Apparent randomness and chaotic non-linearity: change over time in the essays of a cohort of philosophy undergraduates

by James Binchy

A dissertation submitted to the University of Limerick for the degree of Ph.D. by research

Supervisors:

Dr Anne O'Keeffe & Prof Michael McCarthy

Submitted to the University of Limerick, July 2013

Abstract

For most degree programmes in third-level education, the primary form of assessment is by written work submitted by the student to an assessor, either through formal, time-limited examinations or take-home essays.

This research examines a sample of the take-home essays from a selection of students in a single cohort within a degree programme in Mary Immaculate College, Limerick. This research focuses on philosophy modules taken by the students, and in particular a set of essays written for a single assessor. These essays form a corpus of 94 texts submitted at six different points over the course of a degree programme.

By looking at the use and distribution of linguistic items, this research shows that change in the writing of the students displays an apparent randomness and is not linear. Each text within the corpus is unique and each individual writer responds to the influences of genre, task and audience in unique ways. This unique response is because the essay texts are composed through a set of instantial decisions by the writers. It is argued that this uniqueness, apparent randomness and non-linear change is best understood by viewing the change in writing over the course of the degree as a dynamical system that closely approximates that advocated by chaos theory.

Declaration

I hereby declare that this thesis is entirely my own work and has not been submitted for any other awards at this or at any other academic establishment. Where use has been made of the work of other people, it has been fully acknowledged and referenced.

_____

James Binchy, July 2013

Acknowledgements

I would like to say a sincere thank you to my supervisors, Anne and Mike, for their patience, interest and understanding. I would also like to thank them for enlightening, enjoyable and thought-provoking supervision meetings. Without them, this research would not exist.

I would like to express my appreciation and love for my family, Karen, Isabelle and Harry. Their company is a joy unrivalled and their support and understanding has been instrumental in the production of this research. Thank you, Karen. Thank you, Isabelle. Thank you, Harry.

I would like to acknowledge and express appreciation to the Learner Support Unit in MIC for supporting both me and this research. Geraldine and Brian were very willing and able discussants on all areas of this research project and their interest and advice was much valued.

I would like to thank the staff of the Mathematics and Computer Studies Department in MIC for their help with this research. Diarmuid O'Driscoll was a great discussant around correlation and a huge help in the use of Maple software. Niamh Armstrong was a huge aid in enabling the use of Excel and working with long documents. Pat O'Sullivan provided his time for interesting, enlightening and genuinely enjoyable (from my point of view) discussions on Chaos Theory, its meaning and its implications.

Finally, I would like to thank the staff and students of the Philosophy Department in MIC without whom there would have been no data, and with no data there would be no research. Thank you.

Table of Contents

List of Figures

List of Tables

Chapter 1 Introduction

1.1 Introduction

Written assessments, either in take-home or sit-down exam formats, are the most common form of assessment at third-level. Institutions, in a bid to improve student quality, are taking cognisance of the importance of academic writing and are creating formalised structures to improve the writing of their students.

Concurrent to such developments on an institutional level, globally academic writing, for both native speakers of English and speakers of other languages, has grown in emphasis in recent years. Many courses in third-level institutions have mandatory academic writing components, and most third-level institutions provide some form of support for the academic writing needs of their students. Furthermore, there is a proliferation of volumes available to students to provide guidance in the area. Many of these volumes are the academic writing equivalent of self-help books.

In applied linguistics, the area of academic writing is much-studied, and has journals dedicated to this subject, for example the *Journal of English for Academic Purposes*. Much of this material focuses on what instructors should be encouraging their students to do. At times, this advice is based on what professional academic writers do, and the other times, this advice is based on large corpora of academic language generated by student writers or, in some cases, professional academics. In the main, the thrust of such material is both pragmatic and didactic, instructing novice writers to follow set rules in terms of essay structure and language choice. These instructions are based on a view of student writing as a monolithic, static construct. There is no allowance for individuality, nor no focus on dealing with the content matter of these essays.

Overall, the research into academic writing and the instruction manuals this research generates view writing as a firm set of guidelines to be followed by novice writers. The implied notion here is that academic writing is a set genre and by following the conventions of that genre, the student becomes proficient in the use of the said genre.

This study sets out to examine samples of student writing in the Philosophy Department in Mary Immaculate College, Limerick. Instead of taking the sample from one point in time, this study utilises a set of essays written by the same cohort of students for the same lecturer over the course of their degree. This study is unique in this respect. In doing this, the focus is moved from an aggregate, collective view of language to the individual writer. In a way, this is a response to what have become prescriptive notions of academic writing based on a somewhat nebulous majority. Philosophy, like any other academic discipline, has its own conventions. However, unlike some of the other academic disciplines available in Mary Immaculate College, philosophy is not part of the secondary school curriculum in Ireland. For this reason, we can assume that the students taking philosophy in Mary Immaculate College are being introduced to the subject in a formal way for the first time. Furthermore, by looking at the work of students over the period of their degree programme, we are operating out of an implied acceptance that the writing of students changes over time.

To examine these issues, it is best to use empirical data. In the present study, the empirical data chosen is a corpus of essays collected from a cohort of undergraduate students. There are two reasons for choosing this route. The first is that the written product is what is judged and assessed in the third-level context. The degree outcome of a student is dependent on the quality of that written product. The second is that data from written products, in the form of corpora, has been used to inform the prescriptive approach. This study views the written product of the students' work as necessitating further descriptive research.

1.2 Locating the study

This study is based within the field of applied linguistics, and more specifically within the area of corpus linguistics. By choosing to use a corpus methodology, there is an implied preference for empirical data. This study is not about what students should do, might do or are supposed to. It is about what these students actually do in their essays. This is a study of student writing within a specific subject area located within a specific third-level setting.

Traditionally, genre, audience and task have been seen as some of the driving influences behind the creation of written texts (as the review of literature in Chapter 3 will argue). However, there is uniqueness within each text, despite the similarity of context with other texts produced at the same time. We know this because if this uniqueness did not exist, the students would have been accused of plagiarism. The possibility of avoiding plagiarism is dependent on the uniqueness of each text, despite similarities in the context of the production of that text with other texts.

This study recognises that each text in the data used belongs to a genre and that the texts in the corpus used for this study belong to the same genre. Yet, this study argues, in Chapters 2 and 3, that the notion of genre is not sufficient for understanding the individuality of each text nor for understanding how the students change their writing over time. The analysis of the data in Chapters 5, 6 and 7 points to a uniqueness of each text and a non-linearity of change within the cohort. For every instance where it seems as if that the amalgamated cohort change in a certain way, on further inspection it seems that not all individuals change in that way. Based on this, this study argues that the most appropriate way of understanding this change is through the lens of chaos theory. This argument is made in Chapter 8.

1.3 Research questions

The research questions are as follows:

- Main research question:

  Is there a patterned, linear change in the student writing over time?

- Sub-questions:
  1. Is each text unique?
  2. How do writers respond to genre, task and audience?
  3. Is the response to genre, task and audience standardised across writer and/or situation?

To answer these questions, the current research will outline current theories that can be applied to the analysis of change in academic writing (Chapter 2), their limitations and

suggestions for overcoming these limitations (Chapter 3). Then, the data used in the current research and how it was gathered will be outlined (Chapter 4), followed by an analysis of that data (Chapters 5, 6, 7 and 8). The current finishes with a conclusion (Chapter 9).

Chapter 2 will now look in detail at the relevant literature in relation to process-based approaches and genre-based approaches to understanding academic writing.

Chapter 2 Review of the related literature

2.1 Introduction

Since the 1970s, there have been two major approaches to the teaching and researching of academic language in the areas of both native and non-native writing. These are the process approach and genre-based approaches. The process approach, as the name suggests, is concerned with the process of writing and how that process generates the text. Genre-based approaches, again as the name suggests, are concerned with the notion of genre and how this influences writers. In the present study, neither approach is seen as suitable for examining and explaining the changes that occur in the student writing over time.

This chapter outlines the main ideas of both approaches with a particular focus on the implications of these approaches for the understanding of student academic writing and its change over time for particular students. The following chapter, Chapter 3, examines the limitations of both these approaches and uses those limitations to develop a theoretical framework for the present study.

2.2 The process approach

In the 1970s and 1980s, researchers in the area of writing changed their focus from the formal properties of the product produced by the writer to the process the writer engages in to produce the product, the text. Instead of looking at the text produced, researchers concerned themselves with composing-aloud sessions, observation and interviews (Emig, 1971), think-aloud protocols (Perl, 1980), video-taping writers in the act of writing, examining planning and decision making processes during pauses from writing and interviewing writers after revision on their writing and comparing changes in the drafts (Sommers, 1980). The goal of such a change in focus was to ‘replace a

prescriptive pedagogy (select a subject, formulate a thesis, outline, write, proofread) with a descriptive discipline whose members study and teach "process not product"' (Reither, 1985: 620).

Basic Writers

Perl (1979), studying the composing processes of five unskilled college writers at an American university, argues that, prior to her work, little was done to study basic writers and their 'observable and scorable behaviors' in the composing process (1979: 318). One aim of the study was to provide a mechanism for documenting composing processes in a 'standardized, categorical, concise, structural, and diachronic' format (1979: 320). Perl developed a code for what students do in their composing processes. Miscues in students' own reading and writing work were also noted in a standardised format. The discovery that the students' composing processes were consistent, even when the resultant writing appeared to have been done in a haphazard or arbitrary manner, supports the research of Shaughnessy (1977). Perl concludes that writing instructors must look at students' internalized processes to make decisions about instruction.

Building on this, Perl (1980) used Emig's (1971) work on think-aloud protocols to show that the student writers, though often unskilled, had consistent strategies for composing. While the students spent little time on prewriting, there was no indication that this created subsequent problems. Students discovered meaning as they wrote in a process that was recursive, discursive, and decidedly nonlinear. Editing created most problems for the students, as they often over-corrected or began to correct before writing enough to untangle what they wanted to say. Perl argues that these students do know how to write and have stable composing processes. She suggests a 'loosening' of the writing process: 'readying oneself to write, sustaining the flow of writing, shaping the discourse for oneself, readying the discourse for others' as a consideration for writing instruction (1980: 31–32).

Based on a study of undergraduate writers in an American university, Lunsford (1979) claims that basic writers have not attained the level of cognitive development required to succeed at college-level work. Because they have not developed the cognitive ability

to perform tasks that require synthesis and analysis, basic writers have difficulty forming abstract concepts. Lunsford recommends that teachers use various strategies, ranging from grammar and sentence-building activities to essay assignments, to engage students in inferential reasoning rather than in isolated drill exercises and rule memorization. Working in small-group workshops, basic writing students should be allowed to practice analysing, generalising, and then abstracting, all of which are skills that they need to succeed in college.

Basic writing students should be immersed in academic discourse so that they can begin to appropriate it for themselves. Bartholomae (1986) argues that basic writing studies should not concentrate simply on error, but instead there is a need to better understand how basic writers' lack of understanding about constructions of authority and the rules of academic discourse put them at a disadvantage in an arena that values such knowledge. As a result, Bartholomae argues that the basic writer ‘has to invent the university by assembling and mimicking its language’ (1986: 5), often long before the skills of writing in an academic setting are learned.

Bizzell (1982) argues that compositionists form two theoretical camps: those who are outer-directed focus on the social processes that influence language learning and thinking, while those who are inner-directed are interested in universal writing processes and individual capacities. Inner-directed theorists, such as Flower and Hayes (1981), support a linear, cognitive model of writing that separates thought, or planning, from writing, or translating, yet such an approach fails to account for individual knowledge and contextual influences. Outer-directed theorists remain sceptical of all models that claim an understanding of inner processes. Accordingly, outer-directed theorists stress the role of community, ethics, politics, and social interaction in the development of thinking and language. Bizzell (1982) concludes that a synthesis of theories from both camps will offer a fuller understanding of writing.

Developing this, Bizzell (1987) suggests that literacy scholarship is commonly divided into two main schools of thought: those who see the acquisition of literacy as a stage in human cognitive development, and those who question this and focus instead on literacy as social practice. Those who view literacy as social practice demonstrate that literacy ought not be treated monolithically but examined within social and cultural contexts. In applying literacy research to the question of whether American college students are

literate, Bizzell argues for a definition of academic literacy that takes into account its social context and its specific social purposes. It is argued that functional literacy, 'literacy that confers a reasonable degree of education and economic success and political participation' (1987: 135), enables critical reflection on the different relations between social groups and on the educational, economic, and political differences that separate them.

Harrington and Adler-Kassner (1998), in a way that echoes Bizzell (1982, 1987), suggest that basic writing research has taken two perspectives: cognition-based work, which focuses 'on the writers themselves and what happens in the act of composing' (1998: 9), and culture-based work, which 'focuses less on individuals than on a sense of institutional or social culture' (1998: 12). They outline three areas for further investigation: why writers make the decisions they do about their writing, how students define themselves and their work, and how basic writing programs are constructed and administered. The present research is concerned with the first of these questions, namely the decisions writers make, and extends the question to include change over time and the uniqueness in texts the individual decisions lead to. This is done in the present research through the text produced, viewed as an artefact and representation of the sum of these decisions, and in particular the word choices within those texts.

Hyland (2003) argues for the importance of genre approaches, which see ways of writing as purposeful, socially situated responses to particular contexts and communities, to the teaching of L2 academic writing. He sees the process approach to teaching writing as ignoring the social construction of meaning and thus failing to consider the forces outside the writer which help guide purposes, establish relationships, and shape writing. Genre-based pedagogies, however, 'address this deficit by offering students explicit and systematic explanations of the ways language functions in social contexts' (2003: 18).

2.3 Genre-based approaches

Context of situation and the social setting of language

In the view of Gumperz (e.g. 1982) and Hymes (e.g. 1974), linguists had often isolated themselves from other disciplines which sought a deeper understanding of culture. They believed that previous practice by linguists presented a 'fragmented, incomplete understanding of humanity' (Hymes, 1974: vii). For Hymes and Gumperz, the study of language is a multidisciplinary field where not just linguistics but sociology, social anthropology, education and so on are indispensable. Their works show a belief in a study of language that deals with functions as well as structures and in this way the study of language, and language variation, is bound to the socio-cultural setting in which the language functions. While this view sees language as not isolated and away from the participants who use it, unlike Malinowski and Firth, Hymes and Gumperz link language use to the culture of the user (Malinowski (1923) also refers to the context of culture, but does not give it the same importance as Gumperz).

At this time the use of the term register was becoming widely used in linguistic theory for differences in language use (Eggins and Martin, 1997: 237). The term was first introduced to linguistics in 1956 by Reid, who used register to refer to 'text variety' (Ghadessy, 1993: 9). Hymes accepted this use of the term register but regarded it as limited to 'situation-specific use' (1974: 59). Hymes, along with Gumperz and Ferguson, preferred the term *variety* for 'community-wide uses or use in relation to broad domains' (ibid.) while the term genre was associated with traditionally recognized categories such as poem, myth, riddle, tale, and so on, embedded deeply and exclusively within literary theory.

The use of the term register was taken up by structural linguists who undertook much of the research into varieties of language use in the 1960s and 1970s (Ferguson, 1996: 173). At this time, it was assumed that 'for most material purposes register and genre are synonymous' (Hasan, 1978: 230). Much of this work took the form of surface level linguistic description within and across different varieties of language use and was referred to as 'register analysis'. Most register analysis occurred within the tradition of Hallidayian 'systemic grammar' which outlines the contextual categories of field, tenor and mode as three variables in register, or functional linguistic variation. Field refers to the social action (e.g., an auction, asking for an item in a shop), tenor to the role

relationships of the participants (e.g., mother-child, teacher-student) and mode to the channel of communication (e.g., written versus spoken, planned versus unplanned discourse) (Halliday, 1978: 122). This model of context is related closely to the linguistic system through 'the functional components of the semantics' (ibid.) and is generally referred to as 'functional systemic' linguistics. (See below for a further discussion of functional systemic linguistics).

With this reworking of Halliday's model of context has also arisen the question of how genre relates to register. By now the concept of genre had assumed a role outside the traditional category of literary theory and was particularly relevant to the 'sociolinguists' assumptions about the conventionalization of variation' (Ferguson, 1996: 171). Within 'modern' linguistics, the concept of genre has been extended to include classes of language use and communication in everyday life. In contrast to Hasan (1978), many writers have sought to provide a theoretical distinction between the two concepts of genre and register (e.g., Gregory and Carroll, 1978; Ventola, 1984; Martin, 1992) and have suggested that two layers of context are needed, with a new level of context called genre posited above the register variables of field, tenor and mode. This notion of genre refers to staged, goal-oriented social processes mapped out through various combinations of field, tenor and mode. In this way, genre has typically become associated with communicative social purpose. Some researchers (e.g., Allison, 1999) have in fact argued that this attention to genre goes back to Hymes's work on 'speech events', which he described as 'activities…that are directly governed by rules or norms for the use of speech' (cited in Allison, 1999: 144) and so has been there for quite a long while, if only in theory.

While register has typically been defined by formal properties of grammar and lexis in the early Hallidayian tradition (e.g., Halliday, McIntosh and Strevens, 1964), genre has been particularly linked to concerns with communicative and social purposes among user groups forming 'discourse communities' (Swales, 1990: 58). While some researchers work exclusively with one term or the other (e.g., Kay and Dudley-Evans (1998) and Paltridge (1994, 1996) both use the term genre), many see the two concepts as complementary. In register and genre theory (R&GT), Eggins and Martin (1997: 251), in a reworking of Halliday's framework, see register and genre variation as 'two realizational planes in a social semiotic view of text.' In other words, 'text is both the

realization of types of context, and the enactment of what matters to cultural members in situations' (1997: 251). They apply R&GT to discourse analysis in an attempt to theorize how discourses or texts are like and unlike each other, and why. Furthermore, register and genre have often been used interchangeably, particularly in sociolinguistics. For instance, in his 1995 book Dimensions of Register Variation, Biber uses the term register in a way similar to his use of the term genre in his previous work (e.g., 1988). Therefore, register and genre often amount to the same thing: the study of language variation according to its various uses in different contexts.

Kress (1993) identifies two main approaches to genre. The first approach is that proposed by Martin and Rothery (Martin, 1993; Martin and Rothery, 1993) who, building on the work of Ventola in service encounters, see genre as covering everything there is to know linguistically about a text, and this can, in turn be accounted for by ideological context (Kress: 1993: 32). In this model, all aspects of the structure of the text as a whole and of the subsections within the text are viewed in terms of the task that is being performed by or through the text. The second approach is that of Kress himself and focuses on 'the structural features of the specific social occasion in which the text has been produced' and sees these features as 'giving rise to particular configurations of linguistic factors in the text which are realisations of, or reflect, these social relations or structures' (1993: 33). Both approaches, however, recognise that generic form stems from the action of social agents in particular social situations, that genre is socially situated.

Summary of Genre

Bhatia (2004) argues that regardless of whether genre is defined as typification of rhetorical action (eg Miller, 1984), regularities of staged, goal-oriented social processes (eg Martin, 1993) or consistency of communicative purposes (eg Bhatia, 1993), genre analysis is 'the study of situated linguistic behaviour in institutionalized academic or professional settings' (2004: 22). He continues by outlining six points that all theories of genre accept:

> Genres are recognizable communicative events which are characterized by communicative purposes that are identifiable and understood by members of the community in which they occur;

> Genres are both highly structured and conventionalized. They constrain choices available with regard to discourse structure and lexio-grammatical features;
>
> Members of the community have a greater understanding of the genre than apprentices, new members or outsiders;
>
> Although genres are conventionalized, expert members can exploit this to express both private and organizational intentions;
>
> Genres are reflections of disciplinary and organizational cultures and they focus on social actions embedded within institutional practices;
>
> All genres have an integrity of their own, identifiable by textual, discursive and contextual factors.
>
> (Bhatia, 2004: 23)

2.4 Genre and academic writing

Essays, reports, oral presentations and exams are a common form of assessment in most third level institutions and students are expected to make evident the transition from second level to third level in their writing. In an empirical study of pass grade essays, O'Keeffe (2000) has identified deficiencies in terms of syntax, lexis and style. She points out that in addition to writing skills, students 'also need to become accustomed to the 'culture' of this genre of writing. Some students intuitively sense this 'culture', picking up implicit tacit knowledge as they progress but … some do not' (2000: 167). Part of this 'culture' that students are expected to become part of involves the mastery of the required register, or the relationship between language features and their context (McCarthy 1998: 26). In academic writing, the level of language is formal and certain lexical items and syntactic structures are more frequently used (see Carter and McCarthy, 1997: 115, Biber, Conrad and Reppen, 1998:135). The expected formality of academic writing limits the student in terms of the words, expressions, and structures appropriate for use. In addition, students are limited by the expectations of the particular discipline in which the student is writing. Barrass (1995: 1) states that many students perform below their ability not because of low motivation or lack of effort, but because they do not pay enough attention to improving their competence in communicating their thoughts through writing. The case is often made that the ability to replicate the communicative norms of the individual discipline is central to this competence.

While Kress (1993) sees two main approaches to genre, others (for example Johns, 2002; Hyland, 2003), building on the work of Hyon (1996), identify three broad, overlapping schools of genre theory. Hyon (1996) identifies three traditions of research concerning genre, namely English for Specific Purposes (ESP), North American New Rhetoric studies and Australian systemic functional linguistics. She examines these three traditions and then evaluates their merits with regard to the teaching of academic reading and writing to non-native speakers of English.

English for Specific Purposes (ESP) analysis

'Researchers in ESP have been interested in genre as a tool for analysing and teaching the spoken and written language required of non-native speakers in academic and professional settings' (Hyon, 1996: 695). In ESP research, genres are viewed as oral and written texts distinguishable according to both their formal properties and their communicative purpose within their social context. Hyon (ibid) argues that in this tradition researchers pay attention to detailed formal characteristics while ignoring the functions of texts and their contexts.

New Rhetoric Studies

New rhetoric studies concerns L1 teaching and encompasses rhetoric, composition studies and professional writing. Rhetoric is seen as 'the art or the discipline that deals with the use of discourse, either written or spoken, to inform or persuade or motivate an audience, whether that audience is made up of one person or a group of persons' (Corbett and Connors, 1999: 1). Where ESP analysis has focused on the formal properties of genres, new rhetoric studies 'focused more on the situational contexts in which genres occur than on their forms and have placed special emphases on the social purposes, or actions, that these genres fulfil within these situations' (Hyon, 1996: 696).

Australian Genre Theories

The approach to genre in this tradition is based on systemic functional linguistics, a theory of language concerning itself with the relationship between language and social settings (Halliday, 1978; Halliday, 1985). The language forms used are assumed to be influenced by the social context consisting of field, tenor and mode. In general, the Australian school is similar to the ESP approach in that the main concern is the linguistics features characteristic of different genres.

Within functional systemic linguistics, the three contextual parameters of field, tenor and mode have been incorporated into a much larger system where a model of language can be closely related to the organization of context. The model of language put forward by Halliday proposes that an understanding of a text requires an interpretation of three 'metafunctions' (Halliday and Hasan, 1989: 45). The first metafunction is the *ideational* which is 'the learning or thinking function' (ibid: 45). This concerns 'the process being referred to, the participants in these processes and the circumstances…associated with them' (ibid.). The second metafunction is referred to as the *interpersonal* and is described as 'the doing function'. This involves the recognition of 'the speech function, the type of offer, command, statement or question, and the attitudes and judgements embodied in it' (ibid.). The third metafunction is the *textual* which is how the ideational and interpersonal functions are organized into a coherent text. This is how one part of a text relates to every other part (what comes first or last; what is implied rather than actually said, etc.).

Halliday (1978) points out that the contextual features of field, tenor and mode are reflected in the three modes of meaning of language described above. The field is expressed through the ideational function, the tenor through the interpersonal function and the mode through the textual function. According to Eggins and Martin (1997: 239), 'British contextualism is the only tradition that suggests this kind of direct correlation between the functional organization of language and the organization of context.' The three situational and contextual categories of field, mode and tenor, (tenor later became style), have been variously refined and redefined (e.g., Gregory and Carroll, 1978; Martin, 1992; Eggins and Martin, 1997). Martin (1992), for example, offers a description of the mode of a situation in terms of two distance continua; a continuum of spatial distance which influences immediate feedback and a continuum of

experiential distance which concerns the distance between language and the event in which it is involved. However, what is of importance here is that field, tenor and mode are an elaboration of Malinowski's (1923) context of situation:

> … the context of situation, the context in which the text unfolds, is encapsulated in the text, not in a kind of piecemeal fashion, nor at the other extreme in any mechanical way, but through a systematic relationship between the social environment on the one hand, and the functional organisation of language on the other.
>
> (Halliday and Hasan, 1989: 11)

Overall, there seems to be some agreement in the three schools concerning genre, namely that social context plays a role in shaping a text (reminiscent of Malinowski, Firth etc.), while the nature and influence of this context is disputed. However, all three schools of thought aim to use a genre model of research and hence genre-based pedagogical materials to teach students to become more successful when composing academic and work-related texts.

The genre approach to analysing language 'emphasises the cultural and social dimensions which enter into the formation and constitution of language and texts' (Kress, 1993: 23). Hence Hyland (2003) argues that genre is socially situated and the features of a similar group of texts depend on the social context of their creation and use. Therefore, if students are writing to answer similar tasks in a similar setting, their products are going to be of the same genre. As genres are 'rhetorical actions that writers draw on to respond to perceived repeated situations; they are choices which reflect effective ways of getting things done in familiar contexts' (Hyland, 2003: 22), the 'regularity of the situation will give rise to regularities in the texts which are produced in that situation' (Kress, 1993: 27). The expected regularities in the texts produced are examined in the analysis-based chapters in the present study and the notion of such regularity is subsequently questioned.

Communicative Purpose

A lot of work on genre has drawn on definitions of genre relating to communicative social purpose. Martin explains that 'genres are how things get done, when language is used to accomplish them…the term genre…embraces each of the linguistically realized activity types which comprise so much of our culture' (1985: 250). Biber has also

given a definition of genre which corresponds to that given by Martin and Swales. He defines genre as 'text categorizations made on the basis of external criteria relating to author/speaker purpose' and 'readily distinguished by mature speakers of a language' (1988: 68). Swales offers that 'a genre comprises a class of communicative events, the members of which share some set of communicative purposes' (1990: 58), a definition taken up by Bhatia and extended to where a genre is a 'communicative event characterized by a set of communicative purpose(s)' (1993: 13). However, Askehave and Swales (2001) warn of the problem of uncertainty in determining communicative purpose and highlight this uncertainty with reference to three examples, shopping lists, short response letters to recommendations and company brochures. Even in shopping lists, it is shown that there we can have the 'uncomfortable position – at least for all those who stress the categorizing role of communicative purpose – of having identical or near identical texts fulfilling rather different communicative purposes' (Askehave and Swales, 2001: 201). In the present discussion, what is of particular note is the difficulty and uncertainty in identifying the communicative purpose(s) of a text.

2.5 Disciplinary differences in academic writing

While some have presented academic writing as a specific type of writing, for example Coxhead (2002) gives an academic wordlist drawn from the similarities of different disciplines and Biber et al (1999) use academic written language as one of the four classifications in their grammar, others have concentrated on the disciplinary differences in academic writing. Bazerman (1981), MacDonald (1987) and Maimon (1983) have all argued that there are disciplinary constraints in academic writing. Silver (2003) has shown that the word 'evidently' is used differently in history and economics texts, Charles (2003), in a study of the construction of stance through nouns in materials science and politics, has argued that there are differences in distribution and use in both disciplines, and Tucker (2003) demonstrates that research articles in art history use evaluation techniques differently to academic discourse in other fields of study. Silver (2003), Charles (2003) and Tucker (2003) all attribute the differences across disciplines to the research practices and the construction of knowledge in each discipline and suggest that these are not homogeneous across the academic spectrum.

Clyne (1987) investigated differences between linguistic and sociological texts written by English and German speakers. Differences between the English and German texts occur in relation to linearity, symmetry, hierarchy and continuity as well as the position of definitions and advance organisers and the integration of data. He suggests that this is due to, among other factors, varying attitudes to knowledge and content. Furthermore, Charles (2003) argues that the task the writer undertakes is what creates the structure of the text and dictates the language used. This is similar to Freedman (1987) who, in a study of how undergraduate law students acquire a new genre of writing, concluded that students did not learn the new genre from models of writing and an analysis of these models but from attempting to complete the task set to the students. She argues that the students created the complex syntax evident in her data through the 'interplay between the kinds of thinking necessitated by the question and the discipline, on the one hand, and the persuasive strategies and formal structures appropriate for communicating the insights so derived, on the other' (Freedman, 1987: 106). In other words, some features of the genre are created by the students in the actual process of writing the assignment and not through conscious imitation of models.

Stapleton (2002) proposes that research into writing pedagogy often concentrates on form, including and at times highlighting voice and identity, and tends to ignore content (for example, Cadman, 1997; Ivanic 1998; Tang and John, 1999; Hirvela and Belcher, 2001; Hyland, 2001; Ivanic and Camps, 2001; Matsuda, 2001). He argues that the quality of the content, the level of abstraction, the sophistication of the argumentation, the originality, or the creativity of the student can play a role in whether an assignment is viewed favourably or not by the assessor. Interestingly, if the student is concentrating on completing the task assigned instead of concentrating on fulfilling a generic blueprint as Freedman (1987) argues that they do, these features should occur through successful negotiation of the task.

Two very important points become evident from the above:

- There are differences between academic disciplines which are as a result of varying attitudes to knowledge and content;
- These differences also come from the attempt to complete a task.

How students learn academic genres

Freedman (1987) argues that students did not learn a new genre from models of writing and an analysis of these models. Instead the genre is reformulated by the students according to the grade received. She proposes the following model for learning a new genre:

> The learners approach the task with a 'dimly felt sense' of the new genre that they are attempting. They begin composing by focusing on the specific content to be embodied in this genre. In the course of the composing, this 'dimly felt sense' of the genre is both formulated and modifies as this *sense* the composing processes, and the unfolding text interrelate and modify each other. On the basis of external feedback (the grade assigned), the learners either confirm or modify their map of the genre.
>
> (Freedman 1987: 102)

While the students modify their map of the genre, they do this on the basis of failure or success. Failure or success is judged on the grade assigned by the assessor. Freedman argues that the students do not take into account the comments by the assessor in this modification, merely the grade. The 'dimly felt sense' of what constitutes the genre is created or developed from past and current reading, previous writings, the language used by professors, teaching assistants and textbooks, and the explicit instructions given concerning the assignment (Freedman 1987: 104).

One student in Freedman's (1987) study felt that a session with a writing clinic before writing her essay was not worthwhile as they discussed general academic structures and not anything specific to the task at hand.

> Students have acquired the new genre – not through intuiting its rules receptively, on the basis of reading and exposure to appropriate models but rather actively by performing – in fact creating the genre incidentally in their struggle for meaning (a struggle which, we must remember, is in part shaped by and waged against the pressure of already familiar forms of the language).
>
> (Freedman 1987: 111)

If the students are in a process of constant modification of their map of the genre, then we should be able to see in the current study linguistic changes in their writing over time. Furthermore, differences should be more apparent when the task changes, for example changing from a discussion on Socrates to Aquinas' arguments for the existence of god, as Freedman argues that the syntax is a 'result of the interplay between the kinds of thinking necessitated by the question and the discipline, on the one hand, and the persuasive strategies and formal structures appropriate for communicating the

insights so derived, on the other' (Freedman, 1987: 106). Some features of the genre are created in the actual process of writing when the students use language appropriate to the task.

In addition, if students are in a constant modification process, then we would expect an improvement over time in the quality of their writing. However, in general within third-level institutions, the grades assigned for each module across an institution seem to stay proportionally the same and adhere to a similar distribution. The number of As, Bs and Cs remain constant for each module. This suggests that those whose modifications are better will receive better grades as we have to assume that the assessor expects as higher level to the previous assignment on each new assignment.

2.6 Summary

In this chapter, we have examined the process approach and genre-based approaches to academic writing. The process approach is primarily concerned with the steps a writer takes in the production of a text. These steps are non-linear and recursive, and their continual realisation involves the self-actualisation of the writer and hence the production of a text. Genre-based approaches are focused on the finished product, the text, and the linguistic features contained within the text and see those features as a direct result of the social space in which the text was created in tandem with the communicative purpose of that text.

In the present study, neither approach allows us to investigate the research questions as outlined in Chapter 1:

- Is there a patterned, linear change in the student writing over time?
    - Is each text unique?
    - How do writers respond to genre, task and audience?
    - Is the response to genre, task and audience standardised across writer and/or situation?

In essence, the process approach is not appropriate due to its focus on the writer and not on the text while genre-based approaches are not appropriate due to the focus on a text

as a pre-standardised entity before the writer has started composing the text. The main research question, concerning language change, could be modelled as either a random (stochastic) process or as chaotic (underlyingly deterministic) process. In the current research, the latter approach is taken.

The following chapter, Chapter 3, critiques these two approaches in further detail and from that critique builds a fuller theoretical framework more suited to answering the present research questions.

Chapter 3 Theoretical framework

3.1 Introduction

In this chapter, a theoretical framework for the analysis of change within a corpus of undergraduate philosophy essays is developed. To begin with, we re-examine the current approaches to the research of academic writing as outlined in the previous chapter and identify drawbacks with these approaches for the present study. Then, we examine how the present study sees academic writing, namely as an activity that involves a writer, a context of situation and a text and the interplay between these concepts. From the discussion of writer, context of situation and text that follows, we identify how these concepts interact and how such a framework is suited to answering the research questions outlined in Chapter 1 concerning change and uniqueness in undergraduate academic writing, namely:

- Main research question:

  Is there a patterned, linear change in the student writing over time?

- Sub-questions:
    1. Is each text unique?
    2. How do writers respond to genre, task and audience?
    3. Is the response to genre, task and audience standardised across writer and/or situation?

3.2 The limitations of the process approach

As we have seen in Chapter 2, the process approach to academic writing focuses on the process that writers go through in order to create a completed text. This approach sees writing as a means of self-expression and ultimately self-actualisation. It must be noted at this point that all approaches to academic writing acknowledge some form of process in the creation of an academic text, usually centred around planning, drafting and

revising. Grabe & Kaplan (1996) have outlined four stages in the history of writing process approaches, the expressive, the cognitive, the social and the discourse community stages. While each stage has given us useful and important research, a number of concerns still remain with regard to the process approach.

The major problem with the expressive stage of process research is the underlying assumption that the writer already has all the intellectual resources for writing and all they need is an appropriate outlet. It ignores the context of writing (Grabe & Kaplan 1996: 89). With regard to the cognitive stage, it unfortunately does not take into consideration the fact that writers are not likely to have uniform processing preferences and cognitive abilities (see the discussion in Section 3.4 below with regard to cognitive differences in writers). Instead, as Grabe and Kaplan (1996: 92) argue, 'writing involves numerous processing options, and different writers will approach the task employing different processing strategies'.

The essential point of the social context approach is that writing can only be understood from the perspective of a social context and not as a product of a single individual. However, it is argued here that while writing takes place in a social context, not to acknowledge the role of the individual is a mistake. Coulthard (2004), while writing about authorship and forensic linguistics, points to the linguistic uniqueness of an individual. He argues, building on the work of Halliday, McIntosh and Strevens (1964) and Abercrombie (1969), that:

> '… every native speaker has their own distinct and individual version of the language they speak and write, their own idiolect, and [there is] the assumption that this idiolect will manifest itself through distinctive and idiosyncratic choices in texts'. (Coulthard 2004: 231-32)

And, more relevantly for academic writing,

> '… it is expected that any two writers writing on the same topic, even if intending to express very similar meanings, will choose an overlapping, but by no means identical, set of lexico-grammatical items to do so ' (Coulthard 2004: 434).

In Chapters 5, 6, 7 and 8 of the current research, empirical evidence will be used to show a sometimes overlapping, but ultimately unique, set of lexico-grammatical items used by writers writing on the same topic. However, evidence will also be presented to

show that idiolect, as manifested in the written work of the students, is not static. Idiolect itself is changing in a dynamic fashion and manifests itself in unique ways at any particular point in time. The distinctive and idiosyncratic choices are as much instantial and dynamic as pre-planned and static.

If we focus solely on the social context, we are ignoring the role of the individual, the cognitive differences between individuals, the differing approaches to writing process of individuals and the basic concept of idiolect. However, Hyland argues, seemingly in contrast to the social context approach, that 'process represents writing as a decontextualised skill by foregrounding the writer as an isolated individual struggling to express personal meanings' (Hyland 2003: 18). In a way, it could be argued that the process approach to writing sees the skill as decontextualized. However, viewing the writer as an individual is not necessarily inappropriate. It does not follow that seeing the writer as an individual necessarily negates contextual factors, nor does it follow that writing should not involve the expression of personal meanings. More likely, each individual writer, shaped by a multitude of contexts and factors, reacts to their interpretation of the context of situation in a different way to other writers, and this reaction is particular to a point in time.

As Collins (1995: 5) contends, an unquestioned acceptance of the process approach to teaching writing may fail novice writers because of the myths that inform the implicit instruction in this paradigm that 'writing development is natural and that teaching is primarily the facilitation of development'. Collins (1995:5) also notes that process literature promotes a structuralist, binary approach to writing instruction and recommends a more poststructuralist appreciation of 'differences among discourses'.

Although there has been some criticism of the process approach from scholars (for example, Hyland 2003; Collins 1995; Grabe & Kaplan 1996), in the context of the present study, there are two main shortcomings to such an approach. The first is that the process approach tends to take no, or very little, account of the product, the text, that is produced in this process of writing. It is as if the process and the product are not related. This is not the case. Texts are produced and received in a social space and while examination of this space, such as research into discourse communities, and the writer operating in this space, such as studies into the writing process of individuals, is welcome, detailed analysis of the texts produced can provide important information.

Secondly, the process approach takes no account of academic content – in this case philosophical content. This is a major oversight. If we view the academic writing in an undergraduate setting as primarily an attempt to display knowledge of the subject area for assessment purposes, we cannot conceive a text divorced from the message it contains. For these reasons, the present study, therefore, focuses on the actual texts produced by the students. Language choice and content choice are inextricably linked.

The present study views the written text as an artefact surviving in time and space. The text was created by a unique writer to respond to a task in a way the author deemed appropriate. The decision of appropriateness is not a single decision, but rather a much larger set of instantial decisions. It is the complete sum of these decisions that results in the final text.

3.3 The limitations of the genre approach

Bhatia (2004) argues that regardless of whether genre is defined as typification of rhetorical action (for example, Miller 1984), regularities of staged, goal-oriented social processes (for example, Martin 1992) or consistency of communicative purposes (for example Bhatia 1993), genre analysis is 'the study of situated linguistic behaviour in institutionalized academic or professional settings' (2004: 22). He continues by outlining six points that all theories of genre accept:

> - Genres are recognizable communicative events which are characterized by communicative purposes that are identifiable and understood by members of the community in which they occur
> - Genres are both highly structured and conventionalized. They constrain choices available with regard to discourse structure and lexio-grammatical features
> - Members of the community have a greater understanding of the genre than apprentices, new members or outsiders
> - Although genres are conventionalized, expert members can exploit this to express both private and organizational intentions
> - Genres are reflections of disciplinary and organizational cultures and they focus on social actions embedded within institutional practices
> - All genres have an integrity of their own, identifiable by textual, discursive and contextual factors
>
> (Bhatia 2004: 23)

What is striking about such a synopsis of genre is the prominence of the notion of community. In the present study, it is difficult to position the student writers in a community of the academy. These students are, initially, first year undergraduate students. While they are writing essays for assessment purposes, it would be amiss to see these students as part of the wider academic or philosophical communities because they have no knowledge of those communities, no power within such communities and are simply taking part in a course. The students can be seen as part of a localised community whereby they attend lectures and tutorials1 together and possibly discuss among themselves the philosophical content covered in the course. It may, however, be beneficial to see the assessors and lecturers on this course as part of a community. They are philosophers but yet do not expect the students to write like they do. When the students interact with the lecturers/assessors, they are interacting with a member of the philosophical community. This does not make the students active participants within that community.

There are some other problems with trying to apply a genre-based approach to the present study. The genre-based approach takes no account of the individual who produces a text. Instead, genre-based approaches focus on the text itself, which is a positive, but focus on the text not as the product of a person but a product of some ideal prototype that is to be followed and the social space in which this product is created. While the text is produced within a social context, we cannot assume that this context creates the text. In a way, the writer is almost excluded from the text creation. While context does influence text, 'context … gets 'into' text by influencing the words and structures that text-producers use' (Eggins & Martin 1997: 232), the writer controls how the context influences text.

Furthermore, genre-based approaches base their demarcation criteria on discourse structure and lexico-grammatical choices. No account is taken of the content. A philosophical essay is not a philosophical essay without philosophical content. In addition, the reason for an undergraduate writing an essay is to display appropriate knowledge for assessment purposes. Any approach that does not take meaning and

1 In MIC, a tutorial is a formal discussion class with a smaller number of students, typically 10. Tutorials are expected to be more interactive than large-group lectures and often expand on the content of a lecture.

content into account is lacking. In a study of mistakes in undergraduate student writing, Taylor (1988: 64) conclude that 'the deficiencies in students' writing … are in some measure due to confusions or vagueness about content'.

A problem with using a genre-based approach for the analysis of undergraduate essays is the identification of the genre. While all the essays are set in the same situation and should have the same communicative purpose, it is still unclear as to what genre they belong (see Martin (1992) for problems of interpreting communicative purpose as the speaker/writer intended it and also Askehave & Swales (2001) for problems in using communicative purpose in as a means of identifying genres). On one hand, the essays are part of an academic genre because they are situated in an academic setting, but on the other hand, major differences have been shown with regard to writing across disciplines and this does not lend itself to the positioning of philosophical essays beside essays from other disciplines. Added to this is the further complication between expert and novice. The students are obviously novices in philosophy when they start their degree programme, but we can hardly say that they are novices when they finish their study, but they still remain apart from post-graduate students and professional academics in relation to their writing.

In addition to problems with genre identification and genre boundaries, the tendency of genre approaches to focus on similarities between groups of texts, while useful in other research, can lead to a homogenizing of what is presented to learners of academic writing. This can result in the learners receiving an artificially narrow description of what is permitted within the genre. When students are composing their own writing for assessment, each has their own style and preference of writing. By focusing on what the majority do, as genre-based approaches tend to do, we can create a situation where students are encouraged to follow a format that may not suit their individuality.

Overall, the argument here is not that process approaches and genre-based approaches are not valuable research perspectives, but that when it comes to academic writing, they can be misleading. Genre naturally focuses on similarities between texts and the temptation is to extrapolate these similarities into rules to be given to students. However, arguing that the majority behave linguistically in a certain way does not necessarily generate an argument that all students should be urged to behave in this way. Furthermore, genre-based approaches have limitations when looking at the same

individuals over time due to the problems of genre identification and genre boundaries. Process approaches, while at times focusing on individual preferences during the composing process, fail to systematically analyse the language used in texts.

3.4 Development of a framework

The current framework is based on the concepts of writer, context of situation and of text. The reason for the focus on these three areas is that the present study is based on texts written with a similar context of situation by the same cohort of writers. Chapter 4 outlines in more detail the collection and organisation of the corpus used in the present study. For now, it is sufficient to note that 17 students composed 94 essays at 6 different points in their degree programme, all in the same subject area of philosophy.

The writer

Although the writer's role is often marginalised in research with emphasis instead placed on the contextual situation of the act of writing, the genre of the text or the process involved in the composing of the text, the current study sees the writer as the principal agent in the creation of a text. No text exists without the writer. While other notions such as genre and context play a role in the shaping of a text, it is the writer that is of paramount importance when it comes to creating a text. The writer is a unique individual with their own personality, learning style and concept of the world in which they live. Due to this uniqueness in the writer we have uniqueness in the texts produced. The uniqueness in every writer is explored below.

Heikkila and Lonka (2006) outline three traditions in the study of university students learning, namely students' approaches to learning, self-regulated learning and cognitive strategies. While such areas of research are concerned with learning as a whole and not exclusively related to the writing process, some of this work is relevant here.

In the student approach to learning tradition, Marton and Saljo (1976) showed that when reading texts, different students used different approaches to process the text material; surface and deep. A deep approach is concerned with the fundamental message of the

material while a surface approach is more concerned with surface features and remembering the text word for word. Entwistle and Ramsden (1982) and Biggs (1987) added a third approach called strategic or achieving. Students who adopt this approach work very hard, select a strategy in order to maximize academic success and are very concerned with assessment practices (Heikkila and Lonka, 2006: 100). Watkins (2001) showed that academic achievement is related to student approaches and that both a deep approach and the achieving approach are associated with better academic outcomes. Various approaches to learning within the general academic setting (lectures, library, etc) must result in different writing patterns.

In the self-regulated learning approach, there is an emphasis on actual cognitive processing, and motivational, affective and contextual factors are included. A student who regulates their learning is capable of setting task-related, reasonable goals, taking responsibility for their learning and maintain motivation (Heikkila and Lonka, 2006: 101). However, not all students are self-regulating learners.

