation in South African higher education. *Africanus, 40*(1), 6-20.

Luckett, K., & Hunma, A. (2014). Making gazes explicit: Facilitating epistemic access in the humanities. *Higher Education, 67*(2), 183-198.

Luo, N., & Hyland, K. (2017). Intervention and revision: Expertise and interaction in text mediation. *Written Communication*, *34*(4), 414-440.

MacDonald, S.P. (1992). A method for analyzing sentence-level differences in disciplinary knowledge making. *Written Communication, 9*(4), 533-569.

Macnaught, L., Maton, K., Martin, J.R., & Matruglio, E. (2013). Jointly constructing semantic waves: Implications for teacher training. *Linguistics and Education, 24*, 50-63.

Maher, P. (2016). The use of semi-technical vocabulary to understand the epistemology of a disciplinary field. *Journal of English for Academic Purposes, 22*, 92-108.

Martin, J.L. (2016). Musicality and musicianship: Specialization in jazz studies. In K. Maton, S. Hood, & S. Shay (Eds.), *Knowledge-building: Educational studies in Legitimation Code Theory* (pp. 193-213). London: Routledge.

Martin, J.R. (2007). Constructing knowledge: A functional linguistic perspective. In F. Christie & J.R. Martin (Eds.), *Language, knowledge and pedagogy: Functional linguistic and sociological perspectives* (pp. 34-64). New York: Continuum.

Martin, J.R. (2013). Embedded literacy: Knowledge as meaning. *Linguistics and Education, 24*, 23-37.

Martin, J.R., Maton, K., & Doran, Y.J. (Eds.). (2020). *Accessing academic discourse: Systemic Functional Linguistics and Legitimation Code Theory*. London: Routledge.

Martin, J.R., Maton, K., & Doran, Y.J. (2020). Academic discourse: An inter-disciplinary dialogue. In J.R. Martin, K. Maton & Y.J. Doran (Eds.), *Accessing academic discourse: Systemic Functional Linguistics and Legitimation Code Theory* (pp. 1-31). London: Routledge.

Martin, J.R., & Rose, D. (2008). *Genre relations: Mapping culture*. London: Equinox.

Martinez, I.A. (2003). Aspects of theme in the method and discussion sections of biology journal articles in English. *Journal of English for Academic Purposes, 2*(2)*, 103-123*.

Martinez, I.A. (2005). Native and non-native writers' use of first person pronouns in the different sections of biology research articles in English. *Journal of Second Language Writing, 14*(3), 174-190.

Martinez, I.A., Beck, S.C., & Panza, C.B. (2009). Academic vocabulary in agriculture research articles: A corpus-based study. *English for Specific Purposes, 28*(3), 183-198.

Maton, K. (2005a). A question of autonomy: Bourdieu's field approach and higher education policy. *Journal of Education Policy, 20*(6), 687-704.

Maton, K. (2005b). *The field of higher education: Sociology of reproduction, transformation, change and the conditions of emergence for cultural studies* [Unpublished doctoral thesis]. University of Cambridge.

Maton, K. (2007). Knowledge-knower structures in intellectual and educational fields. In F. Christie & J.R. Martin (Eds.), *Language, knowledge and pedagogy: Functional linguistic and sociological perspectives* (pp. 87-108). New York: Continuum.

Maton, K. (2010). Analysing knowledge claims and practices: Languages of legitimation. In K. Maton & R. Moore (Eds.), *Social Realism, knowledge and the sociology of education: Coalitions of the mind* (pp. 35-59). Continuum: London.

Maton, K. (2013). Making semantic waves: A key to cumulative knowledge building. *Linguistics and Education, 24*(1), 8-22.

Maton, K. (2014a). *Knowledge and knowers: Towards a realist sociology of education*. London: Routledge.

Maton, K. (2014b). A tall order? Legitimation Code Theory for academic language and learning. *Journal of Academic Language & Learning, 8*(3), A34-A48.

Maton, K. (2014c). LCT Centre for Knowledge-Building. (2018, April 17). *Karl Maton: Heretical knowledge - Knowledge-blindness in education* [Video]. Youtube. https://www.youtube.com/watch?v=cn9cNZo_bq4

Maton, K. (2016a). Legitimation Code Theory: Building knowledge about knowledge building. In K. Maton, S. Hood, & S. Shay (Eds.), *Knowledge-building: Educational studies in Legitimation Code Theory* (pp. 1-23). London: Routledge.

Maton, K. (2016b). Starting points: Resources and architectural glossary. In K. Maton, S. Hood, & S. Shay (Eds.), *Knowledge-building: Educational studies in Legitimation Code Theory* (pp. 233-243). London: Routledge.

Maton, K. (2018). Thinking like Bourdieu: Completing the mental revolution with Legitimation Code Theory. In J. Albright, D. Hartman, & J. Widin (Eds.), *Bourdieu's field theory and the social sciences* (pp. 249-268). Singapore: Palgrave Macmillan.

Maton, K., & Chen, R.T.H. (2020). Specialization codes: Knowledge, knowers and student success. In J.R. Martin, K. Maton & Y.J. Doran (Eds.), *Accessing academic discourse: Systemic Functional Linguistics and Legitimation Code Theory* (pp. 35-58). London: Routledge.

Maton, K., & Doran, Y.J. (2017a). Semantic density: A translation device for revealing complexity of knowledge practices through discourse, part 1–wording. *Onomazein*, March, 46-76.

Maton, K., & Doran, Y.J. (2017b). Systemic functional linguistics and code theory. In T. Bartlett & G. O'Grady (Eds.), *The Routledge Handbook of Systemic Functional Linguistics* (pp. 629-642). London: Routledge.

Maton, K., & Howard, S.K. (2016). LCT in mixed-methods research: Evolving an instrument for quantitative data. In K. Maton, S. Hood, & S. Shay (Eds.), *Knowledge-building: Educational studies in Legitimation Code Theory* (pp. 49-71). London; New York: Routledge.

Maton, K., & Howard, S. K. (2018) Taking autonomy tours: A key to integrative knowledge-building, *LCT Centre Occasional Paper 1* (June): 1–35.

Maton, K., & Howard, S. K. (2020). Autonomy: the next phase of dialogue between systemic functional linguistics and Legitimation Code Theory. *Journal of World Languages*, 1-21.

Maton, K., Martin, J.R., & Doran, Y.J. (Eds.). (2021). *Teaching science: Knowledge, language, pedagogy*. London: Routledge.

Maton, K., & Muller, J. (2007). A sociology for the transmission of knowledges. In F. Christie, & J. R. Martin (Eds.). *Language, knowledge and pedagogy: Functional linguistic and sociological perspectives* (pp. 14-33). London; New York: Continuum.

Matruglio, E., Maton, K., & Martin, J.R. (2013). Time travel: The role of temporality in enabling semantic waves in secondary school teaching. *Linguistics and Education, 24*, 38-49.

McCarthy, L.P. (1987). A stranger in strange lands: A college student writing across the curriculum. *Research in the Teaching of English, 21*, 233-265.

McEnery, T., & Hardie, A. (2012). *Corpus linguistics: Method, theory and practice*. Cambridge, UK: Cambridge 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.

McNamara, M.S. (2010). What lies beneath? The underlying principles structuring the field of academic nursing in Ireland. *Journal of Professional Nursing, 26*(6), 377-384.