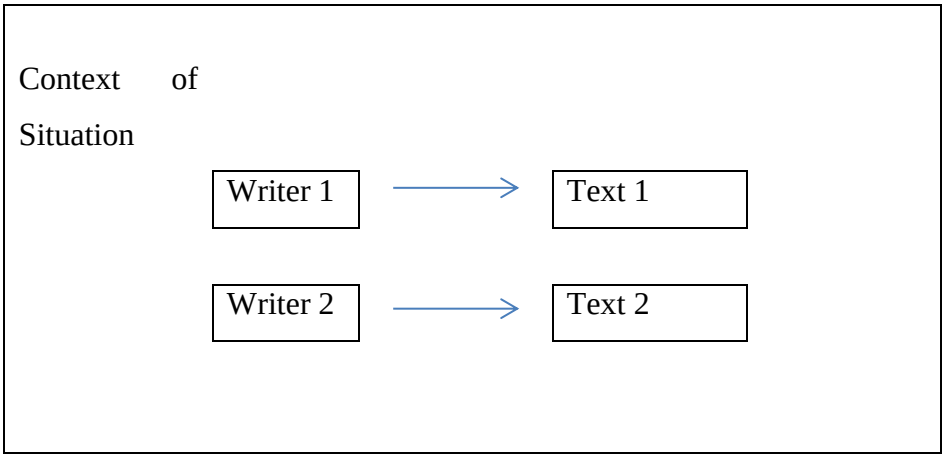
In the cognitive strategy tradition, the term cognitive strategy 'refers to the cognitive, affective and behaviourial process people apply to achieve their goals and to evaluate the outcomes of their actions' (Heikkila and Lonka, 2006: 102). Eronen et al (1998) have shown that there are strategies mainly used in academic settings by students, namely illusionary optimism, defensive pessimism and self-handicapping. Illusionary optimism is where students have high expectations based on previous success and while they take credit for success, are likely to blame others or situational factors for any failure. Students using a defensive-pessimism strategy have low expectations and are generally anxious about their performance. Self-handicappers fear potential failure and engage in task-irrelevant behaviour to create excuses for possible, and likely, failure. Eronen et al (1998) have shown that students using self-handicapping strategies are least successful in their studies while students using illusionary optimistic strategies were the most successful in the long term.

Heikkila and Lonka note that all three traditions share basic assumptions that derive from cognitive psychology: 'they all emphasise that students' expectations, prior experiences and beliefs are unique filters that colour the way they perceive events' (2006: 103). What is of great interest to us is that students perceive events differently even though those events may be similar. They have 'unique filters' that would lead us

to assume that each student views the task of essay writing in their own unique way and as experiences, expectations and beliefs change, we can also assume that the effects of these 'unique filters' change with time. Furthermore, Heikkila and Lonka (2006: 104) go on to argue that 'the learning context is not an objective entity, but, rather, it is perceived, observed or interpretated by the students'. As writing is produced within this learning context, it too is subject to individual interpretation.

As shown, each writer has their own personality type and learning style. This affects a number of issues regarding writing an essay, namely, how the task is approached, how the deadline affects the outcome, how the writer views their audience, what content the writer deems relevant and how the writer views the context of situation. We cannot know everything about the writer or quantify the person into distinct categories. In fact, while some studies into the process of writing academic essays claim to be able to say what the writer is thinking while they compose the text and at different stages throughout the composing process, it is more likely that such an internal activity as thought processes cannot be fully understood nor verbalized by a research informant and therefore cannot be fully explored by a research project. In the present study, the salient point, as shown in Figure 3.1, is that each writer is unique and this gives us unique texts.

Figure 3.1 Writer and text



While there are some similarities between writer 1 and writer 2 and between text 1 and text 2, both writers are unique and both texts are unique. It is the commonality between writers, and hence between texts, that leads to the similarities and conventions

associated with genre-based theories, and it is the individuality in writers that leads to the uniqueness in texts. In addition, the writer does not exist solely as a writer, but as a person engaged in many discourse roles as they interact with their world. The language choices of an individual when writing an academic piece of work is influenced not just by their interaction with that academic sphere of discourse, but by the sum of their language experiences.

While each writer is a unique individual, it would not be prudent to assume that each writer is the same throughout a research project that spans a number of years. In fact, each writer is not a static entity who views the philosophical content, the context and the text in a continuously uniform way. Instead, the writer is constantly changing in themselves and in how they view the outside world. In the first semester data in the present study, each writer is assumed to be new to the discipline of philosophy. While some may have experience of philosophy in an informal way, none of the students have attended formal classes in philosophy nor have they written and submitted a formal assignment in the discipline. However, each student is drawing on their own personal schemata and experience of writing other essays to aid them in the composing of their first philosophy essay.

As the student progresses through the institution, their perception of appropriate content and appropriate language to convey this content is changing. Furthermore, a student does not exist solely as a writer but as a person inhabiting the world and society in general. The language choices, therefore, are influenced not just by academia, but by all their worldly experience. Generally, the student is expected to be producing an improved written work in their final year compared to their first year. Therefore, while genre-based approaches would put all the student essays into one category of academic writing, in particular philosophical undergraduate academic writing, this research recognizes that the writer, the product, and the process to create this product, are not completely similar throughout the semesters. Figure 3.2 shows a static writer producing a number of texts, while Figure 3.3 shows a more accurate diagram of the same writer at different points in time producing a number of texts.

Figure 3.2 Writer as static entity

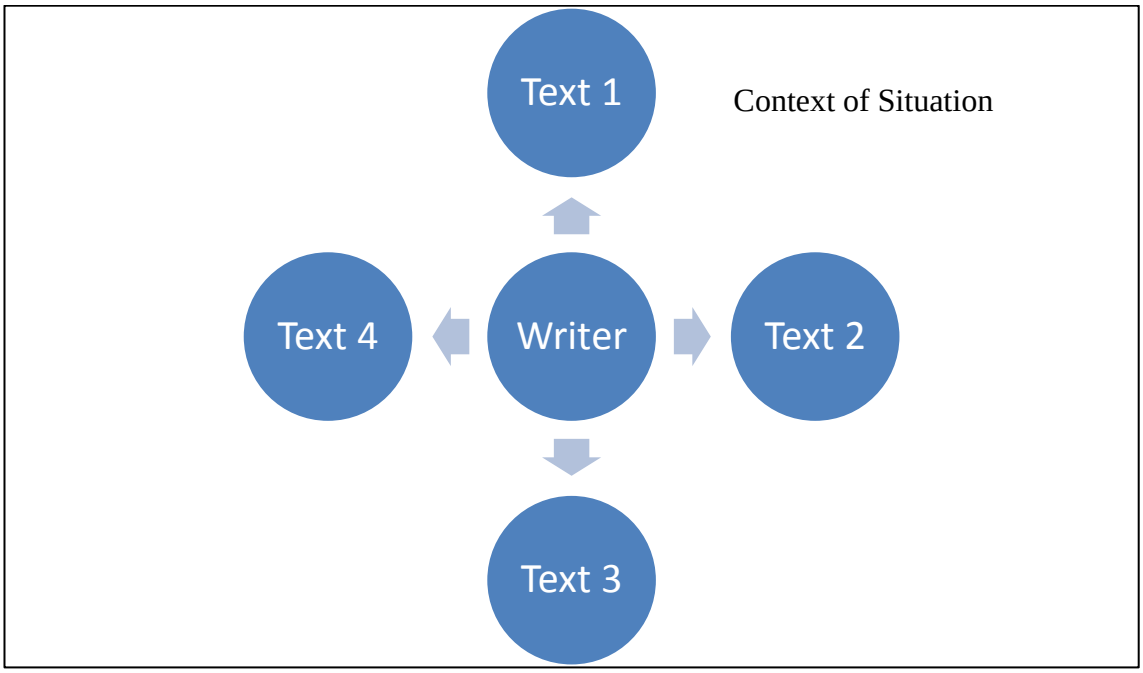
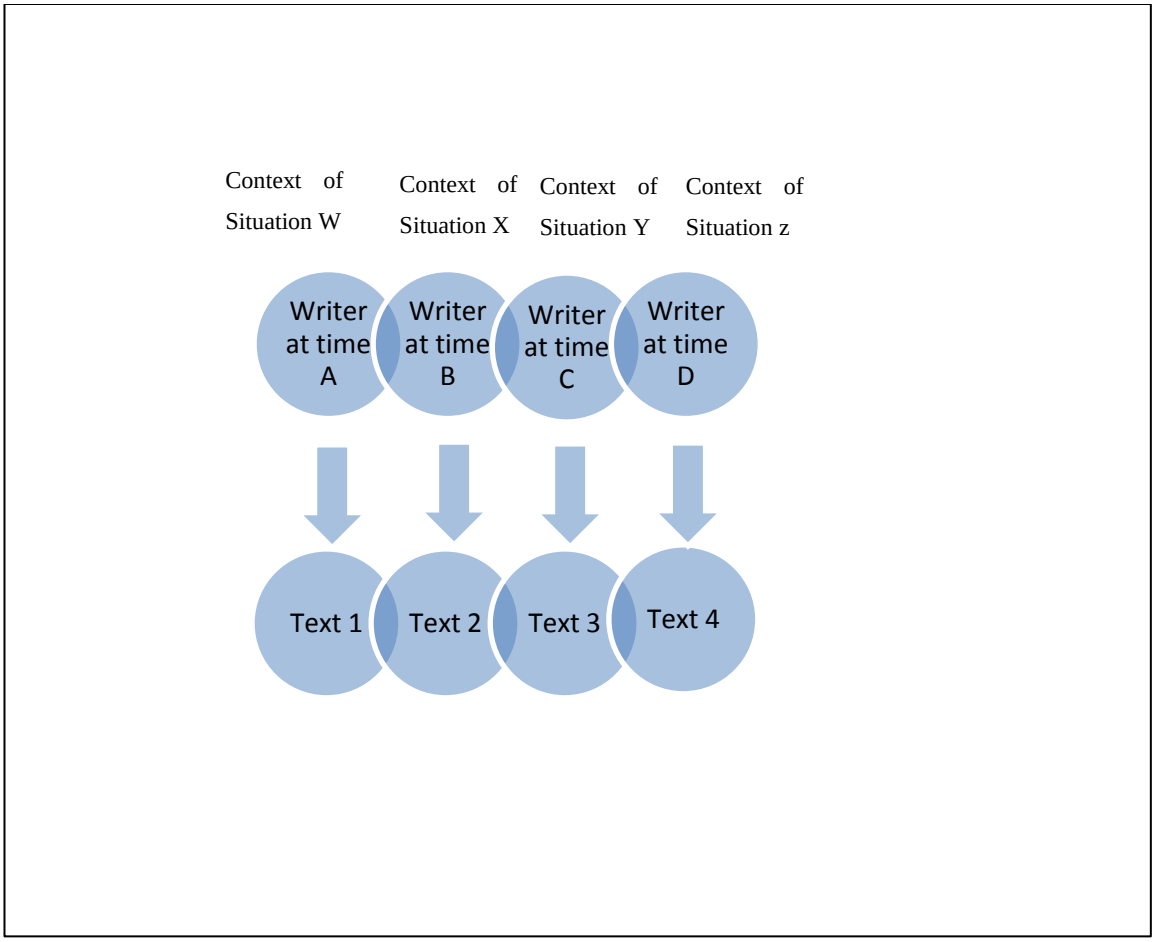


Figure 3.3 Writer as dynamic



While the student may start in the first semester as a complete novice, in the second semester they are slightly more experienced and by their final year it would be hard to argue that they are a novice. However, while a student in their final year cannot be viewed as a novice in relation to first-year students, some researchers view a final-year student as a novice in relation to professional academic writers (see Hyland, 2006 for a further discussion on student writing and professional writing). This is especially true if we view, as we do in the present study, student writing as a type of writing in its own right as opposed to a lower form of writing that is expected to mirror norms found in professional articles. Were we to ask a student to do the same essay as they did in the first semester at any other time after this, even if they did not receive any feedback or grade on the original work, the new text would not be the same. It may be similar but it would not be the same. An example of this can be seen when a writer, be they undergraduate or postgraduate, loses written work due to a computer problem. When they rewrite the work, it is not the same as the original, even though the context and the genre remain the same.

The writers in the present study are all undergraduate students and, as stated above, are not static entities in relation to either their writing or their knowledge of their subject are (philosophy in the present study); their proficiency and expertise changes as time progresses. Bartholomae (1986) argues that each time a student writes, they are reinventing the university in their thoughts to tackle the writing process and this reinvention may not be the same each time the student tackles a new writing task. Taylor also points to each essay being a unique task, but for a different reason when he argues that 'each essay or assignment is indeed something of 'a new beginning' for many undergraduates because sense must be imposed upon unfamiliar and often intellectually challenging material' (1988: 58). Furthermore, even at a given point in time, the writers in the current data are not a homogenous entity. Some are better than others at the skill of writing, some have a better understanding of the philosophical content and the writers are awarded a range of grades by the assessor. It is important to take note of this when examining the data. While there will undoubtedly be similarities between texts, there will also be differences and also there will be different levels of success in relation to the text and the grade received.

If we take into account the fact that there are differences between an individual writer over time and differences between each writer at a given time, it points us in the direction of differences between texts. Furthermore, the production of a single essay is characterised by a set of instantial decisions, both in terms of content and the lexical realisation of that content. Previous research, especially genre-based research, into academic writing tended to focus on similarities between texts, especially as it was assumed that the texts being studied were produced in the same context. However, the present study focuses on both the similarities and differences between texts.

Context of Situation

Although writing about spoken language, Malinowski (1923) introduces the concept of context of situation to the analysis of language. He argues that the study of language cannot be divorced from the context and culture in which it occurs. He comes to the conclusion that 'language in its primitive forms ought to be regarded and studied against the background of human activities and as a mode of human behaviour in practical matters'. To begin with, Malinowski introduced context of situation in relation to primitive languages only, but later (Malinowski, 1935: 58) revised this, noting that:

> This was an error, and a serious error at that. Between the savage use of words and the most abstract and theoretical one there is only a difference of a degree. Ultimately all the meaning of all words is derived from bodily experience.

While Malinowski confined this context of situation to the analysis of specific texts, Firth (1935) broadened the scope of context of situation and came to the view that linguistics was the study of meaning and meaning could only be found through looking at the function in a context. The best practical application of this is probably Mitchell's (1957) study of buying and selling in North Africa. Mitchell identifies how the context, namely auction, market and shop transactions, influence the language used by the participants, where 'a text is a kind of a snowball, and every lexical item and every collocation in it is part of its own context, in the wider sense of this term' (Mitchell, 1957: 186). He also identifies stages within these situations, showing at the same time that each situation has its own unique set of stages and that 'by adhering to the principle that meaning must be sought in use, we are able at the situational level to make a

systematic classification of material on the basis of correlations between texts and their environments' (Mitchell, 1975: 168).

Halliday (1978) also takes up the notion of context of situation and, while acknowledging the work of Malinowski, traces the concept of the situation being the environment in which a text comes to life back to Wegener (1885). Halliday acknowledges that the context of situation may be removed from the what is happening around the act of speaking or writing, but at the same time argues that we need to view situation in a more abstract way again and 'conceive of it not as situation but as situation type … It is a constellation of meanings deriving from the semiotic system that constitutes the culture.' (1978: 109). Although he argues for more abstract notions of situation, Halliday also recognizes the uniqueness of situation for every text:

> A text is embedded in a context of situation. The context of situation of any text is an instance of a generalized social context or situation type. The situation type is not an inventory of ongoing sights and sounds but a semiotic structure; it is the ecological matrix that is constitutive of the text.
>
> (Halliday, 1978: 122)

Discourse Communities

According to Bizzell (1986), novice writers at third level education experience a clash between their home dialects and the language of college; novice writers experience a clash between the discourse forms and genres of their worlds prior to college and the discourse forms and genres of formal college writing; and novice writers experience problems arising from their lack of cognitive development (as measured by the developmental schemes of Piaget or Perry). These three reductive theories can be synthesized into a comprehensive view by means of the notion of discourse community. What basic writers experience is a profound clash of world views. While the discourse communities from which basic writers emerge have not been studied in sufficient depth, it seems certain that the world view favored by the academy will challenge that of basic writers new to college. The western notion of the academy requires a skeptical, questioning frame of mind (what Perry calls a world with 'no Absolutes') and a rational choice of beliefs (what Perry calls 'Commitments'), rather than unquestioning faith. We must note here that this is a Western notion of the Academy and not a globally uniform notion.

Swales (1988), building on the work of Bizzell (1982) and Faigley (1985) among others, discusses and refines the notion of discourse communities, which he distinguishes from speech communities. Speech communities are centripetal (they pull people in), while discourse communities are centrifugal (they set people, or parts of people, apart) (Swales 1988: 212). A discourse community, apart from comprising defining components such as commonality of interest, public goals, purposeful interchange of information and feedback, it also has a forum, a term which Swales attributes to Herrington (1985). Fora can consist of bulletins, meetings, conferences, telephones calls, emails and websites and so on. Via these fora, discourse communities develop and continue to develop discourse expectations which 'may involve the appropriacy of topics, the form, function and positioning of discoursal elements' (Swales 1988: 212). These discoursal expectations, according to Swales, create the genres that articulate the operations of the discourse community. He sees the resultant genres as properties of discourse communities and, as such, they serve as social binding agents to hold together a critical mass of members, via a forum or fora.

As we have seen, since the work of Malinowski, discourse has been seen as a social phenomenon operating within a context. Discourse community stresses the social context in which language occurs. This is especially true in the theory and analysis of written language as the notion of discourse community 'embraces the rhetorical concern with social interchange (discourse) with the situation or context (community)' (Killingsworth, 1992: 110). However, when Malinowski introduced context of situation, it seems as if the context was seen as local. Discourse communities are also generally operationalised in local terms (eg Beaufort, 1997; Pogner, 2003; Woodward-Kron, 2004). Killingsworth (1992: 121), however, differentiates between local and global communities, and concludes that:

> … I would say, in sum, that most people stand between two kinds of discourse communities: local discourse communities, groups of readers and writers who habitually work together in companies, colleges, departments, neighbourhoods, government agencies, or other groups defined by specific demographic features; and global discourse communities, groups of writers and readers defined exclusively by a commitment to particular kinds of discourse practices and preferences, regardless of where and with whom they work.

Before using the concept of discourse community to investigate the language used in an institutional site of composing, Beaufort (1997) outlines two problems with the concept. The first issue relates to the boundaries of communities: where or how can we isolate

one community from another and the features of one discourse community could overlap with another (Beaufort, 1997: 488). However, she, as a result of her study, comes to the conclusion that 'no institution operates in complete isolation from the larger culture or separate from other institutions … discourse communities in fact overlap with each other, each having unique features to itself and overlapping features shared in common with other discourse communities' (1997: 493). The second problem is more of a pragmatic one: 'positing norms for communities of writers and readers begs for its antithesis – what Harris (1989) alludes to as tensions and discontinuities in the writing practices of any community' (Beaufort, 1997: 488).

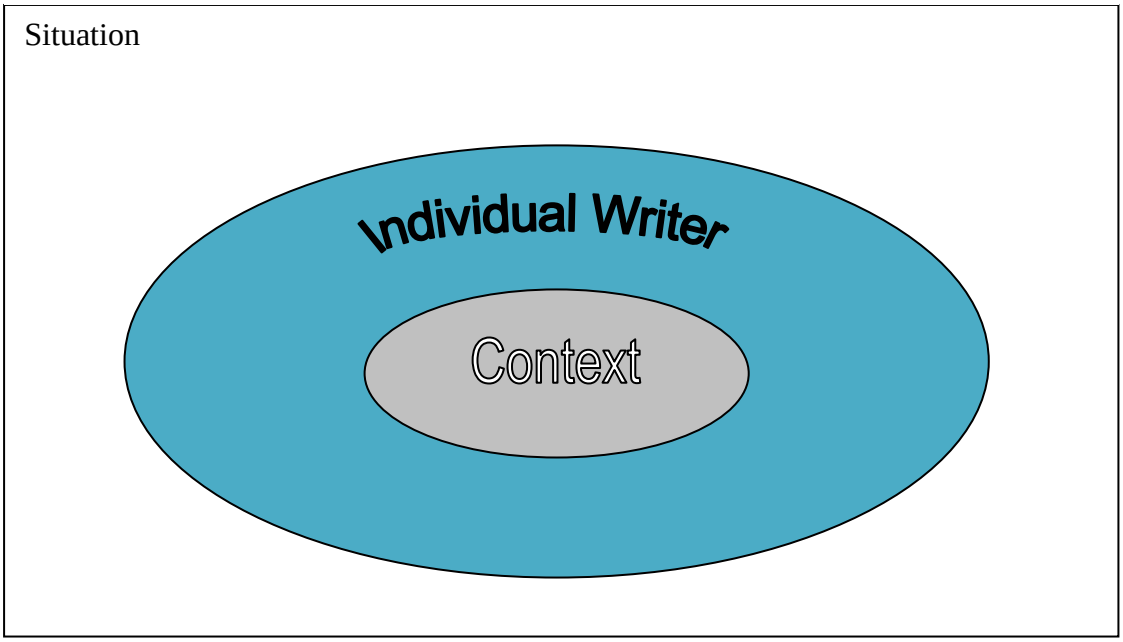
More pertinent to the present study is the work of Woodward-Kron (2002, 2004) which centres around undergraduate student writing and the discourse community these students operate in. She argues that the academic discourse community members share knowledge about the textual practices of a discipline, and these practices are intrinsic to the ways disciplinary knowledge is constructed. If this is the case, then students are not part of the discourse community as they do not share knowledge of the textual practices of a discipline at the beginning of their study and most do not acquire this fully as they progress through their studies.

Context of situation as subjective reality

Within Malinowski's (1923) concept of context of situation, there are in fact two notions at play. The first is the notion of situation, the second is the context of that situation. In this research, situation is seen as an objective reality in which the language production takes place. In the current case, the situation is writing an undergraduate philosophy essay. This situation is the same for all the individuals in the corpus for each task, and therefore the situation for all the texts written in the first semester is the same, the situation for all the texts produced in the second semester is the same and so on. However, the context of this situation is not an objective reality, but instead is subjective and dependent on personal interpretation. As we have seen, every writer is unique and, therefore, for each text produced, the context of situation differs. Depending on the individual writer, their personality types, learning styles, current personal situation, individual time constraints and current events both related to the situation and unrelated, each student has a different concept of the types of content required for the

task and also on how that content is to be realized through the use of language. For this reason, we can conceptualise the context of any given writing task as internal to the writer as this context exists uniquely within the individual. No two writers will see the same situation in the exact same way.

Figure 3.4 Context as internal to the writer



In previous research, it has often been assumed that the context of composing essays is static and uniform, that this context affects language choice, which in turn affects the way the text is received. However, it should be noted that content is more important than context in relation to language choice when it comes to undergraduate essays. Language choices, while influenced by the perceived context of situation, are based primarily on what the writer wants to say, on the meaning the writer wishes to convey through the medium of their written essay. This is the most important consideration when a writer composes an essay:

> We do not, in fact, first decide what we want to say, independently of the setting, and then dress it up in a garb that is appropriate to it in the context, as some writers on language and language events seem to assume. The 'content' is part of the total planning that takes place.
>
> (Halliday, 1978: 33)

Most research into academic writing does not take into account the denotational meaning of the language used. Instead, there is an emphasis on the language used by the writer, or in some cases, the language that should have been used by the writer, and this other language is compared to generic norms. This approach denies the basic function of a student writing an essay, namely to convey to the assessor a meaning concerning the task set to them. While language choice can affect the perception of the reader, the most important part of academic writing remains the content. One reason for ignoring content could be an assumption that the students understand the content and their success, or lack of success, depends solely on their adherence to generic norms and the language choices, in terms of individual words and phrases and organisation. Another reason for the lack of interest in content could be that most research is carried out by teachers of academic writing, or at least those who have some concern with the teaching of academic writing, and not subject matter experts. For those concerned with the teaching of academic writing, there is an obvious reluctance to argue that writing style and language choice plays at best a small part in the production of a successful text as to do so would make the teaching of academic writing irrelevant and thus make the positions of the researchers within the institutions redundant.

If we argue that the context of situation is unique to each individual at a point in time, it follows that the audience, or in this case, the assessor, and genre may be viewed differently by each writer over time. Therefore, despite the fact that the assessor is the same for each of the 94 essays submitted at six different points in time over the course of a degree programme, the perception of the assessor changes for each student over time. In addition, the relevance the students give to the assessor when composing texts may also change over time. While the student perception of the assessor is changing as the semesters pass by, the assessor is also changing. Their attitude to the subject, their expectations and their knowledge are also evolving. Although each of the 94 texts is an essay, the writers' concepts of what an essay entails is also changing. Furthermore, the assessor's perception of the individual student that they are marking is changing (essays are not marked anonymously, therefore, the assessor is aware of which student composed each essay).

The text

In the present study, each text is viewed as unique. There are similarities between each text and another written by a different writer, and similarities between every text composed by the same writer, and yet each text remains unique in that it was composed by an individual at a particular point in time, for a particular contextualised task, and is therefore unique to that individual at the time. The analysis chapters of the current research provide empirical evidence for uniqueness of each text in the corpus, both within each point in time and across points in time. There are, however, similarities between texts on a lexico-grammatical level as well as differences, and there are also similarities between the texts which can be exploited for the purpose of creating a theoretical framework. Furthermore, Zipf (1935) has argued that all texts will share similar properties in the distribution of lexical items within the texts.

The primary function of each text is to convey philosophical meaning from the writer to the reader. Each writer has some power over what this content is going to be, but, at the same time, somebody else has set the task for the writer. The writer is therefore constrained in terms of content in a number of ways. Firstly, the topic of the essay is determined by someone external to the writer. Secondly, how that topic is to be discussed is also determined through reading lists and teaching contact. Thirdly, the writing task is subject to a deadline. However, the text remains the product of the writer.

In genre-based approaches to researching academic writing, no account is taken of the individual in the process of creating a text. While it has been argued here that each text produced within the same task, within the same social setting and within the same time frame is unique, genre-based approaches focus on the similarities of texts. Although similarities between texts can be found (similarities between totally unrelated texts can also be found), each text produced is in fact unique. The reason each text is unique is because it is produced by a unique individual at a point in time. The influencing factors of genre, task and audience has a unique, complex inter-related effect at any given instance. The individual creates the text as they perceive it should be created. No student deliberately writes a bad essay. Process approaches tend to take no account of the product, the text, that is produced in this process of writing. It is as if the process and the outcome are not related. This is not the case as texts are produced in a social

space and are a result of individual activity that occurs during the process of their creation.

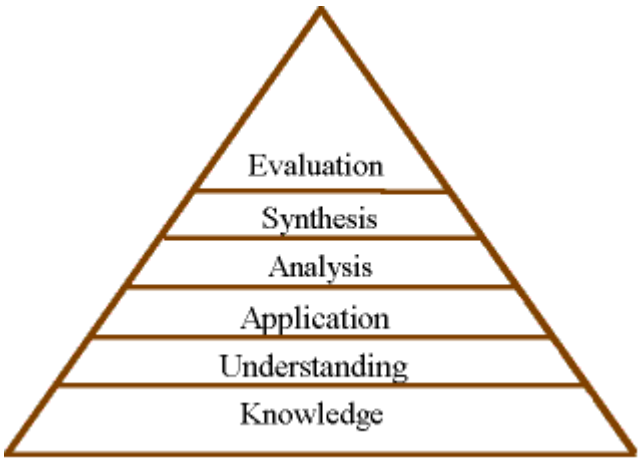
It is also argued here that texts are composed of meaning which is realized through the use of language. As Halliday argues, 'a text … is a semantic unit, which is not composed of sentences but is realized in sentences' (1978: 135). If we view the act of undergraduate academic writing as an attempt by a student to best approximate and then transmit their knowledge of the topic, then we cannot conceive a text divorced from the message it contains. No theory of genre is equipped to integrate content and meaning into its analysis. Content and meaning are vital parts of communication and communication primarily imparts meaning. When students write academic essays, they are attempting to communicate a meaning to the assessor. The grade they receive is not for the use of language nor adherence to the genre, but instead for the level of understanding and analysis of the content material they display (in some cases, inappropriate style may result in students being 'marked down'; however, inappropriate meaning can result in students failing). This understanding is conveyed through language, however; language is the tool and not the objective of the assessment task. If the language is grammatically correct, written in the appropriate register, generically aligned to the expectations of the assessor and organised into a coherent whole, and yet the content is inappropriate, or simply incorrect, then the student will not receive a good grade.

Types of knowledge and content

Bloom (1956) developed a classification of levels of intellectual behaviour in learning. This taxonomy contained three domains which overlap: the affective, the psychomotor and the cognitive. The affective domain concerns interpersonal relations and emotions and is demonstrated by behaviours indicating attitudes of awareness, interest, attention, concern, and responsibility, ability to listen and respond in interactions with others, and ability to demonstrate those attitudinal characteristics or values which are appropriate to the test situation and the field of study. The psychomotor domain relates to fine and gross motor skills and is demonstrated by physical skills; coordination, dexterity, manipulation, grace, strength, speed; actions which demonstrate the fine motor skills such as use of precision instruments or tools, or actions which evidence gross motor

skills such as the use of the body in dance or athletic performance. Of particular relevance here is the cognitive domain, within which Bloom (1956) identified six levels: knowledge, understanding, application, analysis, synthesis and evaluation. Knowledge is regarded as the lowest level and the levels become increasingly more complex and abstract as we move towards the highest level, evaluation.

Figure 3.5 Bloom's Taxonomy



Knowledge is defined as the remembering of previously learned material and is demonstrated by the observation and recall of information, knowledge of dates, events, places, knowledge of major ideas and mastery of the subject matter. It represents the lowest level of learning outcomes in the cognitive domain. Understanding is the ability to grasp the meaning of material. This may be shown by translating material from one form to another (words to numbers), by interpreting material (explaining or summarizing), and by estimating future trends (predicting consequences or effects). These learning outcomes go one step beyond the simple remembering of material, and represent the lowest level of understanding. Application refers to the ability to use learned material in new and concrete situations. This may include the application of such things as rules, methods, concepts, principles, laws, and theories. This requires a higher level of comprehension than that under understanding.

Analysis refers to the ability to break down material into its component parts so that its organizational structure may be understood. This may include the identification of parts, analysis of the relationship between parts, and recognition of the organizational

principles involved. This represents a higher intellectual level than understanding and application because it requires an understanding of both the content and the structural form of the material. Synthesis is the ability to put parts together to form a new whole. This may involve the production of a unique communication (theme or speech), a plan of operations (research proposal), or a set of abstract relations (scheme for classifying information). This area stresses creative behaviours, with major emphasis on the formulation of new patterns or structure. Evaluation is concerned with the ability to judge the value of material (statement, novel, poem, research report) for a given purpose. The judgments are to be based on definite criteria. These may be internal criteria (organization) or external criteria (relevance to the purpose) and the student may determine the criteria or be given them. This level is the highest in the cognitive hierarchy because it contains elements of all the other categories, plus conscious value judgments based on clearly defined criteria.

The reason the notion of different types of knowledge, as exemplified by Bloom's taxonomy, is essential to the current research is to acknowledge that different types of knowledge are more highly regarded than others. Although there may be linguistic markers associated with each of the levels in Bloom's taxonomy, for the students in the corpus in the current research, the appearance of such linguistic markers is not, in itself, sufficient. The assessor, as part of the wider philosophical community as well as rooted within the institution, must also say that displayed knowledge as appropriate.

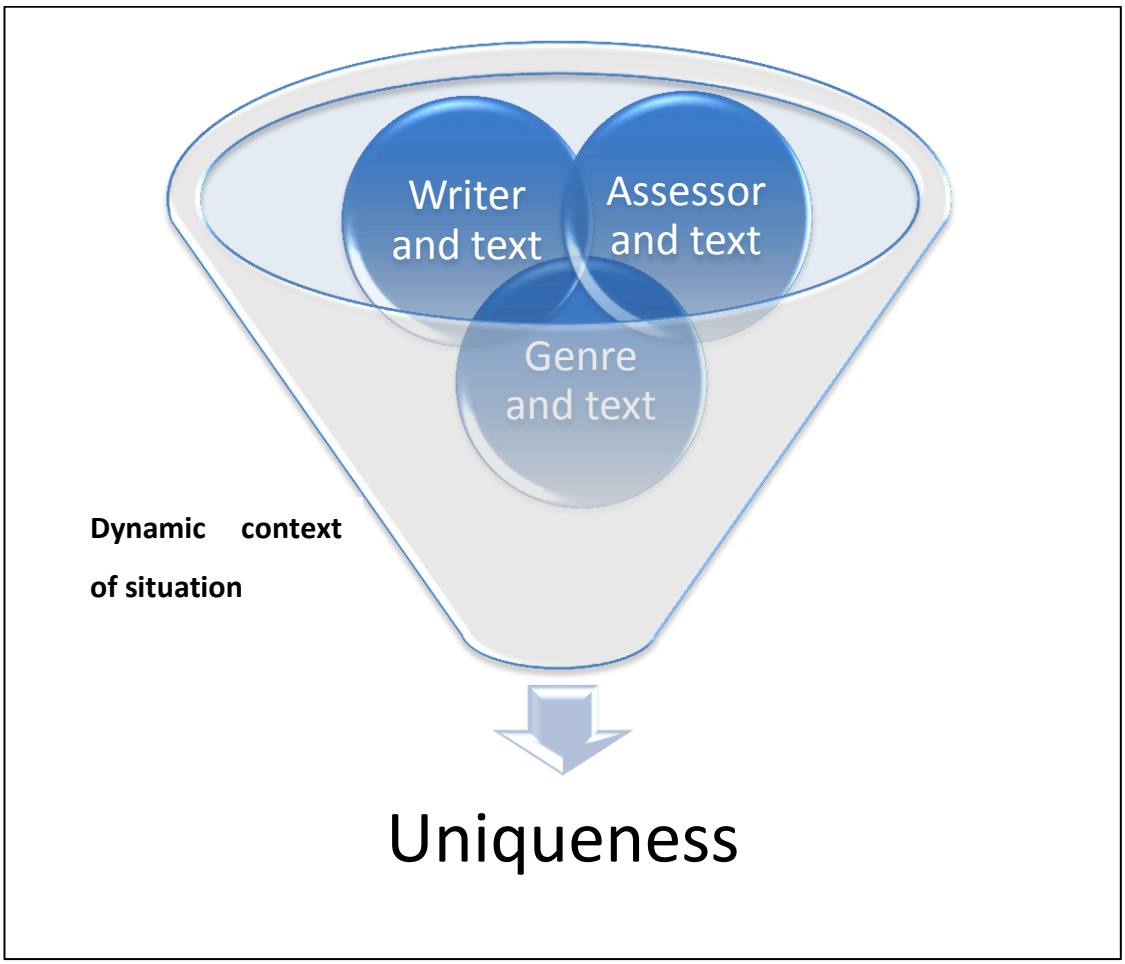
As we have seen, the text is composed of language chosen by the writer to convey meaning. The language on its own does not determine the grade received by the essay. Instead, the content and meaning contained within the essay determine the grade. Some of the language choices made by the writer are due to individual preference and the individual's perception of the language required based upon their unique interpretation of the social context coupled with their own personality. As we have seen, Coulthard (2004) argues that each person has a unique idiolect and this affects their language choices when writing essays (although the notion of individualised language is important to author attribution and plagiarism detection, such is not the focus of the present study). Furthermore, Halliday points to the importance of meaning to language: 'Language is the ability to 'mean' in the situation types, or social contexts, that are generated by the culture' (1978: 34). Choices of philosophical content are also

dependent on the individual, but this is what the assessor is looking for. When researching academic writing of undergraduates, we are at a disadvantage as we cannot determine the quality of the content of the essay. In short, content gets the grade, inappropriate language does not affect grade unless the inappropriate language obscures the meaning intended by the writer.

3.5 Summary of relationship between writer, context of situation and text

A summary of the relationship between the writer, the context of situation and the text is shown below in Figure 3.6.

Figure 3.6 Relationship between writer, genre, assessor and text



In Figure 3.6:

- Text is produced in social space;

- Social space is not objective;
- The interpretation of the context of situation by the writer, including the assessor and genre, is particular to the writer at any given time;
- The assessor is dynamic;
- Subjective interpretation of social space influences text;
- Denotional meaning influences text;
- Denotional meaning as conveyed through text is more important that interpersonal meaning in relation to success of text;
- However, interpersonal meaning is essential to the communication of denotional meaning;
- Linguistic choices are unique to that individual at that point in time. (At another time, the individual still relies on these factors to create the text but these factors may not be the same as the previous time of composing);
- Linguistic choices are a result of a combination of the individual and
    - their personality;
    - their schemata;
    - their interpretation of the context of situation, including the assessor;
    - the meaning they wish to convey.

The core argument in relation to the writer, context of situation and text is that these notions are dynamic, not static. Furthermore, they are continually interacting with each other to create unique and instantial influences on the language used. It is for this reason that change occurs in the writing patterns of undergraduate students and there is a uniqueness in every text. It is only by understanding the dynamic change within those factors that allows us to investigate the change in student writing.

3.6 Conclusion

This chapter set out to develop a theoretical framework for the analysis of change in a corpus of undergraduate philosophy essays. To do this, the limitations of process-based and genre-based approaches to academic writing were identified. A framework was then developed, taking into consideration the limitations with both approaches, based around the notions of writer, context of situation and text.

To support the argument that each text is unique and differs from every other text, despite similarities in the situation of composing, this research will examine differences and similarities between texts. It will be shown that for every pattern of similarities between texts, there are also texts which do not adhere to this pattern and yet are successful. It will also be shown that individual students write texts with dissimilar patterns at different points in time. Each individual is engaged in a process of change, but this change is non-linear. Every essay is, in its own right, a perfectly-formed text that is unique yet has a relationship with all other essays in the corpus. However, under further investigation, the corpus as a whole, and each text within the corpus, seems random and chaotic in the words that are used, resulting from a set of instantial decisions.

The present study is an analysis of the pattern of change in student writing. This analysis is based on empirical data constructed from the essays written by a sample from a cohort of undergraduate philosophy students. The next chapter, Chapter 4, outlines the collection of those texts and the methodology used in their analysis.

Chapter 4 Data gathering and methodology

4.1 Introduction

This study sets out to investigate whether there is a patterned, linear change in student writing over time. The sub-research questions concern the uniqueness of each text, the response to genre, task and audience within the texts and the standardisation of this response. As we have seen in Chapters 2 and 3, the writing as process and writing as genre approaches are not sufficient for understanding how writing changes over time. The current study conceptualises the written text as a response to a combination of factors including genre, task and audience as a unique construct at any given point in time. These three notions are inherent in the context of situation, which is a dynamically-changing subjective perception of an objective reality.

This is an empirical study based on the essays of a cohort of students written during their degree in Mary Immaculate College, Limerick. Mary Immaculate College, founded in 1898, is a university-level College of Education and the Liberal Arts, academically linked with the University of Limerick. The College has a student population of around 3,200 and offers undergraduate programmes in Education, Liberal Arts and Early Childhood Care and Education, as well as a range of postgraduate programmes at Diploma, MA and PhD levels.

This chapter outlines the collection of these essays to form the Mary Immaculate Corpus of Undergraduate Philosophy Essays (MICUPE), their organisation into six essay iterations according to time of submission and the methodology used to investigate the research question and the sub-questions. It then outlines an analytical framework for the analysis of this data.

4.2 Description of the Mary Immaculate Corpus of Undergraduate Philosophy Essays (MICUPE)

This section outlines the collection and organisation of the 94 essay texts that comprise MICUPE.

Collection

The texts in this corpus were collected in Mary Immaculate College, Limerick, Ireland. The writers of these texts were students undertaking a Bachelor of Education (BEd) degree programme. In this three-year degree programme, students study a range of modules that are prescribed to them. They also take a Liberal Arts subject to degree level from a choice of 10 subjects. In the first year of their degree, they undertake two of the Liberal Arts subjects, one of which must be the Irish language. After first year, the student must continue with one of those 10 subjects to degree level2. There are also students who are studying for a BA in Liberal Arts enrolled in the same modules as those students studying for a BEd degree. The students in the corpus were studying in Mary Immaculate College from September 2001 until May 2004 and chose to study philosophy as their Liberal Arts subject.

The essays in this corpus all come from the subject of philosophy. The reason philosophy was selected is because philosophy as a subject is not taught in the secondary-school system. This means that in semester 1 of year 1, no student has studied philosophy in any formal way prior to entering Mary Immaculate College. If the present research had been based on a subject that was available in the second-level educational system, some students may have had experience of that subject while others may not and they would have had varying levels of attainment in their teminal second level exam, the Leaving Certificate. As philosophy is not part of the second-level curriculum, therefore, the input for each student remains constant, as does their antecedent experience and attainment in relation to the discipline. Furthermore, as all students scored in a similar grade range in their final second-level school exams, the

2 The structure of the BEd degree changed to a four-year degree programme in 2012/13 and students no longer study a Liberal Arts major.

Leaving Certificate3, there is as much uniformity as possible in the starting point (this does not mean that the starting points are the same, and the effect of differences in initial conditions is discussed further in Chapter 8).

In order to create the corpus, permission was obtained from the students and from the lecturer responsible for the relevant modules. Not all students in the philosophy degree programme gave permission for their work to be used, and not all lecturers were interested in having their modules be part of the study. Although it had been suggested that the intellectual property rights of the essays belong to the College, which would mean that technically permission would not be needed from the student or the lecturer once the institution agreed, I did not want to use any texts where either the writer or the assessor was not willing to give permission for them to be used. For this reason, permission was granted by the lecturer and each one of the students for their work to be included in this research project.

Transfer to electronic format

The essays were only available in hardcopy (see below for details on transfer to machine-readable format). Those hardcopies resided with the lecturer. At times, although I had permission from the students, particular essays could not be located. The collection agreements with the students willing to be part of this study was that the texts would be collected from the lecturer so that there would be absolutely no disruption to the students themselves. Possible reasons for the texts not being available include being sent to an external examiner, not being submitted in the first instance or simply being misplaced by the lecturer, post marking. The corpus is based on essays written by 17 students. In some instances, not all students are represented in totality for the reasons just outlined. This means that some students have 6 essays in the corpus, while others have 5 essays. No student has less than 5 essays in the corpus.

3 In Ireland, there is a Central Applications Office to which all students seeking a place in Higher Education apply, listing their degree programme choices in rank order. The Higher Education institutions have an entry point cut-off based on a quota system and students are then allocated their choice of programme relative to their overall Leaving Certificate points.

As each essay existed in hardcopy, they had to be transferred to electronic format in order to allow computer-aided analysis of the texts. This was done by using a scanner in tandem with Optical Character Recognition (OCR) software. Subsequently, the new electronic version of each text was checked against the hardcopy to eliminate errors in the transferring process. However, in semester 1 of year 1, the students were permitted to submit hand-written essays should they so desire. In these instances, the essays were transferred to machine-readable format through typing. Again, the electronic version was checked against the hardcopy to ensure accuracy.

The included texts

Some essays contained a separate title page with details such as student identification number, student name, module code and essay title. Other essays included these details at the top of the first page of the essay. To protect the identity of the students and to have consistency of data, such details were omitted from the final corpus. Therefore, within MICUPE, all essays start with the first word of paragraph 1. Some essays included a bibliography or reference list. This list does not inform the present analysis and was therefore omitted from the corpus. The section of the essays included for analysis run from the first word of the first paragraph to the final word of the final paragraph.

Most essays contained footnotes with biographical references. The number indicating the insertion of a footnote with a biographical reference remains in the electronic version of the texts used for analysis. However, the biographical details have been removed from the corpus. In a number of instances, footnotes were used to provide extra information on the content. The text of these footnotes is included in the analysis.

At times, the writers used font effects such as bold, underline and italics. As the corpus was collected, these effects were coded so that they would be identifiable when analysing the data. However, such effects are not part of the current analysis and are therefore not referred to. The present study utilises examples from the student texts where appropriate. In these instances, if a particular feature is under discussion, that feature is presented in bold. This use of bold is to help make the analysis clear, it is not a representation of an original use of bold by the student.

In some instances, although the essay was printed from a word processor, the writer has made changes with a pen to some of the words, for example, to correct a spelling mistake. The final version submitted to the lecturer, including those changes, is the version used for analysis.

Tasks and submission times

For the BEd students whose essays comprise the data, the degree programme was three years long. Each academic year was (and remains) divided into two semesters which means that in total students studied philosophy for six semesters. In the first year, students studied one module of philosophy per semester. In the second and third years, students studied two modules of philosophy per semester. This gives a total of 10 modules in philosophy over the course of the degree programme for these students.

All students represented in MICUPE were enrolled in the BEd degree programme and were, by choice, studying philosophy as their Liberal Arts component of that programme. To preserve the real identities of these students, they are referred to as Student 1, Student 2, Student 3 etc. The student labels identified here remain constant throughout the corpus. Therefore, Student 1 in the first semester is still Student 1 in the last semester.

In order to control the contextual factors as much as possible, all essays used in MICUPE were written for the same assessor. Therefore, the audience for each essay remains constant. That assessor was a lecturer in the Philosophy Department at Mary Immaculate College and had been part of that department for over 30 years. As already stated, a student taking philosophy to degree level undertakes 10 modules and the assessor/lecturer was responsible for three of these modules. Those modules were delivered in semester 1, semester 4 and semester 6. Semester 1 is the first semester of the first year. Semester 4 is the second semester of the second year. Semester 6 is the final semester of the degree programme. It must be noted that in Semesters 1 and 6 the essay work submitted by the student does not constitute their final grade in the module. In both of these semesters there was also a formal written exam at the end of the semester and these exams do not form part of the current corpus. In semester 4, the

students had to write four essays for the module. Together, these four essays constituted the final module grade and all the four essays are included in the corpus.

The modules

In semester 1, the module under study was entitled *Classical Greek Philosophy*. Students were given a choice of essay titles and had to write an essay on one of these titles. Those essay titles and the students who chose to write to those titles are shown below in Table 4.1. This set of essays is referred to as essay iteration 1.

Table 4.1 Essay titles for iteration 1

Title	**Student**
Socrates committed suicide. Discuss.	1, 6, 7, 11, 12, 17
Socrates was a clever orator. Discuss.	2
Socrates was a martyr for the truth. Discuss.	3, 4, 5, 8, 10, 13, 15
Evaluate Socrates' arguments in his own defence.	9, 16
Socrates was guilty as charged. Discuss.	14

In semester 4, the module under study was called *Metaphysics*. Instead of one longer essay, the students had to submit four shorter essay at four different points throughout the semester. With these shorter essays, students had no choice of title. The title was assigned to them by the lecturer. However, not all students were given the same titles. Students in different tutorial groups were given different titles at each of the four submission points. Furthermore, due to the reasons mentioned above, not all 17 students are represented in these essay iterations. However, the essays were submitted sequentially throughout the semester. These four essays are referred to as essay iterations 1, 2, 3 and 4.

Essay iteration 2 was submitted in semester 4 of the degree programme. The titles of the essays and the corresponding students for this essay iteration are shown below in Table 4.2.

Table 4.2 Essay titles for iteration 2

Title	**Student**
Explain Aquinas' account of being.	6, 8, 9, 10, 11, 12, 13, 15, 16, 17
Outline Parmenides' argument for the nature of 'is'.	1, 2, 3, 4, 5
What are the characteristics of being according to Parmenides and why?	7

Essay iteration 3 was submitted in semester 4, two weeks after essay iteration 2 was submitted. The titles of the essays and this corresponding students for essay iteration 3 are shown below in Table 4.3.

Table 4.3 Essay titles for iteration 3

Title	**Student**
Lonergan: the dynamic aspect of knowing.	6, 8, 10, 11, 12, 13, 14, 15, 16, 17
Lonergan's account of insight.	1, 2, 3, 4, 5
Explain the process from question to insight and explain inverse insight.	7

Essay iteration 4 was also submitted in semester 4 of the degree programme and this submission occurred two weeks after the submission of essay iteration 3. The titles of the essays and the corresponding students for essay iteration 4 are shown below in Table 4.4.

Table 4.4 Essay titles for iteration 4

Title	**Student**
Dramatic bias.	6, 8, 9, 10, 11, 12, 13, 14, 16, 17
Relate experiential and pure conjugates to primary and secondary qualities.	1, 2, 3, 4, 5
Lonergan on the intellectual component of common sense.	7

Essay iteration 5 was the final essay submitted in the *Metaphysics* module in semester 4 and was submitted two weeks after essay iteration 4. The titles of the essays and the corresponding students for essay iteration 5 are shown below in Table 4.5.

Table 4.5 Essay titles for iteration 5

Title	**Student**
Distinguish the act of reflection that leads to judgement from the act of enquiry that leads to insight.	6, 8, 9, 10, 13, 14, 15, 17
Explain the dialectic of spontaneity and intelligence as it operates an individual bias, group bias and the longer cycle.	1, 2, 4
Lonergan on things, conjugates, bodies and biological consciousness.	7

The module in semester 6, the final semester of the degree programme, was entitled the *Philosophy of God and Religion*. All 17 students in the corpus are represented in this semester, and this semester is referred to as essay iteration 6. In this essay iteration, the students were given a range of essay titles and were asked to pick one of these titles. The titles, and the students who chose those titles, are represented below in Table 4.6.

Table 4.6 Essay titles for iteration 6

Title	**Student**
Evaluate the argument in the first three of the five ways of Aquinas.	1, 3, 5, 6, 8, 9, 11, 12, 13, 14, 15, 16, 17
Critically consider Lonergan's epistemological approach to God.	2, 10
The significance of hope.	4,
The ontological mystery: is there a reality beyond thought?	7,

As a sample of student writing over the degree programme, MICUPE allows us to look at two different rates of change. We can look at change over the course of the full degree as we have representations from semesters 1, 4 and 6. This means we have an essay from the very beginning of the degree programme, an essay from the very end of the degree programme and a set of essays from the middle of the degree programme. The corpus, therefore, is in itself a representative of different rates of time intervals between essay iterations and because of this, inherent in the analysis of change in the present study is a shorter time-span and a longer time-span. We have four different essay submissions over the course of the semester 4. Unfortunately, in semester 4 not all students are represented in each of the four submissions. Some students are only represented in three of the four essays. Of the 17 students represented in the corpus, 9 have four essays represented in semester 4 while 8 students have three essays represented in semester 4.

Text length

Table 4.7 below outlines the mean text length for each of the six essay iterations.

Table 4.7 Mean text length by sub-corpus

Essay iteration	**Total tokens**	**Mean per text**	**Lowest per text**	**Highest per text**
Essay 1	30,090	1,770	1,146	2,635
Essay 2	14,685	917.8	420	1,640
Essay 3	20,894	1,305.9	820	1,984
Essay 4	17,323	1,154.9	762	2,236
Essay 5	13,071	1,005.5	587	1,750
Essay 6	37,511	2,206.5	1,456	5,596

Naming conventions

As already stated, each student has been given and number from 1 to 17 and these numbers remained constant for those students throughout the collection of the data. Within the data, there are six essay iterations represented and these have been referred to as iterations 1 to 6. Each essay had been graded prior to being included in the corpus. The grades available to the assessor were, ranging from highest to lowest:

A1, A2, B1, B2, B3, C1, C2, C3, D1, D2, F and *NG* (NG means No Grade)

In the modular system, the module grade is an amalgamation of all the assessment grades within that module. For a module, the *D1* and *D2* grades are considered to be compensating fail grades. This means that the student has not passed but should their grades be sufficient in other modules, they can progress to the next year. *F* and *NG* grades, however, are fail grades and a student with these grades cannot progress to the next year without repeating those assessments. However, non-pass grades in an individual module do not mean that the student must repeat the assessment. It is only the amalgamated module grade that can mean this.

The three factors of student, essay iteration and essay grade have being included in the naming of each text. They have been used in that order. For example, *S07.3.A2* as a text name means that the text was written by Student 7 in essay iteration 3 and it was graded at an A2 grade, and *S15.6.B1* was written by Student 15 in essay iteration 6 and was awarded a B1 grade.

This naming convention is consistent both in the examples referred to in the present research and in the filenames included in the concordances presented in Chapters 5 to 8 (see below for a description of concordances).

Word frequencies, concordancing and distribution plots

For word frequencies, concordancing and distribution plots, the software Wordsmith Tools™ was used to aid the analysis of the data in the majority of cases. For two figures, Figure 6.7 Word Sketch of *Socrates* and Figure 6.8 *Socrates* + believe, SketchEngine™ was used. The notions of word frequencies, concordancing and distribution plots are explained below.

In order to explain the notions of word frequencies, concordancing and distribution plots, it is best to work with an example. Below is an extract from text S02.3.A2. As we have seen, the name of the text means that this text was written by Student 2 in essay iteration 3 and was graded as an A2.

Example 4.1

> As mentioned, both a positive and negative element is required for an inverse insight. The first example gives is called 'Irrational (negative) numbers (positive)'. The example of the number between 0 and 1 is referred to as the 'Non-counTable (negative) multitude (positive)'. The positive aspect of the example regarding Newton's theory is that a body continues to move at a uniform rate is a straight line. The negation is that the continuance of the constant velocity depends not on the action of external force but on the absence of such action. And finally, while the negation in the Special Theory of Relativity is the word 'invariant', the positive object consists in the data in its mathematical expressions.
>
> (S02.3.A2)

Taking Example 4.1 as a corpus, we can generate a word frequency list for that corpus. An extract from that word frequency list is shown below in Figure 4.1.

Figure 4.1 Sample frequency list based on an extract from S02.3.A2 using Wordsmith Tools™

N	Word	Freq.	%	Texts	%
1	THE	16	13.45	1	100.00
2	IS	7	5.88	1	100.00
3	OF	6	5.04	1	100.00
4	POSITIVE	5	4.20	1	100.00
5	A	4	3.36	1	100.00
6	AND	3	2.52	1	100.00
7	EXAMPLE	3	2.52	1	100.00
8	IN	3	2.52	1	100.00
9	NEGATIVE	3	2.52	1	100.00
10	ACTION	2	1.68	1	100.00
11	AS	2	1.68	1	100.00
12	NEGATION	2	1.68	1	100.00
13	ON	2	1.68	1	100.00
14	THAT	2	1.68	1	100.00
15	THEORY	2	1.68	1	100.00
16	TO	2	1.68	1	100.00

The column labelled *N* contains a list of numbers starting with 1. These numbers indicate the rank order of the word shown in the adjacent *Word* column. In Figure 4.1, *THE* is the most frequent word in position one and *A* is the fifth most frequent word as it is in position 5. The next column to the right is labelled *Freq.* And this column indicates the raw frequency of the adjacent word, which in the case of *THE* indicates that this word occurs 16 times in the extract. The column labelled *Texts* contains the number of texts that the word appears in. In the current example, since we used one text, this number is 1 for all words. The two columns labelled *%* are indicative of the raw frequency and number of texts for that particular word expressed as a percentage of the corpus. In the case of *THE*, this means that 13.45% of the words are *THE* and it appears in 100% of the texts. The meaning that can be attributed to the frequency lists is discussed in detail in Chapter 5.

In the current research, the words represented in the frequency list are referred to as *types*. In example 4.1, the type *negative* is in position 9 and the type *negation* is in position 12. They have a combined frequency of 5. In the current research, each single occurrence is referred to as a *token*, and therefore *negative* and *negation* account for five tokens.

The current research does not use the term of *lemma*.

> A lemma is a set of related words that consist of the stem form and inflected forms that are all the same part of speech. So, *approach*, *approaches*, *approached*, *approaching* would all be members of the same lemma because they all have the same stem, include only the stem and inflected forms, and are all verbs.
>
> (Nation and Meara, 2002)

As will be shown in Chapter 6, differences in inflection indicate differences in meaning and use. For this reason, the current research does not lemmatise when counting types.

Again, taking Example 4.1 as a corpus, we can concordance any of the types that are present within the corpus. The type *example* has a frequency of 3 in the extract, as seen in Figure 4.1 above. A sample concordance for this type is shown below in Figure 4.2.

Figure 4.2 Concordance of *example* using Wordsmith Tools™

N	Concordance	File
1	for an inverse insight. The first example gives is called 'Irrational	S02.3.A2 (extract).
2	(negative) numbers (positive)'. The example of the number between 0	S02.3.A2 (extract).
3	The positive aspect of the example regarding Newton's theory	S02.3.A2 (extract).

Figure 4.2 shows a concordance for the type *example*. This means that all instances of this type in the corpus are displayed in the *Concordance* column along with extracts from the text that occur before and after this type in the original setting. To the right of the concordance is the *File* column and this tells us which file the concordance line came from. On the left, there is a column labelled *N*. These numbers are not indicative of frequency or importance, but the current research does use them as line numbers for reference where appropriate. For example, in line 3, the writer refers to *Newton's theory*.

Based on the concordance of a type, we can create a distribution plot for that type. The distribution plot tells us in which parts of the text that type occurs in. The distribution plot for *example* from the concordance shown in Figure 4.2 is given in Figure 4.3 below.

Figure 4.3 Distribution plot for *example* using Wordsmith Tools™

N	File	Words	Hits	per 1,000	Plot
1	S02.3.A2 (extract).txt	117	3	25.64	

Figure 4.3 shows the distribution plot for the type *example*. Moving from left to right, the columns indicate a line reference number (*N*), the relevant file name (*File*), the total number of tokens in that file (*Words*), how many of those tokens are the type *example* (*Hits*), and the number of occurrences in relation to the number of tokens expressed as a number out of 1,000 (*per 1,000*). The column labelled *Plot* then indicates the relative positions within the text of the type example. These positions are marked with a red band (|). In Figure 4.3, the type example occurs 3 times in the text S02.3.A2 out of a total of 117 tokens and those occurrences are all in the first half of the text.

4.3 Building an analytical framework

The research questions being addressed in the study, as outlined in Chapter 1, are:

- Main research question:

  Is there a patterned, linear change in the student writing over time?

- Sub-questions:
    1. Is each text unique?
    2. How do writers respond to genre, task and audience?
    3. Is the response to genre, task and audience standardised across writer and/or situation?

In order to investigate these questions, an analytical framework was constructed based on the concepts of writer, context of situation and text as explored in the theoretical framework in Chapter 3 and the nature of the empirical data as outlined in the present chapter. There can be two different approaches to analysing a corpus comprising of a collection of texts: analysing whole texts or analysing a corpus through normalising phenomena.

The analysis, which is presented in the next four chapters, is summarised below.

Chapter 5 Type frequency and distribution in the corpus

The frequency list for MICUPE is used as the starting point for investigating the corpus, showing that 14 types appear in all 94 essays and that the philosophical content is carried the whole way down the list. Similarities and differences between the six essay iterations are investigated in line with the notion of change as outlined in the theoretical framework. Frequency is compared to distribution across the texts and while there is a discernible pattern in frequency and distribution, there are some outliers. These outliers include function words such as *the, in* and *of,* the pronouns *I, we* and *you,* and names of the authors of the relevant primary texts. There is a set of words that are used only once in the corpus, and these words in conjunction with the frequency lists are used to argue for the uniqueness of each text. The outliers in relation to frequency and distribution form the basis of the analysis in Chapter 6.

Chapter 6 Outliers in frequency and distribution

The outliers as identified in Chapter 5 are investigated in further detail in Chapter 6. The function words occurring in all texts, in particular *the,* are seen as indicative of a response to the concerns of genre. The pronouns *I, we* and *you,* due to the interactive nature, are seen as a response to the concerns of audience. The names of the authors of the relevant primary texts are seen as a response to a task. Together, these features are seen as part of the response by the writer to the dynamic context of situation. In line with the theoretical framework, these three features are shown to occur uniquely in each text and are representative of a unique, instantial reaction to the concerns of genre, audience and task. This chapter further establishes the uniqueness of each text, investigates how the writers respond to the concerns of genre, audience and task and shows that each response to these concerns is unique.

Chapter 7 Single-use and multi-use types

Chapter 7 further investigates the uniqueness of each text by examining the distribution of types across the essay iterations and across the texts themselves. This chapter uses the notion of the building of the corpus, and its composition at various times in its construction, to highlight the relativity of uniqueness through single-use and multi-use types. The instantiality of choice of types is empirically outlined and used to reinforce the concepts outlined in the theoretical framework. This chapter demonstrates that the responses to genre, audience and task are not standardised across writer and are not standardised across situation.

Chapter 8 Change as a chaotic dynamical system

Chapter 8 takes the uniqueness and apparent randomness established in Chapters 5, 6 and 7 and shows that one way of understanding the non-linear change that occurs is by viewing this change as a chaotic dynamical system. The justification for claiming that this change is a chaotic dynamical system is shown empirically by establishing that each writer has a unique starting point, the essay iterations do not display a linear change in the features identified in Chapter 5 and analysed in Chapter 6 and that what seems a similar cause has in fact a disproportionate effect.

We will now begin the analysis with an investigation of the frequency and distribution of types in MICUPE.

Chapter 5 Type frequency and distribution in the corpus

5.1 Introduction

The orthodox route for corpus analysis has been to begin by looking at word frequency when investigating a corpus. 'Word lists by themselves are often best seen as a first approach to a corpus. It is by processing the words and looking at the most frequent of them that one can get a rough idea of the kinds of topics being explored, the wealth of vocabulary being used' (Scott, 2010: 148). Evison (2010) points to use of frequency data as a basic technique in corpus studies, of value to the researcher both in itself and as the basis for comparison of the data with other frequency lists generated from other corpora. With corpora that are more focused in their design, comparing frequency lists, either in terms of rank order or in terms of keyness, is often used to reveal an identifying signature of frequency patterns that in some way distinguishes one context of language use from another (Hunston, 2002; McEnery, Xiao & Tono, 2006). Although frequency lists are advocated as a starting point for research, intended to be explored further through concordancing and detailed analysis of the corpus texts, there are some questions relating to their relevance for some forms of corpus research.

In MICUPE, we have a very interesting textual situation. Within the corpus as a whole, the audience for each written text remains the same, and these texts are written by the same 17 students. Furthermore, all the texts come from the same discipline. Within the overall corpus, there are 6 sub-corpora which relate to the essay assessments of a degree programme. Within these essay assessments, not only are the audience and discipline the same, so is the module title. Therefore, many of the factors that are considered an influence on a text, such as genre, audience, individual writer and task, are controlled in MICUPE (see Chapter 4 for details on the composition of MICUPE). This allows for some in-depth analysis of frequencies within the corpus.

For all 17 students in the corpus, there is a natural standardisation of the factors that are thought to influence language. All students are roughly the same age at each point. They are writing their essays in response to similar tasks on similar topics at any given time. Therefore, the genre that they are utilising is, in an objective sense, uniform. Furthermore, the fact that the audience in each case is the same lecturer, the lecturer is also constant.

Taking into account the composition of MICUPE (see Chapter 4 for more details), the benefits of and meanings attributed to a frequency list can be problematic. The first problem is that frequency does not always correlate with importance, either within texts or within the corpus. A higher frequency for a word does not imply that that word is more important to the set of texts than a word with a lower frequency since each word used was deemed, by the writer, necessary at that particular point. The second problem is that there is a homogenising effect on frequency when texts are aggregated together which often does not reflect what happens in the sub-corpora or indeed individual texts. The third problem is that the frequency list of all tokens in the corpus counted by type, dependent as it is on the instantial choices (see Carter and McCarthy, 1988) made by the writers, is unique to the particular corpus it represents.

This chapter investigates the frequency list of the MICUPE corpus with a view to examining:

a) various sections of the frequency list;

b) comparisons between the six sub-corpora that constitution MICUPE;

c) raw frequency correlated with number of texts.

The purpose of this is, on the one hand, to gain an insight into MICUPE and the words that comprise it, and on the other hand, identify features within the corpus that may be worthy of further investigation. Furthermore, this also allows us to investigate one of the research questions as outlined in Chapter 1, namely the sub-question concerning the uniqueness of each text.

5.2 Frequency lists

Introduction

Of the total of 134,289 tokens in MICUPE, there are 7,493 types in the whole corpus. The concepts of type and token were introduced in Chapter 4. To re-cap on this, a token is a single occurrence of any word and a type is a word counted once regardless of frequency. Example 5.1 below is used to illustrate this.

Example 5.1

> Socrates also stood charged with impiety; that he refused to recognise the official Gods of the state and that he introduced new Gods. (S14.1.B1)

In example 5.1, there are 23 tokens and within those tokens there are 19 types. This is because the types *that*, *he*, *the* and *Gods* appear twice each in the sentence.

Although a word list can be a useful introduction to these types, a word list negates some meaning from the original context. We can say for certain that the use of two different types indicates two different meanings. Even when the two different types are related to each other, the meaning inherent in both is different. However, this does not mean that the use of similar types indicates a similar meaning. In fact, meaning is dependent on context or cotext (Sinclair, 1991) and that context or cotext varies for each individual use.

As with any wordlist, the items within it carry some meaning, but they do not constitute a closed, static meaning unit. A fuller meaning is dependent on the context of the word use. Which words are used in which context gives us a unique meaning. No word has the same instantiated meaning in context of use in any two situations – for example, *he* could refer to Socrates a number of times, but this is a referent, not a meaning.

Example 5.2

> ‘**He** tries to explain that it is those people that have given such false impressions of him and that in fact **he** takes "no interest in these things "(19D).’ (S01.1.B3)

In the sentence in Example 5.2, the word *he* is used twice. In both these instances, *he* refers to Socrates. Although Socrates is one person, that person at one stage *tries to explain* and at another stage *takes no interest*. *He*, therefore, refers to Socrates in both instances, yet Socrates is in the text to generate a different meaning as indicated by its surrounding context. Furthermore, the first *he* is a part of the sentence that summarises a reading in the author's own words and the second *he*, along with *takes*, is a part of the sentence that introduces a direct quote.

A variation in meaning of similar types is not limited to pronouns. A function word may perform the same grammatical function in a variety of instances; however, the inherent meaning is unique to a particular instance. Even a word like *the* can have different meanings.

Example 5.3

> 'We assume everything will go out of existence at the same time this may not be i.e. just because one daffodil stops being doesn't mean that all **the** daffodils in **the** world cease to exist as but **the** earth from which they sprang would continue to exist.' (S17.6.B3)

In example 5.3, the type *the* is present three times in the same sentence. Grammatically, each instance of *the* fulfils a determining function, yet the nouns that this word determines differ in the three instances. *Daffodils*, *world* and *earth* are not the same and therefore, although the type *the* is repeated, the inherent meaning in each instance is intrinsically linked to other words in its context and therefore each instance of use points to a different referent.

Interestingly, in Example 5.3, the types *world* and *earth* both occur. In this case, they are not synonymous. However, in other cases, these two types may be synonymous, yet, despite this they would still encode different meanings from each other.

Accepting that meaning is particular to its instantial use does not negate the usefulness of wordlists. It is certainly of note that 17 students create 94 texts and all the meanings within those texts by using only 7,493 types. Therefore, the writers choose the appropriate words, in their instantial estimation, to best express the meaning they wish to communicate to their reader. Those word choices, at times, overlap with previous or future choices by the same student and with the word choices of other students. The overlap gives an indication of a shared concern, at least in relation to non-function

words but also possibly including function words. That concern may be in relation to the topic, content, audience, perceived expectations etc. As we will see both in the current chapter and subsequent chapters, the effects and strength of those concerns are not uniform and it may be a simplification to attribute any lexical realisation of attempted communication by the writer as a linear result of a singular cause or concern.

MICUPE frequency List

Table 5.1 The 30 most frequent words in MICUPE

N	Word	Freq.	Texts		N	Word	Freq.	Texts
1	THE	7,919	94		**16**	ARE	945	92
2	OF	4,754	94		**17**	WE	944	87
3	TO	4,416	94		**18**	HIS	877	75
4	IS	3,760	94		**19**	I	827	83
5	AND	3,112	94		**20**	BEING	786	74
6	THAT	2,800	94		**21**	AN	779	91
7	A	2,640	94		**22**	BY	751	93
8	IN	2,568	94		**23**	HAVE	729	94
9	IT	1,715	94		**24**	WAS	716	79
10	HE	1,625	83		**25**	ONE	698	91
11	BE	1,622	94		**26**	THERE	693	87
12	THIS	1,573	92		**27**	SOCRATES	691	18
13	AS	1,403	94		**28**	CAN	659	90
14	NOT	1,232	94		**29**	ON	658	93
15	FOR	1,131	94		**30**	WHICH	623	89

Table 5.1 shows that 30 most frequent words in the corpus. The most striking aspect of the frequency list is that, apart from the word *Socrates*, there seems to be no indication in the 30 most frequent words that this list comes from a set of philosophy essays. The words are, in the main, function words that could appear in most written amalgamated corpora.

What is also striking is that even among the most frequent words, the number of words occurring in all the texts is quite limited. In fact, only 14 words appear in all 94 essay texts. These words, in frequency order, are:

the, of, to, is, and, that, a, in, it, be, as, not, for, have[4].

These 14 words are function words. No content word appears in all 94 texts. In fact, when sorting not according to raw frequency but according to number of texts a word appears in, the most common lexical items appear in 74 individual texts. In fact, there are two lexical items that appear in 74 texts. These are represented in Table 5.2 below.

Table 5.2 Lexical items appearing in most texts

Word	**Freq.**	**%**	**Texts**	**%**
KNOW	287	0.21	74	78.72
TIME	246	0.18	74	78.72

The raw frequency value for a lexical item simply illustrates how often that item appears in the corpus. As our corpus is a collection of essay texts, the distribution of a lexical item across texts also becomes important when looking at student patterns. An example of a discordance between the frequency of a word and the number of texts that word appears in is the word *he*. In raw frequency terms, this word is the 10th most common word in the corpus. However *he* appears in 83 out of the 94 essays. This creates a conflict of meaning. On the one hand, it is the 10th most common word in the corpus appearing 1,625 times, which, given the overall size of the corpus, would suggest that the type *he* appears more frequently than once every hundred words. On the other hand, even though it appears more than every hundred words on average, 11 texts were written without this type. Furthermore, *have* is in position 23 in the overall frequency list and on average appears slightly more frequently than one in every 200 words. It is just over half as common as *he*. However *have* appears in all 94 texts. This means that no student chose to write an essay without the word *have* while 11 students

[4] *have* could be both lexical or functional

chose to write an essay without the word *he*, but the instances of *he* in the corpus are more than double those of *have*.

The use of the words *know* and *time* are in contrast to the word *Socrates* as seen in Table 5.1 above. This word *Socrates* is in position 27 in relation to raw frequency in the corpus, and yet it only appears in 18 texts. The words *know* and *time* are less than 50% as frequent in the context of the overall corpus and raw frequency, and would therefore appear not to be as important if we were to rely solely on a frequency count. On the other hand, these words could be seen to be more important for philosophy as they span a greater range of texts. The commonality of the word *Socrates* is not surprising given that he is the main topic for discussion in all the essays in essay iteration 1. Having said that, the use of the word *Socrates* is not uniform across all students and we will investigate this further in the following chapter. Furthermore, the topics of the other 5 essays do not populate the same area of the distribution graph, suggesting that the unique distribution of *Socrates* is not simply attributable to it being connected with a task as the other tasks do not generate a similar distribution.

Comparisons with BAWE

For an academic written corpus, the words *we* and *I* seem to be more frequent than expected in positions 17 and 19 respectively. In BAWE (The British Academic Written Corpus) containing '2761 pieces of proficient assessed student writing' (BAWE, 2011), which is just short of 7 million words, *we* appears in position 52 and *I* in position 48. The relative positions in BAWE for each word in the 30 most frequent MICUPE words are presented in Table 5.3 below. The reasons for differences between the current data and BAWE are not the focus of the current study.

Prepositions such as *by* and *on* are quite frequent in MICUPE, being at positions 22 and 29 respectively in the raw frequency list. These two prepositions are interesting, however, as they both appear in 93 out of the 94 texts. This means that in one instance for each of these words, a text was created without the writer needing, or choosing, to include it. The text without *by* is S11.4c.B2 and the highest number of *by* in one text is 48; the text without *on* is S07.4a.A1. At a cursory glance, the grades received by these essays are quite high, with the A1 grade being unsurpassable. In addition, those students

did not write any other texts without those words. Furthermore, these two prepositions appear further up the frequency list in BAWE as compared to MICUPE.

Table 5.3 MICUPE wordlist of 30 most frequent words with BAWE frequency positions

N	Word	BAWE Position		N	Word	BAWE Position
1	THE	1		**16**	ARE	13
2	OF	2		**17**	WE	52
3	TO	4		**18**	HIS	45
4	IS	7		**19**	I	48
5	AND	3		**20**	BEING	72
6	THAT	8		**21**	AN	23
7	A	6		**22**	BY	16
8	IN	5		**23**	HAVE	24
9	IT	15		**24**	WAS	18
10	HE	64		**25**	ONE	41
11	BE	11		**26**	THERE	40
12	THIS	19		**27**	SOCRATES	N/A
13	AS	10		**28**	CAN	25
14	NOT	20		**29**	ON	17
15	FOR	12		**30**	WHICH	22

Further sections of the MICUPE frequency list

While the order and frequency counts of these words are unique to MICUPE, the set of lexical items in Table 5.1 does not mark the corpus as being comprised of undergraduate philosophical essays. However, we cannot say that these function words are non-philosophical. They are used in many domains, one of which is philosophy. The philosophy texts could not have been created without these words. It is likely, though, that the relative distribution is influenced in some way by the context. As we move down through the frequency count (see Tables 5.4 and 5.5 below), the philosophical nature of the corpus becomes more apparent, although the undergraduate essay nature of the corpus is less apparent.

The differences between written genres can often be realised in frequency lists. Evison (2010: 125-126) advocates comparing frequency lists as a way of gaining insights into a corpus. This can, undoubtedly, be useful for signalling some similarities and some differences between corpora. However, such insights are solely created in relation to another corpus, and miss out on features pertaining to the original corpus (it must be noted here that Evison (2010) does not advocate comparing frequency lists as the sole method of exploring a corpus, but as a common and useful method). The appearance or absence of words, or features, in another corpus in no way gives an indication of their importance to the original corpus. For example, *the* is the most common word in MICUPE and in BAWE. Sharing a similar position in both frequency lists would probably result in *the* scoring a low keyness in the comparison of both corpora. However, it is extremely difficult to imagine a philosophical text of the length of the texts in MICUPE without the word *the*. In addition, it would be difficult to make a claim that certain words are solely philosophical. Some words may be more likely to appear in a philosophy essay but that does not make them exclusively philosophical in nature. It does, however, suggest that as a meaning-making exercise in a particular context, some words have a greater appeal to the writer than other words.

Table 5.4 MICUPE frequency list from position 31 to position 50

N	**Word**	**Freq.**	**Texts**		**N**	**Word**	**Freq.**	**Texts**
31	WITH	622	90		**41**	AQUINAS	493	26
32	WOULD	604	90		**42**	THEY	492	92
33	OR	588	90		**43**	EXISTENCE	461	39
34	IF	587	93		**44**	AT	452	85
35	FROM	536	91		**45**	YOU	451	63
36	GOD	534	42		**46**	HAS	438	82
37	BUT	511	87		**47**	SOMETHING	427	71
38	WHAT	501	91		**48**	NO	396	84
39	INSIGHT	499	43		**49**	WHEN	384	83
40	ALL	493	89		**50**	SO	383	87

Taking the next 20 words in a raw frequency list, we see a more philosophical nature in the lexical items chosen by the students. Types such as *god*, *insight*, *Aquinas* and *existence* are lexical choices influenced in some way by the subject matter under discussion and have a philosophical significance. In a broad sense, it can be argued that these types, although not the most common in the corpus, are more philosophical in nature than those ranked higher in the frequency list (Figure 5.1). Although this section of the frequency list may contain more types linked to philosophy, as a set of types, they are not predominantly exclusively philosophical in nature. However, this does not mean that other words that appear have not, in some way, been influenced by the subject under discussion. In addition, the word *if*, at position 34 in MICUPE and 65 in BAWE, although a function word, indicates hypothetical contexts which are often used in philosophical arguments. *If* appears in 93 out of the 94 essays. The essay without *if* is S02.5.A2. Text S02.5.A2 is from essay iteration 5 and there are 13 texts in total in essay iteration 5. In each of the other 12, *if* is present, ranging from one occurrence in one text to 12 occurrences in another. Furthermore, in each of their other five essays in MICUPE, Student 2 uses the word *if*. They use the word *if* 8, 8, 4,2, 20 times in essay iterations 1, 2, 3, 4 and 6 respectively.

Moving further down the frequency list, the words that can be considered philosophical in nature appear more frequently. It is not that these words are exclusively philosophical, but are lexically matched to the content. In the section of the list below, 100 to 130, the list is more lexically dense and gives a sense of some of the content contained in the essays.

This section of the frequency list displayed in Table 5.5 shows an affinity with the tasks given to the students. Words such as *argument*, *death*, *pain*, *insights*, *questions*, *knowledge* and *judgement* in themselves give an insight into the subject matter of some of the texts. Similar to the words that appear in other sections of the frequency list, these words are by no means exclusively philosophical in nature. If we were to accept a linear-type cause and effect scenario, we could attribute the appearance of these words completely to the task set by the lecturer.

Table 5.5 MICUPE frequency list from position 101 to position 130

N	Word	Freq.	Texts		N	Word	Freq.	Texts
101	BETWEEN	175	66		**116**	DIFFERENT	155	54
102	LIFE	175	45		**117**	THING	155	58
103	PEOPLE	175	49		**118**	OWN	153	66
104	ARGUMENT	171	40		**119**	INSIGHTS	152	28
105	DEATH	171	24		**120**	EVERYTHING	150	41
106	SENSE	171	58		**121**	QUESTIONS	150	42
107	ANOTHER	168	69		**122**	UP	150	66
108	SAME	167	64		**123**	PERSON	149	46
109	MOTION	165	19		**124**	THAN	149	60
110	MORE	163	77		**125**	KNOWLEDGE	148	52
111	FACT	160	66		**126**	PROCESS	146	36
112	ANY	158	62		**127**	MAN	145	46
113	HAD	157	59		**128**	MANY	145	62
114	THROUGH	157	65		**129**	JUDGEMENT	144	14
115	TWO	156	63		**130**	LIKE	142	61

The section of the frequency list shown in Table 5.5 gives a sense of philosophy, and gives more of a sense than the sections represented in Tables 5.1 and 5.4. This raises a question concerning what can be deemed important for writing philosophy. Although the types contained in Table 5.1 are more frequent in the corpus than those in Tables 5.4 and 5.5, they cannot be argued to be seen as more integral to the meaning of an essay than the types in Tables 5.4 and 5.5. The philosophical meaning is, in a sense, carried through a combination of all the words used in the corpus. In any essay, the writer deemed every word used as useful, or necessary, to convey the meaning they wished to communicate. In that context, a word that also appears in a multitude of other instances cannot be deemed more integral than a word that does not. In fact, the relevance of frequency to either the writer or the reader is questionable in what is essentially a piece of writing for assessment. Frequency does not determine philosophical appropriateness. Furthermore, appropriateness is context-bound, not bound by frequency, and the context can be instantial and unique.

The end section of the frequency list for MICUPE consists of words that appear only once in the corpus. There are 3,011 of these words and these will be examined in more detail in Chapter 7.

Frequency in the sub-corpora

In this chapter, we have taken MICUPE as a fixed entity, not recognising that it has been constructed from texts written at separate points in time. At a particular point in time, a set of students create a set of essays, in this case essay iteration 1. The same set of students creates more sets of essays at other particular times, essay iterations 2-6. It is the amalgamation of all these essays created at a particular point in time that come together to create MICUPE. The frequency lists in Tables 5.1, 5.3 and 5.4 above are then representative of the totality. This section will examine how the totality as seen above relates to six sub-corpora, as delineated by essay iteration, that comprise the full extent of MICUPE.

On examining Table 5.6, it can be seen that there are differences between the six sub-corpora. There are differences in relation to the frequency order in which words occur. These differences have an effect on the overall frequency list for MICUPE. For example, *Socrates* appears in position seven in essay iteration 1 and is not in the top 30 in any of the other 5 essay iterations. It is in position 27 in the overall corpus. *He* is in position 3 in essay iteration 1, position 23 in essay iteration 2 and position 24 in essay iteration 6. This gives a position of 10 in the overall corpus.

In each of the six essay iterations, the name of a philosopher is in the top 30 words. *Socrates*, *Parmenides*, *Aquinas* and *Lonergan* appear in essay iterations 1-6. However, their relative frequencies vary and only *Socrates* is in the top 30 words in the overall corpus. *Parmenides* is in position 22 in essay iteration 2 but ends up in position 156 in the overall corpus. *Aquinas* is in position 24 in essay iteration 2 and position 14 in essay iteration 6 and is in position 42 of the overall corpus. *Lonergan* is in position 25 in essay iteration 3, 29 in essay iteration 4 and 24 in essay iteration 5 and is in position 59 in the overall corpus.

Table 5.6 The 30 most frequent words in the sub-corpora of MICUPE

	Essay 1	**Essay 2**	**Essay 3**	**Essay 4**	**Essay 5**	**Essay 6**
1	THE	THE	THE	THE	THE	THE
2	TO	IS	OF	OF	OF	OF
3	HE	OF	TO	TO	TO	TO
4	OF	TO	IS	AND	IS	IS
5	AND	THAT	A	IS	A	AND
6	THAT	AND	AND	A	AND	THAT
7	SOCRATES	IT	IN	IN	THAT	IN
8	HIS	BEING	THAT	THAT	IN	A
9	IN	A	IT	AS	I	BE
10	IS	IN	THIS	THIS	THIS	IT
11	A	BE	WE	ARE	INSIGHT	THIS
12	WAS	AS	BE	BE	IT	GOD
13	THIS	NOT	INSIGHT	IT	ARE	AS
14	NOT	ESSENCE	ONE	NOT	BE	AQUINAS
15	AS	EXISTENCE	ARE	WE	JUDGEMENT	NOT
16	FOR	THIS	AS	AN	AS	FOR
17	TRUTH	THERE	AN	OUR	WE	THERE
18	IT	FOR	FOR	FOR	HAVE	BEING
19	I	WE	I	CAN	ONE	EXISTENCE
20	BY	ARE	NOT	INSIGHT	FOR	CAUSE
21	BE	SOMETHING	CAN	WITH	AN	FIRST
22	WOULD	PARMENIDES	AT	OR	ON	WE
23	HIM	HE	THERE	WHICH	CONDITIONS	AN
24	HAVE	AQUINAS	YOU	INSIGHTS	LONERGAN	HE
25	DEATH	WHAT	LONERGAN	FROM	CAN	BY
26	WITH	CAN	WHEN	MAY	NOT	ARE
27	JURY	EXIST	ON	ONE	QUESTIONS	ON
28	ON	WAS	IF	THEY	IF	WHICH
29	BUT	CANNOT	OR	LONERGAN	KNOW	SOMETHING
30	SO	OR	BY	BY	OR	IF

Interestingly, *the* is the most frequent word in each essay iteration, and also the most frequent word in the corpus. *I* features in the top 30 words in essay iterations 1, 3 and 5

in position and 19, 19 and 9 respectively. Because of this, *I* appears in position 19 in the overall corpus, despite not being in the top 30 words in essay iterations 2, 4 and 6.

We is in the top 30 words in essay iterations 2, 3, 4, 5 and 6 in positions ranging from 11 in essay iteration 3 to position 20 to in essay iteration 6, giving it position 17 in the overall corpus. *We* is not in the top 30 words in semester 1. Despite the frequency of the pronouns *I*, *we* and *he*, the pronoun *you* is not as common, as *you* appears in position 24 in essay iteration 3 but not in the top 30 words of any other essay. If we see essays as having been constructed through a series of instantial choices which are influenced in some way by the contextual factors surrounding that essay, the use of these pronouns gives rise to some anomalies. In Chapter 6, we will investigate these pronouns to determine whether the changes in their use are attributable to specific influences and whether the changes in use are uniform across all students represented in the corpus.

The meaning of frequency lists

The examination of frequency lists for MICUPE as a whole, its sub-corpora and comparisons to BAWE point to the homogenising effect of a corpus. There is a difference between the top 30 most frequent types in the sub-corpora and the top 30 most frequent types in MICUPE. The sub-corpora have more lexical items that carry content than the overall frequency list. Words such as *truth*, *insight*, *existence*, *god*, *death*, and *cause* give a sense of the subject matter being discussed. These words are common in their respective semesters, yet not common, relatively speaking, in the overall corpus. It seems that if the collection of essays is controlled in terms of time and topic, content words are more likely to be closer to the top of the frequency list. On the other hand, if the corpus is made up of more disparate topics, even when the discipline and assessor are controlled, more function type words are pushed towards the top of the list. This, then, creates a conflict because when the content words come towards the top it would seem that generalisability suffers because such words are so dependent on the context at a particular point in time, yet when the function words gravitate towards the top of the list a fractioning of representiveness occurs as these words may not have been at the top either in individual essays or within the sub-corpora.

Regardless of whether the words are in the top 30 or the top 100 most frequent, or even whether they only appear once in the corpus, they are in the corpus because a particular writer at a particular point in time deemed it appropriate to use the said words in order to communicate a meaning to the reader. In one way, the fact that other writers use these words in the same task is not important as the writer is writing on their own and their text exists on its own. What the other writers are doing has no bearing on what one particular writer is doing. However, all the writers are writing in similar contexts and similar situations and they have similar influences. It is possible that these similar influences dictate the multi-use of words. We must, however, remember that the multi-use of words does not necessarily mean the multi-use of meanings. Similar words can be used to create unique meanings. At any given point in any of the texts, the writer is making choices based on their perception of optimal communication. Those choices, while subjected to varying degrees of influence, are in reality instantial decisions made by the writer in order to communicate with the reader. It is more a case of 'best fit' than 'must fit'. At such points, the frequency of a word is not a consideration for the writer. It may be an influence, since the writer may be aware that such a word is appropriate to research texts due to encountering the word frequently in their readings. Even if this were the case, where frequency of words in the genre can be argued to be an influence on a writer in a roundabout way, it still remains just an influence, not a determining factor. Influences work in different ways on different people. There is also a set of competing influences. There are also co-interacting influences.

5.3 Distribution of types

Introduction

Comparing wordlists generated in different corpora can be a way of finding out what is unique about a corpus. This, however, lets us know what is unique about the corpus from an external perspective, but does not tell us what is unique about individual texts or smaller groups of texts within a corpus. Plotting frequency against textual distribution allows us to see such uniqueness. Furthermore, comparing wordlists from different corpora, due to the nature of an external perspective, does not allow us to define either corpus as it stands, but only to define it in relation to another corpus.

For some words, there is a disparity between the raw frequency and distribution of that word across the texts. In a specialist corpus such as MICUPE, distribution across the texts is as meaningful as raw frequency. In Table 5.1, it can be seen that the word *Socrates* appears 691 times in the corpus. It is more common than the words *can*, *on* and *which*, which occur 659, 658 and 623 times respectively. The 691 occurrences of *Socrates* are distributed across 18 texts while *can, on* and *which* appear in 90, 93 and 89 different texts respectively. Such disparity between frequency and distribution is not limited to the word *Socrates*.

Table 5.7 Distribution of *argument, death* and *sense*

N	Word	Freq.	%	Texts	%
104	ARGUMENT	171	0.13	40	42.55
105	DEATH	171	0.13	24	25.53
106	SENSE	171	0.13	58	61.7

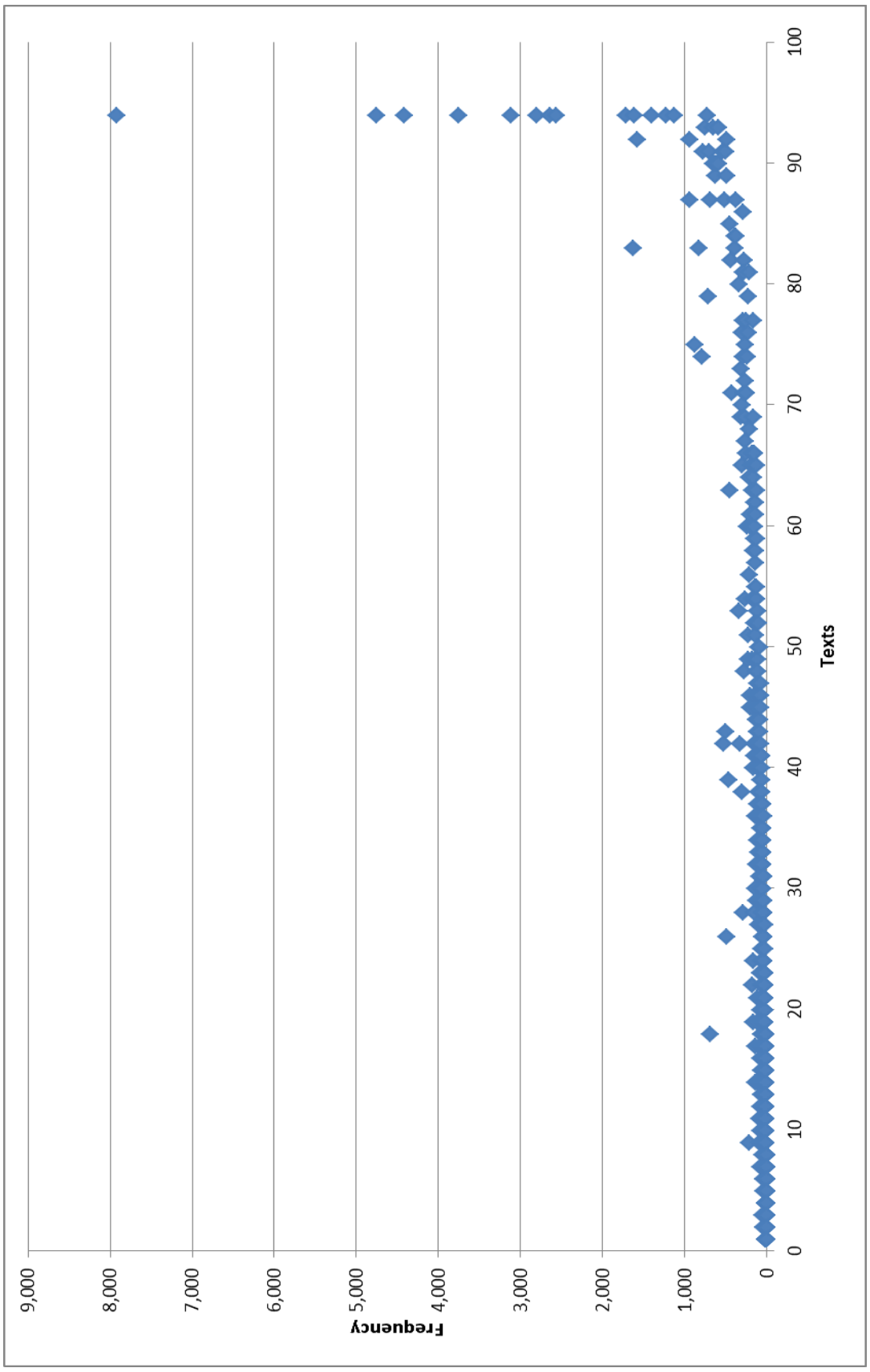
The three types in Table 5.7, *argument, death* and *sense*, appear in the corpus 171 times each. There are 103 lexical items more frequent that these items. Although these words have the same raw frequency in the same corpus, their distribution is very different in that they appear in 40, 24 and 58 texts respectively. This means that *sense* appears in over twice as many texts as *death*, yet has the same total word-count. In relation to the types *Socrates*, *argument*, *death* and *sense*, there is a discrepancy between the frequency and textual distribution of these lexical items. Such a discrepancy is not limited to these types.

The following section examines word-frequency and text distribution in the corpus as a whole.

Word Frequency and distribution in MICUPE

Figure 5.1 below shows the text distribution of the 7,493 types in MICUPE plotted against their frequency value. This is followed by a discussion of Figure 5.1.

Figure 5.1 Word frequency and text distribution in MICUPE



Discussion

There are 7,493 types in 94 texts. 14 types appear in each of the 94 texts in the corpus with varying frequencies, ranging from *have* at 729 occurrences to *the* at 7,919 occurrences. The graph in Figure 5.1 clearly shows 9 types that are more frequent than *have* in terms of overall frequency but whose distribution across texts is less well-spread. The 9 types more frequent in a raw count than *have* but less frequent in terms of amount of texts they appear in are *he*, *this*, *are*, *we*, *his*, *I*, *being*, *an* and *by*.