Merton, R.K. (1968). The Matthew effect in science. *Science, 159*(3810), 56-63.

Miller, C.R. (1984). Genre as social action. *Quarterly Journal of Speech, 70*(2), 151-167.

Monbec, L. (2018). Designing an EAP curriculum for transfer: A focus on knowledge. *Journal of Academic Language & Learning, 12*(2), A88-A101.

Montemayor-Borsinger, A. (2003). A comparison of thematic options in novice and expert research writing. *Estudios de Lingüística Aplicada,* (37), 37-51.

Moore, R. (2013). *Basil Bernstein: The thinker and the field*. London; New York: Routledge.

Moreno, A.I., & Swales, J.M. (2018). Strengthening move analysis methodology towards bridging the function-form gap. *English for Specific Purposes, 50*, 40-63.

Mousavi, A., & Moini, M.R. (2014). A corpus study of shell nouns in published research articles of education. *Procedia - Social and Behavioral Sciences, 98*, 1282-1289.

Muller, J. (2007). On splitting hairs: Hierarchy, knowledge and the school curriculum. In F. Christie, & J. R. Martin (Eds.). *Language, knowledge and pedagogy: Functional linguistic and sociological perspectives* (pp. 65-86). London; New York: Continuum.

Myers, G. (1985). Texts as knowledge claims: The social construction of two biology articles. *Social Studies of Science, 15*(4), 593-630.

Myers, G. (1989). The pragmatics of politeness in scientific articles. *Applied Linguistics, 10*(1), 1-35.

Myers, G. (1992). "In this paper we report…": Speech acts and scientific facts. *Journal of Pragmatics, 17*(4), 295-313.

Nation, I.S.P. (2001). *Learning vocabulary in another language*. Cambridge, UK: Cambridge University Press.

Nation, I.S.P. (2013). *Learning vocabulary in another language* (2nd ed.). Cambridge, UK: Cambridge University Press.

Nation, I.S.P., & Chung, T.M. (2003). Technical vocabulary in specialised texts. *Reading in a Foreign Language, 15*(2), 103-116.

Nicol, A., Van Dissen, R. J., Stirling, M. W., & Gerstenberger, M. C. (2016). Completeness of the paleoseismic active-fault record in New Zealand. *Seismological Research Letters, 87*(6), 1299-1310.

North, S. (2005). Disciplinary variation in the use of Theme in undergraduate essays. *Applied Linguistics, 26*(3), 431-452.

Okamura, A. (2008). Citation forms in scientific texts: Similarities and differences in L1 and L2 professional writing. *Nordic Journal of English Studies, 7*(3), 61-81.

Parkin, F. (1979). *Marxism and class theory: A bourgeois critique.* London: Tavistock.

Paul, D. (2000). In citing chaos: A study of the rhetorical use of citations. *Journal of Business and Technical Communication, 14*(2), 185-222.

Peacock, M. (2002). Communicative moves in the discussion section of research articles. *System, 30*(4), 479-497.

Peacock, M. (2006). A cross-disciplinary comparison of boosting in research articles. *Corpora, 1*(1), 61-84.

Peacock, M. (2011). The structure of the methods section in research articles across eight disciplines. *Asian ESP Journal, 7*(2), 97-124.

Pecorari, D. (2006). Visible and occluded citation features in postgraduate second-language writing. *English for Specific Purposes, 25*(1), 4-29.

Pennycook, A. (1994). The politics of pronouns. *ELT Journal, 48*(2), 173-178.

Penrose, A.M., & Katz, S.B. (2010). Writing in the sciences: Exploring conventions of scientific discourse (3rd ed.). New York: Longman.

Pérez-Llantada, C. (2013). The Article of the future: Strategies for genre stability and change. *English for Specific Purposes, 32*, 221-235.

Pérez-Llantada, C. (2016). How is the digital medium shaping research genres? Some cross-disciplinary trends. *ESP Today, 4*(1), 22-42.

Perrucci, R., Perrucci, C., & Subramaniam, M. (2019). Publications in four sociology journals, 1960-2010: The role of discipline demographics and journal mission. *Sociological Focus, 52*(3), 171-185.

Pilson, M.E.Q. (2013). *An introduction to the chemistry of the sea* (2nd ed.). New York: Cambridge University Press.

Poulet, C. (2016). Knowledge and knowers in tacit pedagogic contexts: Freemasonry in France. In K. Maton, S. Hood, & S. Shay (Eds.), *Knowledge-building: Educational studies in Legitimation Code Theory* (pp. 214-230). London; New York: Routledge.

Prior, P.A. (1998). *Writing/disciplinarity: A sociohistoric account of literate activity in the academy*. Mahwah, NJ: Lawrence Erlbaum.

Raab, T., & Frodeman, R. (2002). What is it like to be a geologist? A phenomenology of geology and its epistemological implications. *Philosophy & Geography, 5*(1), 69-81.

Rampton, B., Tusting, K., Maybin, J., Barwell, R., Creese, A., & Lytra, V. (2004). *UK linguistic ethnography: A discussion paper*. Retrieved from http://www.lancs.ac.uk/fss/organisations/lingethn/documents/

Rashidi, N., & Ghaedi, H. (2013). Thematic analysis of the conclusion sections of TEFL and geology research articles. *Studies in Applied Linguistics, 1*(7), 29-52.

Rau, G., & Shih, Y.S. (2021). Evaluation of Cohen's kappa and other measures of inter-rater agreement for genre analysis and other nominal data. *Journal of English for Academic Purposes, 53*, Article101026.

Reed, C. (2008). *Earth science: Decade by decade*. New York: Infobase Publishing.

Ren, J. (2021). Variability and functions of lexical bundles in research articles of applied linguistics and pharmaceutical sciences. *Journal of English for Academic Purposes, 50*, Article 100968.

Rey, P.F. (n.d.). *Introduction to structural geology*. Retrieved from http://www.geosci.usyd.edu.au/users/prey/Patrice_Intro_to_SG.pdf

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

Samraj, B. (2004). Discourse features of the student-produced academic research paper: Variations across disciplinary courses. *Journal of English for Academic Purposes, 3*(1), 5-22.

Samraj, B. (2014). Move structure. In K.P. Schneider & A. Barron (Eds.), *Pragmatics of discourse* (pp. 385-405). Berlin, Germany: Mouton de Gruyter.

Samraj, B. (2016). Research articles. In K. Hyland & P. Shaw (Eds.), *The Routledge handbook of English for Academic Purposes* (pp. 403-415). London; New York: Routledge.

Schmitt, N. (2010). *Researching vocabulary: A vocabulary research manual.* Basingstoke, UK: Palgrave Macmillan.

Shapin, S. (1984). Pump and circumstance: Robert Boyle's literary technology. *Social Studies of Science, 14*(4), 481-520.

Shay, S. (2013). Conceptualizing curriculum differentiation in higher education: A sociology of knowledge point of view. *British Journal of Sociology of Education, 34*(4), 563-582.

Shay, S. (2016). Curricula at the boundaries. *Higher Education, 71*(6), 767-779.

Shearer, P.M. (2009). *Introduction to seismology* (2nd ed.). New York: Cambridge University Press.

Shehzad, W. (2010). Announcements of the principle findings and value addition in Computer Science research papers. *Ibérica, 19*(1), 97-118.

Sheldon, E. (2009). From one *I* to another: Discursive construction of self-representation in English and Castilian Spanish research articles. *English for Specific Purposes, 28*(4), 251-265.

Skelton, J. (1997). The representation of truth in academic medical writing. *Applied Linguistics*, *18*(2), 121-140.

Sinclair, J. (2004). Corpus and text – Basic principles. In M. Wynne (Ed.), *Developing linguistic corpora: A guide to good practice*. Retrieved from http://icar.univ-lyon2.fr/ecole_thematique/contaci/documents/Baude/wynne.pdf

Singh, P. (2015). The knowledge paradox: Bernstein, Bourdieu, and beyond. *British Journal of Sociology of Education, 36*(3), 487-494.

Smagorinsky, P. (2008). The method as conceptual epicentre in constructing social science research reports. *Written Communication*, *25*(3), 389-411.

Spack, R. (1988). Initiating ESL students into the academic discourse community: How far should we go? *TESOL Quarterly, 22*(1), 29-51.

Starfield, S. (2001). 'I'll go with the group': Rethinking 'discourse community' in EAP. In J. Flowerdew & M. Peacock (Eds.), *Research perspectives on English for academic purposes* (pp. 132-147). Cambridge, UK: Cambridge University Press.

Starfield, S. (2014). Current and future directions in English for specific purposes research. *Revue francaise de linguistique appliquee, 19*(1), 9-14.

Starfield, S. (2015). First person singular: Negotiating identity in academic writing in English. In D.N. Djenar, A. Mahboob, & K. Cruikshank (Eds.), *Language and identity across modes of communication* (pp. 249-262). Berlin, Germany: Walter De Gruyter.

Starfield, S., & Ravelli, L.J. (2006). "The writing of this thesis was a process that I could not explore with the positivistic detachment of the classical sociologist": Self and structure in new humanities research theses. *Journal of English for Academic Purposes, 5*(3), 222-243.

Street, B.V. (2013). *Social literacies: Critical approaches to literacy in development, ethnography and education*. London; New York: Routledge.

Strevens, P. (1988). ESP after twenty years: A re-appraisal. In M.L. Tickoo (Ed.), *ESP: State of the art* (pp. 1-13). Singapore: SEAMEO Regional Language Centre.

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

Swales, J.M. (1997). English as Tyrannosaurus rex. *World Englishes, 16*(3), 373-382.

Swales, J.M. (2001). EAP-related linguistic research: An intellectual history. In J. Flowerdew & M. Peacock (Eds.), *Research perspectives on English for academic purposes* (pp. 42-54). Cambridge, UK: Cambridge University Press.

Swales, J.M. (2002). Integrated and fragmented worlds: EAP materials and corpus linguistics. In J. Flowerdew (Ed.), *Academic discourse* (pp. 150-164). London: Longman.

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

Swales, J.M. (2019). The futures of EAP genre studies: A personal viewpoint. *Journal of English for Academic Purposes, 38*, 75-82.

Swales, J.M., & Feak, C.B. (2004). *Academic writing for graduate student: Essential tasks and skills* (2nd ed.). Ann Arbor, MI: The University of Michigan Press.

Swales, J.M., and Feak, C.B. (2012). *Academic writing for graduate students: Essential tasks and skills* (3rd ed.). Ann Arbor, MI: The University of Michigan Press.

Szenes, E., Tilakaratna, N., & Maton, K. (2015). The knowledge practices of critical thinking. In M. Davies, & R. Barnett (Eds.), *The Palgrave handbook of critical thinking in higher education* (pp. 573-591). New York: Palgrave Macmillan.

Tang, R., & John, S. (1999). The 'I' in identity: Exploring writer identity in student academic writing through the first person pronoun. *English for Specific Purposes, 18*, S23-S39.

Tann, K., & Scott, A. (2021). Bridging disciplinary knowledge: The challenge of integrating EAP in business education. *Higher Education, 81*(3), 453-470.

Tarone, E., Dwyer, S., Gilette, S., & Icke, V. (1981). On the use of the passive in two astrophysics journal papers. In J.M. Swales (1985), *Episodes in ESP: A source and reference book on the development of English for science and technology* (pp. 191-205). Oxford, UK: Pergamon.

Tarone, E., Dwyer, S., Gilette, S., & Icke, V. (1998). On the use of the passive in two astrophysics journal papers: With extensions to other languages and other fields. *English for Specific Purposes, 17*(1), 113-132.

Teddlie, C., & Yu, F. (2007). Mixed methods sampling: A typology with examples. *Journal of Mixed Methods Research, 1*(1), 77-100.

Thompson, G. (2013). *Introducing functional grammar* (3rd ed.). London; New York: Routledge.

Thompson, G., & Ye, Y. (1991). Evaluation in the reporting verbs used in academic papers. *Applied Linguistics, 12*(4), 365-382.

Thompson, P. (2002). Manifesting intertextuality in the PhD thesis. *Revista Canaria de Estudios Ingleses, (*44), 97-114.

Thompson, P., & Tribble, C. (2001). Looking at citations: Using corpora in English for Academic Purposes. *Language Learning & Technology, 5*(3), 91-105.

Tognini-Bonelli, E. (2001). *Corpus linguistics at work*. Amsterdam, the Netherlands: J. Benjamins.

Toulmin, S.E. (2003). *The uses of argument* (Updated ed.). Cambridge, UK: Cambridge University Press.

Tse, P., & Hyland, K. (2010). Claiming territory: Relative clauses in journal descriptions. *Journal of Pragmatics, 42*(7), 1880-1889.

Tusting, K., McCulloch, S., Bhatt, I., Hamilton, M., & Barton, D. (2019). *Academics writing: The dynamics of knowledge creation*. London; New York: Routledge.

Tyler, W. (2014). Legitimation Code Theory and the social semiotics of pedagogic discourse: A Bernsteinian critique. Retrieved January 30, 2022, from https://www.researchgate.net/publication/

271207986_Legitimation_Code_Theory_and_the_Social_Semiotics_of_Pedagogic_Discourse_a_Bernsteinian_Critique

Vahed, A., McKenna, S., & Singh, S. (2016). Linking the "know-that" and "know-how" knowledge through games: A quest to evolve the future for science and engineering education. *Higher Education, 71*(6), 781-790.

Valipouri, L., & Nassaji, H. (2013). A corpus-based study of academic vocabulary in chemistry research articles. *Journal of English for Academic Purposes, 12*(4), 248-263.

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.

Van Krieken, R., Habibis, B., Smith, P., Hutchins, B., Martin, G., & Maton, K. (2014). *Sociology: Themes and perspectives* (5th ed.). Sydney: Pearson.

Varttala, T. (2001). *Hedging in scientifically oriented discourse. Exploring variation according to discipline and intended audience*. Tampere University Press.

Vassileva, I. (2001). Commitment and detachment in English and Bulgarian academic writing. *English for Specific Purposes, 20*(1), 83-102.

Vergaro, C. (2015). Ways of asserting. English assertive nouns between linguistics and the philosophy of language. *Journal of Pragmatics, 84*, 1-17.