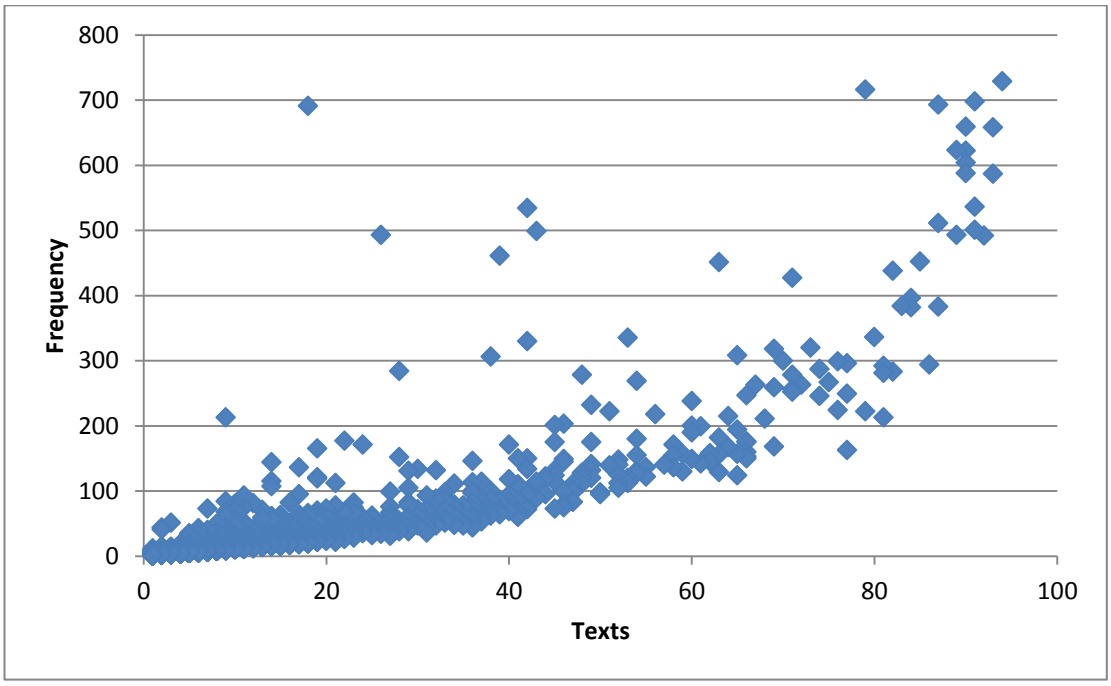
4,096 types appear in more than 1 text. This equates to slightly less than 55% of all the types in the corpus. 3,398 types are limited to one text only. This equates to slightly more than 45% of the types in the corpus. 3,011 of these types appear once only. The point on the graph (1,1), indicating one occurrence and one text, represents 3,011 types. 387 types are limited to one text but appear more than once in that text. The fact that 3,398 types appear in one text only and 3,011 of these appear only once, points to the uniqueness of each text. In the theoretical framework within the present research, it was posited that this uniqueness within the texts is a result of a set of instantial decisions influenced by a unique realisation of a complex interaction between the features in the context of situation, the writer and the text. Single-use types are elaborated on below and along with multi-use types will be examined in greater detail in Chapter 7. Furthermore, types that are shared across a range of texts do not necessarily indicate a similarity in use, and this is investigated further in Chapter 6.

The Correlation Coefficient (Pearson's r) calculated on the basis of correlation between frequency and number of texts is 0.534. The Correlation Coefficient is a number between -1 and 1 which gives an indication of the strength of the relationship between two sets of values, in this case the relationship between the word frequency and number of texts. The current value of 0.534 indicates a medium to weak positive correlation between frequency and text distribution. When we look at the correlation coefficient, it is the high frequency of words that appear in all texts that negatively influence the correlation coefficient. If we omit the 14 words that appear in 94 texts, and calculate the correlation coefficient using the remaining 7,479 types and their distribution, the correlation coefficient increases to 0.779. This tells us that some of those 14 words are in fact outliers. Visually, it would seem that 13 out of those 14 words do not fit the

curve, and this is borne out when we include *have* and exclude the other 13 words, the coefficient increases very slightly to 0.781.

Overall, the curve created by the individual points is an exponential and is well-defined. There is an outlier at 18 texts (*Socrates*) and another at 26 (*Aquinas*). There is a group of outliers around the 40-text mark and another group between 75 and 85 texts, including *he*, *his*, *I* and *we*. The nature of the curve is easier to see in Figure 5.2 below where the 22 most frequent words in the corpus have been omitted. This does not suggest that those words are being ignored. They have been omitted in order to let us magnify the graph. The y-axis representing frequency has been magnified to show a maximum of 800 occurrences. While Figure 5.1 above gives us the overall picture, the scale of the Y-axis at 9000 occurrences is too large to see the actual spread.

Figure 5.2 Word frequency and text distribution in MICUPE without 22 most frequent words



At this magnification, the curve itself is clearer. Also, after increasing the zoom by taking out the 22 most frequent words, it becomes apparent that there is a correlation between frequency and text distribution, but this is not strong and there are a number of outliers. Firstly, lexical items with a frequency between 100 and 200 occurrences range

from a distribution over 10 to 80 texts with a clustering between 30 and 70. Secondly, lexical items appearing in 70 or more texts have a frequency ranging from 200 to just over 700. There is a trend whereby the frequency of types occurring across 70 or more texts increase dramatically, but not linearly or universally, in frequency as they appear in more texts. Theoretically, a type that occurs 94 times in the corpus could appear in 94 texts, appearing once in each text. This is not the case. In fact, the only type that appears in 94 times in the corpus is *else* and this word occurs in 40 texts only. Table 5.8 below outlines the distribution of types across the six essay iterations in MICUPE.

Table 5.8 Distribution of types across the essay iterations

	Essay 1	**Essay 2**	**Essay 3**	**Essay 4**	**Essay 5**	**Essay 6**
No. Of Types	3,378	1,660	2,409	2,429	1,852	3,390
Types in all texts	53	26	28	29	21	45
Types in more than 1 text	1,574	715	1,140	1,087	714	1,470
Types in 1 text only	1,804	945	1,269	1,342	1,138	1,920
Types with 1 occurrence	1,622	805	1,079	1,199	945	1,632

If we look at the number of types in each semester, what seems like a similarity between essay iterations 1 and 6 emerges. In essay iteration 1 there are 3,378 types across the 17 texts and in essay iteration 6 there are 3,390 types across the 17 texts, giving a difference of just 12 types between the total number of types used in each semester. This number refers to frequency only, and may not reflect any similarities in the semesters between the actual types used. Furthermore, both these essay iterations have a very similar number of types that occur only once, showing a difference of just 10 types occurring once between these essays. However, essay iteration 1 shows more types in all texts and more types that appear in two or more texts. Essay iteration 6 has a greater number of types that are limited to one text only. Given the similarities between these essays in relation to number of types with one occurrence, this implies that essay

iteration 6 has a greater number of types that are limited to one text only and appear more than once in that text. Furthermore, in Tables 4.1 and 4.6 in Chapter 4, there was a difference between the essay iteration in terms of essay title choice. In essay iteration 1, the students wrote to five essay titles, and the highest number of students writing to the same essay title is seven. In essay iteration 6, the students wrote to four essay titles and 13 of the 17 students wrote to the exact same essay title (See section 4.2 for details of essay titles).

Although there is a broad similarity between essay iteration 1 and essay iteration 6 despite differences in the task, as seen in Table 5.8, essay iterations 2-5 display less similarity. Essay iteration 2 has the fewest number of types but does not have the lowest number of types in all texts and types in more than one text. Essay iteration 4 has the highest number of types and the highest number of types in all texts, yet does not have the highest number of types in more than one text.

Each essay iteration has more types that are limited to one text only than types that occur in more than one text. However, essay iteration 3 has more types that occur in more than one text than types with a singular occurrence. All the other essays have more types with a singular occurrence than types that appear in more than one text.

Overall, Table 5.8 suggests that the frequencies of types across the essays shows some difference. This difference will be investigated further in Chapter 7.

In the overall corpus, there are 7,493 types used across 94 texts. This tells us immediately that the overall number of types in the whole corpus is not the sum of the types in essay iterations 1 to 6. There is an overlap between the iterations. In fact, were the number of types in each semester unique to that semester, there would be 15,118 types in the corpus. Obviously, that is not the case. Another way of putting this might be by measuring each semester as a percentage of the overall corpus:

Based on Table 5.9, we can see essay iterations 1 and 6 as containing the highest percentage of types and essay iterations 2 and 5 as containing the lowest percentage of types. This will be discussed in further detail in Chapter 7. In the present chapter, it is sufficient to note that each essay iteration contains a different number of types as compared to the other essay iterations.

Table 5.9 Number of types per semester as a percentage of the overall corpus

Sub-corpus	Number of types	Percentage of overall types
Essay 1	3,378	45.07%
Essay 2	1,660	22.15%
Essay 3	2,409	32.15%
Essay 4	2,429	32.41%
Essay 5	1,852	24.72%
Essay 6	3,390	45.23%
MICUPE	7,493	100%

Single use of types

There are two ways in which types can be considered as single-use within the corpus. One way is where the type only appears once in the whole corpus. The other way is where a type appears more than once in the corpus but is limited to only 1 of the 94 texts in MICUPE. In relation to the types that occur in 1 text only and the types that appear one time only, there are differences again between the essay iterations. These are summarised in Table 5.10:

Table 5.10 Single-use types

	Types in 1 text only	Types that appear 1 time only	Types that appear more than once but in 1 text only
Essay 1	1,804	1,662	182
Essay 2	945	805	140
Essay 3	1,296	1,079	217
Essay 4	1,342	1,199	143
Essay 5	1,138	945	193
Essay 6	1,920	1,632	288
MICUPE	3,398	3,011	387

When this data is viewed in Table 5.10, the most striking thing is that MICUPE is not simply the sum of the six sub-corpora in terms of number of types. If we were to add together the number of types that appear one time only as compared within each essay iteration, there should be 7,322 single-use types in the corpus. However, this is not the case since there are, in fact, 3,011 single-use types in MICUPE. Words that are, in one sub-corpus, limited to either 1 text or 1 occurrence, are no longer so when compared to the whole corpus. This suggests that the use of types is dependent on instantial choices. Students writing the same essay in the same time and space for the same audience do not all choose to use the same types. Furthermore, as the degree programme progresses over three years, there is a re-use of the types that were only used once in a separate essay iteration. Single-use types will be investigated in further detail in Chapter 7, but for the present chapter it is important to note that there is a degree of uniqueness both within the sub-corpora and MICUPE. Furthermore, this degree of uniqueness is carried down to the 94 texts themselves, as displayed in Figure 5.3 below.

Figure 5.3 Single-use types per text

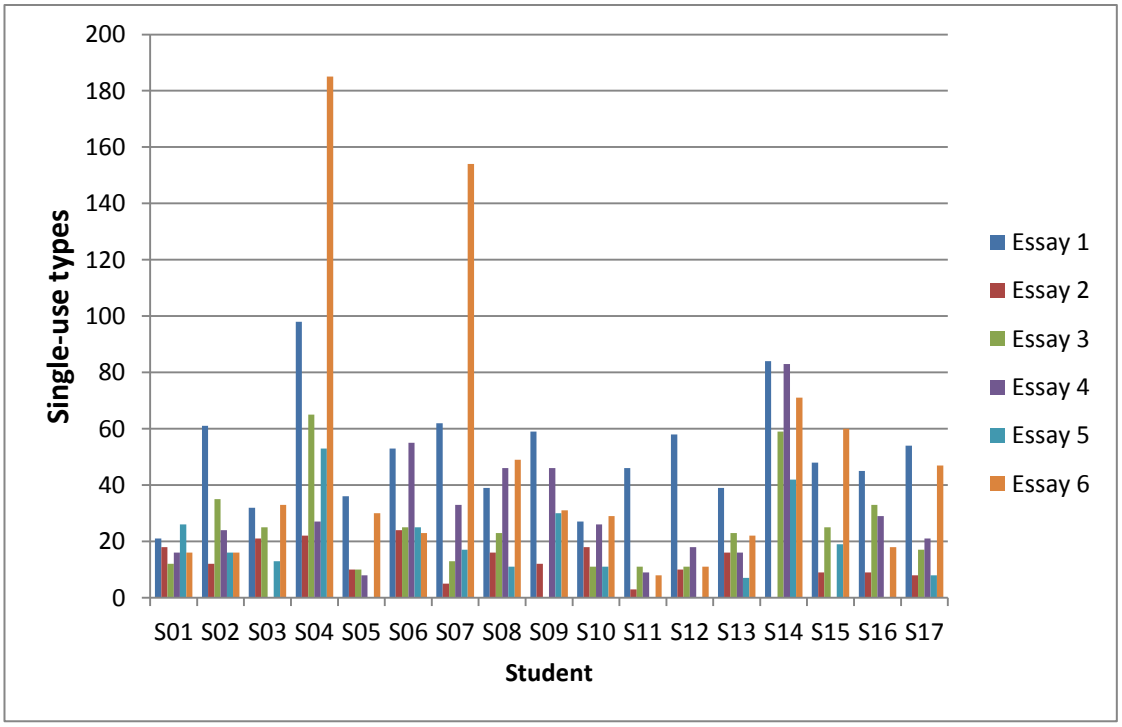
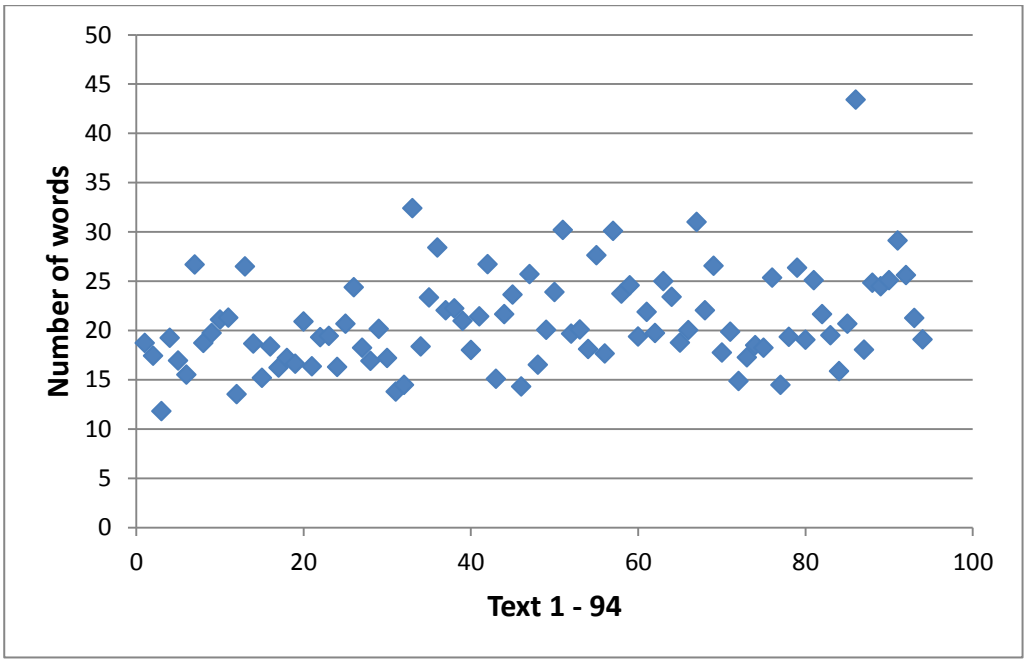


Figure 5.3 shows the distribution of the 3,011 single-use types according to each text. In Figure 5.1 and Figure 5.2, these 3,011 types occupy the (1,1) position. Each one of the 94 texts has a set of types that are used only once in that text are not used in any other text in the corpus. The text with the lowest number of single-use types is S11.2.B2 which has 3, and the text with the highest number of single-use types is S04.6.B1 which has 185. No one text is composed completely of types that are used in other texts, again highlighting the uniqueness of each text brought about by the set of instantial decisions made by the writer.

So far in this chapter, we have seen that only 14 types appear in all 94 texts. Essay iteration 1 has 53 types in all 17 texts. Each of the other five essay iterations has fewer types spanning all texts in that iteration. Each iteration has types that appear in no other iteration, and at the same time, each iteration has types that only appear once in that iteration and again once in another iteration. Furthermore, each individual text contains types that no order text in the corpus contains. In addition, Figure 5.4 below shows the mean sentence length for each of the texts and visually represents variety across the texts.

Figure 5.4 Mean sentence length for the 94 texts in MICUPE



From Figure 5.4, not only can we say that each text uses different types, but also organises them into sentences of differing lengths, again reinforcing the notion of uniqueness inherent in each text.

5.4 Summary and conclusion

This chapter examined:

a) various sections of the frequency list;

b) comparisons between the six sub-corpora that constitution MICUPE;

c) raw frequency correlated with number of texts.

Through this examination, this chapter has answered one of the research questions as outlined in Chapter 1, namely the question concerning the uniqueness of each text. In terms of use of types, an analysis of the empirical data has shown each text to be unique, the reasons for which were outlined in the theoretical framework in Chapter 3.

In relation to the various sections of the frequency list, we have seen that the subject matter that the writers are addressing in the essays is carried the whole way down this list. The top of the frequency list tends to be populated with function words, and those words that can be considered more philosophical in nature tend to appear further down the list. The choices made by the writers are not random. We can say this because each essay exists as a text and is comprehensible to a reader as an essay. This would suggest that, regardless of position in a frequency list, each word performs a necessary function as deemed by the writer in that time and space of composing the text.

The fact that only 14 words appear in all texts suggests a localised concern when composing as opposed to a ubiquitously shared set of lexical items that are reused in a linear and regular pattern. The students are composing similar tasks in the same genre for the same lecturer at the same time and have had a similar set of lectures and tutorials as well as similar reading material. Despite this, there are 3,011 types that appear only once in the whole corpus. Furthermore, a type that appears only once in an essay

iteration may appear in a separate essay iteration. This, again, points to the notion of instantial choices as outlined in the theoretical framework.

In relation to the comparisons between the six sub-corpora, we have seen that the frequency lists for each essay iteration are different. In addition, the overall frequency list is different to each of these lists. That difference is not limited to choice of type and order in which they appear. There are also differences in the number of types that appear in all texts and the number of types that are limited to one text only. This points to the homogenising effect of a corpus. If we extend this idea, we have to acknowledge that even within an essay iteration, the frequency lists for each individual essay are unique. Again, it must be noted that this is within a similarity of context.

When we examine the correlation between frequency and distribution across texts we note that a general pattern exists but that pattern is not without outliers. There are words that appear in the majority of texts, yet are absent from a minority. There are words that appear in a minority of texts, yet are absent from the majority. This gives the impression of randomness, however, we know that the appearance of words is not random.

In this chapter, we have seen that word frequency lists can give an insight into a corpus, but frequency by itself is not necessarily an appropriate determiner of the importance of a word. In fact, each word used, be it used many times or just once, is necessary at that point in time, hence its occurrence. These frequency lists, generated at a corpus level, are representative of the corpus as a whole and may not, in fact, mirror the sub-corpora they are comprised of. The homogenisation of the texts into sub-corpora are further homogenisation into a corpus give us a rich set of data on the one hand, yet on the other hand result in the loss of the notion of the text. This text is what was created by the writer and interpreted by the reader. The writers also use a number of types to achieve this that do not otherwise appear in the corpus.

We have also seen that raw frequency can be supplemented by distribution across texts. For parts of the corpus as a whole, there are correlations between the frequency of words and their spread across texts. For other parts, this correlation is less strong.

Given this, it is proposed to examine, in Chapter 7, how the tension between text, sub-corpus and corpus arises as the corpus is created. Following this, single-use and single-text types are investigated as an indicator of uniqueness.[5]

The outliers within MICUPE that we have identified are;

- function words such as *the*, *in*, *of*, and *for* that appear in each of the 94 essay texts
- the pronouns *I*, *we* and *you*
- the names of the authors of the primary texts for each essay such as *Socrates*, *Aquinas* and *Lonergan*

These three sets of words will form the basis of the following chapter and their analysis will be used to investigate the research questions concerning genre, task and audience.

[5] It is not being claimed that single-use and single-text types are the sole indicator of uniqueness.

Chapter 6 Outliers in frequency and distribution

6.1 Introduction

Chapter 5 identified three features within the corpus that stand out either due to their distribution versus frequency or their relative position in the frequency list compared to that of BAWE. These features can be considered to be indicative of certain influences that help to create an essay text. The purpose of this chapter is to examine the effect of some of the influences on the texts and to investigate differences across and within the texts despite the similarities in task, genre and audience. This will help answer some of the research questions outlined in Chapter 1, namely by further establishing the uniqueness of the texts and investigating how writers respond to genre, task and audience.

The 14 words identified as occurring at least once in each of the 94 texts can be seen as an indicator of a response to generic concerns. This is particularly true for the most frequent of these words such as *the*, *of* and *to* which tend to be used for nominalisations. Biber (2006: 76) argues that 'because they are so compact, prepositional phrases are often used in sequence, resulting in highly complex noun phrases with multiple modifiers'. This form of nominalisation is a feature of academic writing, and for this reason, we can attribute the use of these words to generic concerns. Below are two examples of complex noun phrases.

Example 6.1

> Therefore, although **the existence of God as prime mover and as efficient and exemplary cause of the world** can be established by reason alone, the full meaning of "God" can come only from faith. (S05.6.B1)

Example 6.2

> 'Already' refers to **the anticipation of a biological consciousness which recognises a body that could be used as a means to satisfy its appetites**. (S07.5.A1)

In examples 6.1 and 6.2, words such as *the*, *of*, *and*, *which*, *that* and *to* are used by the writers to create complex noun phrases.

The names of the philosophers, such as *Socrates* and *Aquinas*, were shown to display a disparity, compared to other types, between their frequency and distribution across texts. In each of the essay titles the students had to choose from or were given in each essay iteration, the name of a philosopher is either identified explicitly or implied through the subject matter for discussion. Due to this, we can assume that the use of these philosophers' names within the texts is a response to the task set for the student writers.

The pronouns *I*, *you* and *we* were shown in Chapter 5 to occur with differing relative frequencies across the six essay iterations. Furthermore, these pronouns, when taken in the context of the overall corpus, were shown to occur more frequently in MICUPE than BAWE. In the current chapter, we can assume that the pronoun *I* is in some way indicative of the writer while the pronouns *you* and *we* include the reader. Together, their use gives the overall impression of interaction between the writer and the reader. These pronouns can also be used in a generic sense, and that generic use can also be seen as interaction between the writer and the reader. The discussion in the current chapter includes both generic and non-generic uses of these pronouns.

Attributing these features to the concerns of genre, task, audience and writer is a generalised way of looking at the function of these words. Furthermore, these words are not the only realisations of the said influences and it must be noted that, within a text, every word is a reaction to all these influences at the same time. It would be a mistake to assume that these features are the only realisations of those mentioned influences and it would also be a mistake to assume that the appearance of these features in different texts is in some way representative of a similar response by different writers to those influences.

6.2 Analysing the features

14 words in all texts

This section analyses the 14 words that appear in all texts to show that there is a variation in their frequency across these texts. *The* is taken as an example to demonstrate the unevenness both in frequency and distribution within the texts. *The* is seen as an indicator of a response to the genre since *the* is extremely common in academic writing (see Biber et al 1999: 267 for a quantitative distribution of *the* in four registers including academic prose).

Figure 6.1 Frequency and distribution of words in all texts

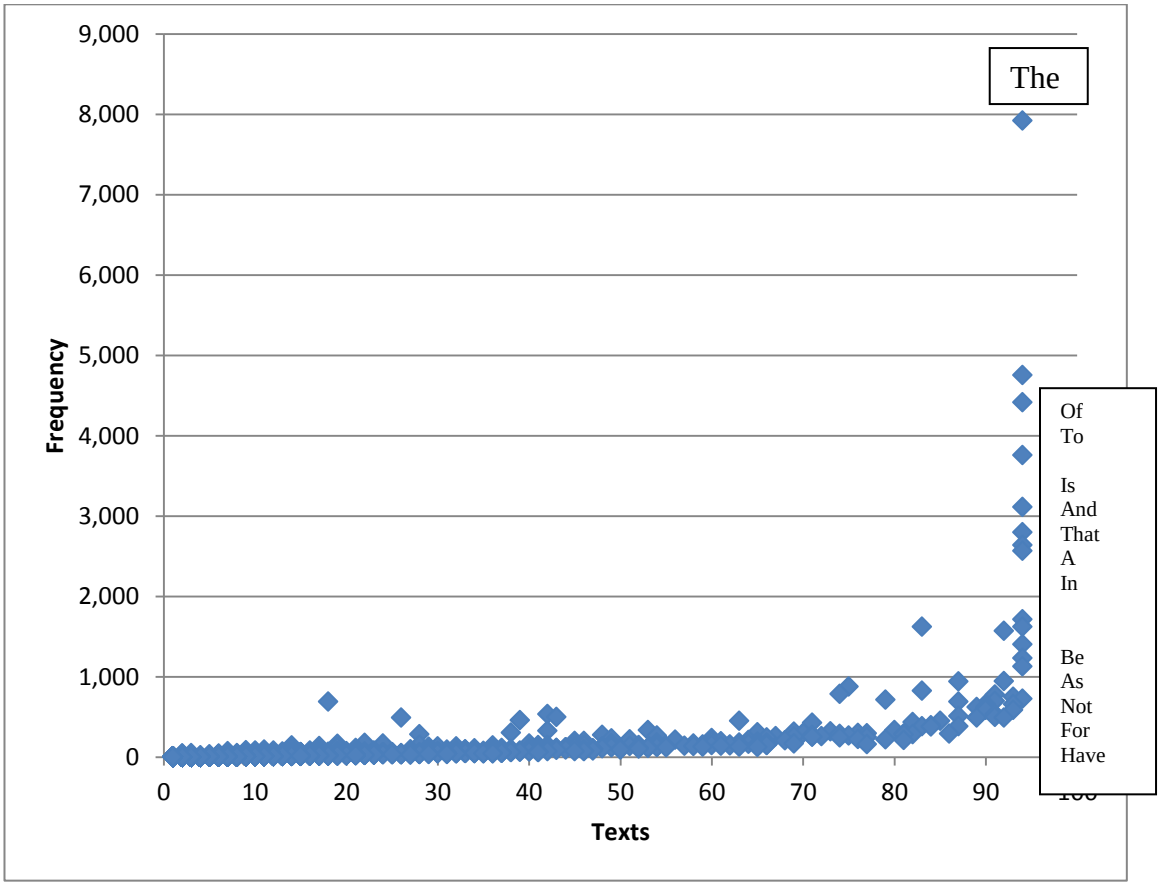


Figure 6.1 displays the frequencies of the words that appear in all 94 essay texts. What is of immediate interest is that there is a large span in the frequencies of these words. *have* is the lowest frequency of the 14 with 751 occurrences, while *the* is the most

frequent of the 14 with 7,919 occurrences. Furthermore, it can be seen that there are a range of words that have higher frequencies than those that appear in all texts yet those words with higher frequencies appear in fewer texts. This would suggest that raw frequency on its own does not correlate to distribution across texts. From this, we can speculate that the frequency list is the realisation of a set of instantial choices made by the writers at particular points in time. The reason words appear in all texts is, simply, because the writers chose to put those words in their texts. Of the current sample, no essay was written without these words.

Below is the distribution for each of the 14 words that appear in all 94 essay texts.

Table 6.1 Frequency data for the words that appear in all texts

Word	Total	Average per text	Highest per text	Lowest per text
the	7,919	84.24	310	18
of	4,754	50.57	163	8
to	4,416	46.98	153	7
is	3,760	46.98	211	4
and	3,112	33.1	87	6
that	2,800	29.78	139	8
A	2,640	28.06	101	3
in	2,568	27.32	117	7
it	1,715	18.25	93	2
be	1,622	17.26	84	2
as	1,403	14.93	56	2
not	1,232	13.10	113	1
for	1,131	12.03	36	1
have	729	7.75	53	1

Table 6.1 above gives the total, average, highest per text lowest per text occurrences of each of the 14 words that appear in all 94 texts in the corpus. They are displayed in

descending order according to total frequency in the corpus. In addition to the range of overall frequency of these words as seen in Figure 6.1 above, there is also a range from highest to lowest frequency for each individual word. For example, the word *the* appears 310 times in one text but only 18 times in another text and the word *of* appears 163 times in one text yet only 8 times in another text. Despite the relative importance of these words in noun phrases and prepositional phrases to modify the head noun, their distribution is not uniform.

It is also interesting that while the words are organised according to overall frequency in a descending fashion, the 'highest per text' column does not follow a linear descending pattern. For example, *of* appears in the corpus overall just short of 1,000 times more frequently than the word *is*, yet the highest number of the word *of* in a single text is 163 while the highest number of the word *is* in a single text is 211. A similar non-linear pattern exists in the 'lowest per text' column. Again, this points to a non-uniform distribution of these words.

The words *it*, *be*, *as*, *not*, *for* and *have* are limited to one or two occurrences in at least one text. This raises an interesting question. There are words that do not appear high up the frequency list for the whole corpus, yet appear more frequently in certain texts than some of these 14 words that span all texts. In the context of the overall corpus, these words are far more frequent. In the context of an individualised text, a set of which comprise the corpus, these words are less frequent.

If the 'highest per text' counts occurred in one single text, then this text would have a combined frequency for these 14 words of 1,716. In fact, the text with the highest amalgamated frequencies of these words has 1,697 in total. This text is S15.6.B1 which also has the largest number of tokens for an individual text in the corpus at 5,542. There is not much of a difference between the figure of 1,697 and 1,716. However, this one text has more than 2000 tokens more than the next largest text in the corpus, yet this one text does not have the highest frequency of each of the 14 words identified in Chapter 5. This would suggest that the frequencies of these words are not solely linked to word count.

On the other hand, if the 'lowest per text' counts occurred in one single text, then that text would have a combined frequency for these 14 words of 70. This is not the case.

The text with the lowest combined frequency of these 14 words actually has 130. This tells us that the distribution of the lowest frequencies of these words per text does not occur in the same text. This supports the notion that the frequencies of these words are not solely linked to word count.

Distribution within the texts

There is a tendency for these 14 words to co-occur in the texts. Figure 6.2 below shows the collocates of a concordance for the word *the*.

Figure 6.2 Collocates of *the*

N	Word	L5	L4	L3	L2	L1	Centre	R1	R2	R3	R4	R5
1	THE	403	425	726	62	3	7,910	3	62	726	425	403
2	OF	200	241	197	78	861			1,855	406	139	222
3	TO	176	181	216	348	385			190	136	164	205
4	IS	169	209	208	93	280		4	306	214	208	221
5	AND	140	120	155	153	217			311	97	245	208
6	IN	114	137	111	47	601			114	95	119	117
7	THAT	110	130	95	96	203			360	130	119	111
8	A	102	134	155	18				30	189	113	112
9	HE	107	106	100	107			1	52	116	56	58
10	AS	79	100	58	26	138			78	54	57	65
11	FOR	47	54	44	53	223		1	83	35	38	48
12	BE	48	43	105	33	52		1	7	97	66	81
13	THIS	82	104	85	73	9		1	8	70	36	38
14	IT	64	65	55	80	4		2	28	76	67	62
15	ON	32	39	23	17	218		1	24	24	40	40
16	NOT	42	51	65	69	24		1	10	64	58	65

Figure 6.2 shows the co-occurrence patterns of the word *the*. 12 of the remaining 13 words appear in the top 16 collocates of *the*. The one word that does not appear is *have* and this word appears in position 27. The strongest collocate of *have* is *the*. Furthermore, looking at the 'left 3' and the 'right 3' columns in particular, it can be seen that the word *the* has a tendency to occur in its own vicinity (example 6.1 and 6.2 above

also included *the* occurring in its own vicinity). Here is a sentence from S04.3.B3 showing an example of a selection of these words co-occurring:

Example 6.3

> **The** point **of the** postulate **is that the** same laws **of** mathematical expression apply **to the** apple butt regardless **of** what frame **of** reference **it is in** (eg car, road). (SO4.3.B3)

There are 29 words in the above sentence. 14 of those words are words that also occur in every text in the corpus. This means that roughly 48% of the words in the sentence also occur in each one of the 93 other texts in MICUPE. The total frequency of these 14 words in the corpus is 39,801. The total number of tokens in the corpus is 134,289. This equates to 29.64% of all the tokens in the corpus are one of these 14 words. In other words, these words appear more frequently than one in four.

If we look at the four-word strings in the corpus, we get a total of 119,499 types of string. 6,699 of these strings appear more than once. However, if we omit the strings that contain at least one of these 14 words, there are 26,374-word strings remaining, of which 723 appear more than once. This tells us that despite their ubiquity and the appearance of occurring more frequently than one in four, there are over 26,000 types of string in the corpus that do not contain these words.

If we extend the length of string to a 10-word string, there are 125,233 types of 10-word string in the corpus, of which 744 appear more than once. Again, if we omit those strings that contain one of the 14 words that appear in each and every text and account for just short of 30% of all tokens in the corpus, there are 2,365 10-word strings left, of which one appears twice and all the others appear once. Therefore, despite the high frequency of these words, and their seeming importance for writing philosophy essays at undergraduate level due to this frequency and distribution across all 94 texts, there are still areas of at least 10 words in a row that do not contain one of these 14 words. In essence, this suggests that while these words are in the main necessary for writing the essays, they are not obligatory for every single part of the essay. If they are not

obligatory for every pattern the essay, it seems reasonable to suggest that where they are used it is the result of an instantial choices made by the writer.

The

Taking the corpus as a whole, *the* is the most frequent word. In Table 6.2 below, the 10 most frequent words in the corpus are shown.

Table 6.2 10 most frequent words in MICUPE

N	Word	Freq.	%	Texts	%
1	THE	7,919	5.89	94	100.00
2	OF	4,754	3.53	94	100.00
3	TO	4,416	3.28	94	100.00
4	IS	3,760	2.79	94	100.00
5	AND	3,112	2.31	94	100.00
6	THAT	2,800	2.08	94	100.00
7	A	2,640	1.96	94	100.00
8	IN	2,568	1.91	94	100.00
9	IT	1,715	1.27	94	100.00
10	HE	1,625	1.21	83	88.30

The occurs 7,919 times, accounting for approximately 5.89% of the words in the corpus. *The* appears in all 94 texts, and consequently appears in each of the six sub-corpora.

Table 6.3 The 5 most frequent words in the 6 essay iterations

N	Essay 1	Essay 2	Essay 3	Essay 4	Essay 5	Essay 6
1	THE	THE	THE	THE	THE	THE
2	TO	IS	OF	OF	OF	OF
3	HE	OF	TO	TO	TO	TO
4	OF	TO	IS	AND	IS	IS
5	AND	THAT	A	IS	A	AND

In fact, as can be seen in Table 6.3, *the* is the most common word in each of the six sub-corpora. Naturally, if *the* is the most frequent word in each sub-corpus, it must be the most frequent word in the whole corpus.

Despite *the* being the most common word in each sub-corpus and the corpus as a whole, when we look at individual essays, *the* is not necessarily the most common word in that essay. In fact, there are 13 texts in which *the* is not the most common word. There are essay texts in five of the six sub-corpora that do not have *the* as the most frequent word. In the semester six essay, each of the 17 texts have *the* as the most common word. In Table 6.4 below, the essays where *the* is not the most common word are identified along with the words that are more common in that particular essay than *the*.

Table 6.4 Essays where *the* is not the most frequent type

Student	**Essay**	**Frequency order**
1	1	He – That - The
1	4	To – The
1	5	To – The
2	2	Is – Being – of – The
3	2	Is – To – That – The
3	3	To – The
4	2	Is - The
5	2	Is – The
7	2	Is – The
7	4	Of – The
7	5	Of – The
9	2	Is – The
15	2	Is – The

It is noticeable that some students are represented in Table 6.4 above more than once. Student 1 and Student 7 use other types more commonly than *the* in three of the six essay iterations. It is also noticeable that some essay iterations are there more than once.

Essay iteration 2 is there a total of six times. For two of the essay texts, from Students 2 and 3, *the* is the fourth most common word in that text. However, if we were to create a new corpus from these 13 texts, the word *the* is the most frequent word in that corpus due to the fact that there is a variety of words that are more common than *the* across each of the 13 texts.

Distribution within the texts

Figure 6.3 below shows an excerpt from the distribution plot (see Chapter 4 for an explanation of a distribution plot) for the word *the*.

Figure 6.3 Distribution plot for *the*

N	File	Words	Hits	per 1,000	Dispersion	Plot
39	S05.3.B3.txt	1,219	68	55.78	0.896	
40	S06.3.B1.txt	1,487	106	71.28	0.868	
41	S07.3.A2.txt	828	53	64.01	0.939	
42	S08.3.B2.txt	1,584	119	75.13	0.899	
43	S10.3.A2.txt	1,046	77	73.61	0.858	
44	S11.3.B1.txt	901	75	83.24	0.888	
45	S12.3.B2.txt	1,135	78	68.72	0.895	
46	S13.3.B1.txt	1,482	117	78.95	0.904	
47	S14.3.A1.txt	1,990	127	63.82	0.893	
48	S15.3.B1.txt	1,814	165	90.96	0.947	
49	S16.3.B1.txt	1,334	115	86.21	0.901	
50	S17.3.B1.txt	1,027	67	65.24	0.862	
51	S01.4.B2.txt	1,069	40	37.42	0.866	
52	S02.4.A1.txt	1,131	68	60.12	0.918	
53	S04.4.B1.txt	778	45	57.84	0.785	
54	S05.4.B1.txt	757	38	50.20	0.831	
55	S06.4.A2.txt	1,458	80	54.87	0.907	
56	S07.4.B1.txt	830	43	51.81	0.801	
57	S08.4.B2.txt	1,223	95	77.68	0.843	
58	S09.4.B2.txt	1,626	100	61.50	0.849	
59	S10.4.A2.txt	982	47	47.86	0.810	
60	S11.4.B2.txt	792	48	60.61	0.828	
61	S12.4.A2.txt	1,278	71	55.56	0.847	
62	S13.4.A2.txt	1,036	75	72.39	0.852	
63	S14.4.A2.txt	2,206	143	64.82	0.823	
64	S16.4.B2.txt	1,050	55	52.38	0.833	
65	S17.4.C3.txt	1,017	87	85.55	0.929	
66	S01.5.B2.txt	1,003	37	36.89	0.815	
67	S02.5.A2.txt	760	56	73.68	0.792	
68	S03.5.B3.txt	615	40	65.04	0.901	
69	S04.5.B1.txt	1,125	52	46.22	0.868	

Looking at the distribution plot for the word *the*, we can see a variety in the number of hits per text, ranging from 37 in line 66 to 165 in line 48. There is also a variety in the 'per 1000' column, ranging from 36.89 to 90.96. The distribution plot gives a visual representation of these differences. It can be seen that in some instances the distribution of the word *the* within a text string occur close together whereas in other instances the distribution is more dispersed. In addition, it is also noticeable that there are strings within some of the texts where the word *the* does not occur at all.

Overall, the distribution of the 14 types across the texts is not as uniform as the position they hold in the overall frequency list would suggest, and if we take the appearance of these types as part of a reaction to genre, we have seen how writers respond to genre and how that response is not uniform. Within the texts themselves, there is a lack of uniformity in distribution, as evidenced by the number of 10-word strings that do not contain these 14 words. Furthermore, the distribution of the most frequent word in the corpus, *the*, is not uniform. It is the most frequent word in the corpus, and the most frequent word in each of the six sub-corpora. However, *the* is not the most frequent word in 14 individual texts. In addition, its distribution across and within the texts is not uniform and gives the impression of randomness. However, we can be certain that the distribution of these words is not random since they fulfil particular functions and were chosen by the writers in those instances that they were used.

Philosophers as task indicators

In Chapter 5, we saw that the word *Socrates* stood out as having a high frequency but a narrow distribution across texts. Other philosophers such as *Aquinas*, *Lonergan* and *Parmenides* also stood out as being positioned at various distances from the general curve. These philosophers are, in a way, representative of the task that each essay is responding to. We could argue that *Socrates* appears in all the texts in semester one because he is the subject of the tasks in semester one. The same can be said for the other philosophers.

Not all the essay titles include the name of a philosopher. For some titles, an idea or subject matter is specified which means that the name of the particular philosopher associated with that idea is implied (See Chapter 4 for details on all essay titles). In the

current section, a general overview of the use of the names of philosophers is given. Subsequently, emphasis is put on essays 1, 2 and 6 as these are the essays that explicitly mention the name of a philosopher in their titles.

Figure 6.4 Frequency and distribution of *Socrates*, *Aquinas* and *Lonergan*

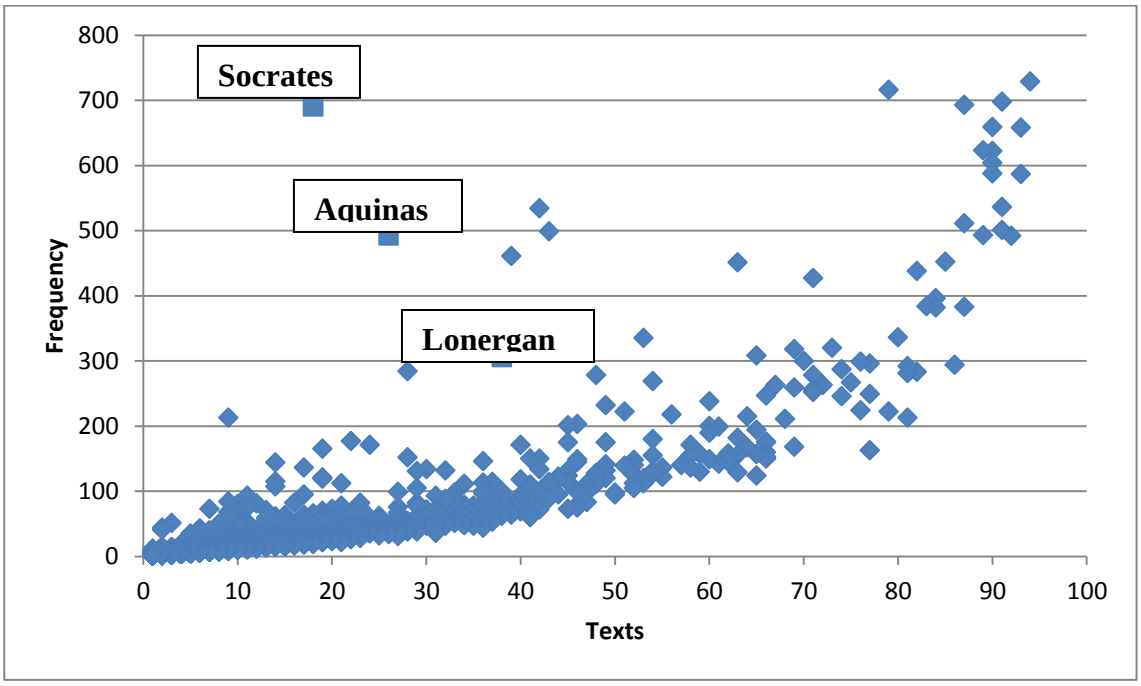
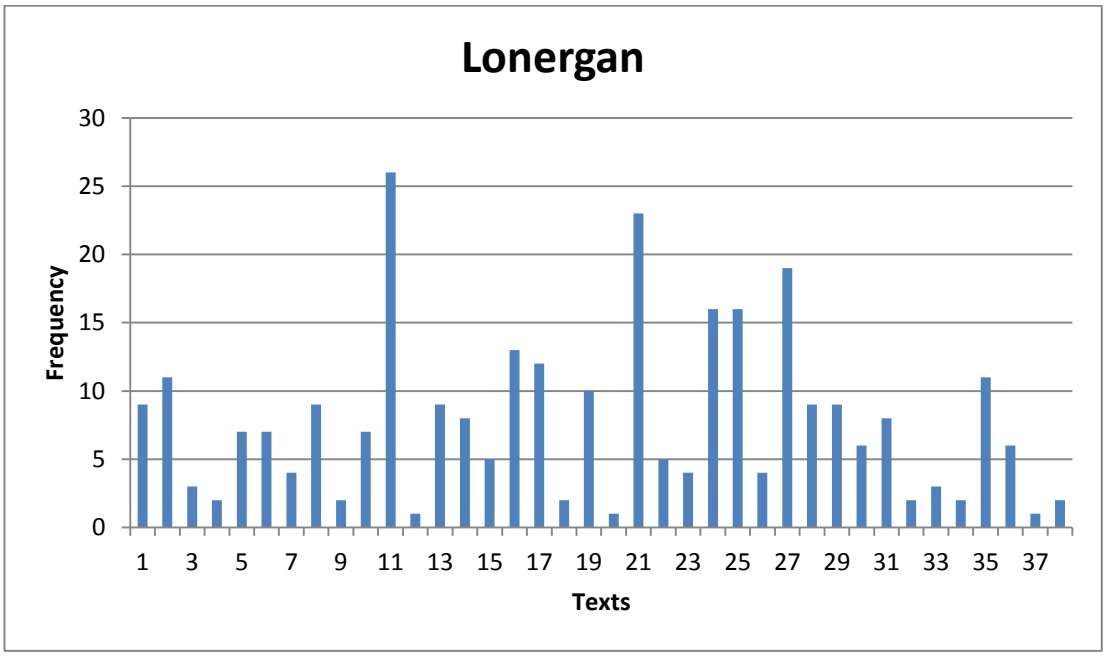


Figure 6.4 shows the positioning of *Socrates*, *Aquinas* and *Lonergan* in the scatterplot of frequency against number of texts. All three philosophers are positioned away from the general curve, with *Socrates* showing the furthest distance from this curve. As stated above, these philosophers can be seen as indicative of the influence of task since they represent the key texts being discussed in the relevant essays and are either mentioned are implied in the essay titles the students had to choose from.