Vladimirou, D. (2007). 'I suggest that we need more research': Personal reference in linguistics journal articles. In C. Gabrielatos, R. Slessor, & J.W. Unger (Eds.), *Papers from the Lancaster University Postgraduate Conference in Linguistics and Language Teaching: Vol. 1. Papers from LAEL PG 2006* (pp. 139–57). Retrieved from http://www.lancaster.ac.uk/fass/events/laelpgconference/papers/v01/Vladimirou.pdf

Vongpumivitch, V., Huang, J., & Chang, Y. (2009). Frequency analysis of the words in the Academic Word List (AWL) and non-AWL content words in applied linguistics research papers. *English for Specific Purposes, 28*(1), 33-41.

Vygotsky, L. (1986). *Thought and language* (A. Kozulin, Trans.). Cambridge (MA); London (UK): The MIT Press. (Original work published 1934)

Wacquant, L.J.D. (1989). Towards a reflexive sociology: A workshop with Pierre Bourdieu. *Sociological Theory, 7*(1), 26-63.

Wang, J., Liang, S., & Ge, G. (2008). Establishment of a medical academic word list. *English for Specific Purposes, 27*(4), 442-458.

Ware, M. & Mabe, M. (2015) *The STM Report: An overview of scientific and scholarly journal publishing*. University of Nebraska – Lincoln. Retrieved from http://digitalcommons.unl.edu/cgi/viewcontent.cgi?article=1008&context=scholcom

Web of Science. (n.d.). https://apps.webofknowledge.com/

Weninger, C., & Williams, J.P. (in press). Goffman and sociolinguistics. In M. H. Jacobsen & G. Smith (Eds.), *Routledge International Handbook of Goffman Studies*. London; New York: Routledge.

White, H.D. (2004). Citation analysis and discourse analysis revisited. *Applied Linguistics, 25*(1), 89-116.

Widdowson, H.G. (1983). *Learning purpose and language use*. Oxford, UK: Oxford University Press.

Williams, I.A. (1999). Results sections of medical articles: Analysis of rhetorical categories for pedagogical purposes. *English for Specific Purposes, 18*(4), 347-366.

Wolff, K. (2017). Engineering problem-solving knowledge: The impact of context. *Journal of Education and Work, 30*(8), 840-853.

Wolff, K. (2020). Researching the engineering theory-practice divide in industrial problem solving. *European Journal of Engineering Education, 45*(2), 181-195.

Xia, S.A. (2020). Genre analysis in the digital era: Developments and challenges. *ESP Today, 8*(1), 141-159.

Xue, G., & Nation, I.S.P. (1984). A university word list. *Language Learning and Communication, 3*(2), 215-229.

Young, B. (2014). What do engineering researchers cite? A citation analysis study of sixteen engineering journals. Retrieved from http://www.istl.org/14-winter/refereed4.html

Young, M.F.D. (1971). Introduction. In M.F.D. Young (Ed.), *Knowledge and control: New directions for the sociology of education* (pp. 1-17). Collier-Macmillan.

Young, M.F.D. (2008). *Bringing knowledge back in: From social constructivism to social realism in the sociology of education*. London; New York: Routledge.

Young, M.F.D. (2010). Why educators must differentiate knowledge from experience. *The Journal of the Pacific Circle Consortium for Education, 22*(1), 9-20.

Young, M.F.D. (2020). Knowledge and the sociology of education. *Acta Paedagogica Vilnensia*, *44*, 10-17.

Zamel, V. (1998). Questioning academic discourse. In V. Zamel & R. Spack (Eds.), *Negotiating academic literacies: Teaching and learning across languages and cultures* (pp. 187-197). Mahwah, NJ: Lawrence Erlbaum.

Appendices

Appendix I: A list of journals from which the corpus research articles were sourced

Journal sources for marine chemistry RAs	No. of RAs
Applied Geochemistry	1
Biogeochemistry	1
Biogeosciences	9
Chemical Geology	3
Deep-Sea Research II	2
Earth and Planetary Science Letters	7
Estuarine, Coastal and Shelf Science	6
Geobiology	1
Geochimica et Cosmochimica Acta	8
Geophysical Research Letters	2
Global Biogeochemical Cycles	4
ICES Journal of Marine Science	1
Journal of Geophysical Research: Oceans	1
Journal of Sea Research	1
Marine Chemistry	5
Marine Ecology Progress Series	2
Ocean Science	1
Paleogeography, Paleoclimatology, Paleoecology	2
Precambrian Research	1
Quaternary Science Reviews	2

Journal sources for seismology RAs	No. of RAs
Acta Geophysica	1
Arabian Journal of Geosciences	1
Bulletin of the Seismological Society of America	3

Comptes Rendus Geoscience	1
Earth and Planetary Science Letters	3
Earth, Planets and Space	1
Geochemistry, Geophysics, Geosystems	1
Geofisica Internacional	1
Geophysical Journal International	8
Geophysical Research Letters	3
Geotectonics	1
Journal of African Earth Sciences	2
Journal of Asian Earth Sciences	2
Journal of Geodynamics	1
Journal of Geophysical Research: Earth Surface	1
Journal of Geophysical Research: Solid Earth	8
Journal of Seismology	4
Journal of Volcanology and Geothermal Research	2
Natural Hazards	1
Natural Hazards and Earth System Sciences	1
Physics and Chemistry of the Earth	1
Polar Science	1
Pure and Applied Geophysics	1
Russian Geology and Geophysics	1
Russian Journal of Pacific Geology	1
Seismological Research Letters	1
Swiss Journal of Geosciences	1
Tectonophysics	6
Terrestrial, Atmospheric and Ocean Sciences	1

Appendix II: The coding sheet used to identify ranking clauses

Identifying clauses: Guiding principles

1. Orthographic sentences (starting with a capital letter and ending with a full stop) correspond to grammatical clause simplexes and clause complexes.
2. A verbal group realizing Process is an obligatory clause element.
3. A process may be realized by a complex verbal group (e.g., **succeed in changing/ has been sleeping/could have been being watched/seem to have disappeared/ was found to be/tends to be**). The secondary group is always non-finite; it may be perfective (with or without *to*) or imperfective (participle). "A verbal group nexus is intermediate between a clause nexus and a verbal group: a verbal group construes a single event, and a clause nexus construes two distinct processes; but a verbal group nexus construes a single process consisting of two events" (Halliday and Matthiessen, 2013, p. 588).

- … responsible [[for **helping** Chen **keep** his promise [[to clean up graft]]]]. ||| [Text 13]
- … [[that **are** actually **helping** us to **get on** with the business [[of increasing understanding]]]] [Text 32]

- I **saw** it **being done**.

- Causative (**make sb do; force sb to do; cause sb to do; get sb to do; have sb to do; require sb to do; let sb do; allow sb to do; permit sb to do; oblige sb to do; encourage sb to do**): an alternative realisation of effective agency: an Initiator brings about the action done by the Actor): *Mary **made** John **roll** the ball*. In other words, "the causative verbal group complex is thus an alternative realization of the feature of 'effective' agency: an additional participant is introduced into the clause through the expansion of the verbal group realizing the Process" (Halliday and Matthiessen, 2014, p. 579).

John	made	the ball	roll
	Pro-		-cess
	α		×β

Mary	made	John	roll	the ball
	Pro-		-cess	
	α		×β	

- Causative: ||| This strong support **has enabled** us **to execute** these missions || without taxing our already-stressed readiness and modernization accounts. ||| [Text 115]

- Causative: Mary **reminded** John **to do** it.

- Causative: Mary **wanted/expected/predicted** John **to go**.