In essays 3, 4 and 5 the implied philosopher is Bernard Lonergan. The essays are an exploration of an idea contained in his work (See Chapter 4 for details of the essay titles). Figure 6.5 below displays the frequencies of the word *Lonergan* in these essay texts where he is mentioned at the least once.

Figure 6.5 Distribution of *Lonergan* in essay iterations 3-5



There are 38 texts that include the word *Lonergan* out of a total of 44 essays in the three iterations combined. This means that six essays do not refer to the name of this philosopher while the other 38 essays do. Furthermore, the essays that include the word *Lonergan* do so in various frequencies ranging from 1 per essay to 26 per essay. This tells us that the same essay titles have different effects on each writer since they refer directly to *Lonergan* with a variety of frequency. However, this is indicative of an oblique effect of task since the word *Lonergan* is not actually in the essay title. To see the effect of task in a more direct manner, we will examine the essays that were written which had a philosopher named directly in the title.

Socrates

Socrates is the subject of all the essays in the first iteration and is mentioned in the title of all essays in this iteration. No essay in this iteration is written without this word. *Socrates* is also mentioned in one other essay written approximately one and a half years later although he is not the subject of the essay. The frequencies per text of the

type *Socrates* in the 17 essays in essay iteration 1 and the one text in essay iteration 2 (SO2.2.A2) that includes the word *Socrates* are detailed in Table 6.5 below.

Table 6.5 Distribution of *Socrates* among the texts

File	**Words**	**Hits**	**per 1,000**
S01.1.B3.txt	1,475	24	16.27
S02.1.B1.txt	2,456	66	26.87
S03.1.B3.txt	1,147	25	21.80
S04.1.B1.txt	2,665	54	20.26
S05.1.A2.txt	1,588	43	27.08
S06.1.A2.txt	1,810	35	19.34
S07.1.A1.txt	1,568	16	10.20
S08.1.C1.txt	1,489	46	30.89
S09.1.C1.txt	1,690	48	28.40
S10.1.A2.txt	1,549	27	17.43
S11.1.B2.txt	1,626	45	27.68
S12.1.B3.txt	1,579	40	25.33
S13.1.C1.txt	1,688	39	23.10
S14.1.B1.txt	2,125	52	24.47
S15.1.B1.txt	2,067	54	26.12
S16.1.B2.txt	1,804	28	15.52
S17.1.B2.txt	2,055	48	23.36
S04.2.A2.txt	1,013	1	0.99

In Table 6.5, we can see the distribution of the word *Socrates* in the corpus. As all the essays in essay iteration 1 concern *Socrates*, it is no surprise that this word appears in each of those texts. The word *Socrates* also appears in one other text, written approximately a year and a half after this first essay. Task has a quantifiable, empirical effect in that no essay in this semester is written without the word *Socrates* but there is a variance in the frequencies of this word.

There is a large spread in the number of instances of *Socrates* despite the fact that the students are completing similar tasks. They range from 16 instances in the essay written by Student 7 to 66 instances in the essay written by Student 2. This, in itself, raises an issue. As *Socrates* is the topic of all the essays, it would seem reasonable to suggest that this word appears in the texts as a direct consequence of task. Saying this, then, raises further issues. The first concerns the labelling of certain words as task-related and excluding others from this label. The second concerns the effect of that task on each individual writer.

Words as task-related

Figure 6.6 below is an extract taken from a concordance of *Socrates*.

Figure 6.6 An extract from a concordance of *Socrates*

N	Concordance	File
233	truth you would ultimately do the good. Socrates cared more for his soul than he	S12.1.B3.txt
234	truth and goodness as vital in one's life. Socrates life was modelled on that of his	S14.1.B1.txt
235	quo in Athens. It must also be noted that Socrates states in the Apology that these	S14.1.B1.txt
236	know the good, they will do the good. 4 Socrates young followers were living	S14.1.B1.txt
237	or after death" (Ibid: 62). So the truth that Socrates has been preaching is really	S05.1.A2.txt
238	and so wanted rid of him. Again, Socrates was in the wrong place at the	S14.1.B1.txt
239	and the perfection of the soul?" (29 D) Socrates also states in this passage that	S04.1.B1.txt
240	believe, therefore it is true. To establish if Socrates did in fact die for the truth, I	S13.1.C1.txt
241	wrong doings. I personally believe that Socrates was not guilt either of charge	S14.1.B1.txt
242	true. In his cross-examination of Meletus, Socrates leads him into saying that all	S16.1.B2.txt
243	earlier prejudice. It is easy to admire Socrates for the way that he outsmarted	S11.1.B2.txt
244	the jury and of course with meletus. Now Socrates is more unpopular than ever	S17.1.B2.txt
245	distill the truth is once again highlighted. Socrates is dragged towards defining	S08.1.C1.txt
246	unjustly, like a good man or a bad one". Socrates believed that in all his actions	S12.1.B3.txt
247	for, such as his quest for truth. When Socrates was convicted in the trial it was	S05.1.A2.txt
248	would dispute the Oracle. Even though Socrates was charged with introducing	S14.1.B1.txt
249	in and can only be admired for it. Socrates was a skilled speaker, some	S06.1.A2.txt
250	afraid to expose but were eager to stop. Socrates was only guilty of displaying an	S14.1.B1.txt
251	to his downfall. Throughout the apology Socrates shows complete disregard for	S17.1.B2.txt

Taking this set of concordance lines, chosen at random, illustrates the issue with labelling words as task-related. It would be difficult in the extreme to argue that any

other word in the set of concordance lines, or in the whole corpus for that matter, is not related to the task. In the instance of essay iteration 1, the word *Socrates* comes directly from the task. However, everything that is said about Socrates also comes from the task. Every word in the 17 essays in essay iteration 1 concerns, directly or indirectly, *Socrates*. Furthermore, every word in every essay in the corpus comes from the task. This, then, creates an issue as although we can still identify the word *Socrates* as task-related in essay iteration 1, we must also label every other word in that iteration as task-related as all of the texts are a response to a task. All the words within those texts are a response to the task. If one word were omitted from a text, or if one word were added to a text, it would be a different response to the task. Therefore, all the types seen in the frequency list in the previous chapter are a response to the task. No text was created using only the words given in the essay title. The writers must say something extra in the whole text about the task than merely repeating those words given in the essay title.

Effect of task on the writer

Although each student is doing a similar writing task, the effect of that task on each individual piece of writing is unique. We know this because:

- each essay contains a set of types that are in no other essay (see Chapter 5)
- each essay uses a different number of tokens to any other essay (see the 'Words' column in Table 6.5 above)
- the essays contain differing frequencies of the word *Socrates* (see Table 6.5 above)
- this is despite similar tasks.

A *word sketch* 'is an automatic corpus-derived summary of a word's grammatical and collocational behaviour' (Kilgarriff et al, 2010: 372). A *word sketch* for *Socrates,* generated using SkecthEngine™ software, is shown below in Figure 6.7.

Figure 6.7 Word Sketch of Socrates

Socrates *(noun)* MICUPE1 freq = 691 (4538.9 per million)

object_of	105	1.3
accuse	4	10.02
charge	3	9.5
blame	2	9.23
prosecute	2	9.22
sentence	2	9.21
annoy	2	9.18
see	6	9.16
convict	2	9.09
avoid	2	9.07
regard	3	8.99
think	3	8.51
die	2	8.45
question	2	8.41
find	3	8.37
call	2	8.2
believe	2	7.48
make	2	6.88
be	11	5.45

subject_of	486	8.2
believe	23	9.97
try	10	9.17
do	24	9.13
begin	9	9.09
be	143	9.08
continue	8	8.94
use	11	8.91
have	21	8.74
show	7	8.65
tell	6	8.45
manage	5	8.36
make	9	8.29
say	6	8.18
seem	5	8.17
go	6	8.09
refuse	4	8.05
question	4	7.84
devote	3	7.62
fail	3	7.61
decide	3	7.59
face	3	7.58
commit	3	7.56
see	4	7.5
feel	3	7.49
want	3	7.48

adj_subject_of	35	4.1
guilty	15	11.8
willing	3	10.83
truthful	2	10.42
aware	3	10.11

predicate_of	49	9.5
martyr	23	12.16
orator	2	10.12
man	9	9.48

possessed	31	5.9
style	2	10.39
Plato	2	9.71
skill	2	9.64
lack	2	9.61
argument	4	7.95

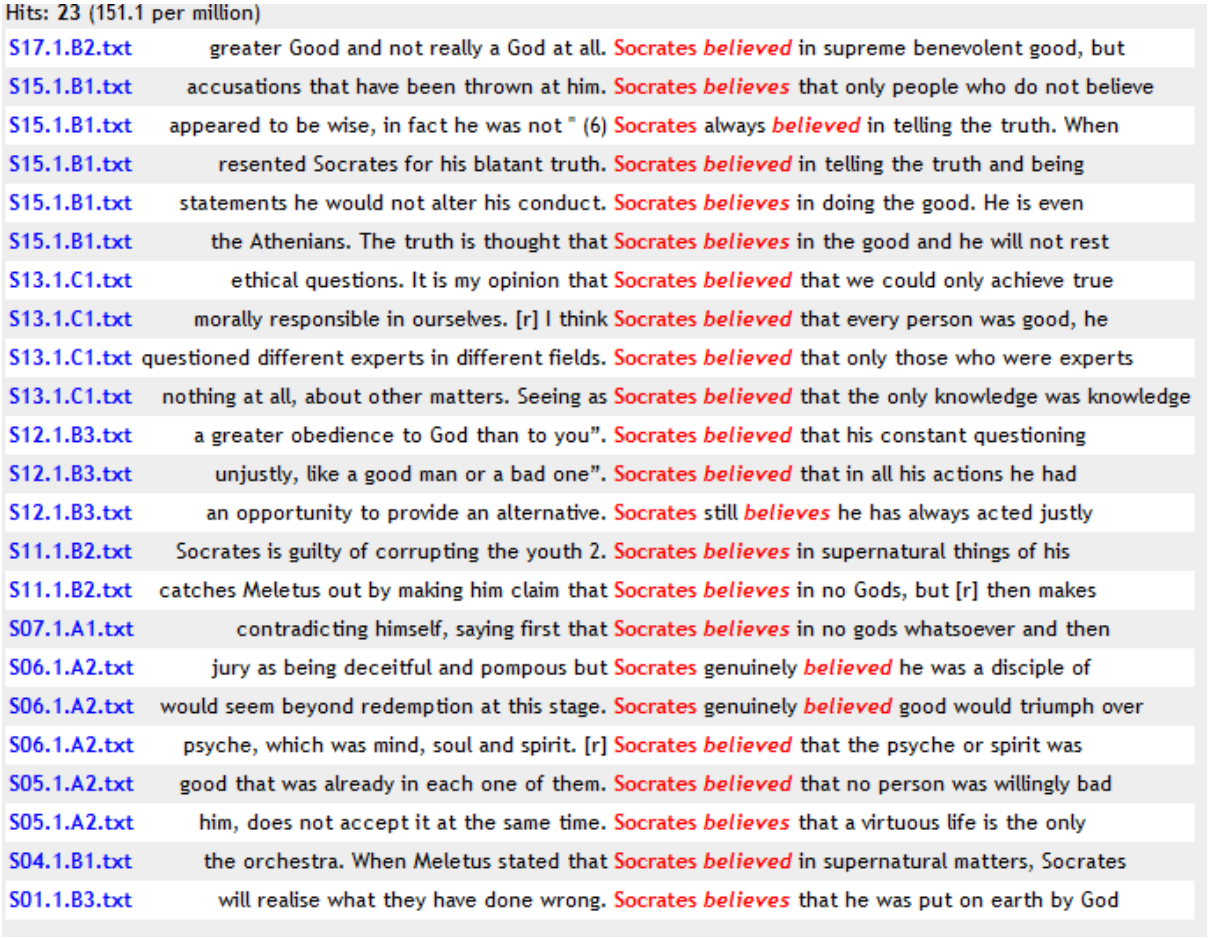
In the *word sketch* for *Socrates* in Figure 6.7, there are 9 tables displaying information about the use of this type. The order within the tables is governed by statistical significance, not raw frequency. Their meaning and an example are shown below in Table 6.6.

Table 6.6 *Word sketch* explained

Code	**Meaning**	**Example**
object_of	Where *Socrates* is the object of a verb	… these confused victims **blamed Socrates** for … (S14.1.B1)
subject_of	Where *Socrates* is the subject of a verb	**Socrates begins** his defence by … (S17.1.B2)
adj_subject_of	Where *Socrates* is the subject of an adjective complement	**Socrates** is **guilty of corrupting** … (S11.1.B2)
predicate_of	Where *Socrates* is a predicate of a noun or noun phrase	**Socrates** is a **martyr** for the truth … (S15.1.B1)
possessed	Where *Socrates* is used in the possessive form	They were enrages by **Socrates' lack** of repentance (S06.1.A2)

The word sketch for *Socrates* above in Figure 6.7 shows us something interesting. We have already seen that there are variances in the frequency of the occurrence per text of this word. The word sketch demonstrates that there are also variations in the use of this word. It is the object of, the subject of, and the predicate of a range of other types. Not only do frequencies vary, and when added together impact on the word frequency list, as seen in Chapter 5, but so do individual uses. *Socrates* is not used in the same way in every instance. If we select *believe* as 'subject-of', we get further individuality of use.

Figure 6.8 *Socrates* + believe



Hits: 23 (151.1 per million)

S17.1.B2.txt	greater Good and not really a God at all. Socrates ***believed*** in supreme benevolent good, but
S15.1.B1.txt	accusations that have been thrown at him. Socrates ***believes*** that only people who do not believe
S15.1.B1.txt	appeared to be wise, in fact he was not " (6) Socrates always ***believed*** in telling the truth. When
S15.1.B1.txt	resented Socrates for his blatant truth. Socrates ***believed*** in telling the truth and being
S15.1.B1.txt	statements he would not alter his conduct. Socrates ***believes*** in doing the good. He is even
S15.1.B1.txt	the Athenians. The truth is thought that Socrates ***believes*** in the good and he will not rest
S13.1.C1.txt	ethical questions. It is my opinion that Socrates ***believed*** that we could only achieve true
S13.1.C1.txt	morally responsible in ourselves. [r] I think Socrates ***believed*** that every person was good, he
S13.1.C1.txt	questioned different experts in different fields. Socrates ***believed*** that only those who were experts
S13.1.C1.txt	nothing at all, about other matters. Seeing as Socrates ***believed*** that the only knowledge was knowledge
S12.1.B3.txt	a greater obedience to God than to you". Socrates ***believed*** that his constant questioning
S12.1.B3.txt	unjustly, like a good man or a bad one". Socrates ***believed*** that in all his actions he had
S12.1.B3.txt	an opportunity to provide an alternative. Socrates still ***believes*** he has always acted justly
S11.1.B2.txt	Socrates is guilty of corrupting the youth 2. Socrates ***believes*** in supernatural things of his
S11.1.B2.txt	catches Meletus out by making him claim that Socrates ***believes*** in no Gods, but [r] then makes
S07.1.A1.txt	contradicting himself, saying first that Socrates ***believes*** in no gods whatsoever and then
S06.1.A2.txt	jury as being deceitful and pompous but Socrates genuinely ***believed*** he was a disciple of
S06.1.A2.txt	would seem beyond redemption at this stage. Socrates genuinely ***believed*** good would triumph over
S06.1.A2.txt	psyche, which was mind, soul and spirit. [r] Socrates ***believed*** that the psyche or spirit was
S05.1.A2.txt	good that was already in each one of them. Socrates ***believed*** that no person was willingly bad
S05.1.A2.txt	him, does not accept it at the same time. Socrates ***believes*** that a virtuous life is the only
S04.1.B1.txt	the orchestra. When Meletus stated that Socrates ***believed*** in supernatural matters, Socrates
S01.1.B3.txt	will realise what they have done wrong. Socrates ***believes*** that he was put on earth by God

Figure 6.8, generated using SkecthEngine™ software, shows the concordances where *Socrates* is the subject of the verb *believe*. Looking qualitatively deeper into the instances of *Socrates* and its co-occurrence with *believe*, we see again that there are individual choices made by the writers at particular points. For example, some students do not use the word *believe* in conjunction with *Socrates*. 10 students used *believe*, which means that seven students did not use this word in conjunction with *Socrates*. Again, at the micro-level, this is indicative of instantial localised choices made by individual writers.

Furthermore, despite the clear demarcation of time in relation to Socrates as being a philosopher who lived over 2000 years ago and has died, some writers refer to Socrates in the present (*believes*), often called the *historical present*, while others refer to him in the past (*believed*). For Student 11, Socrates believes in supernatural things, while for

Student 4 he believed in supernatural matters. For Students 15, 12 and 5, Socrates both *believes* and *believed* in the same text. Furthermore, in four instances, students deemed it appropriate to modify that belief with either *always*, *still* or *genuinely*. For some students, Socrates believes only one thing, for others e.g. Student 15, Socrates believes five things. We can assume that the writer inserts the belief of Socrates as they themselves believe this is necessary to complete the task. Furthermore, we can also assume that the essays are not merely a response to the task and the writers are also aware of generic expectations.

Specific Tasks

In semester one, the students were asked to choose one essay title from a set range of options. The essay titles of the texts in the corpus are as follows:

- Evaluate Socrates' arguments in his own defence.
- Socrates was guilty as charged. Discuss.
- Socrates was a clever orator. Discuss.
- Socrates was a martyr for the truth. Discuss.
- Socrates committed suicide. Discuss.

Seven of the students chose to do the essay entitled 'Socrates was a martyr for the truth. Discuss.' If we look at those seven essays only, we will be looking at essays that write to the exact same essay title with the exact same task direction at the same point in time for the same audience. As this is the case, we are excluding the influence of small differences in the task that may be important when comparing students writing to a different essay title about Socrates.

We can break this essay titled down into its eight words and count the instances of each word in each essay. This is represented in Table 6.7 below.

Table 6.7 Frequency counts for each word in the title

	Socrates	**was**	**a**	**martyr**	**for**	**the**	**truth**	**Discuss**
S03	25	10	22	7	14	53	25	0
S04	54	52	49	7	29	149	33	2
S05	43	34	32	6	24	83	27	0
S08	46	17	21	6	16	97	42	0
S10	27	31	24	8	24	97	33	0
S13	39	30	18	3	29	90	23	0
S15	54	29	44	11	27	121	25	0

Out of these seven students, only one uses the word *discuss*. Its use in context is shown below:

Example 6.4

> 'In this essay I will try to prove that he was a martyr by examining eight points which I have noted in my research.
>
> Item 1 a fundamental reason to believe that Socrates was a martyr comes from the fact that he was given the chance, even though he had been found guilty, to renounce his philosophy in return for his life. Socrates declined to accept this offer.
>
> Item 2 I will examine Socrates pursuit of the truth in discussion with Meno later in the essay.
>
> Item 3 I will examine further evidence of his attachment to the truth in his discussion with Laches, the army general.
>
> Item 4 His refusal to arrest Leon of Salamis and put him to death at the risk of being executed himself is direct evidence of his willingness to die for the truth.
>
> Item 5 I will examine his views on unpopularity, which shows his persistent and shocking ability to question common conventions.
>
> Item 6 I will **discuss** what Socrates saw as his single duty in life - to expose the people of Athens to the truth.
>
> Item 7 I will show the falsity of Meletus' claims, through Socrates defence. This evidence reinforces his martyrdom.
>
> Item 8 I will **discuss** the aftermath of his death, showing the realisation by the people of Athens that Socrates was indeed a martyr for the truth.' (S04.1.B1)

This extract is from the essay written by Student 4 and is the second paragraph. It is the only essay out of the seven that contains the word *discuss*. It is also the only essay out of the seven that gives a detailed list of items that will be examined in the essay. In the original text, each one of the items 1-8 is given in what appears to be a separate paragraph. As part of these items, as well as *discuss*, the student will also *examine* and *show*. Another student, Student 10, wrote an essay introduction that has some similarity to that of Student 4:

Example 6.5

> I believe that Socrates was a martyr for the truth and I intend to show this in my essay by these main points. Firstly, he dedicated his life to his search for the truth. Secondly, even though he was a poor man he never charged anyone who listened to his philosophy. Thirdly, his disgust at the Sophists and how they lived their lives shows that he was dedicated to the truth. Fourthly, even though he had many enemies and he knew he was placing his life in danger he carried on. Fifthly, he told the jury in "The Apology" that if they were to offer him an acquittal on the grounds that he would no longer philosophize and continue on his quest for the truth that he would refuse. Sixthly, he did not lie during the apology or use persuasive language during his trial. And finally when choosing his punishment he decided not to choose banishment as he wanted to continue his quest for the truth in the capital where it would be most successful, and therefore was executed. (S10.1.A2)

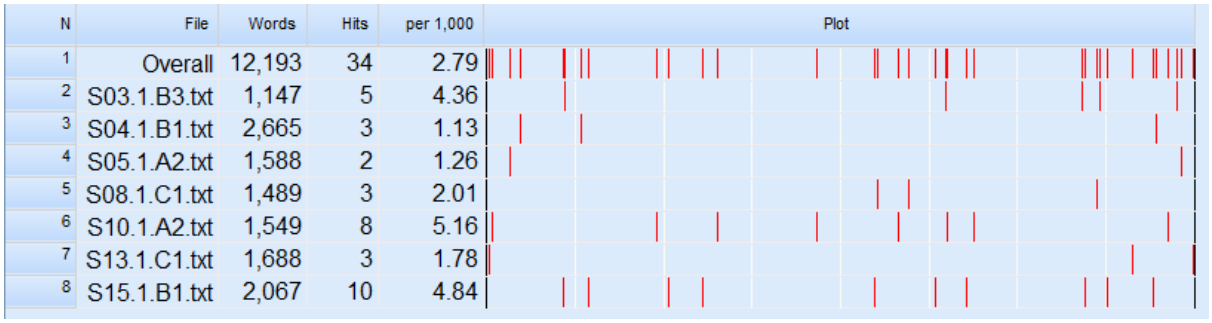
The list given by Student 10 is not a list that explicitly states what will be covered in the essay. It is, instead, a form of synopsis for the logical argument in reaching the conclusion that Socrates was a martyr for the truth. This synopsis, however, although not explicitly stated as such, does serve as an outline of the points that the writer will address in the essay.

Despite the students writing to the same essay title, in Table 6.7 it can be seen that the word *Socrates* varies in the number of occurrences per text from 25 to 54. *Martyr* varies from 3 to 11. *Truth* varies from 23 to 42. Interestingly, despite the variances in individual instances, each student uses the word *truth* significantly more often than the word *martyr*. Furthermore, in the main, students use the word *Socrates* more often than the word *truth*. However, Student 10 does not follow this pattern.

The words *was*, *a*, *for* and *the* also display different patterns within each essay despite the fact that the students are answering the exact same essay title. This raises another issue in relation to the influence of task.

No student writes the essay without using all the other words of the essay title apart from *discuss* at some point in their essay. This tells us that task has an effect on the words used in an essay. One other student, writing to a different essay title, uses the word *martyr* in their essay. This is Student 6, who is writing an essay titled 'Socrates committed suicide. Discuss'. Five other students wrote that essay without using the word *martyr* although Student 6 uses the word *martyr* twice for the same essay title. The word *truth* appears in eight other essays[6]. However, the most occurrences per essay in those 8 essays is 11 compared to the least in this essay being 23. It does raise the issue though of how many of the *truth* in the essay would have been there even if it was not mentioned in the title and how many are there because of the title. Interestingly, in the collocates list for the word *the*, an extract of which was shown in Figure 6.2 above, the most frequent word in the 'right 1' column is *truth* occurring 175 times. The full phrase 'martyr for the truth' only appears in the essays taking this title and it appears in all seven of these essays. Here are the details.

Figure 6.9 Plot for *martyr for the truth*



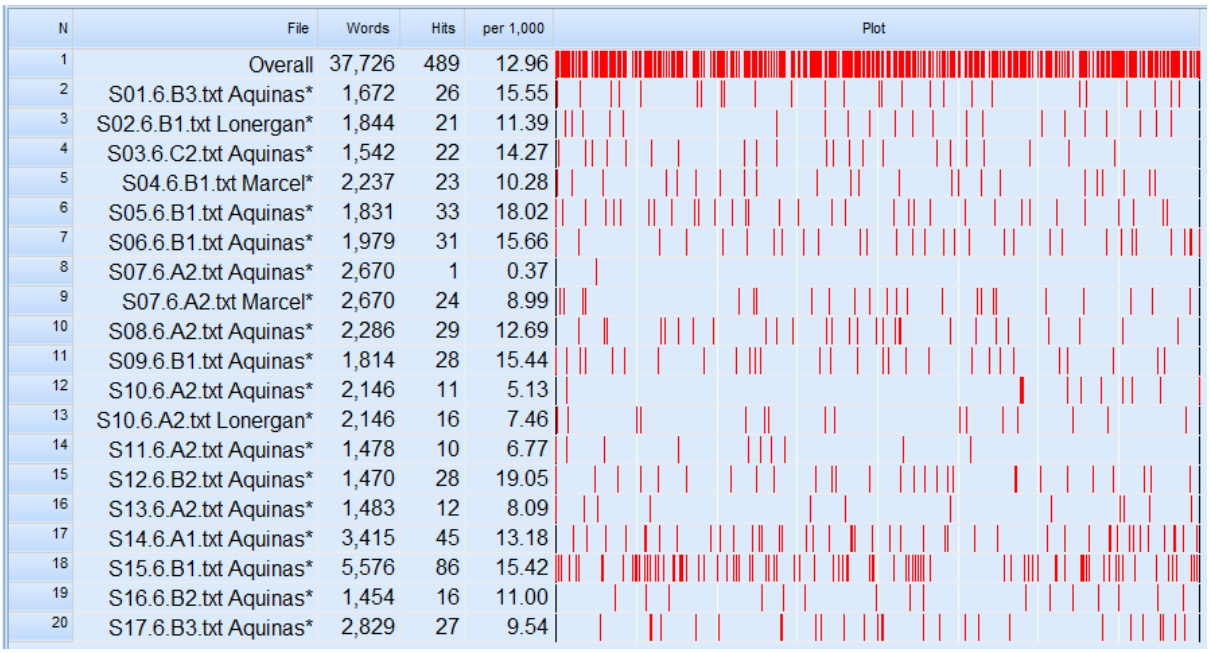
N	File	Words	Hits	per 1,000	Plot
1	Overall	12,193	34	2.79	
2	S03.1.B3.txt	1,147	5	4.36	
3	S04.1.B1.txt	2,665	3	1.13	
4	S05.1.A2.txt	1,588	2	1.26	
5	S08.1.C1.txt	1,489	3	2.01	
6	S10.1.A2.txt	1,549	8	5.16	
7	S13.1.C1.txt	1,688	3	1.78	
8	S15.1.B1.txt	2,067	10	4.84	

Despite the fact that all seven students are writing to the same essay title, the phrase *martyr for the truth* from that title ranges from 2 occurrences in the essay of Student 5 to 10 occurrences in the essay of Student 15. In addition, there is a variance in the relative positioning in the texts of this phrase. Some of the writers have the phrase near the beginning of the essay. Others, such as Student 3 and Student 15, do not use the

6 This means that 15 of the 17 texts in this essay iteration use the word *truth*. If 15 students used this word, 2 students did not.

phrase until slightly further into the essay. Student 8, on the other hand, does not use the phrase until over half way through their essay. There is a greater uniformity closer to the end of the essay as six of the seven students use the phrase in this position. Overall, however, we can clearly see here the uneven effect of the task on each written text. This uniqueness is evident, also, in essay iteration 6 where a distribution plot for *Aquinas*, *Marcel* and *Lonergan* is presented in Figure 6.10 below (for details on essay titles, see Chapter 4).

Figure 6.10 Plot for *Aquinas*, *Marcel* and *Lonergan* in essay iteration 6



N	File	Words	Hits	per 1,000	Plot
1	Overall	37,726	489	12.96	
2	S01.6.B3.txt Aquinas*	1,672	26	15.55	
3	S02.6.B1.txt Lonergan*	1,844	21	11.39	
4	S03.6.C2.txt Aquinas*	1,542	22	14.27	
5	S04.6.B1.txt Marcel*	2,237	23	10.28	
6	S05.6.B1.txt Aquinas*	1,831	33	18.02	
7	S06.6.B1.txt Aquinas*	1,979	31	15.66	
8	S07.6.A2.txt Aquinas*	2,670	1	0.37	
9	S07.6.A2.txt Marcel*	2,670	24	8.99	
10	S08.6.A2.txt Aquinas*	2,286	29	12.69	
11	S09.6.B1.txt Aquinas*	1,814	28	15.44	
12	S10.6.A2.txt Aquinas*	2,146	11	5.13	
13	S10.6.A2.txt Lonergan*	2,146	16	7.46	
14	S11.6.A2.txt Aquinas*	1,478	10	6.77	
15	S12.6.B2.txt Aquinas*	1,470	28	19.05	
16	S13.6.A2.txt Aquinas*	1,483	12	8.09	
17	S14.6.A1.txt Aquinas*	3,415	45	13.18	
18	S15.6.B1.txt Aquinas*	5,576	86	15.42	
19	S16.6.B2.txt Aquinas*	1,454	16	11.00	
20	S17.6.B3.txt Aquinas*	2,829	27	9.54	

Looking at the distribution of the philosophers in the final essay in Figure 6.10, a similar pattern to that of *Socrates* emerges. Student 15 uses the word *Aquinas* 86 times, while Student 11 users the same word 10 times despite the fact that they are answering the exact same essay question. Student 10 is writing an essay about Lonergan yet still has the word *Aquinas* 11 times in the text. Student 7 is writing about Marcel but refers to *Aquinas* on one occasion. Student 10 uses both *Aquinas* and *Lonergan*. In addition to the uneven frequency and distribution of these philosophers as seen in Figure 6.10, there is also a variance in the way that these philosophers are referred to. In some cases, the writer chooses to use a full name for the philosopher as opposed to just their surname.

This, too, is evidence for instantial choices made by the writers, and is investigated further below.

Full Names

The writers have a choice as to how they refer to the philosopher in question. *Aquinas* can be preceded by *Thomas*, *Lonergan* can be preceded by *Bernard* and *Marcel* can be preceded by *Gabriel*. There are 22 instances in essay iteration 6 where this occurs and those instances are shown in the concordance displayed in Figure 6.11.

Figure 6.11 Instances of full names in essay iteration 6

N	Concordance	File
1	Sir Thomas Aquinas was born around the	S01.6.B3.txt
2	to the Philosophy of St. Thomas Aquinas' not only does man	S03.6.C2.txt
3	I have a study of Gabriel Marcels' 'Metaphysic of Hope'	S04.6.B1.txt
4	Thomas Aquinas was a Christian; as such	S05.6.B1.txt
5	based upon reason. St. Thomas Aquinas saw that belief in God	S05.6.B1.txt
6	would cease to be. Saint Thomas Aquinas' argument was that (11)	S05.6.B1.txt
7	of the first three ways of Thomas Aquinas, there are inevitably	S05.6.B1.txt
8	The question though of Thomas Aquinas' devotion to God and	S05.6.B1.txt
9	St. Thomas Aquinas was a medieval	S06.6.B1.txt
10	cannot afford to ignore Gabriel Marcel's essay on the same topic.	S07.6.A2.txt
11	to agree with him. * * * Gabriel Marcel was, I think, acutely aware	S07.6.A2.txt
12	It will also look at Thomas Aquinas' influence on Lonergan	S10.6.A2.txt
13	widely linked to that of Thomas Aquinas, especially that of	S10.6.A2.txt
14	Thomas Aquinas sought to prove the	S13.6.A2.txt
15	the first three of St. Thomas Aquinas' Five Ways. It is my	S13.6.A2.txt
16	and knowledge. Thomas Aquinas has provided fro the	S13.6.A2.txt
17	Although Thomas Aquinas offers five ways to prove	S15.6.B1.txt
18	believe that he exists. St Thomas Aquinas was a philosopher who	S16.6.B2.txt
19	the most influential being Thomas Aquinas. He outlined his proof for	S17.6.B3.txt
20	evil is not God but man. Thomas Aquinas argues that God's	S17.6.B3.txt
21	or is its own cause" (Thomas Aquinas Five Ways: Internet) as	S17.6.B3.txt
22	three of the five ways that Thomas Aquinas puts forward as proof for	S17.6.B3.txt

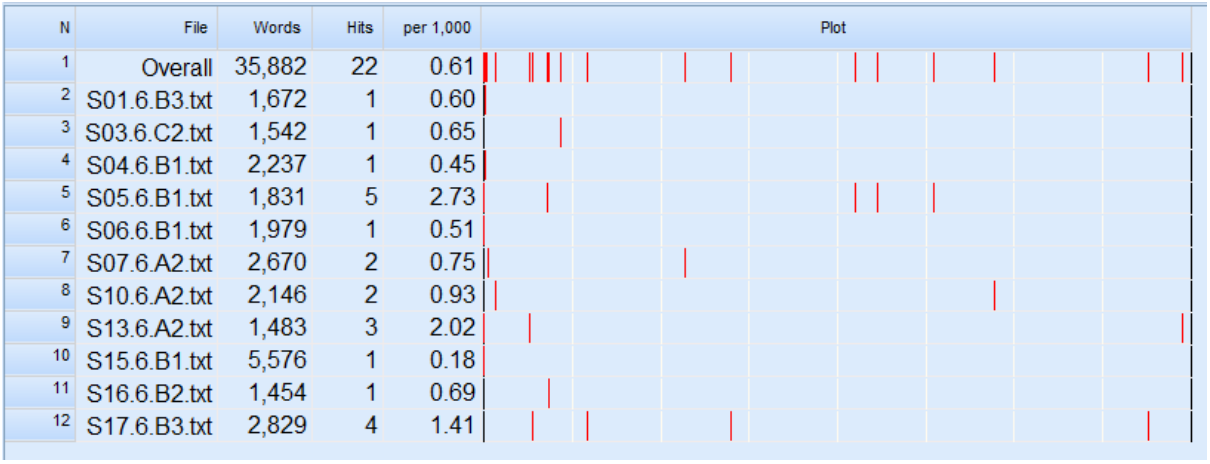
In essay iteration 6, there are 22 examples where the student uses a full name for the philosopher who is the subject of the task as shown in Figure 6.11. Only two of the

philosophers are used with their full name. No writer chooses to use *Bernard* with *Lonergan* despite the fact that *Lonergan* occurs 37 times across two essays. 11 out of the 17 writers used a full form at least once. This means that six writers did not. It is also interesting to note that at times, *Aquinas*, as we have already seen, is referred to *Aquinas*, at other times *Thomas Aquinas*, on five occasions as *St*. or *Saint Thomas Aquinas*, and in one instance as *Sir Thomas Aquinas*. Four of the instances of a full name, with *Sir* and *St*. and without either, are the first words of the essay in question. In another instance, they are very close to the start of the essay, being preceded only by the word *although*. It is worth, at this point, looking at the dispersion of these phrases.

Full titles dispersion

Figure 6.12 shows the dispersion within the texts of the use of full names in essay iteration 6 (the concordance of which is shown in Figure 6.11 above).

Figure 6.12 Dispersion of *Thomas Aquinas* and *Gabriel Marcel*



N	File	Words	Hits	per 1,000	Plot
1	Overall	35,882	22	0.61	
2	S01.6.B3.txt	1,672	1	0.60	
3	S03.6.C2.txt	1,542	1	0.65	
4	S04.6.B1.txt	2,237	1	0.45	
5	S05.6.B1.txt	1,831	5	2.73	
6	S06.6.B1.txt	1,979	1	0.51	
7	S07.6.A2.txt	2,670	2	0.75	
8	S10.6.A2.txt	2,146	2	0.93	
9	S13.6.A2.txt	1,483	3	2.02	
10	S15.6.B1.txt	5,576	1	0.18	
11	S16.6.B2.txt	1,454	1	0.69	
12	S17.6.B3.txt	2,829	4	1.41	

In Figure 6.12, we can see that in the main, the writers use the full name only once. We can also see that in these cases this use of the full name is near the start of the essay. There are some students who do use the full phrase more than once, and the essay with the most frequent use of a full name is that of Student 5.

Student 5 uses a full name for *Aquinas* on five occasions. This is out of a total of 33 instances of Aquinas in that essay text. In three instances, he is referred to as *Thomas*

Aquinas. In one instance he is referred to as *St. Thomas Aquinas*, and in another instance as *Saint Thomas Aquinas*. Between each instance of the full name in this essay, he is referred to by just the surname *Aquinas* at least once. In three instances, the full name is the first mention of *Aquinas* in that paragraph, including that at the very start of the essay. In two instances, he has previously been referred to as *Aquinas* within the same paragraph prior to the use of the full name. What this seems to suggest is that the use of a full name is not linked to a singular deterministic cause. Instead, it seems to be more of an instantial choice, around whether to use a full name and what form that full name should take.

Philosopher as first word

In Figure 6.11 above, we saw that in some instances the full name of the philosopher occurred at the beginning of the essay. In this section, we will examine the name of a philosopher in the initial position, i.e. as the first word or phrase in the text, in the four essay iterations in semester 4. This is a further investigation of uniformity of response to a uniform task. These four essay iterations span a total of 6 weeks in submission times. In the 60 texts in essays 2-5, all written in Semester 4 of a 6 semester degree, there are 534 instances of the words *Lonergan*, *Aquinas* and *Parmenides* combined. These instances are spread across 54 texts, which means that six texts do not use these words despite writing to the same essay titles. The highest mention of anyone philosopher in a single text is 26, while the lowest is 1. In 15 of these cases, the particular philosopher is the first word or phrase of the essay.

In Figure 6.13, we see the names of philosophers in the initial position in essay iterations 2-5. This phenomenon occurs across a range of students and also across the full range of essay iterations. Three students (Students 6, 14 and 17) begin two of their essays with the name of the philosopher, while six students do it on one occasion. It happens six times in essay iteration 2, three times in essay iteration 3, twice in essay iteration 4 and once in essay iteration 5. Furthermore, in line with what we have seen in the previous section, four students begin their essay with the full name of a philosopher.

Figure 6.13 Philosophers in initial position in a text in essay iterations 2-5

N	Concordance	File
1	Parmenides was pre-socratic and lived	S04.2.A2.txt
2	St. Thomas Aquinas was a medieval theologian and	S06.2.B1.txt
3	Aquinas subscribes to Aristotle's mode	S09.2.B1.txt
4	Aquinas agrees with Aristotle in saying	S11.2.B2.txt
5	Aquinas was a great admirer of Aristotle	S16.2.B2.txt
6	St Thomas Aquinas read Aristotle's works from the	S17.2.B3.txt
7	Bernard Lonergan was one of the leading thinkers	S06.3.B1.txt
8	Lonergan's theory of discovering and	S12.3.B2.txt
9	Bernard Lonergan devised the heuristic structure	S17.3.B1.txt
10	Lonergan on the Intellectual Component	S07.4.B1.txt
11	Lonergan (1990: 257) states that	S14.4.A2.txt
12	Lonergan describes the process of	S14.5.A1.txt

Again, we are seeing the individual responses to the task with only a selection of students on a selection of locations choosing to begin the essay with the name of the philosopher in question. It must be noted here that although the name of the philosopher is the first word or phrase in the essay, this does not necessarily mean that this was the first word or phrase written by the student. This is because some students may edit or redraft their essays and we have no way of knowing in the present study what the first word or phrase was written by the student for that essay. However, we can be certain that in these instances shown in Figure 6.13, the name of the philosopher is the first word or phrase read by the intended audience.

In terms of what directly follows the name of that philosopher, it can clearly be seen that each essay is different. However, there are some echoes for one particular student of a pattern that we have already seen in this chapter. Below are two extracts from two essays written by Student 6. Both essays concern the philosopher Aquinas and they were written approximately one year apart.

Example 6.6

> St. Thomas Aquinas was a medieval theologian and philosopher. He was greatly influenced by the philosophy of Aristotle . (S06.2.B1)

Example 6.7

> St. Thomas Aquinas was a medieval theologian and philosopher. Indeed he is regarded as one of the greatest Christian philosophers ever to have lived. He was greatly influenced by the philosophy of Aristotle. (S06.6.B1)

The first sentence of both essays is the exact same. The second sentence of the text from essay iteration 2 is the exact same as the third sentence of the text from essay iteration 6. In essay iteration 6, the student inserted a new sentence concerning *Aquinas* being a *great Christian philosopher*. It is not possible to say exactly why that student decided to insert a different second sentence one year later. However, we can speculate that it may be due to the task of essay iteration 6 which concerns the existence of God and therefore the Christianity of this philosopher becomes important to the task. Yet, we can also speculate that the reason the second sentence is now present is due to the fact that this student has studied philosophy for an extra year compared to when essay iteration 2 was written and did not hold this opinion when essay iteration 2 was written. The reason could be neither of these, and an amalgamation of these or an amalgamation of these with another reason.

It is noteworthy that this student uses similar sentences for two different essay titles written over a year apart. The student may have had a copy of the text from essay iteration 2 to hand when writing essay iteration 6 and this may account for the similarity. It is also possible that the start of both essays was written independently and the similarity between them is indicative of the personal style of the writer. Yet, it is also noteworthy that other students do not follow this pattern and instead show no similarities between the starts of different essay iterations. Furthermore, this particular student, Student 6, begins their essay in iteration 4 with a reference to previous essays:

Example 6.8

> **As discussed in previous essays** Lonergan focussed on the central role, which the act of insight plays in mathematical and scientific investigations. (S06.4.A2)

Example 6.8 shows that this student sees a connection between the essays they write for this particular audience and gives the impression of the continuation of an on-going discussion.

What is of interest to us here is that students receiving similar tasks begin the written product in different ways. In the 60 essays in essay iterations 2-5, 45 essays did not start with the name of a philosopher while 15 essays did so. Although we can say that there is some similarity with the students in that they start with the name of a philosopher, by word 4 of the essay all sentences are different. This is indicative of unique responses to the task set be the assessor.

The pronouns *I, you* and *we*

Tang and John (1999), building on the work of Ivanic (1988), examined 27 first year undergraduate essays collected from students studying English Language at the National University of Singapore. They investigated how the identities of the writers are revealed through the use of the first person pronouns (*I, me, my mine, we, us, our* and *ours*). While Ivanic (1988) identified four aspects of writer identity (autobiographical self; discoursal self; self as author; and possibilities for self-hood in the socio-cultural and institutional context) and suggested that these could be placed on a continuum, the focus was on societal and discourse roles. Tang and John (1999), however, focused on genre roles, creating a continuum ranging from least powerful authorial presence to most powerful authorial presence. The six steps along the continuum, in order of least to most powerful authorial presence, are representative, guide, architect, recounter of research process, opinion-holder and originator.

The first role Tang and John (1999) identify for the writer's voice is that of the representative. This is when the first person pronoun is used to represent a larger group of people. Tang and John note that this role is usually represented in the plural form and can range from people in general (*the English we know today*) to a more specific discourse community (*We know that all dialects are* ...). They also argue that this is the least powerful role that an individual can front as it 'effectively reduces the reader to a non-entity' (1999: 27). The next role outlined by Tang and John is that of guide through the essay. This is when the writer 'shows the reader through the essay ... locates the reader and writer together in the time and place of the essay, draws the reader's attention to points which are plainly visible or obvious within the essay ... and arrives at a conclusion (destination) that he or she presumes is shared by the reader.' (1999: 27). Here they argue that this particular role is usually realized through the plural as it intends to bring the writer and reader together through the essay.

The next role of the writer's voice in an essay, according to Tang and John (1999), is that of the architect of the essay. In many respects this is close to the guide as the architect voice 'foregrounds the person who writes, organizes, structures, and outlines the material in the essay' (1999: 28). The main difference, they argue, is that this is a more powerful voice and does not state the obvious. The fourth role identified by Tang

and John (1999) is that of recounter of the research process. This is when the writer explains actions they undertook.

The fifth role is that of opinion-holder. This is when the writer 'shares an opinion, view or attitude (for example, by expressing agreement, disagreement or interest) with regard to known information or established facts' (Tang and John 1999: 28). The final role they outline, that of originator, is similar to the opinion-holder, but it is stronger in some ways. In this role, the writer advances knowledge claims in the essay and signals these as new.

Hyland (2002) categorizes the use of personal pronouns in professional academic writing into discourse functions. These are, in order of frequency in Hyland's data, stating a goal/purpose, explaining a procedure, stating results/claims, expressing self-benefits and elaborating an argument. He then compares this to student writing and concludes that students significantly underuse authorial pronouns. The question arises here whether professional academic writing can be seen as a model for undergraduate students and whether it is appropriate or desirable for students to mimic the professional academic.

Pennycook argues that pronouns are ‘very complex and political words, always raising difficult issues of who is being represented’ (1994: 173), and points to their simultaneously inclusive and exclusive nature through naming a self, selves and others. In relation to pronouns, including *I, you* and *we*, different positions are referenced within different discourses.

To begin with, let us remind ourselves of the frequency and distribution of *I, you* and *we* when frequency is plotted against the number of texts in which these pronouns occur. This is shown below in Figure 6.14.