- **Succeed in changing; suffice in triggering; may attempt to circumvent; I'm going to try and attach; I wouldn't know how to behave** (i.e., knowing as in having an ability)**; hasten to add; happen to be lying; forget to return; could start by talking; insist on doing; help to do; help in/with doing;** To distinguish it from several verbal groups, three tests are applied:

- "In order" test for *verb **to** verb*: A complex verbal group doesn't permit "in order" to be inserted within it (cf. The model **was used to analyse** the data. –> **was used** *in order* **to analyse** the data or **To analyse** the data we **used** the model (= two clauses)

and We **tend to use** the model for data analysis. –> ***tend** *in order* **to use** the model or ***To use** the model we **tend** (= one clause))

- "Time reference" test for verb **to** verb: A complex verbal group represents an activity that occurs over a single time frame (i.e., there is no time distinction between the first Process and a subsequent one).
- "Mobility" test for other verbal sequences: Alvin **got up having slept** for five days. –> **Having slept** for five days, Alvin **got up**. (= two clauses)

4. Some clause-like segments occur as part of a larger clause. Such segments are then called *downranked* (*rankshifted* or *embedded)* clauses. They have the form of a clause, but function as a participant. Downranked clauses can occur only with nominal groups (as subjects or complements or as parts of subjects or complements, often as qualifiers, which are almost always embedded) or as parts of adverbial groups, but never within verbal groups.

Examples of downranked (rankshifted or embedded) clauses:

Embedding in a nominal group (including adjectives)

a) As a postmodifier/qualifier of a subject

- ||| The only person **[[who was kind to him at all]]** was the Skin Horse, || who had lived longer in the nursery than any of the others. ||| [Text 28]
- The decision [by the Arbitration Commission] **[[to award the 2.3 percent pay increase]]** is unfortunate, but it was to be expected. [ACE_B]

b) As a subject

- **[[What is beautiful]]** is also in some respects ugly; **[[what is just]]** is in some respects unjust. (Bloor & Bloor, 2013, p. 172)
- **[[That Aristotle appealed to such principles]]** is not surprising. (Bloor & Bloor, 2013, p. 172)
- You find some humorous proverbs, for instance, and the humor is that **[[whoever made these proverbs]]** was not going around the world with his eyes closed. [Text 16]

c) As a postposed subject

- When a clause contains an anticipatory *it* and an embedded clause, both *it* and the embedded clause can be the Subject.
- *It* is this latter question **[[*which is often ignored]]***.
- ||| It's not nice **[[what's happening to him]]** || but he is creating the situation just as much as [[what they are]], || because he's been caught || drinking on the job, || (whisper) which is no good you know. ||| [Text 71]
- ||| My dear Ellen: Your kind letter of October 19th, 1917, has just been received || and it seems nice **[[to hear something of the family]]** || and I shall be thankful to you for further communications. ||| [Text 111]

- **Predicated Theme**

- ||| **It was because of the protracted delays [[caused by litigation]]** [[that land ceiling laws were put in the Ninth Schedule of the Constitution || to give them immunity [[from being challenged in the courts]]]]. ||| [KOHL_B]
- However, **it is the presence of Mn– Fe oxides on sediments throughout the subterranean estuary** [[that resulted in increased partitioning between the adsorbed and dissolved phase]] (MChem 49)

d) As a (postmodifier/qualifier of a) complement

1. As Postmodifiers of Complements

This method is suitable [[for analysing the collected samples]].
He was successful [[in applying for grants]].
[this is ok I think. In SFL, adjectives are taken as nominal groups or part of, so they can be followed by postmodifiers.]

2. As Adjuncts (as they seem to be mobile, whereas Postmodifiers cannot be moved to the beginning of the clause) [dubious. As the complements still belong more to the adjectives.]

This method is suitable for [[analysing the collected samples]].
He was successful in [[applying for grants]].

3. Ranking clauses [better not]

This method is suitable || for analysing the collected samples.
He was successful || in applying for grants.

- Our knowledge is insufficient **[[to determine X]]**.
- They took **[[what they wanted]]**. (Bloor & Bloor, 2013, p. 173)
- I'm trying to do **[[what you asked]]**. (Butt, Fahey, Feez, & Spinks, 2012, p. 201)
- You will enjoy **[[meeting your fellow members]]**. (Bloor & Bloor, 2013, p. 173)
- The last thing [[we needed]] was **[[for you to wake the baby]]**. (Butt, Fahey, Feez, & Spinks, 2012, p. 201)
- ||| The worsening concentration of global corporate power over our government has turned that government frequently against its own people, || denying its people their sovereignty **[[to shape their future]]**. ||| [Text 174]
- I have kind of an eclectic mind so I get interested in lots of different things, and I generally get very focused on **[[whatever it is I'm interested in]]**. [Text 17]
- ready **[[to die for the cause]]**; quick **[[to reply]]**; glad **[[to oblige]]**; eager **[[to help]]**; happy **[[to be of service]]** (Bloor & Bloor, 2013, p. 176)

e) As a postmodifier/qualifier of an adjunct

Adjuncts provide more description. They are grammatically optional.
Adjuncts are circumstantial (place, time, manner, associated participants), conjunctive, and modal. Adjuncts can be realised by an NG, a PP, or an adverb.
A crude probe for distinguishing between Adjunct and Postmodifier: an Adjunct can (usually) be moved; a Postmodifier cannot.

- We have to recognize that the opportunity for doing such harm is monumental if the exercise is not conducted with very close attention to **[[[what works already || and therefore does not need fixing or extra coordinating]]]**. [Text 32]
- The circuit should be tested with some device **[[which reliably indicates the presence of mains voltage]]**. (Bloor & Bloor, 2013, p. 164)
- In modern homes the mains switch and the fuses are contained in a box **[[called a consumer unit]]**. (Bloor & Bloor, 2013, p. 169)
- On the Friday morning, we moved from **[[writing in the cosines of the cottage]]** to **[[lunching in the dampness of Wentworth Falls]]**. (Butt, Fahey, Feez, & Spinks, 2012, p. 201)
- … that this idea started from **[[people saying** [[Arthur's sword was magic…]]]]… (Butt, Fahey, Feez, & Spinks, 2012, p. 209)

Embedding in an adverbial group

A rankshifted clause cannot be equated with the whole adverbial group; it can only be a part of it.
With expressions such as *more, as; too much, such, not ... enough, so.*

- ||| He left Weeks **as quickly [[as he could]]**. ||| [Of Human Bondage]
- ||| If the planned global change programs are **as successful [[as they promise to be]]**, || they are going to create many **more problems** for the policy and management community **[[than they solve]]**, at least in the short run. ||| [Text 32]
- I'm **as certain** of it [[**as if** his name were written all over his face]]
- **as grimly [[as if his life depended on it]]**
- ||| We need to get that message across at least **as much [[as we need to be concerned [[[with getting the FY 1989 budget secured || or getting a congressional hearing on immediate solutions to immediate problems]]]]]**. ||| [Text 32]
- **much more easily [[than you would have expected]]**
- The actual formation of shale is somewhat **more complex** [[**than** indicated in Table 4-3]]. [Text 68]
- ||| The abundance of shale is **somewhat less [[than is predicted from the abundance of clay-forming silicate minerals]]**, ||| [Text 68]
- Another survivor, soaked, wide-eyed with shock and **too distressed** [[**to give** his name]], said 'We were having a wonderful time when it turned into a nightmare.' [Text 30]
- **too quickly [[for us to see [[what was happening]]]]**
- **too cleverly [[for anyone to imitate]]**
- Within the vortex, temperatures become cold **enough** [[**to form** stratospheric ice crystals]]. [Text 33]
- **not long enough [[to find my way around]]**
- Then he told us anecdotes of how he had gone across the Channel when it was **so rough** [[**that** the passengers had to be tied into their berths, and he and the captain were the only two living souls on board who were not ill]]. [Text: Three Men in a Boat]
- ||| His face grew very flushed || and his little body was **so hot [[[that it burned the Rabbit || when he held him close]]]**. ||| [Text 28]
- **so soon [[that no one was ready]]**
- **so sad [[that I cried]]**

5. Use the following probes to look for rankshifted clauses (especially with mental processes):

    1) Is subject possible?
    A ranking clause cannot be a participant (subject), while a rankshifted can in the passive variant.
    *They reasoned that the plume indicated an eruption. –> *That the plume indicated an eruption was reasoned by them.*
    They accepted [[that the plume indicated an eruption]]. –> That the plume indicated an eruption was accepted by them.
    2) Is direct thought possible?

A ranking clause can be rephrased as a direct thought/utterance:
They reasoned, "The plume indicates an eruption".
**They accepted, "The plume indicates an eruption".*

3) Is "the fact" possible?
Only rankshifted clauses can postmodify *"the fact"*:
**They reasoned the fact that the plume indicated an eruption.*
They accepted the fact that the plume indicated an eruption.

4) Is "focus" possible?
It + be + focus + non-defining relative clause
Ranking clauses cannot be the focus in cleft constructions.
**It is that the plume indicated an eruption that they reasoned.*
It is that the plume indicated an eruption that they accepted.

5) Replacing with "so" (ranking) or "it" (rankshifted)? (This is the weakest test!!!)
They reasoned so.
They accepted it.

Projection

"There is a grammatical difference between the projections with verbs and with nouns, however, in that the clauses projected by the verbs are dependent clauses, whereas those which occur with the nouns are embedded" (Bloor and Bloor, 2013, p. 209).

Verbal-projection of locutions (what is said)		**Mental-projection of ideas (what is thought)**	
Verb	**Noun**	**Verb**	**Noun**
Assert	Assertion		
Claim	Claim	Assume	Assumption
Confess	Confession	Believe	Belief
Declare	Declaration	Conceive	Conception
Deny	Denial	Conclude	Conclusion
Insist	Insistence	Desire	Desire
Proclaim	Proclamation	Hypothesise	Hypothesis
Promise	Promise	Imagine	
Say		Speculate	Speculation
State	Statement	Suppose	Supposition
Swear		Theorise	Theory
Tell	Tale?	Think	Thought

Vow	Vow		View
		Wish	Wish

Verbs serving as Process in 'verbal' clauses used to quote (Halliday & Matthiessen, 2014, p. 514)

	Proposition	**Proposal**
(1) general member	*say*	*say*
(2) verbs specific to speech function; (a) giving	(a) statement: *tell (+ Receiver), remark, observe, point out, report, announce*	(a) *offers: suggest, offer, threaten* ('offer: undesirable'), *vow* ('offer: sacred'), *promise* ('offer: desirable'), *agree ('*offer: in response*')*
(b) demanding	(b) questions: *ask, demand, inquire, query*	(b) commands: *call, order, request, tell, propose, decide; urge* ('command: persuasive'), *plead ('*command: desperate'), *warn ('*command: undesirable consequences*')*
(3) verbs with additional circumstantial feature: (a)	*reply* ('say in response), *explain* ('say in explanation'), *protest* ('say with reservation'), *continue* ('go on saying'), *add* ('say' in addition'), *interrupt* ('say out of turn'), *warn* ('say: undesirable consequences')	[see (2) above]
(b) manner specifying connotation	*insist* ('say emphatically'), *complain* ('say irritably'), *cry, shout* ('say loudly'), *boast* ('say proudly'), *murmur* ('say sotto voce'), *stammer* ('say with embarrassment'), *enthuse* ('say with approval'), *gush* ('say effusiveness'), *rave* ('say with enthusiasm')	[largely the same as for propositions] *blare, thunder* ('order imperiously'), *moan* ('plead whiningly'), *yell ('*order vociferously*'), fuss ('*order officiously*')*

Verbs serving as Process in 'mental' clauses reporting ideas (Halliday & Matthiessen, 2014, p. 517)

		Proposition		**Proposal**
		statement [indirect declarative clause: (*that) ...]*	**question [indirect interrogative clause:** *whether/if; who, which, when ...*	**[perfective non-finite clause, or modulated indirect declarative clause]**
perceptive				-
cognitive	like' type	*believe, guess, think, know, imagine, double., remember, forget, dream, predict*	*wonder, double; consider; find out, ascertain, check; determine, judge; predict [interrogative/ negative clause:] know, remember [= the*	-
		e.g. she knew || that he'd left	*e.g. she wondered (didn't know) || whether*	
	please' type	*strike, occur to, dawn on*		
		e.g.. it struck her that he'd left		
desiderative	like' type	-	-	*want, would like, wish, intend, plan for, hope*
				e.g. she wanted || him to leave (that he should leave)
emotive		-	-	

Other cases

- The non-finite clause is often introduced by a preposition or prepositional group functioning conjunctively, e.g. *besides*, *apart from*, *instead of*, *other than*, *without*; for example

  - **additive**

    - ||| **Apart from** being amusing || what else does The Nun's Priest's Tale do? ||| [Text 125]
    - ||| **Besides** being gifted with literary talent, || Amir Khusrau was a musician, too. ||| [KOHL_C]
    - ||| Most families are dependent on two salaries coming into the home, || **with** women now constituting almost half the country's workforce. ||| [Text 388]

  - **adversative**
    ||| Until we do that, || the opportunities may come and go || **without** our having a compelling rationale [for pushing commitment and action]. ||| [Text 32]

    - ||| The arrow changed its course || and fell to the ground || **without** harming anyone. ||| [Text 65]

- **replacive**
  ||| **Instead of** finding the perpetrators, || they criminally charged the Earth First! activist, || who was left crippled for life. ||| [Text 214]

- **subtractive**
  ||| We call him a murderer, || but for him there is no way out || **other than** doing the deed. ||| [KOHL_K]

- ||| The observed normal-reverse composite source model of the event [[connected via the linking fault]] could also be an early *M*8 reverse-faulting aftershock [[triggered simultaneously [[as the *M*8 normal-faulting earthquake propagates]]]]. |||

- ||| Residual bathymetry is calculated || by regionally subtracting a spectral average of the trench-normal topography (Figure 11) [*Bassett and Watts*, 2015a, 2015b]. |||

- Soon the **time** came [[**for Kukul** to take his place among the men of his nation]]

- It was **careless of him [[to put** another man's helmet on]]

Assorted individual expressions

- It's an interesting fact [[[that, <<**compared with** other countries, >> Australians are not very heavy drinkers]]]. [Australia fact website]

- **Including** = preposition

- **Considering** = preposition (followed by a noun), conjunction (followed by *that*, *what*, *how*)