Figure 6.14 Frequency and distribution of *I*, *you* and *we*

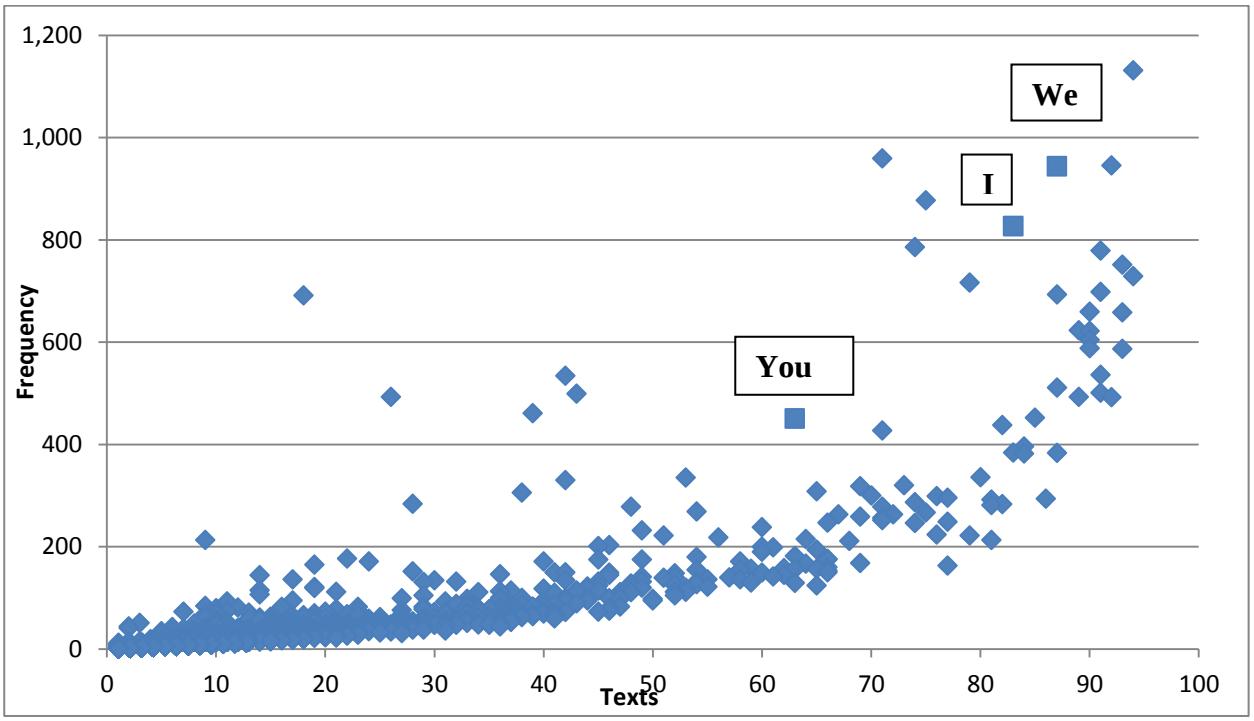
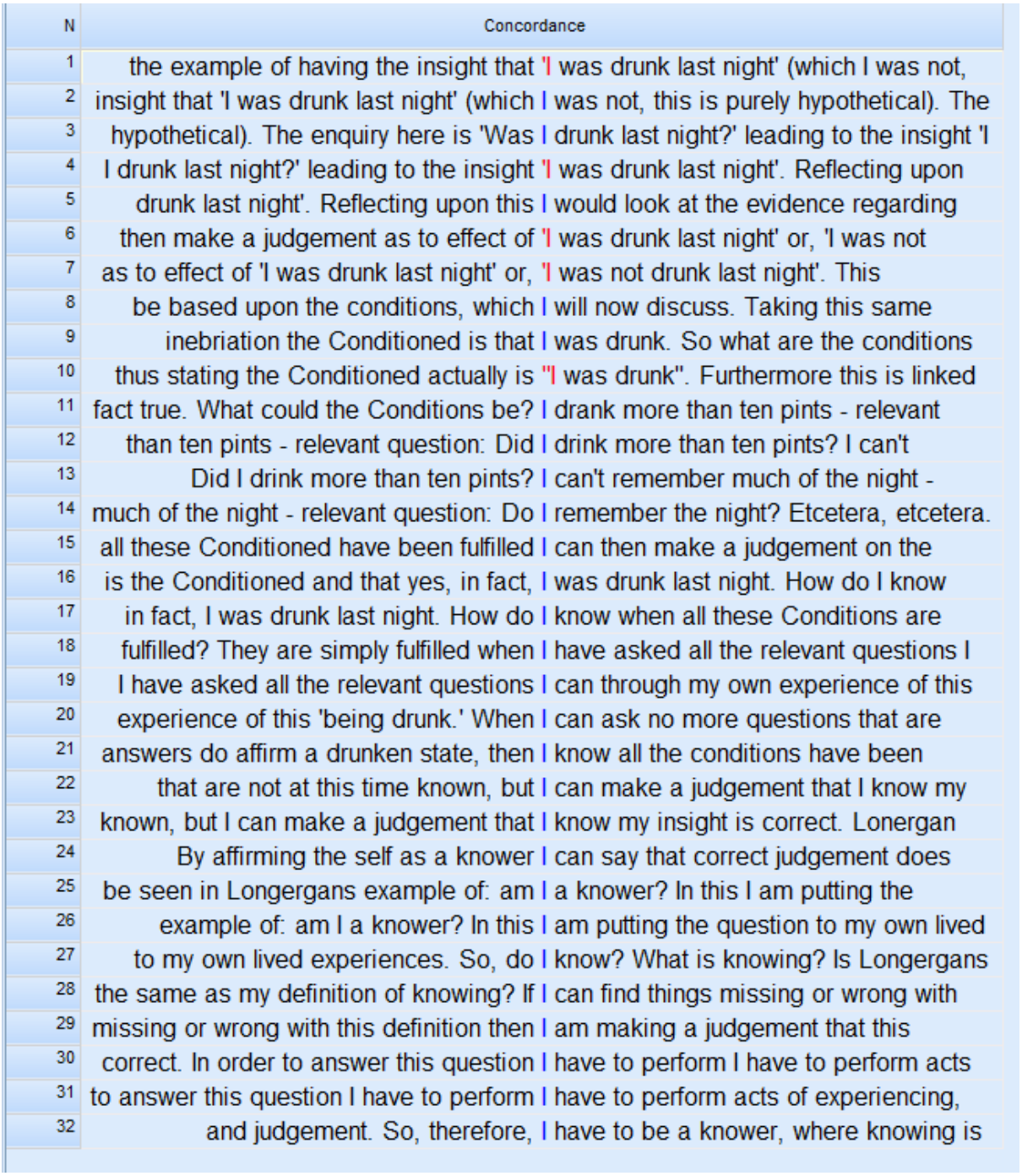


Figure 6.14 shows the frequency and distribution of the pronouns *I*, *we* and *you*. *We* is the most frequent and also appears in the highest number of texts. *You* is the least frequent and appears in the lowest number of texts. The current section examines the three pronouns individually.

I

The text with the highest frequency of the word *I* in the corpus is S13.5.A2. This essay has 22 instances of the word *I* in 590 tokens. The concordance of *I* for this essay is given below in Figure 6.15.

Figure 6.15 *I* in S13.5.A2



N	Concordance
1	the example of having the insight that 'I was drunk last night' (which I was not,
2	insight that 'I was drunk last night' (which I was not, this is purely hypothetical). The
3	hypothetical). The enquiry here is 'Was I drunk last night?' leading to the insight 'I
4	I drunk last night?' leading to the insight 'I was drunk last night'. Reflecting upon
5	drunk last night'. Reflecting upon this I would look at the evidence regarding
6	then make a judgement as to effect of 'I was drunk last night' or, 'I was not
7	as to effect of 'I was drunk last night' or, 'I was not drunk last night'. This
8	be based upon the conditions, which I will now discuss. Taking this same
9	inebriation the Conditioned is that I was drunk. So what are the conditions
10	thus stating the Conditioned actually is "I was drunk". Furthermore this is linked
11	fact true. What could the Conditions be? I drank more than ten pints - relevant
12	than ten pints - relevant question: Did I drink more than ten pints? I can't
13	Did I drink more than ten pints? I can't remember much of the night -
14	much of the night - relevant question: Do I remember the night? Etcetera, etcetera.
15	all these Conditioned have been fulfilled I can then make a judgement on the
16	is the Conditioned and that yes, in fact, I was drunk last night. How do I know
17	in fact, I was drunk last night. How do I know when all these Conditions are
18	fulfilled? They are simply fulfilled when I have asked all the relevant questions I
19	I have asked all the relevant questions I can through my own experience of this
20	experience of this 'being drunk.' When I can ask no more questions that are
21	answers do affirm a drunken state, then I know all the conditions have been
22	that are not at this time known, but I can make a judgement that I know my
23	known, but I can make a judgement that I know my insight is correct. Lonergan
24	By affirming the self as a knower I can say that correct judgement does
25	be seen in Longergans example of: am I a knower? In this I am putting the
26	example of: am I a knower? In this I am putting the question to my own lived
27	to my own lived experiences. So, do I know? What is knowing? Is Longergans
28	the same as my definition of knowing? If I can find things missing or wrong with
29	missing or wrong with this definition then I am making a judgement that this
30	correct. In order to answer this question I have to perform I have to perform acts
31	to answer this question I have to perform I have to perform acts of experiencing,
32	and judgement. So, therefore, I have to be a knower, where knowing is

As can be seen above, at times the *I* in these instances refers to a hypothetical, generic person used by the writer in order to create an example to illustrate meaning. The first example is contained in lines 1-16 which illustrate the hypothetical situation of the writer having had too much to drink the previous night and uses the example to illustrate the content they were asked to discuss in the task, namely judgement:

Example 6.9

> Taking the example of having the insight that '**I** was drunk last night' (which **I** was not, this is purely hypothetical). The enquiry here is 'Was **I** drunk last night?' leading to the insight '**I** was drunk last night'. Reflecting upon this **I** would look at the evidence regarding this - which lead onto the conditions and conditioned - weigh it up and then make a judgement as to effect of '**I** was drunk last night' or, '**I** was not drunk last night'. This judgement would either be that this insight is not correct. This would be based upon the conditions, which **I** will now discuss. (S13.5.A2)

In Example 6.9, the writer themselves points to the hypothetical nature of *I*, pointing out that *this is purely hypothetical*. However, the preceding *I, which I was not*, does refer to the writer in a real non-hypothetical sense.

Line 8 in figure 6.15, the last sentence in Example 6.9, is an exception to this where the writer uses the phrase *which I will now discuss* to perform a narrative function within the text and introduce the next area to be covered in the essay. The remaining instances of *I* in this essay are also generic and concern the other areas the essay was to consider, namely insight and knowing.

When looking qualitatively at the use of *I*, it becomes apparent, in the first instance, that the context and surrounding words are what indicate the intended meaning of the writer and not the pronouns themselves. At times, the writer is referring to themselves, either as an entity within the text or external to the text. At other times, the writer is referring to an imagined entity that, depending on context, can range from the whole of humanity including the writer to a specific person who is not the writer. It must be remembered that at all times each of these roles is being utilised by the writer and thus carries out a function that they deem to be either necessary or desirable for the production of their essay text and are interactional in some way.

The illustrations of the use of *I* from this essay give us some interesting insights. To begin with, it would seem that the high frequency of *I* in this essay was generated by the choice of the writer to use a personal/hypothetical example to illustrate a point. This use of the word *I* does not in any real sense indicate the writer as an individual. However, it is more personalised than some of the alternative choices available, for example using the phrase ‘a person’, in that it positions the writer as part of this set that the generic *I* covers. Secondly, at a particular point, the writer uses the word *I* in a phrase to declare themselves as the narrator. Again, this could be done in a way which did not use the word *I*. It must also be noted here that the other students had the possibility of using a

similar number of the word *I* in their texts. While the use of *I* by Student 15 is not necessarily an explicit use of the writer's voice, it does signal the choices the writer made both in terms of content and how to convey the content to the reader through the conventions of the genre in order to address the task.

In an overall way, this essay has the appearance quantitatively of containing an amount of an explicitly signalled writer's voice due to the high frequency of the word *I*. On further investigation, it seems that only two instances in the essay are the explicit writer's voice. However, the use of the word *I* for the examples in a generic fashion is interactional by nature and personalises the reflective practice of the writer within the text and at the same time includes the reader within that space.

The text with the second highest frequency of the word *I* in an essay is S15.1.B1. A concordance of the instances of *I* in this text is shown in Figure 6.16. In Figure 6.16, there are 30 instances of the word *I*. However, not one of these instances refers to the writer. Each example of the word *I* is contained within quotation marks and refers to Socrates. In a similar way to the example from Student 13 above, the high frequency of the word *I* is indicative of a set of choices made by the writer. These choices did not concern the insertion of an explicitly signalled writer's voice. Instead, they concern choices around how much quotation to insert into the text and how much of that to quote directly. Furthermore, in a similar way to the example examined from Student 13, all other students had the possibility of creating a text with a similar number of the word *I*. They did not.

Figure 6.16 *I* in text S15.1.B1

N	Concordance
1	the jury that he is not a sophist. "I should certainly become a proud
2	a proud and gentlemanly figure if I understood these things, but in fact
3	these things, but in fact gentleman I do not " (3) Socrates is on a mission
4	Oracle who said he was the wisest. "I am only too conscious that I have no
5	wisest. "I am only too conscious that I have no claims, no wisdom, great or
6	what can he mean by asserting that I am the wisest man in the world ? He
7	wasn't. "And in conversation with him I formed the impression that although
8	wiser than the politician. "Well, I am certainly wiser than this man " (7)
9	do. "At any rate it seems that I am wiser than he is to this small
10	than he is to this small extent, that I do not think that I know what I do not
11	small extent, that I do not think that I know what I do not know " (8)
12	that I do not think that I know what I do not know " (8) Socrates was
13	to find the truth of the Oracle. "I want you to think of my adventures
14	what the implications were. "As I pursued my investigation at the gods
15	you have the true fact , which I present to you without concealment
16	or suppression, great or small. I am fairly certain that this plain
17	that my statements are true, and that I have described correctly to the
18	now or later, you will find that facts as I have just described them" (11) In his
19	you acquit me or not; you know that I am not going to alter my conduct, not
20	going to alter my conduct, not even if I have to die a hundred death " (13)
21	justice so he did not support them. "I thought that it was my duty to face it
22	beliefs. "on this occasion , however I again made it clear, not by my words
23	tells them exactly what he thinks. "But I do not think then that I ought to stoop
24	he thinks. "But I do not think then that I ought to stoop to servility because I
25	I ought to stoop to servility because I was in danger, and I do not regret
26	servility because I was in danger, and I do not regret now the way in which I
27	I do not regret now the way in which I pleased my case " (17) Socrates
28	what the implications were "When I leave this court I shall go away
29	were "When I leave this court I shall go away condemned by you to
30	people in the other world. "Above all I should live to spend my time there,

Taking one more essay text, S02.2.A2, we get a different use of the word *I*.

Figure 6.17 *I* in text S02.2.A2

N	Concordance
1	his reasoning behind his ideas appear satisfactory. I agree fully with the concept that what cannot be
2	what cannot be thought cannot be and vice versa. I would find it hard to accept otherwise, say, for
3	be thought. Dealing with another metaphysical topic, I find it difficult to fault Parmenides' argument
4	Having tried to imagine absolute nothingness, I have found it impossible, as my mind is ever
5	is ever present. Having said that, it is inevitable that I would support Parmenides' argument where he
6	and there always will be something. Therefore, I agree with the characteristics of indestructibility and
7	philosophy are not as convincing to me. Although I agree with the viewpoint that everything 'be's', I
8	I agree with the viewpoint that everything 'be's', I would have to differ on Parrnenides' disbelief in
9	on Parrnenides' disbelief in degrees of being-ness. I do not endorse the idea that there is no difference
10	and also to Aristotle's hierarchy of substances. I believe that in the same way and in the same order,
11	to Parmenides, being is timeless. Although I was slightly confused by this notion when I first
12	Although I was slightly confused by this notion when I first heard it, especially as I was to discover that
13	by this notion when I first heard it, especially as I was to discover that there is no future, I soon
14	as I was to discover that there is no future, I soon understood what was being said. As being
15	if it is going to be in the future" (Parmenides) Overall I am very impressed with Parrnenides' outline for the

In lines 1 to 10 and in line 15 the writer here uses the word *I* to give their own opinion*; I agree fully* …, *I find it hard to accept* …, *I would support* … etc. In lines 11 to 14 the writer is using the word *I* to refer to themselves as somebody outside the text who was engaged in a learning process:

Example 6.10

> Although **I** was slightly confused by this notion when **I** first heard it, especially as **I** was to discover that there is no future, **I** soon understood what was being said. (SO2.2.A2)

Here, they are, on the one hand, outlining the steps they went true in understanding a concept and, on the other hand, introducing and then dismissing a possible criticism of the concept they are writing about.

Explicit roles

At times, the writer explicitly points to the role they inhabit when they insert their own voice into the text. Examples of this can be seen in Figure 6.18.

Figure 6.18 Explicitly signalling roles

N	Concordance	File
1	can the phrase ‘is not’ not be thought? As a child I used to try to imagine that I	S04.2.A2.txt
2	because we were brought into being .As a Christian I believe that I was	S06.2.B1.txt
3	‘I cannot doubt that I doubt that I exist as a conscious thinking being. He	S01.3.C1.txt
4	the typical observer at a match as a male. Once I experienced insight,	S02.3.A2.txt
5	at their own pace through insight. As a conscious being, a female, and	S04.3.B3.txt
6	theoretical knowledge. In my own case as a B.Ed student, there are many who	S14.3.A1.txt
7	causing us to act in such a biased way. As a B.Ed student, it is very apparent	S14.4.A2.txt
8	of the knower. By affirming the self as a knower I can say that correct	S13.5.A2.txt
9	these arguments, as the case may be. As a believer in God, albeit a sceptical	S13.6.A2.txt

In the examples in Figure 6.18, at times the writers explicitly signal the role they inhabit. In the examples in MICUPE, these concern religion, gender, childhood, being and studenthood. This too points to the instantial nature of their choices. Only at certain times do they realise the need to qualify the roles, despite the fact that we must assume that religion, gender and studenthood remain constant for the duration of the period under study. In addition, despite the fact that other students certainly have a gender and are students in the same programme, they show no realisation of this phenomenon.

You

The text with the highest frequency of the word *you* is S10.6.A2. This text has 36 instances of the word *you* in 2,146 tokens. Interestingly, the word *I* is more frequent in the overall corpus than the word *you*, yet the text with the highest incidences of *I* contains 32 occurrences while that with the highest incidences of *you* contains 36 occurrences.

Figure 6.19 Concordance of *you* in S10.6.A2

N	Concordance
1	how their works can be linked. When you understand something it is non-
2	is non- spatial and non- temporal. Say you are trying to understand how to do
3	how to do fractions in maths and you are having trouble understanding it.
4	a little clearer and then it clicks and you understand it. "When I feel "Aha!
5	It is not important how long it took you to understand it anymore or where
6	we try to understand it. For example, you are walking down the street one
7	walking down the street one night and you see a group of people crowding
8	around something in the street. You might think, there is a fight going
9	This is your experience. To be sure you go closer and you see the person
10	To be sure you go closer and you see the person on the ground, with
11	This is your understanding or insight. You then reflect on your insight, to do
12	then reflect on your insight, to do this you ask all the relevant questions, such
13	evidence of a fight etc. And then when you have asked all the relevant
14	have asked all the relevant questions you make a judgement based on this,
15	injured on the ground. But how do you know what questions to ask? Your
16	of the situation implies how well you will know what the relevant
17	are. For example, your dog is sick and you think he might have been poisoned,
18	think he might have been poisoned, if you take him to the vet the vet will know
19	is causing the dog's illness. However if you took your dog to the bakery the
20	All the conditions have been filled when you answer all the relevant questions
21	answer all the relevant questions that you can come up with. You then make
22	questions that you can come up with. You then make an intelligent judgement
23	that there were some questions that you didn't ask or answer but you make
24	that you didn't ask or answer but you make a judgement on what you
25	but you make a judgement on what you know at the time. "Being, then, is
26	strongly is this one of reflection. You are asked to look back on your
27	in to other lessons I teach? Etc", here you are asking all the relevant
28	all the relevant questions and finally you make a judgement based on these
29	way to approach things. It leads you in a positive direction for life as
30	you in a positive direction for life as you are always looking back on
31	back on situations and seeing how you can improve on them for a more
32	bits and pieces as we go through life. If you want to learn how to cook, you
33	life. If you want to learn how to cook, you can't in one flash know everything
34	is to know about cooking, it takes time, you learn it one dish at a time.
35	logical sense as to say that what you know is non-intelligible is like
36	the hob and water must be added, then you have to think, how to heat it, you
37	then you have to think, how to heat it, you turn on the cooker, where does the

Looking at the concordances of *you* from this essay shown in Figure 6.19, we can assume that, similar to some uses of the word *I*, the referent of this word is a hypothetical generic entity. We do not assume that, as in line 19 *if you took your dog to the bakery*, the reader would take the dog to the bakery should that dog be sick. In fact, the reader may not even have a dog. In a similar vein, the reader may not want to learn how to cook (line 32 *If you want to learn how to cook*), although they may be happy with a suggestion that they personally make intelligent judgements (line 22 *You then make an intelligent judgement*) along with all the other referents of the generic *you* in this example.

Figure 6.20 Concordance of *you* in S01.1.B3

N	Concordance	File
1	and throughout this essay I will prove to you that through sarcasm and self	S01.1.B3.txt
2	the issues that I have picked out will lead you too to believe that he did indeed	S01.1.B3.txt
3	place and other such gathering areas. You must understand that the jury consisted	S01.1.B3.txt
4	of only ten or twelve. Due to this fact you would imagine that Socrates would	S01.1.B3.txt
5	himself which will do him no good as you will see at a later stage. After all this he	S01.1.B3.txt
6	my faculties " (29d) and he also states that "you know that I am not going to alter my	S01.1.B3.txt
7	I am really pleading on yours, to save you from misusing the gift of God, by	S01.1.B3.txt
8	by killing him that "vengeance shall fall upon you with a punishment far more painful that	S01.1.B3.txt
9	I hope that from your reading of my essay you too will agree with my choice of title.	S01.1.B3.txt

In contrast, the above instances of the word *you* in Figure 6.20 taken from one text in essay iteration 1 demonstrate a different use of this word. At times *you* is directed at the reader, for example in lines 1, 2 and 9. Lines 3 and 4 refer to the reader again, but in a slightly different way to lines 1, 2 and 9. In lines 1, 2 and 9, the reader is somebody who has set the task and the student is outlining how that reader should think by the end of the essay. In lines 3 and 4, the reader is somebody who is to be influenced by the writer in a certain way so the writer directs them towards what they reader should understand or imagine. In line 5, the reader is somebody engaged in the text who has to be narrated to. Instead of pointing out what the writer will write about later (see above for an example of this happening through the use of the word *I*), the writer is pointing out what the reader will see.

In lines 6, 7 and 8, the word *you* is contained in a quote from Socrates and in this case refers to the jury at the trial of Socrates that took place over 2000 years ago. It does not refer to the reader.

We

I is more common in the overall corpus than the word *we*, yet, in a similar way to *you*, the text with the highest frequency of the word *I* contains 32 instances while the text with the highest frequency of the word *we* contains 43 instances. The text with that highest frequency of the word *we* is S06.3.B3. A sample of 10 concordance lines from that text is given below in Figure 6.21.

Figure 6.21 10 concordance lines of *we* from S06.3.B1

N	Concordance	File
1	Structure. As human beings we are naturally inquisitive. We	S06.3.B1.txt
2	beings we are naturally inquisitive. We have an innate desire and	S06.3.B1.txt
3	an innate desire and ability to think. We are in the constant pursuit of	S06.3.B1.txt
4	pursuit of knowledge. As children we are always asking questions	S06.3.B1.txt
5	us; our knowing is always dynamic; we are always moving on to the	S06.3.B1.txt
6	moving on to the next step."(p.60). We look at things reflectively to help	S06.3.B1.txt
7	things, and to clarify explicitly what we know. Lonergan was concerned	S06.3.B1.txt
8	questions "What is happening when we are knowing?" and the	S06.3.B1.txt
9	unknown .It makes us question what we do not know to systematically	S06.3.B1.txt
10	gain some insight into it .We know that it is unknown, we	S06.3.B1.txt

Above in Figure 6.21 is a selection of 10 instances out of the 43 occurrences of the word *we* in text S06.3.B3. Line 1 gives us an insight into this word as it immediately positions the *we* as part of a set of human beings. Obviously, this set of human beings includes both the reader and the writer but does not refer to either specifically. The innate desire in line 2 and the constant pursuit in line 3 are assumed to be generic to all human beings. The other instances of *we* in this text are of a similar nature, referring to human beings in a generic sense. However, both the reader and the writer are assumed to be part of that set. This phenomenon bears similarity to the generic uses of *I* and *you* already seen.

There are 10 instances of the word *we* in text S17.1.B2. Their use is quite different to that seen in the previous text. Those examples are shown below in Figure 6.22.

Figure 6.22 *we* in S17.1.B2

N	Concordance	File
1	never wrote anything down on paper so we are entirely dependent on the works	S17.1.B2.txt
2	life, trial and eventual death. As a result we are given an image of Socrates, as	S17.1.B2.txt
3	great passion for truth. The main basis we have for learning about the	S17.1.B2.txt
4	is "the apology" written by Plato. Here we meet Socrates as he defends	S17.1.B2.txt
5	an entirely different manner to the way we may view it and as we read "the	S17.1.B2.txt
6	to the way we may view it and as we read "the Apology" we are aware of	S17.1.B2.txt
7	view it and as we read "the Apology" we are aware of this view. It becomes	S17.1.B2.txt
8	life or after death" (19) Also, because we have studied Socrates in some	S17.1.B2.txt
9	have studied Socrates in some detail we are aware of his need to seek the	S17.1.B2.txt
10	through improper self-defence and we must also bear in mind that he once	S17.1.B2.txt

In the instances of *we* shown in Figure 6.22, the *we* is more interactional than those instances shown in Figure 6.21 as it encompasses and narrower set of people and positions those people within the argument within the text and, as a result, attempts to guide the reasoning process of the reader. At times, the *we* seems to refer to philosophers engaged in the task of studying a certain text. At other times, the *we* could be argued to solely include the writer and the reader.

Overall, all three of the pronouns *I*, *you* and *we* can be used in a generic sense and a non-generic sense. For the ones that are used in a generic sense, the writer is making a set of choices as there are other options available to them. For the ones that are used in a non-generic sense, the writer is still making choices as to whether they want to bring the reader into the text or bring themselves into the text to achieve certain discourse functions. It is the sum of these instantial choices that leads these three words to appear in their relative positions on the frequency list. Naturally, should the writers have made different choices at those points in time, these words would appear in different positions on the frequency list.

93 out of the 94 essay texts include at least one of the pronouns *I*, *you* or *we*. The one text that does not is S17.4.C3. However, in this text a generic pronoun that performed an interactional function is used. This pronoun is *one*. The concordance from that text is shown below in Figure 6.23.

Figure 6.23 *one* in S17.4.C3

1	can be explored and argued about. One may argue "from the	S17.4.C3.txt
2	insights that would help aid one develop a balanced and	S17.4.C3.txt
3	and well-rounded viewpoint. As one's mind isn't open to other	S17.4.C3.txt
4	to be a reflection of the type of day one experienced. Sleep is the time	S17.4.C3.txt
5	consciousness they would disturb one's sleep as well as violating ones	S17.4.C3.txt
6	well as addressing the issue of how one may explore a desired insight,	S17.4.C3.txt
7	mind isn't open to other insights one may be said to have a biased	S17.4.C3.txt
8	of the extrovert world leads one to become absorbed in the	S17.4.C3.txt
9	in a type of daydream and robs one of the developments of common	S17.4.C3.txt
10	navigate through the wear and tear one's nerves have endured while	S17.4.C3.txt
11	and are visible to others through one's rejection of further questioning	S17.4.C3.txt
12	the particular insight. However when one excludes an insight they are	S17.4.C3.txt
13	where the insight is unwanted. When one experiences an insight it	S17.4.C3.txt
14	neural demand functions. While one sleeps the ego and the persona	S17.4.C3.txt

The use of the pronoun *one* in Figure 6.22 is in some ways similar to the generic uses of the pronouns *I*, *you* and *we*. For the purpose of the current section on pronouns in the corpus, the discrepancies of this one text are critical to our understanding of the choices made by the writers. We have seen that a writer chooses to make a point in a certain way and that choice then influences the language choices that writer makes. In this essay from Student 17, similar choices are made to use a generic set of people in a hypothetical sense to illustrate a philosophical point. However, this student chooses to use the word *one*, while other students chose to use words such as *I*, *you* and *we*. Although the broad notion of using a hypothetical real-person example (if this is not a contradiction) appears in some texts but not in others, where it does appear its realisation can have various forms. Whether the author highlights the *I*, the *you*, the *we* or the *one* is their own choice after they have made the choice to use this rhetorical device of a hypothetical generic example. Furthermore, although in some ways these four words may be broadly similar, the use of the various types must have a different, however small, effect on the reader.

6.3 Conclusion

This chapter has examined some of the 14 words that appear in all 94 essay texts, the use of philosophers' names and the three pronouns *I*, *you* and *we*. For each one of these, we have clearly seen that in individual texts, the frequency, distribution and use of these words varies. Through examining these words, we have seen that each student responds to similar, or at times the same, task written in the same genre for the same reader at the same six points in time in a unique way. This suggests that the research question posed in Chapter 1 concerning how writers respond to genre, task and audience can be answered in that all students respond uniquely to these concerns, that the complete text is a response to these concerns and not simply certain words, and by focusing on certain indicators of the writer response, we see that each text has a different realisation of these indicators.

If we see the word *the* as functioning with certain propositions to create complex noun phrases and see this as an expected part of the genre, we can clearly conclude that each writer in the corpus responded to generic concerns in different ways. Of course, this is not the only way of responding to generic concerns, nor is it the only generic concern involved when writing an undergraduate essay. However, at this point that is not a constant because it is sufficient to say that the influence of genre on these texts is different. This is in accordance with the theoretical framework as outlined in Chapter 3.

If we see the mentioning of philosophers' names as a response to the task, again we see that each text is unique. Each text addresses the task but does so in a manner that is not replicated in any other text. Furthermore, when the philosopher's names were examined in detail, it became clear that riders be made instantial choices about how to refer to these philosophers. This is in accordance with the theoretical framework as outlined in Chapter 3.

The pronouns *I*, *you* and *we* are used in an interactional way. Although they can be used in a generic sense, this use is still to some degree interactional in that it brings both the reader and the writer into the text. In a similar fashion to the other two features examined in this chapter, these pronouns are uniquely distributed and used in each individual text and give empirical evidence for the uniqueness posited in the theoretical framework of Chapter 3.

Taken together, these three features suggest that the choices that the writers make are based on influences such as task, the genre and audience. However, it is also apparent that these influences have differing effects on each writer at particular points in time. Therefore, returning to the research questions of the current research, the response to genre, task and audience is not standardised either across writer or situation. No essay iteration elicits the same response from two or more students and each student changes how they respond over time. These changes will be investigated in Chapter 8. Furthermore, it seems that the use of these features is the result of instantial choices determined by localised factors within the text in combination with the influence of genre, task and audience.

This chapter has examined three features identified in Chapter 5 as standing out in terms of frequency and distribution when compared to all types in the corpus. The following chapter, Chapter 7, will examine some of the types with much lower frequencies in the overall corpus and investigate how those types are distributed across the essay iterations. By doing this, we will further show the uniqueness of text within and across essay iterations and again highlight the instantiality of writer choice as posited in the theoretical framework as outlined in Chapter 3 of the current research.

Chapter 7 The relativity of single-use and multi-use types

7.1 Introduction

In Chapter 5, we examined distribution and frequency of types. We saw that the frequency and distribution do we not always correlate. Furthermore, there is a set of types in the corpus that are limited to one text and at times, to one occurrence in that text.

In Chapter 6, we examined three sets of words that displayed an unusual distribution when compared to the rest of the types in the corpus or to BAWE. Here, we saw that there is a uniqueness in the distribution within the texts of these types and furthermore there is a uniqueness in the function and meaning of these types which depends on the localised context that is created by instantial choice of the writer within the text.

In both Chapters 5 and 6, there was an apparent randomness to the distribution and use of types. This is despite the essays coming from the same cohort of students in the same discipline written for the same reader. Furthermore, within each essay iteration there was a range of use of types, some of which were used by some or all writers, others that were used by one writer only. In the current chapter, we will investigate all types, including the less-common types, and examine their distribution according to essay iteration and text. Then, we will begin to look at the corpus in a dynamic way as an entity created from the addition of texts. In the next chapter, Chapter 8, we will examine this dynamism in relation to the essay iterations and the individual students.

7.2 Number of types per essay iteration

In MICUPE, there are 7,493 types. We have already seen that these types can be examined, or ranked, according to either frequency or distribution across texts. When

frequency and distribution are correlated, depending on the magnification used in terms of text, sub-corpus or full corpus, some lexical items stand out at some or all of these levels. We also saw that while some words appear in all texts, either within a sub-corpus or within the whole corpus, other words are limited to one text only and also at times one occurrence only.

There are 3,011 words in the corpus that appear only once. Given that such is the case, these words must also be limited to one text only. A further 387 types have more than one occurrence but all occurrences of that type are limited to one text. This gives a total of 3,398 types that are limited to one text only within MICUPE. If a type is limited to one text, it must also be limited to one sub-corpus.

The 6 individual essay iterations use a different number of types and obviously there must be some overlap in types between the semesters. We have already seen in Chapter 5 that there are types that appear across all texts and therefore appear in all essay iterations. Table 7.1 outlines the number of types per essay iteration.

Table 7.1 Types per essay iteration

Iteration	**Number of types**	**Number of texts**	**Number of tokens**
Essay 1	3,378	17	30,426
Essay 2	1,660	16	14,696
Essay 3	2,409	16	20,895
Essay 4	2,429	15	17,323
Essay 5	1,852	13	13,071
Essay 6	3,392	17	37,812

If we look at the number of types per sub-corpus, and add these together, we get a total of 15,120. We know that there are 7,493 types in MICUPE. This gives us an indication that there are some types that share use across the sub-corpora. We saw in Chapters 5 and 6 that there are only 14 types that appear in all texts in MICUPE. There are, obviously, a lot more than 14 types appearing across all semesters but yet not appearing

in all texts within those semesters. In addition, there are types, more than likely, that appear in more than one of the sub-corpora but possibly not in every sub-corpus.

Looking at Table 7.1, in some of the sub-corpora, it would seem that the number of texts dictates the number of types. This is true of essay iterations 1 and 6 where the 17 texts in both iterations have 3,378 and 3,392 types per sub-corpus respectively, which is only a difference of 14. However, essay iterations 2 and 3 also have the same number of texts as each other, in this case 16, yet have a great difference in the number of types used in those texts. In fact, essay iteration 3 uses 749 more types than essay iteration 2. This equates to roughly 45% more types in essay 3 than in essay 2. In addition, essay iterations 3 and 4 are quite similar in number of types with a difference of only 20 types, even though they differ in number of texts and number of tokens. Essay iteration 5 has fewer texts and fewer tokens than essay iteration 2, yet has more types. In this case, the temptation is to attribute the difference in number of types to the difference in the number of tokens. However, as we have just seen, essay iterations 1 and 6 have similar number of types, yet have very different numbers of tokens. It would seem inconsistent to, in one instance, attribute the difference to the number of tokens and at the same time attribute a similarity to a number of texts. Therefore, we must accept at this point that within the sub-corpora there is a variance in the number of types the students choose to use in their essays. This variation, since it does not correlate to either number of texts or number of tokens, must have a different explanation for its occurrence.

We have already established that there are 7,493 types in MICUPE. Keeping this in mind, Table 7.1 above shows us that, within MICUPE, no one essay iteration uses 50% or more of those types. In fact, essay iteration 6 is the essay iteration with the largest number of types, and this iteration uses 45.35% of the types found in the full corpus. Essay iteration 2 is the essay iteration that uses the fewest number of types, and this equates to 22.15% of the types found in the full corpus. In Table 7.2 below, we can see that the percentage of types does not equate to the percentage of words within MICUPE.

Table 7.2 Percentage of types and percentage of tokens

Iteration	% of types	% of tokens
Essay 1	45.07606	22.65711
Essay 2	22.15105	10.94356
Essay 3	32.14572	15.55973
Essay 4	32.4126	12.89979
Essay 5	24.7131	9.733485
Essay 6	45.26288	28.15718

In Table 7.2, no essay iteration shows a perfect match between percentage of tokens in the corpus and the percentage of types used in that sub-corpus. We could not expect an exact match in this regard since the percentage of types is relative and shows an overlap, adding up to just over 200%, whereas the percentage of tokens is fixed proportionately, adding up to 100%. However, there is not a proportional match between types and tokens, with essay iterations 1 and 6 having 45% of types per essay, yet iteration 1 has 22% of tokens but essay iteration 6 has 28% of tokens.

If we reconstitute the corpus in a slightly different way where one of the sub-corpora is omitted, the number of types in the new corpus when subtracted from the overall number of types in MICUPE would indicate how many types are unique to the omitted semester. This is shown in Table 7.3 below.

Table 7.3 Number of types unique to the sub-corpora

Iteration	Number of types when sub-corpus is omitted	Number of types in this sub-corpus only
Essay 1	6,167	1,327
Essay 2	7,179	315
Essay 3	6,939	555
Essay 4	6,823	671
Essay 5	7,151	343
Essay 6	6,408	1,086

Table 7.3 shows us the number of types that are unique to each of the sub-corpora. If we add the number of types that are unique each sub-corpus, we get 4,297 types in the corpus are unique to one essay iteration. This means that just over 57% of the types in the corpus appear in one essay iteration only.

Types per essay

Table 7.4 below shows the number of types in each individual essay text. Blank spaces are left where there is no essay in the corpus for the particular student in an essay iteration.

Table 7.4 Number of types per essay text

	Essay 1	Essay 2	Essay 3	Essay 4	Essay 5	Essay 6
Student 1	420	300	417	370	365	440
Student 2	739	295	506	381	296	430
Student 3	411	359	369		260	431
Student 4	809	344	544	327	483	809
Student 5	500	237	369	252		470
Student 6	588	471	502	513	382	513
Student 7	554	186	295	372	327	907
Student 8	495	303	501	464	268	614
Student 9	604	290		549	392	513
Student 10	463	323	332	391	230	519
Student 11	527	213	336	305		387
Student 12	508	273	359	380		367
Student 13	491	302	431	358	181	387
Student 14	664		633	694	473	744
Student 15	575	238	499		365	858
Student 16	563	258	413	385		387
Student 17	624	237	370	381	206	677

The table above further illustrates the crossover in the use of types. If there was no re-use of types across the texts, there would be 40,843 different types in the corpus. Again, as we already know that there are 7,493 types in the corpus, there is a degree to which the types appearing in one text also appear in other texts.

In Table 7.4, there is not a consistent pattern either across the students or across the essay iterations. Each student displays a range in the number of types they use for each essay iteration. Each essay iteration displays a number of types used by individual students within that iteration. For example, Student 7 has the least number of types per essay of any student in essay iterations 2 and 3 and has the highest number of types per essay in essay iteration 6.

Figure 7.1 Types per student

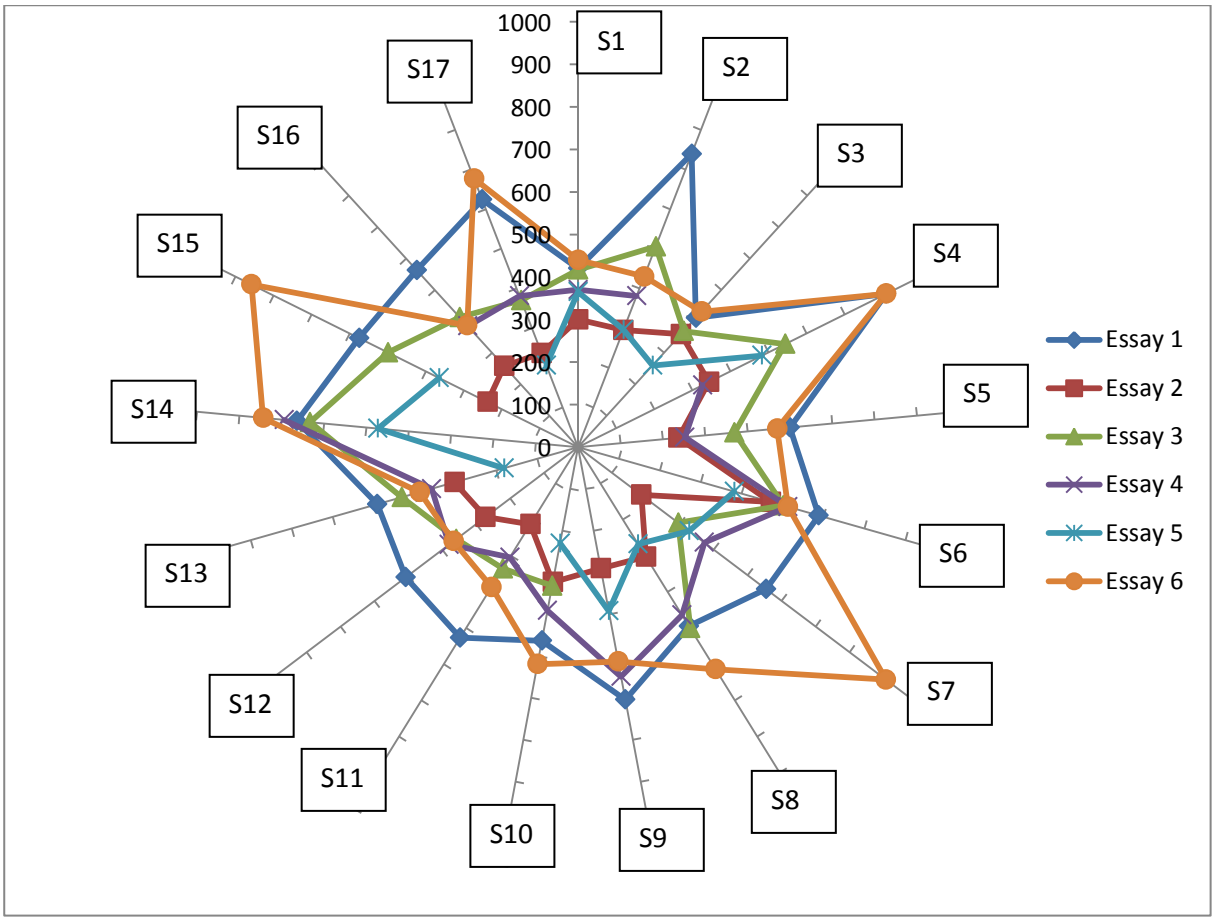


Figure 7.1 shows the number of types for each essay iteration by each of the 17 students in a radar chart. Each spoke within the chart represents a student. The closer to the centre of the chart indicates a lower number of types while the further out the spoke indicates a greater number of types. Some students display a large range in the number of types used in different essay iterations, for example Student 7 and Student 15. It is of

particular note that the text with the lowest number of types in the whole corpus was written by Student 7, and the text with the highest number of types in the whole corpus was also written by Student 7. Other students show a narrower range in the number of types used in different essay iterations, for example Student 1 and Student 3.

For all students, the essay iteration with the highest number of types is either essay iteration 1 or essay iteration 6. Furthermore, the essay iteration with the fewest number of types for all students is either essay iteration 2 or essay iteration 5. However, this apparent pattern is not as fixed as this would suggest. In some cases, essay iterations 1 and 6 are not the two essays with the highest number of types. For Student 2, essay iteration 3 has more types than essay iteration 6 while Student 9 uses more types in essay iteration 4 than in essay iteration 6. In the main, essay iteration 2 has the fewest number of types for the students. However, this is not the case for Students 3, 6, 10, 13 and 17.

The variety shown by each individual student displays a distinct lack of a set pattern. It would seem that in relation to the number of types used in a particular text, the students that are included in this corpus do not fit a stable profile. This suggests that the number of types used to create a text in a set context is not determined solely by a factor inherent and permanent within the writer. The writers are subject to change, as discussed in the theoretical framework of Chapter 3. Furthermore, as seen in the previous chapter, factors such as task, genre and audience also have instantial realisations of their influence. It would seem that the use of types in each individual text is determined by a complex interaction between the writer at a particular point in time and task, genre and audience among other factors. This complex interaction produces a unique text because the interaction is different each time it happens. This notion will be further examined in Chapter 8.

There is further variety in the essay iterations that have the highest number of types as well as the essay iterations with the lowest number of types. Again, it would seem that in relation to the number of types used in a particular text, the essay iterations included in this corpus do not fit a stable profile. This suggests that the number of types used to create the text in a set context is not determined solely by the audience and genre as these remain constant. It is worth noting here that although audience and genre may be constant, this does not mean that they are fixed notions for the writer. The variety of

types within the essay iterations also suggests that task is not a sole factor in determining the number of types used in a particular essay text.

In relation to the number of types that a student uses in order to write a particular essay at a particular point in time, three points become apparent. Firstly, each student does not have a set profile in the number of types that they consistently use. They vary their use of types for each essay iteration. Secondly, each essay iteration does not have a similar effect on each student with regard to the number of types they use. Thirdly, there is no clear pattern in the change of the number of types used as the students progress through the degree programme from essay iteration 1 to essay iteration 6. No student displays a clear pattern of either increasing or decreasing the number of types they use in a linear fashion over the course of these six essay iterations.

In Chapter 6, we saw that a similarity of types across texts does not necessarily indicate a similarity of use for those types. Therefore, we can say that the writers represented in this corpus have made choices about which types to use and what function those types have within the essay texts. These choices seem to be based on instantial decisions as evidenced by the range of types used across students and essay iterations as well as within the work of individual students and within the individual texts of an essay iteration. This variety, and uniqueness, exists despite the similarities of task, genre and audience. We can therefore conclude that in relation to the types used by a writer, task, the genre and audience do not have a set, uniform influence on the texts.

In Chapter 8, we will further investigate the changes over the course of the six essay iterations both as a set of sub-corpora and as a set of essays written by individual students. In the current chapter, we will continue by examining how the corpus grows as each text is added in relation to the number of types within the corpus as a whole. This allows us to examine the effect of each individual text on the corpus as the corpus grows.

7.3 Corpus as dynamic

Generally, a corpus is seen as a static entity unless it is a monitor corpus.

> Corpora which include the time dimension as a design feature are not very common, and are of two kinds. Diachronic corpora present 'snapshots' at intervals of time, usually spanning at least a generation, while monitor corpora are devised so that language change can be plotted as it occurs.
>
> (Tognini-Bonelli, 2010: 22)

MICUPE is not a monitor corpus but, as it was collected over a period of time, it is a diachronic corpus, spanning one cohort of students as they complete a three-year degree programme. Once that time has elapsed and the corpus is fully collected, it has become a static entity containing 94 texts written by 17 different individuals at six different points in time. It is of interest, though, how the static entity is created, changed and homogenised or diversified as each text is added. In the previous chapter, this was seen when the relationship between frequency and distribution varied according to whether we were looking at text, sub-corpus or MICUPE as a whole, which displays the tension that exists between the variously-sized entities. This is especially true in relation to MICUPE since this corpus is a collection of essays written over time by the same cohort of students.

Building the corpus

Figure 7.2 shows the growth in types within the corpus as each text is added. The vertical axis indicates the number of types while the horizontal axis indicates the number of texts. All the graphs labelled 1-93 on the x-axis display temporal information inter sub-corpora but non-temporal information intra sub-corpora.

Figure 7.2 The increase in types in the whole corpus as new texts are added

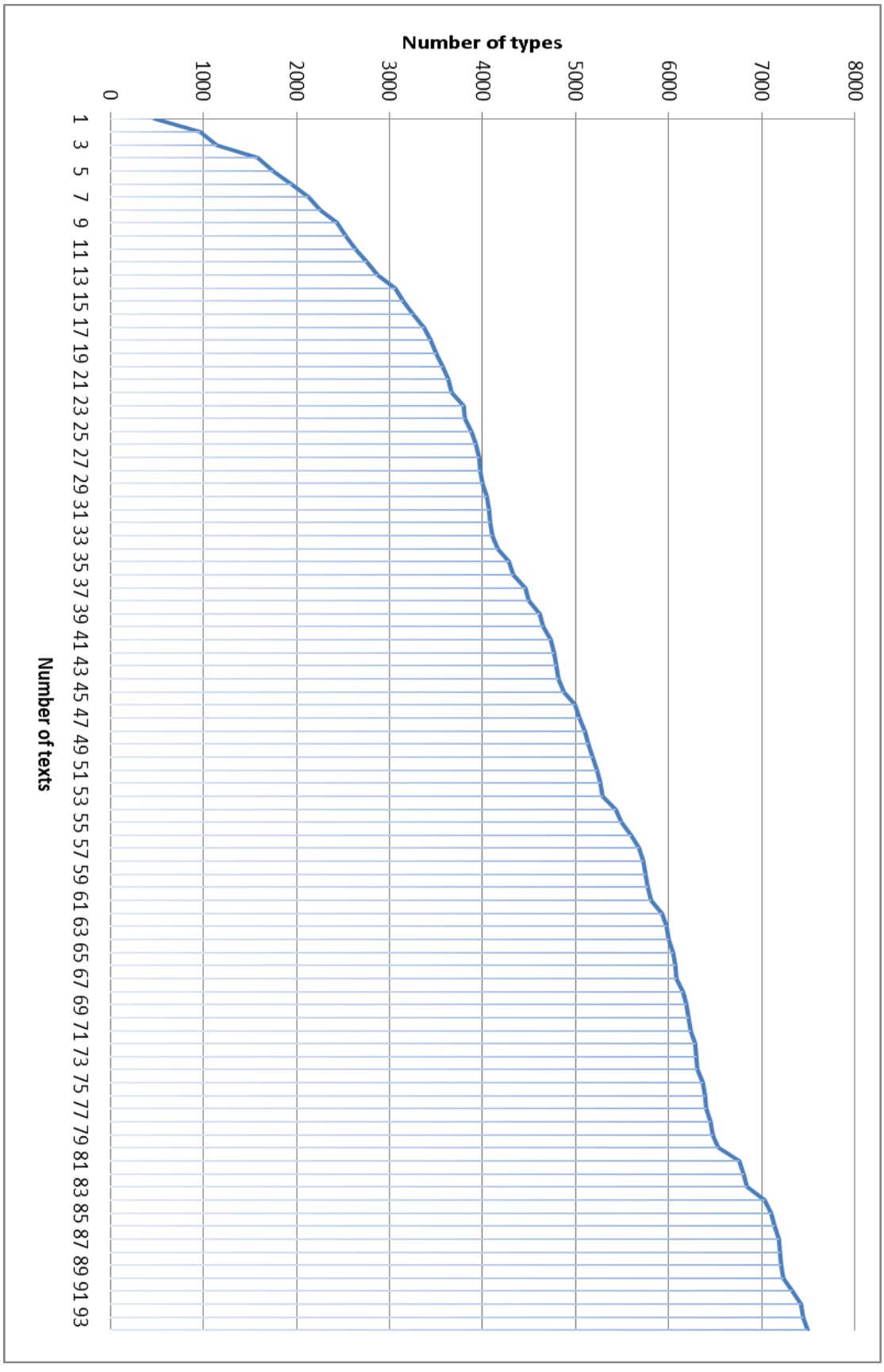


Figure 7.2 shows what happens to the number of types as each essay text is added to the corpus. The texts are added according to essay iteration. The texts from essay iteration 1 are added first, then the texts from essay iteration 2 and so on. For each essay iteration, the texts are added according to student number. For essay iteration 1, the texts are added according to Student 1, Student 2, Student 3 and so on. The same is done for each essay iteration. There is no rationale for deciding that this student is number 1 and this other student number 17. However, the order of students is consistent, meaning that Student 1 is the first student in each essay iteration.

In Figure 7.2, the number of types cannot decrease. This is because once a type is in the corpus it cannot be removed. The number of types can only increase or stay level. The graph here shows a non-linear increase in the number of types. Furthermore, this number does not seem to be reaching a limit. Logically, there must be a limit since the number of possible lexical items is finite. At one point during building of the corpus, around where the frequency of types reaches 4,000, it looks as though the curve is levelling off and reaching a limit. However, that trend does not continue after text 34 and the number of types continues to grow until it reaches 7,493.

In Figure 7.2, there are changes in essay iteration after the following points on the X-axis:

- 17 (end essay 1)
- 33 (end essay 2)
- 49 (end essay 3)
- 64 (end of essay 4)
- 77 (end essay 5)

It is, to an extent, surprising that the change in essay iterations does not correlate with any major change in the trajectory of the graph. In relation to lexical types, on a corpus level at least, the change of topic and the change of expertise of the writer seem to have no more impact on the trajectory of the number of types than the addition of texts from the same iteration. This correlates with the data already seen for the individual students in Figure 7.1. Each text has a distinct and unique effect on the corpus. No one text is composed purely of words that are used by either other writers or the same writer across all essay iterations (as already seen in Chapter 5). Furthermore, since the addition of the new essay iteration does not affect the rate of addition of types to the corpus in a

different way to new texts from the same essay iteration, it would seem that task plays a subtle and integrated role in influencing type selection. As we have already seen in Chapter 6, task does not have a uniform effect.

If we take the data from Figure 7.2, and present it as a rate of change, we can see the individual effect of each text on the number of types as each essay is added. This rate of change is displayed below in Figure 7.3.

Figure 7.3 The rate of change of number of types as each text is added

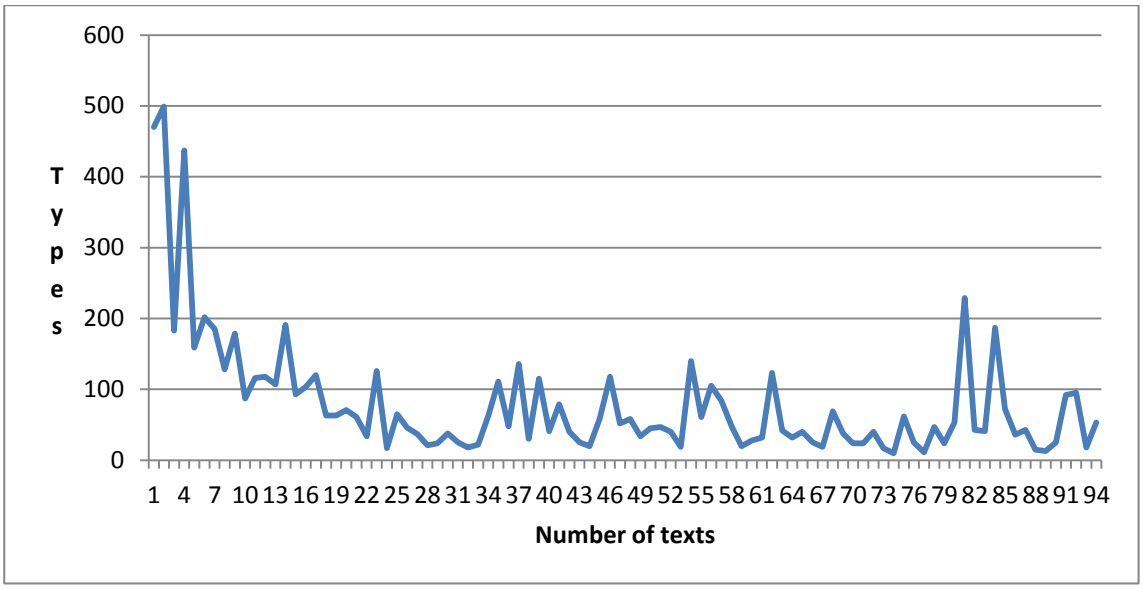


Figure 7.3 outlines the rate of change in the number of types within the corpus as each text is added. To begin with, after the initial growth in number of types as the first texts are added, there is an erratic fall in the rate of change in the number of types for the first 22 texts. After that, as each new text is added, there is an oscillating effect on the number of types. As previously stated, it is not possible for the number of types to actually drop since once the type is in the corpus it cannot be taken out. However, it would be possible for a text to have a contribution of 0 new types if every word that text contained had been used by an amalgamation of all the other previous texts in the corpus. In the current data, this does not happen. In fact, the text with the lowest amount of new types, relative to the previous texts in the corpus and not those that come subsequently, is text S13.5.A2 which contributes 10 new types to the corpus.

The 10 types not used before the addition of this text to the corpus are:

etcetera, Lonergans, pints, conditions, drank, drunken, hypothetical, inebriation, preceded, suspected

S13.5.A2 itself uses 181 types over 590 words. Interestingly, this is the text with the lowest number of tokens in the corpus. Given this, it is probably no surprise that this text contributes only 10 new types in Figure 7.3 above. Yet, there are still new types within the text, despite this text being of a particularly low token-count and being the 75th text into the corpus, 9 of which were also from essay iteration 5. Furthermore, if we remove the canon of work of Student 13 for the first 4 essay iterations, those 10 words remain the only difference in types between this particular essay text and the rest of the corpus as it is built to this point. This means that there are not further types in this particular essay that are limited to the writing of this particular student. Therefore, these 10 types are unique to this text and are unique to this particular student at this point in time. There are no further types in this essay text that are unique to the work of this student only. All other types apart from these 10 have been used by other writers either in essay iteration 5 or previous essay iterations.

A concordance for these types is given below in Figure 7.4.

Figure 7.4 Concordance of unique types in S13.5.A2

N	Concordance	File
1	for the Conditioned to be true? The Conditons can be found by asking all	S13.5.A2.txt
2	true. What could the Conditions be? I drank more than ten pints - relevant	S13.5.A2.txt
3	and all the answers do affirm a drunken state, then I know all the	S13.5.A2.txt
4	question: Do I remember the night? Etcetera, etcetera.... Providing all	S13.5.A2.txt
5	Do I remember the night? Etcetera, etcetera.... Providing all these	S13.5.A2.txt
6	night' (which I was not, this is purely hypothetical). The enquiry here is	S13.5.A2.txt
7	this same example of my suspected inebriation the Conditioned is that I	S13.5.A2.txt
8	does take place. This can be seen in Longergans example of: am I a	S13.5.A2.txt
9	So, do I know? What is knowing? Is Longergans definition of knowing the	S13.5.A2.txt
10	question: Did I drink more than ten pints? I can't remember much of the	S13.5.A2.txt
11	Conditions be? I drank more than ten pints - relevant question: Did I drink	S13.5.A2.txt
12	phase in the process of knowing. It is preceded by the act of inquiry. First in	S13.5.A2.txt
13	Taking this same example of my suspected inebriation the Conditioned	S13.5.A2.txt

Three of these 10 types are used twice while the other seven are only used once. We have already seen examples from this text in the previous chapter when we were examining the use of the word *I*. In that instance, the use of the word *I* was attributed to the hypothetical scenario created by the writer in order to illustrate a point. That hypothetical scenario involves the student (hypothetically) drinking too much alcohol and becoming inebriated. Some of the types around this notion are unique to this text. The other types that are unique to this text are generated by unique spelling. *Lonergans* is missing an apostrophe and *etcetera* is spelled out in full (and realised as a single word) instead of the shortened form. Two other types, *preceded* and *suspected*, while linked to the philosophical content of the text, do not seem to have any straightforward explanation for their uniqueness within the corpus apart from the fact that this student chose those words in this text and no other text uses those words.

That Student 13 in this essay chooses to use the word *etcetera* is also emblematic of an instantial choices made by this writer. We can say this because in three previous essays in the same semester as essay iteration 5, the student users a different form of roughly the same word, namely *etc*. The use of that word is shown in the concordance in Figure 7.6 below.

Figure 7.5 *Etc* in the essays of Student 13

N	Concordance	File
1	a dog would be its flesh, hair, bone, etc. made-up in the form or shape of	S13.2.B1.txt
2	the essence i.e. the flesh, hair, bone, etc. The existence of the dog is 'that it	S13.2.B1.txt
3	about, be it a human, a dog, a cat, etc, they all be. Equivocal predication	S13.2.B1.txt
4	the hands in relation to the clock face, etc. To use an even more simplified	S13.3.B1.txt
5	have a similar colour, shape, feel, etc. All these similarities can be	S13.3.B1.txt
6	size or temperature at the same rate etc. This second type of	S13.3.B1.txt
7	grouping, their angle to the cue ball, etc, prediction is becoming	S13.3.B1.txt
8	cushions would be a causing factor), etc. There are too many unknowns to	S13.3.B1.txt
9	between throw one and two etc. This is just a random sequence -	S13.3.B1.txt
10	of the mind - further questioning, etc.- and this causes an obscurity and	S13.4.A2.txt

Here we can clearly see that this student has previously used the form *etc* a total of 10 times in three previous essays. These three essays were written in the same semester as the essay containing the type *etcetera*. This means that over the course of 6 weeks from the submission of essay iteration 2 to essay iteration 5, this student uses *etc* in three

essays in a row and then *etcetera* in the following one. The use of the word *etcetera* is unique to this particular student in this particular text, while *etc* appears 78 times in total in the corpus spread across 36 texts. Given that *etc* is the accepted form, and that Student 13, along with other students, use this form, it seems as if Student 13 has unlearned the accepted form and replaced it with a new variant.

Example 7.1

> It is in these ways that scotosis remains a unconscious process, in the sense that we do not know that we are doing it, yet it can be affected by the conscious aspect of the mind - further questioning, **etc**.- and this causes an obscurity and confusion in our minds. (S13.4.A2)

Example 7.2

> I drank more than ten pints - relevant question: Did I drink more than ten pints? I can't remember much of the night - relevant question: Do I remember the night? **Etcetera**, **etcetera**.... (S13.5.A2)

Examples 7.1 and 7.2 show the use of *etc* and *etcetera* from two essays submitted by Student 13. These essays were submitted approximately 2 weeks apart. In both examples, the word is used to show that a list is in complete and allows the reader to either recognise that this list is incomplete or supply themselves with further possibilities should they see fit. There are differences in the context between these two examples, but it does not seem as there is anything of note within the textual context that would influence the choice of one form over the other, nor anything to influence the repetition in Example 7.2.

Overall, the erratic nature of the effect in relation to the types of each text in the corpus as shown in Figure 7.3, coupled with the samples of uniqueness of use of types shown in Figure 7.4, seem to suggest that the choices with regard to types writers make are not as constrained as expected. In a similar way to the high frequency types examined in Chapter 6, the growth in types within the corpus seems to suggest the instantiality of choices made by the writers. The following section will examine this in more detail, drawing further on the notion of a dynamic corpus as appropriate.

7.4 Single-use types

What is also of interest when looking at the frequency list for MICUPE is that out of the 7,493 types represented in the corpus, 4,483 occur more than once. This means that 3,011 types appear only once in the corpus. This could be considered to be very small since there are 134,289 words in the corpus. Yet this also means that there is an average of just over 32 types per essay that are unique to that essay within the corpus. Although it can be argued that multi-use words are in some way indicative of a shared space for the writers, their use and meaning within each particular context can vary as already seen in Chapter 6. This variance is not the case with single-use types simply because they appear only once in the corpus. Single-use types are therefore indicative of a uniqueness (in the current research, uniqueness is seen as absolute). We must also be aware that single-use is a relative concept dependent on all other texts in the corpus. Single-use can, however, be also conceptualised within a given essay text. There are types that are used only once in a text but these types are not single-use types within the whole corpus.

Words that appear only once in the corpus are obviously limited to one text only. As we have already seen in Chapter 5, we also need to consider distribution across texts. There are another 387 words which appear more than once in the corpus but whose occurrences are limited to one text. In terms of raw frequency, they range from a frequency of 2 to a frequency of 12 while still being limited to one text. Table 7.5 below outlines the single-use types per essay iteration.

Table 7.5 Single-use types per essay iteration

Iteration	**Types occurring once**	**Types occurring in one text only**	**Number of texts**	**Number of tokens**
Essay 1	1,622	1,804	17	30,426
Essay 2	804	945	16	14,696
Essay 3	1,079	1,269	16	20,895
Essay 4	1,199	1,342	15	17,323
Essay 5	945	1,138	13	13,071
Essay 6	1,632	1,920	17	37,812

If we were to add the number of types occurring only once within each semester, there would be 7,281 single-use types in total in MICUPE. However, in actuality, there are 3,011 types occurring only once in the whole corpus. If we added the number of types that appear in only one text within the semester, the total is 8,418. In actuality, there are 3,298 types in the whole corpus that appear in one text only. In a similar way to the total number of types, as shown in Table 7.1 above, single-occurrence and single-text types display a crossover between the iterations. This means that a) the collection of the larger corpus is negating in some way what happens at a different level of magnification; and b) uniqueness, in terms of word choice, is a relative concept which varies depending on what texts are included.

Furthermore, uniqueness in terms of word selection cannot solely be dependent on task or content since there is a crossover between the semesters. There are types that are unique in a particular essay iteration but not unique in the corpus as a whole. This means that for a particular essay iteration, one writer of the 17 is the only one to use that type, but in another essay iteration, that type could be used a number of times or even just once.

Given the nature of MICUPE, with standardisation in terms of degree subject, audience, genre, context, institution, teaching input and the same sample from the same cohort of students in each iteration, the levels of uniqueness, at least in terms of lexical choices, across the texts indicates the influence of an, as yet, unaccounted for factor. In the theoretical framework, it was posited that the uniqueness of each individual as determined by the billions of interactions inherent within that individual would have to lead to a unique response by an individual even in a repeated situation. If all the factors mentioned above had an equal effect on each student, and were uniform within each semester but varied from semester to semester, the pattern as displayed above would not materialise. Uniqueness within a semester would remain as uniqueness within the corpus. That is not the case. Uniqueness within the semester does not always result in uniqueness within the corpus because that uniqueness is repeated in another iteration, or grouping of iterations, and hence the uniqueness evaporates.

The high-frequency words generate a certain attraction as it may seem that by investigating these words we can see what is common to all texts. However, there is a paradox in that one such commonality to all texts is the uniqueness inherent within the

text. This also merits investigation and this uniqueness is created through the use of lower frequency items as well as unique meanings generated by the individualised use of higher frequency items.

In the corpus, when we see it as a finalised collection of texts, a number of things are clear:

- Only 14 words span all 94 texts;
- There is not an exact correlation between raw frequency and distribution across texts;
- Both frequency and distribution are important;
- The philosophical content is carried right down the frequency list;
- 3,011 types appear only once in the corpus.

Building the sub-corpora

We have already looked at how the corpus was created. It is also helpful to examine some of the essay iterations in more detail and see how they were created. Table 7.6 indicates the changes to the sub-corpus as each text is added.

Table 7.6 The building of the essay iteration 1 sub-corpus

Number of texts in sub-corpus	**Types**	**Tokens**	**Multi-use**	**% multi-use**	**Single-use**	**% single-use**
1	470	1,471	181	38.51	289	61.49
2	969	2,501	409	42.21	560	57.79
3	1152	1,146	508	44.10	644	55.90
4	1589	2,659	704	44.30	885	55.70
5	1748	1,592	805	46.05	943	53.95
6	1950	1,807	913	46.82	1037	53.18
7	2135	1,569	1004	47.03	1131	52.97
8	2263	1,480	1082	47.81	1181	52.19
9	2442	1,692	1180	48.32	1262	51.68
10	2529	1,556	1241	49.07	1288	50.93
11	2645	1,629	1314	49.68	1331	50.32
12	2763	1,580	1382	50.02	1381	49.98
13	2870	1,688	1442	50.24	1428	49.76
14	3061	2,123	1548	50.57	1513	49.43
15	3154	2,045	1607	50.95	1547	49.05
16	3258	1,797	1673	51.35	1585	48.65
17	3378	2,068	1756	51.98	1622	48.02

As stated, the numbering of students was a random process. That means that as the corpus is built, the addition of each student is random. The numbers above would not be the same if we had started with Student 2 instead of Student 1. Be that as it may, this does give a good indication of what happens to the corpus as each text is added. We can clearly see that with the addition of each text the number of types in the corpus increases, and with this both the number of multi-use types and single-use types increase. It must also be noted that the size of the jumps for each column is not uniform. The texts of some students seem to contain more unique words than the texts of other students. These increases are represented visually in Figure 7.6 below.

Figure 7.6 The change in multi-use and single-use types in essay iteration 1

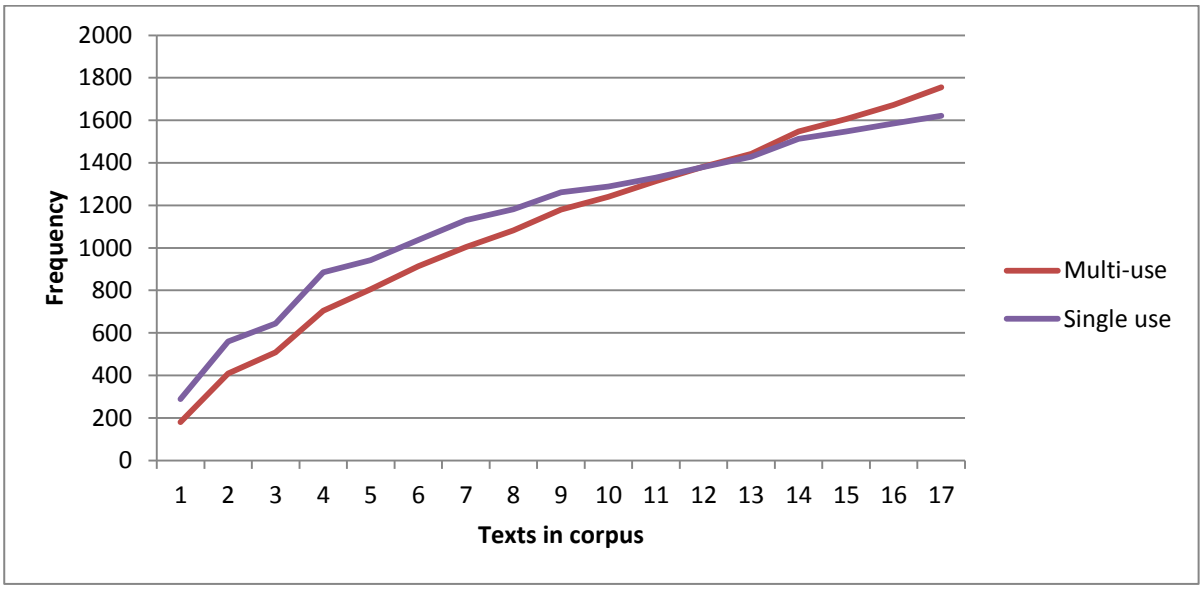
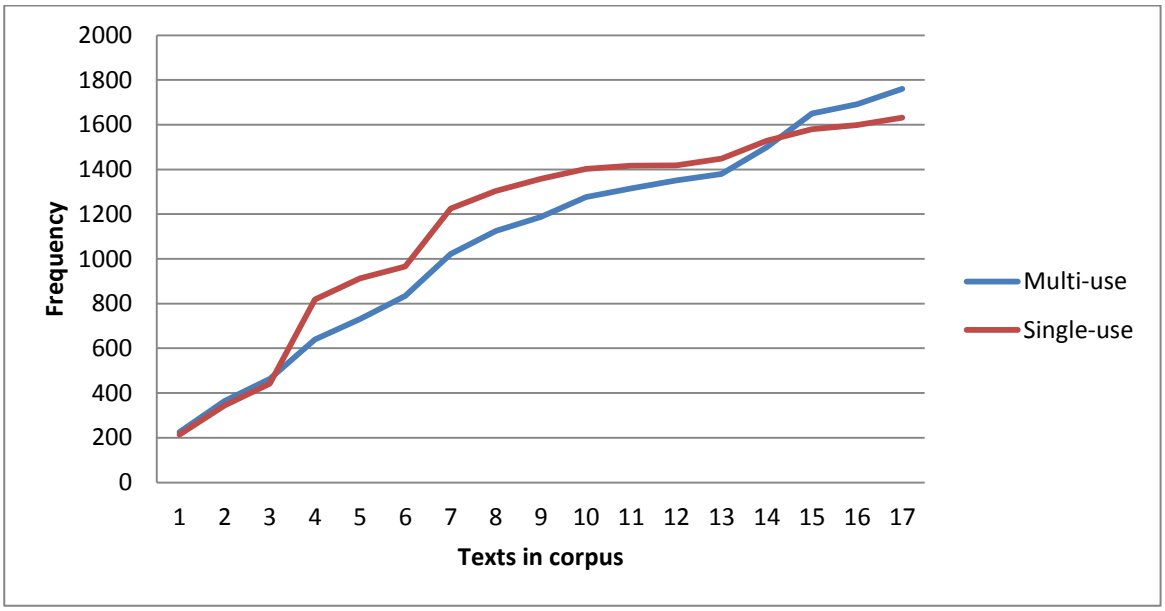


Figure 7.6 shows the change in the corpus as each text is added to essay iteration 1. As the first 12 texts are added, there are more single-use types than multi-use types in the corpus. However, the addition of the essay of Student 13 switches this relationship so that there are more multi-use types then single-use types. The rate of increase of single-use types slows down as each text is added. Yet, it must be noted that at no point does the number of single-use types decrease. Figure 7.7 below examines how essay iteration 6 grows from the addition of each text.

Figure 7.7 The change multi-use and single-use types in essay iteration 6



From Figure 7.7, in essay iteration 6, the change in single-use and multi-use types seems more erratic than that in essay 1. To begin with as the sub-corpus grows, multi-use types are more frequent than single-use types. The addition of text 4 changes this dramatically, showing a large increase in the number of single-use types. As more texts are added to the sub-corpus, the gap between single-use and multi-use narrows, albeit not in a uniform fashion. By the time the text of Student 15 is added, there are now more multi-use types in this sub-corpus than single-use types.

In relation to types used, and their contextual function and meaning, MICUPE is unique. Within the corpus, each iteration is unique. Within each iteration, each text is unique. Within each text, the types are at times unique and at times shared with other texts and within the same text, yet their contextual meaning is unique. This uniqueness at each level of the corpus is despite the fact that the texts were written by the same 17 students studying the same modules within the same degree programme and submitting essays for the same audience. It would seem that, despite being subjected to similar influences, each writer is making a set of instantial choices as to the types they used and how they use them. These choices are repeatedly made on an ongoing basis and because they are new and unique to each time they are made, they give rise to uniqueness within texts, hence the uniqueness within the essay iterations and uniqueness at the corpus level.

Single-use and multiple-use types in the whole corpus

We have seen that taken individually, essay iterations 1 and 6 display different patterns in relation to the distribution of single-use and multi-use types. Figure 7.8 below shows the increase of multi-use and single-use types in the whole corpus as each of the 94 texts is added.

Figure 7.8 Multi-use and single-use types in MICUPE as each text is added

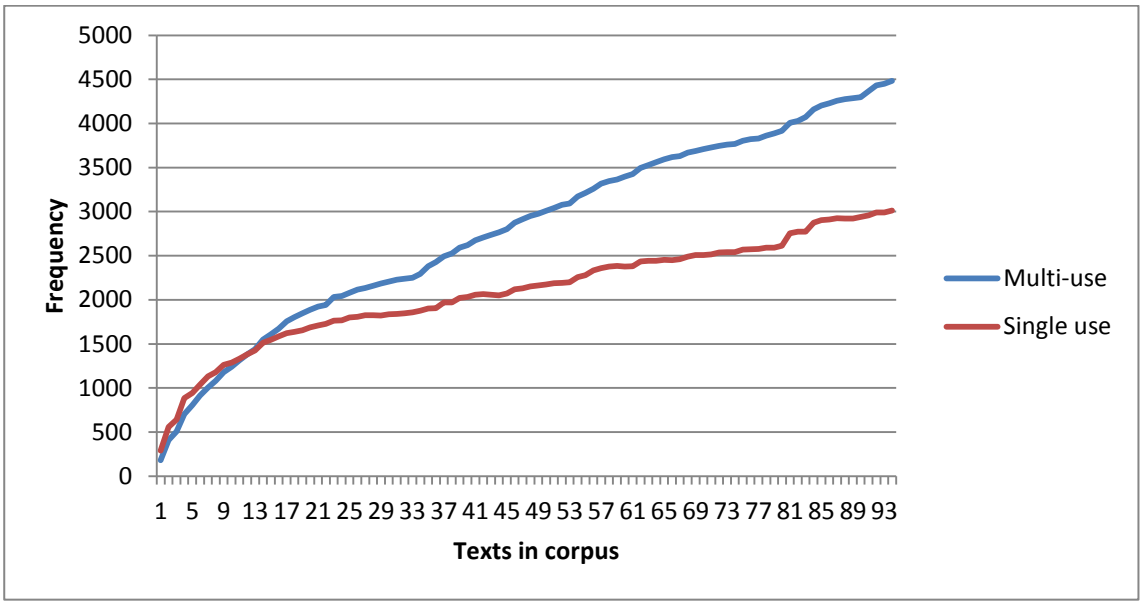
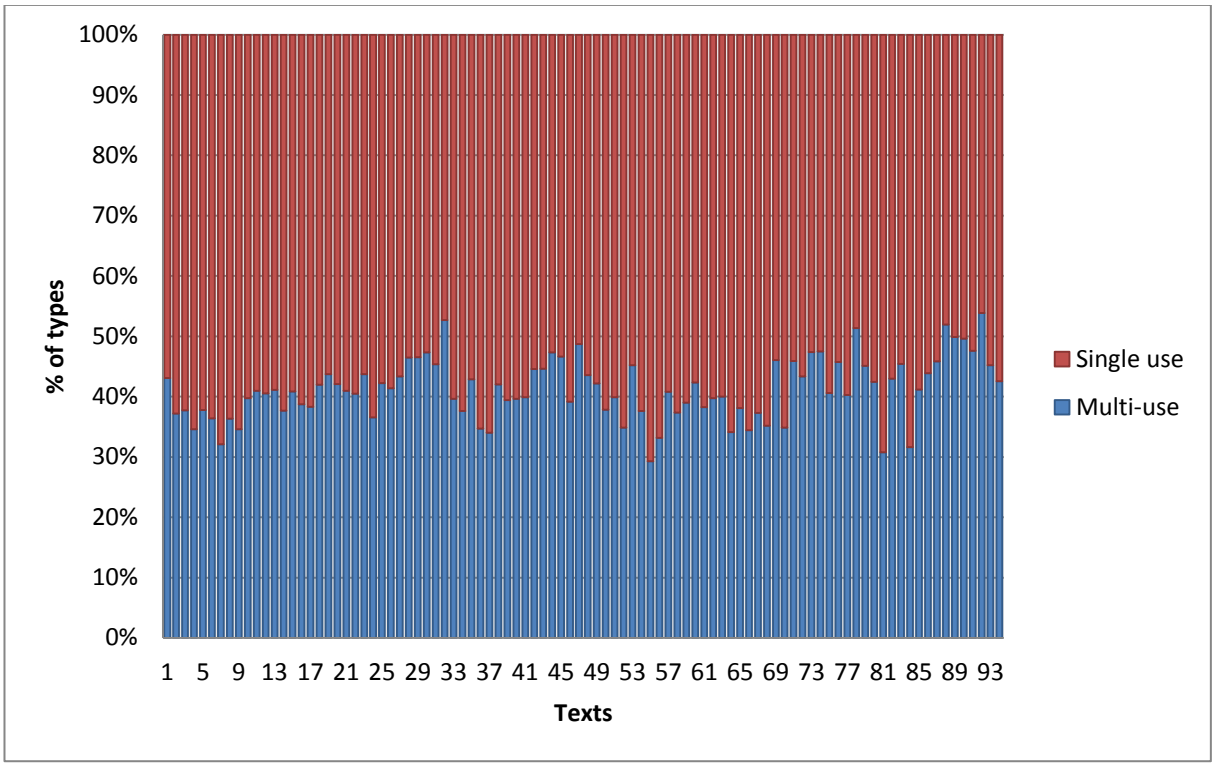


Figure 7.8 raises some interesting points. As the corpus begins to grow, there are more single-use types in the corpus than multi-use types. After the addition of text number 13, this changes. It does not change back. For the remainder of the additions of each text, multi-use types are more frequent than single-use types. If each text were the exact same, there would be no increase in types as each text is added. At the same time, if each text used completely different and unique types, the increase in types would mirror that of the raw word count. In Figure 7.8 we can see that neither multi-use or single-use types are levelling off in their rate of increase. If we were to imagine a hypothetical corpus of an infinite number of essay texts, the composition of the corpus would be such that the multi-use types would reach a limit, possibly bounded by the limits of types available in the language. At the same time, the single-use types would approach zero.

The plot lines for single-use and multi-use types are not the same, and neither of these lines is the same as the growth in types in the overall corpus as seen in Figure 7.2 previously, although it is the sum of both of these that give rise to the overall number of types. Furthermore, this is in contrast to localised notions of single-use and multi-use as evidenced in each individual text. Figure 7.9 below outlines the proportion of single-use types to multi-use types in each of the 94 texts.

Figure 7.9 Single-use and multi-use types per essay



In Figure 7.9, we can see that in the main there are more single-use types in each text than multi-use types. In fact, 90 essay texts have more single-use types than multi-use types while only four texts have more multi-use than single-use types. In essay iterations 1, 3, 4 and 5, no text in the corpus has more multi-use types than single-use types. One text (S16.2.B2) has more multi-use types in essay iteration 2. Three texts (S01.6.B3, S11.6.A2 and S15.6.B1) in essay iteration 6 have more multi-use types. It is of note that these four essays are spread across four different students and four different grades, suggesting that the pattern of the use and re-use of types can vary in any given instance even when the texts are written by the same student. Furthermore, this pattern does not, on its own, determine the grade received by the essay.

There is, therefore, a tension between what happens in relation to types in individual texts and what happens in relation to the types in the corpus as a whole. This is to be expected as one of the major uses of corpora is for lexicographical work and the purpose of corpora in such instances is to generate multiple uses of each type. However, in the current context, this generation of multiple uses of some types is inherently

misleading as in their naturally-occurring context, the majority of types are single-use. This process of making the occurrence of multi-use types more common than single-use types through the addition of texts also happens at the level of the sub-corpora in MICUPE as shown in Table 7.7.

Table 7.7 Percentage of single-use types per essay iteration

Iteration	**Number of types**	**Number of single-use types**	**% of single-use types**
Essay 1	3,378	1,622	48.01%
Essay 2	1,660	804	48.43%
Essay 3	2,409	1,079	44.79%
Essay 4	2,429	1,199	49.36%
Essay 5	1,852	945	51.02%
Essay 6	3,392	1,632	48.11%

In Table 7.7, in all but one of the essay iterations, there are more multi-use types than single-use types. We have already seen how this comes about for essay iterations 1 and 6 in Figures 6.6 and 6.7. Essay iteration 5 is the only essay iteration in the corpus that has more single-use types than multi-use types. Due to the nature of the corpus as explained in Chapter 4, essay iteration 5 only has 13 texts. The lower number of texts may account for this. In essay iteration 1, by the time 13 texts were added to the sub-corpus, there were more multi-use types than single-use types. In contrast, essay iteration 6, previously having had more multi-use types before text 4, has more single-use types at text 13.

In relation to single-use and multi-use types, we have seen that in the main, the texts themselves are predominantly comprised of single-use types but the sub-corpora and corpus representing these texts are predominantly comprised of multi-use types. Table 7.8 below looks at all 94 texts in the corpus grouped according to student.

Table 7.8 Multi-use and single-use types in MICUPE per student

Student	Types	Multi-use	%	Single-use	%
1	1408	657	46.66	751	53.34
2	1654	793	47.94	861	52.06
3*	1194	511	42.80	683	57.20
4	2188	866	39.58	1322	60.42
5*	1168	544	46.58	624	53.42
6	1808	869	48.06	939	51.94
7	1724	677	39.27	1047	60.73
8	1699	786	46.26	913	53.74
9*	1531	668	43.63	863	56.37
10	1406	664	47.23	742	52.77
11*	1183	546	46.15	637	53.85
12*	1250	614	49.12	636	50.88
13	1348	650	48.22	698	51.78
14*	2134	993	46.53	1141	53.47
15*	1646	863	52.43	783	47.57
16*	1314	605	46.04	709	53.96
17	1632	710	43.50	922	56.50

(* based on 5 texts)

In Table 7.8, only one student, Student 15, uses more multi-use types and single-use types over the course of the essay iterations. Student 15 has five texts in the corpus, and in these five texts, in a context localised according to text, there are 783 single-use types. Seven other students have five texts in the corpus and although the proportion of single-use types to multi-use types is above 50% in their localised contexts, Student 15 has a higher raw frequency of single-use types than Students 3, 5, 11, 12 and 16. In fact, Student 15 also has a higher raw frequency of single-use types than three students with 6 texts in the corpus (Students 1, 10 and 13). Student 4 has the highest proportion of single-use types, at just over 60%.

The corpus creates a tension - on the one hand, there exists an implied notion that everybody is different, which is evidenced through a need to generate larger corpora with bigger samples and through the reluctance to use one sample, one user, one text, or one context to generate something that is considered generalisable. On the other, there exists an implied notion that everybody is the same as evidenced by the fact that averages and homogenisation are accepted as universal once the sample issue in the previous point has been resolved. This tension damages the worth of any corpus. The use of empirical data, which using a corpus must be, cannot ignore the unique.

When we add the texts into the corpus, it has the effect of negating what is really happening in each individual text. So, for each student and each text they produce, the single-use types within an essay are as essential to that text as items shared by other language producers in the same or similar contexts.

Change in single-use and multi-use types

We have already seen that each individual text has a unique effect on the corpus, since that text is in itself unique, in relation to the types that are present in the corpus. We have also seen that there are single-use types and multi-use types. The notions of multi-use and single-use are relative. Within an individual text, whether a type is multi-use or single-use is dependent on the other types within that text. Within a sub-corpus of an essay iteration, this notion is dependent on the types used within all texts in the sub-corpus, including being single-use within its text of origin. For a type to be single-use within the corpus as a whole, it is necessary that that type is not used again in the particular text, the particular sub-corpus or the corpus as a whole.

Since the concepts of single-use and multi-use are dependent not just on the individual text but also on all other texts, the addition of the new text not only brings its unique types but also can have the effect of making what was prior to the addition of this text a single-use type into a multi-use type. Figures 7.10 and 7.11 show the rate of change for both single-use types and multi-use types as each one of the 94 texts are added to MICUPE.

Figure 7.10 Change in single-use from addition of 1 text to addition of next.

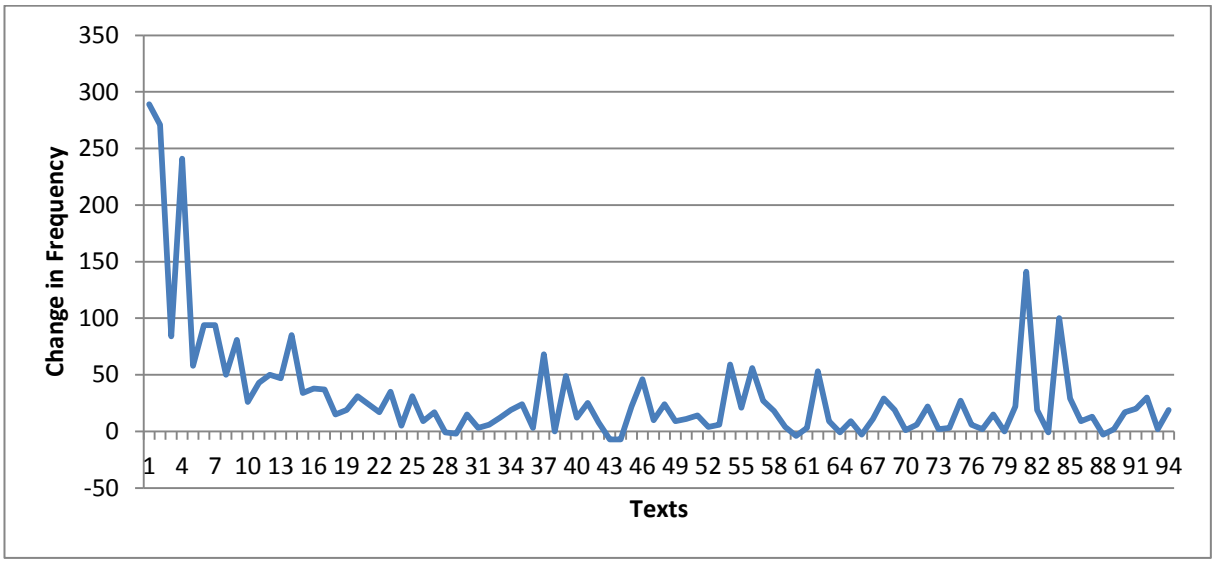
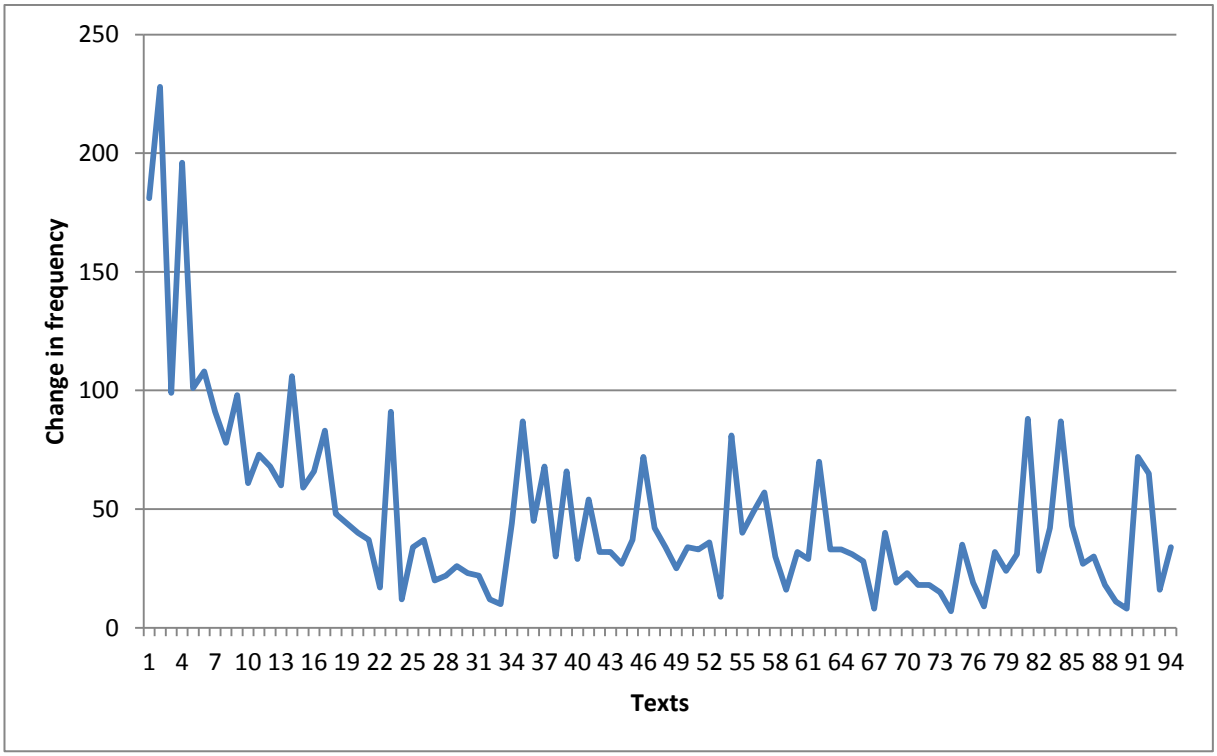


Figure 7.11 Change in multi-use as each text is added

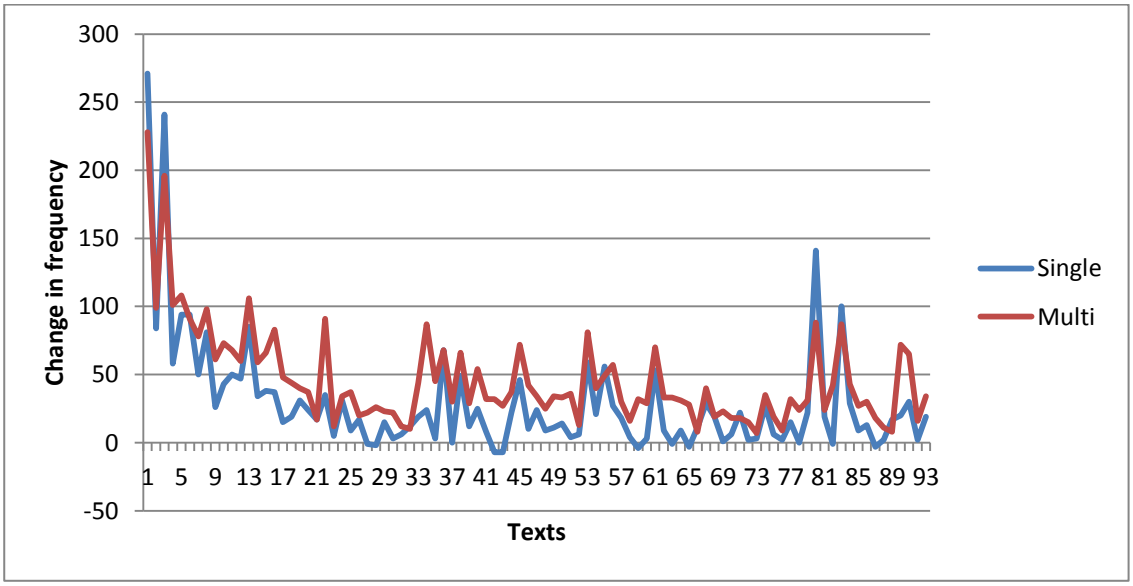


In Figures 7.10 and 7.11, we can see that the rate of change of types within the corpus as each text is added is not linear. In both figures, there is a fall in the rate of change by the addition of text 5. However, that fall was not linear as text 3 brought a major slow-down in the rate of change for both figures. After this fall, there seems to be an erratic, oscillating rate of change as each subsequent text is added. In Figure 7.10, some texts

result in a negative rate of change, for example at texts 28 and 43. When we consider that each text has a set of unique, single-use types, it is surprising that this phenomenon is happening.