- **As opposed to** = phrase

- **Given** their reaction with citrate bicarbonate solution… = preposition

- **Owing to** = preposition

- **Depend on** = phrasal verb

- **Taken together,** the deterioration of relations with Laos and Cambodia worried the Thai military regime. [Wyatt, 284]

- **be able to do something/capable of doing something/be available to do something** = phrases (**no** embedding such as able [[to do something]])

- ||| The disappearance rate of hGH from plasma is multiexponential (Cameron *et al*. 1969) || and therefore the significance of half life estimates of 20 to 30 minutes **is difficult to state.** |||

- ||| Thus much chert is recrystallized, || **making** the origin **difficult to discern**. |||

- it is unclear [[whether this size-abundance response to pCO_2 is monotonic in that arctic system]] (MCHem 15)
- ||| He learned to walk in a certain way, || to have a certain accent, || all **based on** the nose. ||| [Text 17]
- ||| These indicators show us || where we've been || and will help us project future readiness trends [[**based on** current funding and OPTEMPO]] . ||| [Text 115]

Modalizers vs. non-modalizers

- **Assuming** (not a modalizer but a conjunction, although listed as a hedge) that rising CO_3^{2} indicates loss of CO_2 gas from seawater]]…
- **This assumes** (modalizer) that equilibration between CO2[aq] and HCO3 is always faster than losses of CO2 by primary production,
- **An important implication of this is…** (modalizer)
- **The alternative explanation for the very low d13CPOC values is…** (modalizer)
- **Together these observations suggest** (modalizer) that the migrating hybrid earthquakes were **the first indicators** [[[that the lava [[moving through the conduit]] was stalling, || and that the conduit itself was becoming clogged]]] Seis 3
- **This is the first time [[that** left-lateral shear is clearly documented inside the Gulf of Evia rift]] Seis 2
- is in agreement with the observations of Thompson et al. (2006) [[that natural Fe(III) oxides became more crystalline during repeated redox oscillations]]. MChem 45
- but the mean pattern is at least consistent with our claim [[[that non-monotonic responses may be common, || even when sophisticated methods [[to address acclimation]] have been included]]] MChem 15

Participle clauses with their own subject

- The subject is often introduced by ***with*** when the clause expresses accompanying circumstances.

- ||| With him/his taking time off || everyone has to work harder. |||
- A car roared past with smoke pouring from the exhaust.
- With Peter travelling most of the week, the house seems pretty empty.
- The score being level after 90 minutes, a replay will take place.

- Principle 6: Patterns of agnation exist between circumstance types in the clause and the logico-semantic types of relation between clauses in the clause complex.

||| Each day, she prayed **with all her heart**. ||| [Text 65]

is agnate with a (non-finite) clause linked to *each day she prayed* in a clause complex:

||| Each day, she prayed || using all her heart. |||

The phrase *with all her heart* augments the clause circumstantially within the domain of the clause; in contrast, the clause *using all her heart* **expands** the clause, relating it to a full-fledged clause (rather than to a prepositional phrase or adverbial group) and thereby opening up a clause complex. Here are some additional examples of circumstances serving within the domain of the clause with suggested agnate clauses within the domain of the clause complex:

[enhancement: time] **Under his rule**, there was peace throughout the land. ~ **when** he ruled [Text 65]

[enhancement: reason] **because of this child**, he would never become hala uinic. ~ **because** this child lived [Text 65]

[extension: alternation] **In place of the usual expensive, elaborate costumes**, the Motleys created simple but beautiful sets and costumes, made from inexpensive materials often picked up at sales. ~ **instead of** creating the usual expensive, elaborate costumes [Gielgud, 168]

[projection: angle] Torture and sexual violence against prisoners is widespread in jails across the United States, **according to a report published yesterday** ~ **says** a report published yesterday [Text 2]

The pattern of agnation works the other way around too, of course; we can often find circumstances that are agnate to clauses:
[elaboration: role] but after a time, Chirumá was chosen || **to be the new chief**. ~ **as** the new chief [Text 65]

Lexico-grammatical annotation

But, the following embedded clauses do not form a clause complex, because there is no subject or finite [[allowing…]]

||| This meant **[[allowing the Commission to raise charges on these lines to the point [[where they would pay for themselves]]]]** || – which charges would probably be more [[than the traffic could bear anyway]]. |||

Aspects to note

- Verbs vs. nominalizations: if you can put *the* in front of a word, it is a noun.

- Present or past participles can function as modifiers (adjectives) in nominal groups. They are thus not verbs realizing Process.

1. (MChem 2: *The high resolution 3D biogeochemical model presented conserves mass and reproduces the approximate timing and magnitude of the* ***observed*** *seasonal cycles of nutrients (nitrate ammonia, DIP), phytoplankton,* ***dissolved*** *oxygen and* ***dissolved*** *organic carbon, at most stations throughout the estuary quite well.*)

- Adjectives (or participles) as qualifiers following the Thing can be interpreted as Attributes in reduced relational attributive relative clauses, but are not clauses.

    1. (a river fully navigable by large vessels = a river that is fully navigable by large vessels)

        - || The next step is [[to remove some impurities]]. ||| [Text 410]

Appendix III: The coding sheet for hedges and boosters

Neutral (examples)
Assign, on average caveat, comment, described as evidence of/for, emphasise, favour, mean, (should) note, observe, prefer, be understood as,

The passive voice test for hedging: *the organic-rich sediments in the eutrophic estuary accumulate relatively large amounts of P* à *relatively large amounts of P are accumulated*… = no hedging of the verb

Ignore "double" propositions: *The samples had higher pH values, probably due to partial neutralization of...*

Potentially confusing cases (-1)	**Potentially confusing cases (+1/ neutral)**
be/have + adverb: …*were relatively sparse*	
the implication is…	which has implications for…
be (not) a likely (+ noun)	
likely sources are…	
little is known about…	nothing is known about…
has little effect	we observe little change in…
there's little evidence	there's no evidence
a possible explanation is	
have the potential to…	
remains to be explained	
remains open for now	is/are inconclusive
remains to be tested how/whether	remains debated
remains unclear	remains unknown
it should be pointed out, however	trend (v)

	Hedging	Boosting
A	about, according to some earlier studies, admit, admittedly, aims to, not always, allegation, allege(dly), allude, almost, ~~alternative~~, it was anticipated that, it is apparent that (Cao Feng lists this as both a hedge and a booster.), apparently (Cao Feng lists this as a booster.), appear, appearance, appear to, it appears that, it appears then that, it also appears that, appreciable, approximate(ly), approximation, apt to, arguable(y), argue that, be argued that, it can be argued that, it has been argued that, it is possible to argue that, the argument that, around, ascribe, assert(ion), assess (re-), assessment, assume that, it is assumed that, it was assumed that, assuming that (= if), the assumption being that, there is no assurance that, avenue, attempt to, attempted, attribute	absolute(ly), actually, in accord with, always, it is apparent that (Cao Feng lists this as a hedge and a booster.), it is immediately apparent that, it was apparent that, apparently (Varttala lists this as a hedge), assure, assuredly
B	barely, ~~based on the limited data available, based on previous surveys~~, basically, there is reason to believe that, is believed to, broadly, broadly speaking	believe that, I believe, there is every reason to believe, we believe that, we do not believe (Here, *believe* expresses conviction.)
C	~~calculate~~, candidate, caution, the caveat is (= -1), central, it is by no means certain that, it is uncertain, chance, characteristic, charge, claim, it is far from clear, it is not clear, it is not quite clear, it was not clear, what is not clear, it is unclear, it is therefore unclear, close(ly), clue, (un)common(ly), comparatively, conceive, it is conceivable that, conceivably, concept(ualize), we are aware of the concerns expressed, raise concern(s) that, may conclude, does not confirm that, conjecture that, it is further conjectured that, consider, be considered, considerable(y), construct, contend, the contention that, could/ can, it could be that	can (= be able to), cannot/can't/couldn't, categorical(ly), it is certain that, certainly, clear that, it became clear that it is clear that , it is quite clear that, it should be clear, it was clear that, it was becoming clear that, what is clear , what is also clear, what is becoming increasingly clear is, clearly, compelling, complete(ly), comprehensive(ly), conclude that, concludes that, be concluded that, it can be concluded that, it was concluded that, we can conclude that, we concluded that, the conclusion that, conclusive(ly), we can say with confidence that, confident, confirm that, as the results confirm, confirmation, considerable(ly), consistent with, consistently, conspicuous(ly), constantly,
D	deem, deduce, it was deducted that, at least to some degree, it cannot be determined, we have not been able to determine precisely, it is difficult to conclude whether, it is doubtful that	decidedly, decisive(ly), definite(ly), definitive(ly), demonstrate that, as this study has demonstrated, it has been demonstrated that, demonstrating that, deserve(dly), there is no denying that, determine, leaves little doubt that, discovered that, discovery that, distinctively, do/ does/did no doubt, doubtless, without a doubt
E	essentially, estimate(ion), (re-) evaluate(ion), does not record clear evidence, little evidence, it cannot be excluded that, one cannot exclude a possibility that, expect that, be expected that, it was expected that, expectation, a possible explanation, to a certain extent, to some extent, extrapolate	in effect, ~~emphasise that, be emphasised that~~, enhanced (adj.), ensure that, entirely, in essence, essential(ly), especially, establish, provide evidence that, it is evident that, it was evident that, it was therefore evident that, what is evident, evidently, exact(ly), exclusively, exhaustive(ly), explicitly, extensive(ly), extraordinary, extremely

F	fair(ly), feasible, feel, forecast, frequent(ly)	as a matter of fact, in fact, the fact that, this fact, factually, find, found, find that, finding that, the finding that, the well-established finding that, this study's finding that, found that, it was found that, it was also found that, we found that, we also found that, firm(ly), forceful(ly), fully, fundamental(ly)
G	general(ly), given that, greatly, gross, guess	genuine(ly), great
H	hold that, as I hope, the hope, we hope, the hypothesis that, it was hypothesised that, we hypothesise that, hypothetical(ly)	have to, highlight–(consider it neutral), highly
I	idea, ideally, if, imagine, implicate(ion), imply, improbable, be inclined to, less inclined to, inclination, indicate that, indicating that, indication, is indicative of, infer, inference, interpret(ation)	illustrating that, impossible(ly)/not possible, impressive(ly), it is improbable that, inconceivable, incontrovertible(ly), indeed, indispensable(ly), indisputably, inevitable(ly), intensively, invariably
J	judge, just	
K	as far as we know, we do not know whether, to my knowledge, at least to my knowledge, to our knowledge, it is not known,	know
L	by and large, largely, lead us to, this leads us to, do not lead, are also likely to, are highly likely, it is likely that, the least likely to, less likely, be less likely that, be more likely to, makes it more likely, are most likely to, it is still likely that, the likelihood that, little	literally
M	main(ly), maintain, may, it may also be the case that, it may be that, maybe, it may even be that, it may simply be that, might, it might also be the case that, it might be that, moderate(ly), modest(ly), mostly	by no means, manifest(ly), marked(ly), meaningful(ly), must
N	nearly, not necessarily, does not necessarily, unnecessarily, negligible, normal(ly), ~~note (Listed in Farrokhi and Emami.),~~ notion	it is natural that, it was natural that, naturally, necessarily, needless to say, never, notable(ly), noteworthy, noticeable(y)
O	~~observe (Listed in Farrokhi and Emami.)~~, occasionally, offer, offer an explanation, often, ~~only~~, opinion, opportunity, to a first order, ostensibly, overall	obvious, it is obvious that, it was obvious that, this was obvious that, obviously, most obviously, of course, outstanding

P	partially, partly, what I perceive to be, perception, perhaps, plausible, it is plausible that, points to, portend, posit, it is posited that, possibility that, another possibility, the possibility that, it is possible that, it is also possible that, it is equally possible that, it is very possible that, it was possible that, possibly, quite possibly, postulates that, the potential to, potentially, practically, predict, predicted that, prediction, the prediction that, predictive, predominant(ly), premise, presumably, it can still be presumed, prevalent, primarily, principally, probabilistic, probability, it is probable that, it is quite probable that, probably, project, promise, prone to, the proposal that, propose that, we propose that, it was proposed that, proposition, does not provide definitive answers, provided that, putative	particularly, patently, perfect(ly), persuasive(ly), plain(ly), pointed out, powerful, precisely, precision, profound(ly), prominent(ly), proof, prove to
Q	questionable, quite	quite
R	rare(ly), rather*, reasonable(y), recognize, recommend, regard as, regular, relative(ly), ~~report~~ (Listed in Farrokhi and Emami.), reportedly, reputedly, rough(ly) *rather than = instead of (conjunction phrase) *rather, … (sentence connection)	radical(ly), the reality that, real(ly), reinforce, reliable(y), remarkable(ly), revealed, revealing, reveals, rigorous(ly)
S	scenario, see (as), seek, it would seem that, it seemed that, it seems that, seemingly, it remains to be seen / can be seen, seldom, should, sign, slight(ly)–(not always a hedge), small, some, sometimes, somewhat, it is possible to speculate that, be speculated that, it can be speculated that, the speculation that, speculative, suggest that, venture to suggest that, we suggest that, it can be suggested that, it is suggested that, it was suggested that, suggesting that, suggestion that, the suggestion that, suggestive of, suppose, supposedly, supposition, surmise that, suspect that, we suspect that	safe(ly), secure(ly), self-evident, should, show, what the data show is that, it showed that, showing that, shown that, as this study has shown, it has been shown previously that, this project has shown, shows, it shows that, simply, sizeable(ly), stress that, striking(ly), strongly, successfully, superior, to be sure, sure(ly)
T	technically, it is tempting to, tend to, tendency, tenet, tentative(ly), theoretical(ly), theorise, theory, there is reason to think that, we think that, trend, typical(ly)	thorough(ly), total(ly) adj./adv., it is true that, truly
U	unlikelihood that, not likely, unlikely that, unlikely to, are unlikely to, is unlikely to, it is unlikely, it is highly unlikely that, it was unlikely that, (not) usual, usually	unambiguous(ly), unarguably, unavoidable(ly), it is undeniable that, undeniably, underline that, underscore, undoubtedly, unequivocal(ly), uniformly, uniquely, unlimited, unmistakable(ly), unprecedented, unquestionable(y), uphold/ upheld

V	the view that, viewpoint, in the view of many scholars, virtual(ly)	vast(ly), vital(ly)
W	widely, will likely have, wish, would	well-known, will, will not/won't
X		
Y		
Z		