When we merge the rate of change for both single-use types and multi-use types, as in Figure 7.12, there is a distinct lack of a set pattern.

Figure 7.12 Change in single-use and multi-use types as corpus grows



At times, the addition of a text results in a large increase in both the numbers of single-use and multi-use types, for example the addition of text number 80. At other times, the addition of a text results in a decrease in the rate of change for both, for example the addition of text number 23. At other times, the addition of a text results in an increase in multi-use types by the decrease in single-use types for example the addition of text 27.

The addition of text 43 demonstrates the unique effect of each individual text in the corpus. This text is S11.3.B1. Prior to its addition, there are 42 texts in the corpus. They are comprised of all of essay iteration 1, all of essay iteration 2 and 9 texts from essay iteration 3. Prior to the addition of S11.3.B1, there are 4,772 types in the corpus. 2,066 of these are single-use the at this point and 2,706 are multi-use. Text S11.3.B1 has 336 types, 186 of which are single- use types within that text. 19 of those single-use types have not appeared in the corpus to this point. 26 types which were single-use items to

this point are used in text S11.3.B1. The effect of this text on the corpus is a net loss of single-use types. As can be seen in Figure 7.12, this phenomenon is not unique to this text but the appearance of this phenomenon seems erratic and does not fit a set pattern according to each essay iteration nor according to the overall size of the corpus. However, the addition of subsequent texts from this essay iteration does not result in a net loss of single-use types. In fact, the other texts in this essay iteration result in a net gain in single-use types. A concordance of the types that prior to the addition of S11.3.B1 were single-use types in the corpus but become multi-use after the introduction of this text can be found in Figure 7.13.

Figure 7.13 Concordance of single-use types that become multi-use with the addition of S11.3.B1

N	Concordance	File
1	nowadays when juries consist of only ten or twelve. Due to this fact you would imagine	S01.1.B3.txt
2	the juries sympathy for Socrates is fading fast. Socrates then goes on to address the	S01.1.B3.txt
3	of Socrates uncovering the truth in another person's definition of a common sense idea	S04.1.B1.txt
4	he says that he does not believe God would allow a better man to be harmed by one	S07.1.A1.txt
5	choosing his punishment he decided not to choose banishment as he wanted to	S10.1.A2.txt
6	seemingly harmless but which really tested the interlocuter. For example, he	S14.1.B1.txt
7	human knowledge in our endeavour to seek solutions and explanations. Explanations to	S08.2.B3.txt
8	such as "It's a cylindrical shaped object, commonly made from plastic or glass, with	S10.2.A2.txt
9	though it was directly proceeded by the working of Plato and Aristotle, it was still a	S13.2.B1.txt
10	moments. The act of insight does not follow rules as, if this were true, discoveries would	S02.3.A2.txt
11	B.Ed's, insight is highly significant. The new curriculum (1999) bases children's learning	S04.3.B3.txt
12	the example of special relativity, I am no scientist but here's how I see it; I am	S04.3.B3.txt
13	changes somewhat. To conclude, I will give Lonergans' characteristics of insight, and	S04.3.B3.txt
14	understand, to know. Man then is faced with puzzlement and a want to figure it out, to	S05.3.B3.txt
15	can sit in a room without distractions for hours but that doesn't mean one will	S05.3.B3.txt
16	when they do it may be in the most obscure setting or moment in time! Lonergan's next	S05.3.B3.txt
17	a car it would not have the same effect as throwing a paper out of a spacecraft.	S05.3.B3.txt
18	what he meant using a mathematical algebraic problem. He gives the example of	S06.3.B1.txt
19	process. This process not only applies to maths problems but every aspect of life. We	S06.3.B1.txt
20	This process not only applies to maths problems but every aspect of life. We may	S06.3.B1.txt
21	to each other. Scientific measurements of weight, temperature, length et cetera, have	S06.3.B1.txt
22	the penny is no longer floating insight is obtained, crossing the threshold from not	S08.3.B2.txt
23	ball. This is different from any other process governed by concrete insight linking	S08.3.B2.txt
24	and angular momentum and the various surfaces that it may fall on. The possibility of	S08.3.B2.txt
25	out of the sum of tosses. If there is some "systematic factor" favouring one side rather	S08.3.B2.txt
26	the line, there is also an infinite number of angles at which you can hit the cue ball. The	S10.3.A2.txt

Figure 7.13 shows a concordance for the types that were, prior to the addition of text S11.3.B1, single-use types within the corpus but became multi-use types once this text was added. This text was written in essay iteration 3. Nine of the types that were single-

use were previously used in essay iterations 1 and 2. Interestingly, the 8 different texts these types came from were not written by Student 11. Furthermore, it is difficult to attribute a linear deterministic factor to the appearance of these types either in the essay of Student 11 or the previous eight essays they appeared in. The types *twelve, fast, person's, allow, choose, tested, solutions, commonly* and *working* do not lend themselves to linear causal factors such as audience, task or genre. Furthermore, the texts these words appear in show a range of grades.

17 of the types shown in Figure 7.13 were previously used once in essay iteration 3 and span six different texts. It is expected that there would be more types from this essay iteration than previous ones simply because this essay iteration had similar tasks and was written at the same point in time. However, two points must be made here. The first is that while this text uses single-use types from six other texts in this essay iteration, it does not use all the single-use types from those essays, nor does it use single-use types from the three other essays from that of iteration that were in the corpus prior to the addition of the text from Student 13. Secondly, apart from *angles* in the text of Student 10, for each one of the types that were single-use, further texts had been added before that of Student 11 which did not result in these types becoming multi-use.

As we have seen, within the texts themselves, the majority of types tend to be single-use. These single-use types within the texts may or may not be used in another text. In a way, it seems that the re-use of a type is random. However, this apparent randomness only exists when looking at the corpus, as within the texts themselves, the appearance of types is not random. Their appearance is governed by instantial choices made by the writer who is a dynamic construct and is subjected to a complex, dynamical system of influence that is unique in each instance of its realisation. It is for this reason that despite similarities in relation to genre, audience, writer and task, we see the erratic nature of change as the corpus is built in Figure 7.12 and the lack of an explanation for the re-occurrence of the types in Figure 7.13 save that of attributing the corpus to an amalgamated set of instantial decisions.

Overall, this section has shown that each individual text has a unique effect on the corpus. We have seen in previous chapters that similarity of type does not necessarily indicate similarity of use. However, whether a type is a single-use type or a multi-use type is a relative concept dependent on other texts in the corpus. The addition of each

text does not have a set, determined effect on the corpus. Types that were used by some students in previous essay iterations can be used by other students in different essay iterations. The re-use of these types does not always have a linear cause and of these types are not shared by every student.

We have examined what happens to the types in the corpus as the corpus grows. We have looked at those types according to essay iteration, student and individual text. Once the corpus has finished growing, as is now the case with MICUPE, the dynamic aspect no longer holds and the corpus becomes static. The following section will examine how types are shared among the 6 essay iterations once this static corpus is achieved.

7.5 Types across essay iterations

This section examines the sharing of types between the 6 essay iterations. We saw in Chapter 5 that there are 14 types that appear in all 94 essay texts and examined some of these in further detail in Chapter 6. Naturally, if a type appears in all 94 texts, it must also appear in all six essay iterations.

Although there are 14 types that appear in each of the 94 essays, each essay iteration has a different number of types that appear within each essay of that iteration. The number of types in all texts in an essay iteration and the list of those types is shown below in Table 7.9.

As we have already seen in Chapter 5, only 14 types appear in each of the 94 texts. However, within each essay iteration there is a greater number of types that appear in all texts within that iteration. Table 7.9 shows that within the iterations, the number of types that appear in all texts range from 21 to 53. Interestingly, the essay iteration with the fewest texts also has the fewest number of types appearing in all those texts.

Table 7.9 Types in all texts in each essay iteration

Iteration	Number of texts	Number of types in all texts	List of types in all texts
Essay 1	17	53	A, ABOUT, ALL, AN, AND, AS, AT, BE,BY, DEATH, DO, FOR, FROM, GOOD, HAVE, HE, HIM, HIMSELF, HIS, I, IF, IN, IS, IT, LIFE, MAN, MY, NO, NOT, OF, ON, ONE, ONLY, OR, OWN, SO, SOCRATES, SUCH, THAT, THE, THEIR, THEY, THIS, TO, UP, WAS, WAY, WHAT, WHICH, WHO, WITH, WOULD, YOU
Essay 2	16	26	A, AND, ARE, AS, BE, BEING, BUT, BY, CAN, FOR, HAVE, IF, IN, IS, IT, NOT, OF, PARMENIDES, THAT, THE, THERE, THEY, TO, WE, WHAT, WHICH
Essay 3	16	28	A ,ALL, AN, AND, ARE, AS, AT, BE, BY, FOR, FROM, HAVE, IF, IN, IS, IT, NOT, OF, ON, ONE, OR, THAT, THE, THERE, THIS, TO, WE, WOULD
Essay 4	15	29	A, AN, AND, ARE, AS, BE, BUT, CAN, FOR, FROM, HAVE, IF, IN, INSIGHT, IS, IT, NOT, OF, ON, OR, SO, THAT, THE, THESE, THEY, THIS, TO, WITH, WOULD
Essay 5	13	21	A, AND, ARE, AS, BE, BY, CAN, FOR, HAVE, IN, IS, IT, NOT, OF, ON, THAT, THE, THESE, THEY, TO, WHAT
Essay 6	17	45	A, ACT, ALL, ALSO, AN, AND, ARE, AS, AT, BE, BEING, BY, CAN, CANNOT, COULD, EVERYTHING, EXISTENCE, FOR, FROM, GOD, HAVE, HE, IF, IN, IS, IT, ITS, ITSELF, MUST, NO, NOT, OF, ON, ONE, OR, THAT, THE, THERE, THESE, THIS, TO, WE, WHAT, WHICH, WOULD

Each of the 7,493 types in MICUPE were checked against the essay iterations that they occurred in. This data is presented in a two-stage process below. Figure 7.14 shows the types that are in essay iterations 4, 5 and 6. It does not include the types that are shared between these iterations and essay iterations 1, 2 and 3 as these will be added in stage 2.

Figure 7.14 Sharing of types between essay iterations 4, 5 and 6

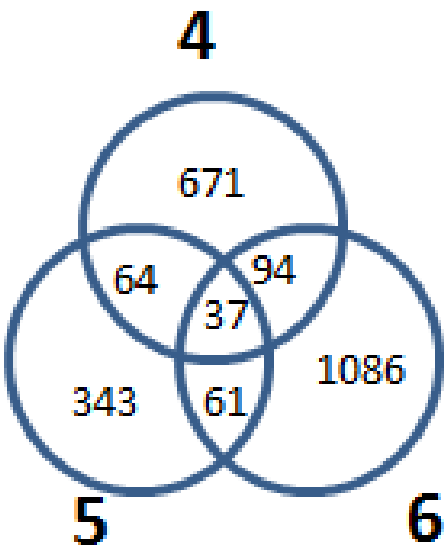
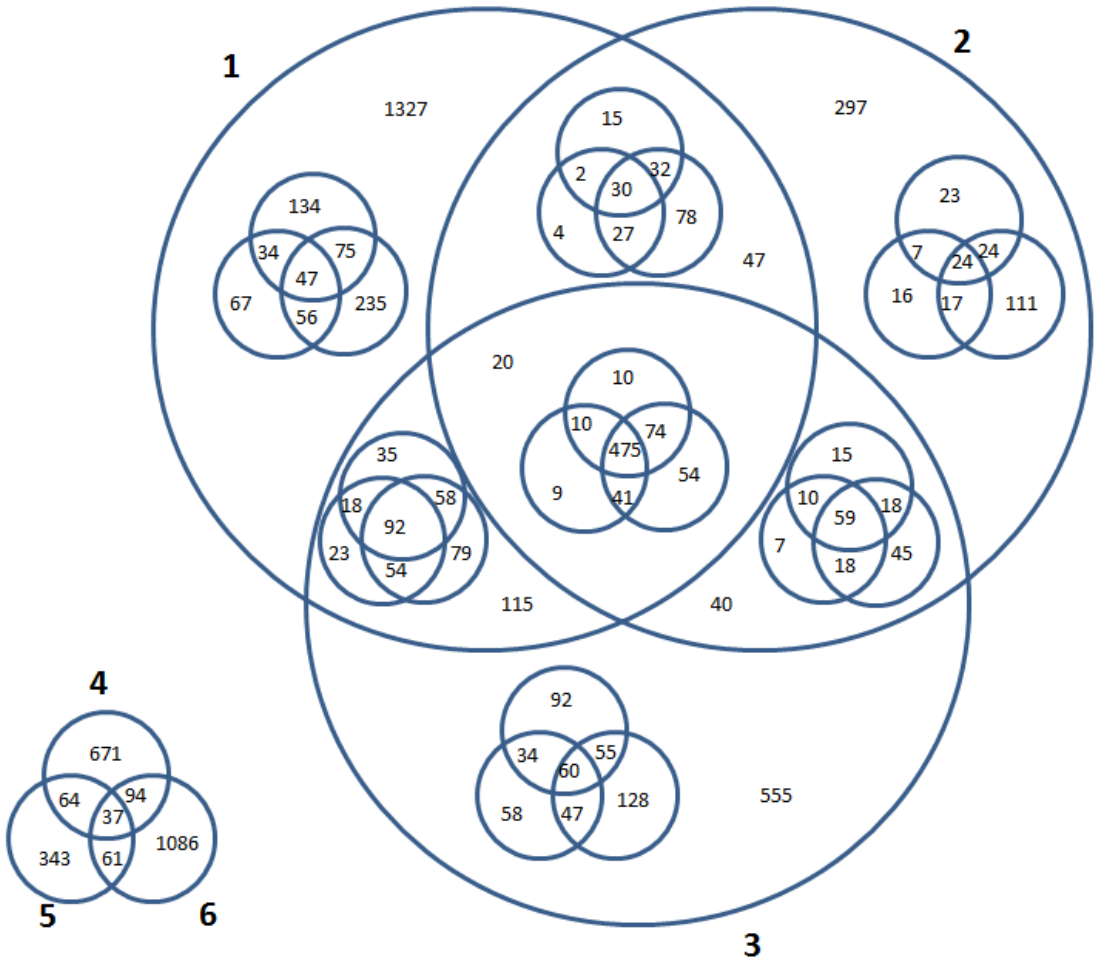


Figure 7.14 shows the intersections of types between the essay iterations 4, 5 and 6. Due to the size of the corpus, it is not possible to fit the actual types into this diagram. Instead, the intersections in the diagram contain numbers which refer to the number of types that appear in those intersections. For example, there are 671 types that appear in essay iteration 4 only, 61 types that appear in iterations 5 and 6 only and 37 types that appear in iterations 4, 5 and 6 only.

This data is then embedded into essay iterations 1, 2 and 3 to give us a broader picture of the sharing of types between the 6 essay iterations. This is represented in Figure 7.15 below. In this figure, the embedded 3-set diagram represents iterations 4, 5 and 6.

Figure 7.15 The sharing of types between essay iterations 1-6



In Figure 7.15[7], the bottom left contains the diagram representing essay iterations 4, 5 and 6 as previously seen in Figure 7.14. Those iterations are also embedded in iterations 1, 2 and 3. For example, there are 1,327 types that appear in iteration 1 only. There are 47 types that appear in iterations 1 and 2 only. There are 20 types that appear in iterations 1, 2 and 3 only. There are 10 types that appear in iterations 1, 2, 3 and 4 only. There are 30 types that appear in iterations 1, 2, 4, 5 and 6 only. There are 475 types that appear in each of the 6 essay iterations.

7 The 6-set nested Venn diagram shown here is adopted from one devised by Nicholas J. Radcliffe as outlined at http://scientificmarketer.com/2010/02/nested-venn-diagram.html.

The six essay iterations interact with each other in 63 ways. There are types in each of those 63 intersections. Even though each essay iteration is written at a particular point in time, deals with a particular subject matter and has a particular set of tasks, there is an apparent randomness in the sharing of types between the iterations. For example, there are nine types that appear in essay iterations 1, 2, 3 and 5, but not in iterations 4 and 6. Those nine types are:

apart, becoming, decided, definitions, discover, distinct, learned, read, research

Together, these nine types have 66 occurrences in the corpus and are spread across 37 texts. The lowest frequency types of these words are *apart* and *becoming* as both have only four instances in the corpus. For each, these four instances are spread across four iterations and it is four different students that use these words. *Decided* is the highest frequency type of these words and is also spread across four different essay iterations but has 12 instances spread across eight texts written by six different students.

There are 235 types that appear in essay iterations 1 and 6 only. There are 1,105 combined instances of these 235 types across both essay iterations. Essay iteration 6 was written nearly 3 years after essay iteration 1. Each of the 34 essays, 17 from iteration 1 and 17 from iteration 6, contain at least one of these words. The word *trial* has the highest frequency of this set, at 49 instances. There are 74 types that appear twice only and those two occurrences are divided between essay iteration 1 and essay iteration 6.

As we have already seen, *the* is the most frequent type in the corpus and the most frequent type occurring in each of the 6 essay iterations. Logically, a type would have to be in the corpus a minimum of 6 times to appear in 6 essay iterations. However, there is no type out of the 475 types in all six essay iterations that appears 6 times and in 6 essay iterations, unlike the intersection of iterations 1, 2, 3 and 5 shown above which has types that only appear 4 times. The lowest frequency types that appear in all 6 essay iterations are *context*, *overall* and *several* which appear a total of eight times each in MICUPE. These 3 types are shown below in Figure 7.16.

Figure 7.16 Concordance of *context*, *overall* and *several*

N	Concordance	File
1	in areas such as business. Such a cultural context is important to answer as to	S14.1.B1.txt
2	has the same meaning no matter what context in which it is used therefore dog is	S17.2.B3.txt
3	we expect an answer of particular context and status. However, if we find that	S02.3.A2.txt
4	pure conjugates both exist equally in the context of time and space. 5) Galileos	S04.4.B1.txt
5	attitudes regarding subjects within such a context. They may even behave irrationally	S14.4.A2.txt
6	would also gain an insight. Knowing the context in order to make a judgement. Of	S08.5.B1.txt
7	Aristotle. I will discuss their viewpoints, in context, as the essay progresses. His	S09.6.B1.txt
8	moved by something else. "Motion" in this context means change or state of change.	S15.6.B1.txt
9	and the life of good secure the psyche. Overall Socrates is a martyr for the truth as	S08.1.C1.txt
10	is going to be in the future" (Parmenides) Overall I am very impressed with	S02.2.A2.txt
11	false illusion that I thought was the case. Overall, I am impressed with Lonergan's	S02.3.A2.txt
12	has some bearing in the figurative sense. Overall it seems to Lonergan 'insight' is an	S03.3.B2.txt
13	The overall aim of Empirical method is to be	S05.4.B1.txt
14	to subjects and subjects to objects. Overall, dramatic bias exists as one	S09.4.B2.txt
15	In its essence, it deals with general bias. Overall, Cosmopolis, sets out to open the	S02.5.A2.txt
16	upsurge in belief in God! In conclusion, my overall evaluation of the arguments in the	S14.6.A1.txt
17	that he is not guilty of this charge by using several analagies. "Is there anyone who	S14.1.B1.txt
18	relationship in each being. This resembles several words in the english language for	S17.2.B3.txt
19	bodies unless one is paralysed. There are several more demonstrations of the	S03.3.B2.txt
20	time and I repressed my feelings. I dreamt several nights that I had a huge row with	S16.4.B2.txt
21	of fragmentary scenes. Here there are several elements lacking, critical reflection,	S17.4.C3.txt
22	that wanting an insight can stimulate several complimentary insights an	S17.4.C3.txt
23	after the insight is given. They must take several questions into account. What	S15.5.B3.txt
24	It does not matter if the intermediate is several or just one. To take away the	S15.6.B1.txt

Figure 7.16 shows a concordance of *context*, *overall* and *several*. As can be seen from the *File* column, each type appears in all 6 essay iterations. In the case of *context*, seven different students use this type within MICUPE. Student 14 uses this type twice, once in essay iteration 1 and once in essay iteration 4. In the case of *overall*, six different students used this type. Student 2 uses *overall* in essay iterations 2, 3 and 5. Interestingly, the instances in iterations 2 and 3 are quite similar (lines 10 and 11), collocating with *impressed with*, although by iteration 3 the writer has included a comma after *overall* and omitted *very*. These two instances are used to introduce the writer's opinion. The instance of *overall* from this writer in iteration 5 (line 15), not having been used in iteration 4, is different to these in iterations 2 and 3 by this writer, setting out a general summary of another writer's argument. Of the three types, *several* has the narrowest spread of students as it is used by five students within the corpus. Student 15 uses *several* in essay iterations 5 and 6. Student 17 uses *several* once in essay iteration 2 and twice in iteration 4. What we are seeing here is that while these 3 types are re-used by some students, there is a randomness about their appearance in the texts. They all span each of the 6 iterations but each can be used in a variety of ways.

This seems to suggest that the choice of types by writers is based on instantial decisions. The number of types and how many iterations they appear in are represented below in Figure 7.17.

Figure 7.17 Number of types and number of iterations

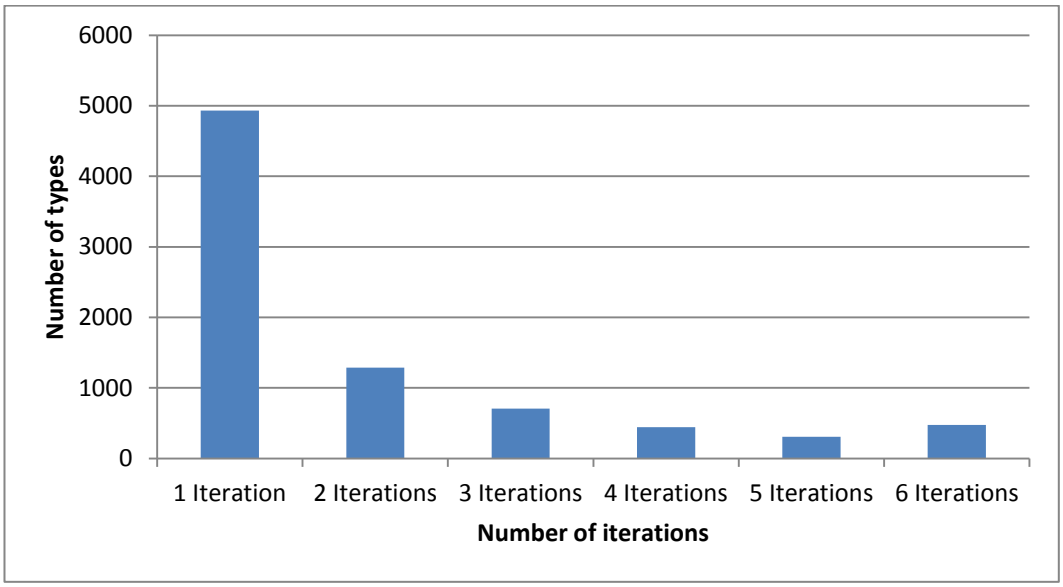


Figure 7.7 shows that the majority of types, in fact over 65% of types, appear in one iteration only. This number drops as we increase the amount of iterations included, dropping to 1,298, 706, 441 and 306 for the number of types appearing in 2, 3, 4 and 5 iterations respectively. Interestingly, the number of types appearing in all 6 iterations increases slightly from those appearing in 4 and 5, to 474 types. This phenomenon of being spread across all categories being preferred to missing from a few was already seen in Chapter 5 where more types appeared in all 94 texts than 93, 92 etc. In fact, the number of types appearing in a number of texts only exceeds 14 when we go back as far as 27 texts. The same phenomenon occurs at the level of the sub-corpora.

The distribution of types across the essay iterations, with each one of the 63 possible combinations of intersection being filled by a minimum of two types, points to an apparent randomness in the use of types. However, this randomness only exists at a level of analysis that is beyond that in which the texts were originally written. Within each individual text, and each part of that text, there is no randomness. In each particular instance of use of any type within the corpus, there is a reason, or set of

reasons, for its inclusion. Writers do not choose words at random and readers do not contribute meaning to those words at random. Despite the homogeneity of student, subjects, context and audience represented in MICUPE, there is uniqueness at every level of the corpus.

7.6 Conclusion

This chapter has shown that writers can change the number of types they choose to use to complete a set task. Furthermore, individual writers completing the same set task can choose to use a variety of types, both in frequency and in meaning. It would seem that it is a set of instantial decisions that give rise to these uses, as posited in Chapter 3 of the current research. At a text level, the writers are displaying great variance in the number of types each individual uses to complete the task. When essays are taken as texts, there are more single-use than multi-use types. When the essays are amalgamated, there are more multi-use types and single-use types. Whether a type is single-use or multi-use does not in any way indicate an importance as in its localised context each token is essential to the text existing as it does.

In relation to the corpus as a whole, this chapter has shown that the addition of each individual text has a unique effect on that corpus. The change in types as the corpus is added to shows an almost erratic change in the corpus. Not only does the addition of each text have a different effect on the corpus, Figure 7.1 shows that groups of texts have different effects and these groupings do not seem to adhere to semester divisions. When the corpus is completed, it has become a unique set of texts with a unique set of types used in unique way. This would hold for any corpus. Once completed, the analysis of the corpus becomes just that, an analysis of a corpus which is a collection of texts. The corpus may be representative of the 17 students writing for one particular lecturer, but in a way is no longer represents each individual text. In fact, the only representation of any individual text is that text itself.

In relation to the research questions outlined in Chapter 1 and the theoretical framework posited in Chapter 3, the present chapter has shown that uniqueness goes the whole way through the corpus and that responses to genre, task and audience are not standardised across writer or situation. This builds on the findings of the two previous chapters.

The following chapter, Chapter 8, will examine the apparent randomness created by instantial writer choice, seen to date in the current research, in further detail. It will argue that this apparent randomness adheres in some way to the mathematical model of Chaos Theory.

Chapter 8 Change as a chaotic dynamical system

8.1 Introduction

In previous chapters of the current research, it has been established that:

1. The existing paradigms for understanding academic writing, process-based approaches and genre-based approaches, are not appropriate for understanding change over time (Chapter 2);
2. The writer, who is a dynamic entity, and the context of situation, which is subjective and changing, interact complexly with each other and result in unique instantial choices by the writer. This is a more appropriate conceptualisation for understanding change over time (Chapter 3);
3. Frequency and distribution do not necessarily correlate across the corpus and each text has a set of types unique to that text, making each text unique (Chapter 5);
4. Task, genre and audience are factors in the creation of a text, but their influence is not uniform and where frequently occurring types are shared across texts, distribution and use varies, even in texts written under similar contexts (Chapter 6);
5. The uniqueness and apparent randomness of the use of types within the corpus is mirrored, and due to, the uniqueness of each sub-corpus, which in turn comes about because of the uniqueness of each text created by unique instantial choices by the writer (Chapter 7);
6. The differences in distribution of types, uniqueness of their occurrence and instantiality of use is despite the texts being produced at the same points in time by the same cohort of students in the same degree programme for the same assessor (Chapter 4).

Points 1-6 above give rise to an apparent randomness with regard to the use of types in the 94 essays. This apparent randomness exists despite a stability in the students, degree subject, institution and audience. To now in the current research, this has been accounted for by reference to instantial decisions made by the writers under subjective interpretations of a dynamic context of situation in tandem with the writer as a dynamic entity. The current chapter investigates the changes in the realisation of those instantial choices over the course of the degree, as represented by essay iterations 1 to 6.

It will be empirically shown that the rates of change of certain features within the texts are non-linear and apparently random. However, we know that the factors such as genre, task and audience do have some effect on every text, albeit a unique effect. We also know that every text in the corpus is unique, although there is some sharing of types between the texts both in relation to the sub-corpora and the corpus as a whole. Since every text is different, we can conclude that there is change over time in the essays. We can also conclude change from the grades awarded to the essays as each essay in this corpus was deemed by the assessor to be successful. No essay has an F (outright fail) grade. At any given point, the degree to which it was deemed successful can vary and the assessor awards a grade, such as A2 or B3, to reflect this. By essay iteration 6, the assessor would expect an improvement in terms of philosophical understanding as compared to essay iteration 1. If the texts in essay iterations 6 did not represent an improvement from the first semester of a degree programme to the final semester of a degree programme, those texts in the final semester would not have been deemed successful. Since each essay is unique and has been graded by the assessor in a manner appropriate to the stage of the degree programme the essay was submitted for, we can say for certain that there is change as the essay iterations are submitted. This change should be evident both at the level of the cohort (the sub-corpora) and that the level of each individual student (the texts).

The current chapter uses the notion of chaos to account for the change over time in the writing of the 17 students represented in the corpus. This is done by looking at a unique starting point for each student, dissimilar outcomes for each student in any given essay iteration and an apparent randomness in these changes despite the similarity in causal factors at any given essay iteration.

Chaos theory

According to one definition, 'chaos theory is the qualitative study of unstable aperiodic behavior in deterministic nonlinear dynamical systems' (Kellert, 1993). As the system is dynamical, it changes over time. Not only do the outcomes of the system change over time, but so does the influence of that system. Although this system is deterministic, it is not predictable. Non-linearity means that the output is not necessarily proportional to the input and the causes themselves are subject to synergistic reactions in which the whole deterministic factor at any given point is not necessarily equal to the sum of its parts. The instability and the aperiodicity referred to mean that the system does not repeat itself. If a system is deterministic, regardless of its complexity, an exact repeat within that system of any iteration would signal that the system must repeat itself.

In recent years, complexity theory, or the study of Complex Adaptive Systems (CAS), has grown in applied linguistics (for example, Larsen-Freeman, 1997; Ellis and Larsen Freeman, 2006; Ellis, 2008; Larsen-Freeman and Cameron, 2008; Seedhouse, 2010, Garcia and Kleifgen, 2010; Larsen-Freeman; 2013, Verspoor, 2013). This theory has been derived from chaos theory and tends to be used in a metaphorical way. Furthermore, applications of CAS tend to be based around one point in time. As the current research is empirically-based and spans 6 different points in time, it will refer to chaos theory, the original mathematical theory, not complexity theory. In doing so, the current research recognises the similarities between chaos theory and complexity theory and notes that in some instances they are considered synonymous.

In a chaotic dynamical system, there is a:

- sensitivity to initial conditions - any difference in the starting point can lead to dissimilar outcomes even under the same determining factors
- convergence around an attractor or strange attractor despite the apparent randomness
- replication of the pattern at different levels of magnification (self-similarity on various scales and levels)

This is because of:

- the self-organisation and adaption of many interacting agents
- surface complexity arising out of deep simplicity
- an interaction of its parts function as a whole, which is more than the sum of its parts

(Summarised from Larsen-Freeman, 1997 and Seedhouse, 2010)

The sensitivity to initial conditions is the cause of apparent randomness and non-linearity in rates of change. This then gives rise to a disproportionate effect from what seems a similar cause. An attractor is the position that the system is moving towards in a non-linear fashion. An attractor is called *strange* if it has a fractal structure and this is often the case when the dynamics are chaotic (Gleick, 1997).

Larsen-Freeman (1997) is often credited with bringing the notion of complexity theory to the study of language-learning, be it first language, second language, or third language. Since then, this notion has been used by various researchers. De Bot *et al*'s (2013: 199) asserted that complexity theory 'can unify and make relevant a number of different 'middle-level' theories on Second Language Acquisition'. Ellis, O'Donnell and Rommer (2013) analysed verb argument construction in the 100-million-word British National Corpus. Beckner *et al*'s (2009: 18) argument that linguistic patterns are:

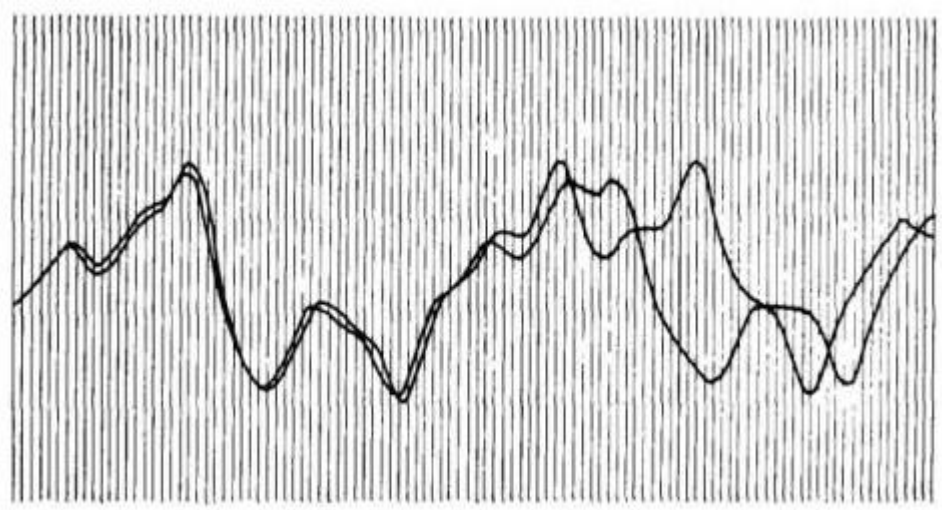
> emergent—synchronic patterns of linguistic organization at numerous levels (phonology, lexis, syntax, semantics, pragmatics, discourse, genre, etc.), dynamic patterns of usage, diachronic patterns of language change (linguistic cycles of grammaticalization, pidginization, creolization, etc.), ontogenetic developmental patterns in child language acquisition, global geopolitical patterns of language growth and decline, dominance and loss, and so forth.

To show that the change in student writing over time can be considered a chaotic dynamical system, this chapter will establish that in relation to the features already examined in the current research, there is a unique starting point, there is an apparent randomness and non-linearity across the sub-corpora and there is an apparent randomness and non-linearity across the students.

8.2 Unique starting point

The notion that a system can be extremely sensitive to initial conditions is often attributed to the work of Edward Lorenz. In his investigations into weather prediction, he noticed that what seemed like an inconsequential difference in starting point gave rise to very different outcomes. In Figure 8.1, there are two lines, one which starts at 0.506 and the other at 0.506127. Both initial points are subjected to the same deterministic system of mathematical equations whereby each iteration is used in the calculation of the next iteration. In Figure 8.1, both systems retain a level of similarity for a period of time. However, by the third and fourth 'humps' the systems no longer bear resemblance to each other.

Figure 8.1 Two weather patterns diverging from almost the same starting point (From Lorenz 1963)

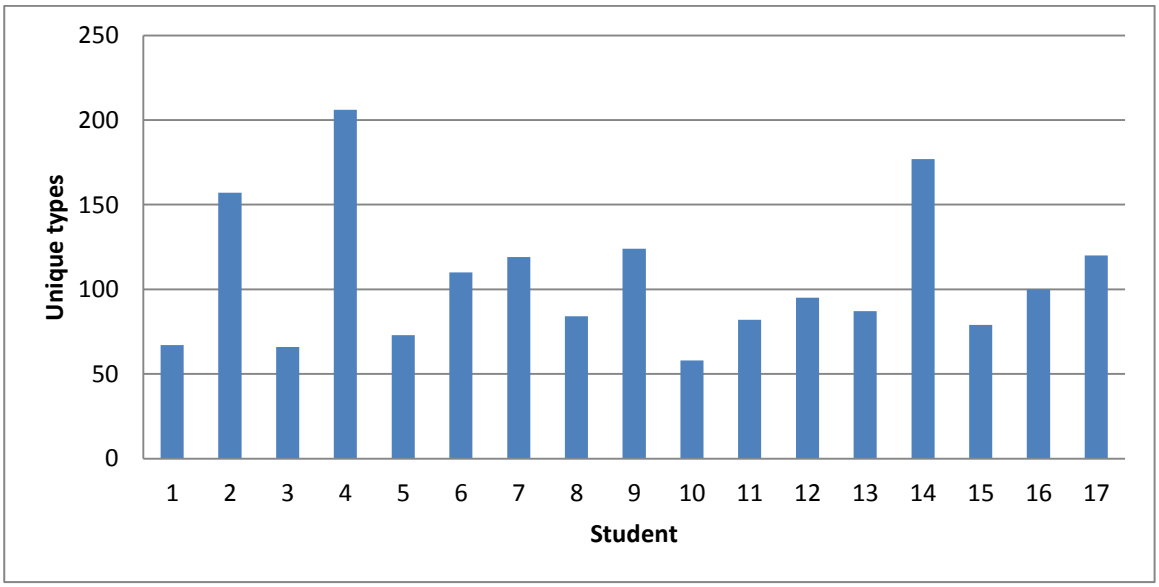


In a linear system, such a discrepancy in starting point would have an effect on a finishing point. However, that discrepancy would remain constant and proportional. In the printout from Lorenz, that proportionality and consistency is not evident. It is because of this that sensitivity to initial conditions is considered one of the factors of a chaotic, complex system. If we are to say that the change in essay writing within the corpus is a chaotic dynamical system, we must first establish that all students are starting at a unique point. It must be remembered that uniqueness is absolute and non-gradable. It is also important to note here that sensitivity to different starting points does not automatically indicate a chaotic dynamical system. This can only be done by examining the starting point plus the other conditions mentioned above.

As we have seen in Chapter 5, there is a set of words that are used only once in the corpus. Therefore, within the confines of the current data, those words are unique to that essay, or in other words unique to that student both within that iteration, but also within all 6 iterations in the corpus. Each word has a different meaning. Even if that word were synonymous with other words in the corpus, its contextual meaning is still different. If a text uses a word that no other text in the corpus uses, then that text must be considered unique within the corpus. Therefore, taking this into account, we can say that if each text has at least one unique word in comparison to the other texts, it is indicative of a unique starting point. We can say this because if a word is unique the text is unique. It is also important to note that this uniqueness is despite the similarities in genre, task, audience and other contextual issues.

Looking at MICUPE, there are 17 texts in essay iteration 1. The bar chart in Figure 8.2 below indicates the actual number of unique words for each of those 17 texts as determined by comparing them to each other. This is not a measure of uniqueness in relation to the full corpus, it is simply a measure of uniqueness in relation to the first iteration. Although not fully relevant here, it is worth noting that this chart simply indicates the number of single-use types. There is also a set of types that are unique to a text but appear in that text more than once.

Figure 8.2 Unique types in the 17 texts of essay iteration 1



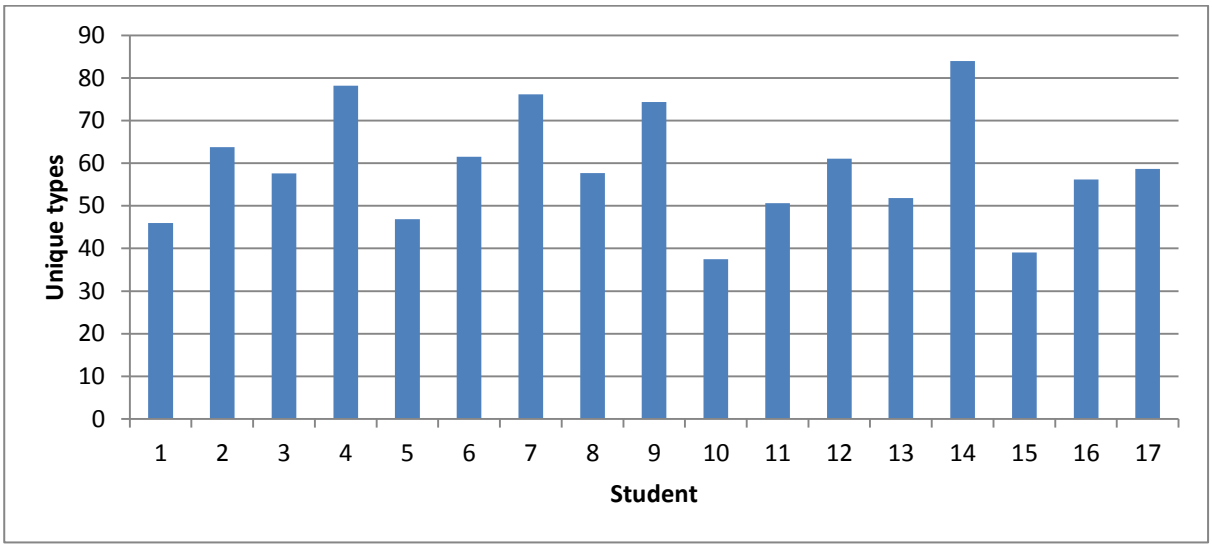
Here we can see that each text has some words that are unique to that text in relation to the other 17 texts in iteration 1. Not only is each text unique in that it had as a set of words that are unique to it, it is also unique in that each text has a unique number of those words. In the case of essay iteration 1, Student 10 has the fewest number of unique words at 58. Student 4 has the most unique words at 206.

In this chapter, in order to facilitate comparisons between the various essay iterations, the features are standardised to per 1,000 words when comparing at the level of the cohort (the sub-corpora). In doing this, it is recognised that the individuality of each text is, in a way, negated. This is because the essays are being produced for the same, or at least similar, task within the same institution for the same audience and in the same time and space. Therefore, the fact that each student uses a different number of tokens despite similar word count guidelines is an indicator of uniqueness. For this reason, when we are discussing differences in texts within the same submission point, we will not standardise to per 1,000 words. Furthermore, when examining the students changing over time in comparison to each other, we will not standardise to per 1,000 words. However, when we compare the iterations as 6 amalgamated sub-corpora, since they may have a different number of texts, it is best to standardise to per 1,000.

Taking into account that each text has a different number of words, for the purposes of illustration, we will convert these raw figures of unique types as seen in Figure 8.2 into a normalised to per 1,000 count.

In the standardisation to per 1,000 words, the effect on the graph is noticeable. Student 14 now has the most unique words at just short of 84 per 1,000. Student 4, who has the highest unique words in a raw count, has the second-highest in the standardised count. Student 10, similar to the raw count, has the fewest unique words at roughly 37.5 per 1,000. There is also a bunching of the data, with the range not appearing as spread out. As stated, the standardisation is necessary to control the different number of texts in the data for the 6 iterations. However, the effect of this standardisation must also be noted. No student wrote exactly 1,000 words. The standardised figure is therefore eliminating some of the instantial choices made by the students.

Figure 8.3 Unique types per text in iteration 1 standardised to per 1,000

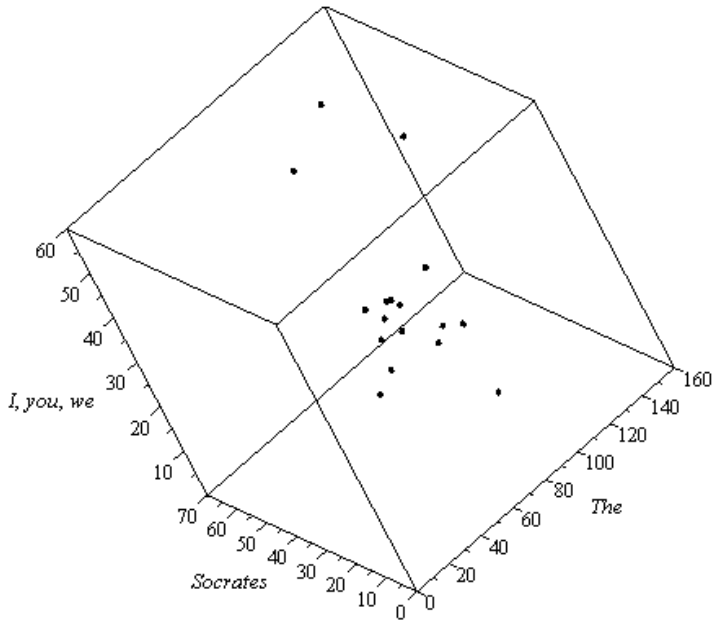


In both Figures 8.2 and 8.3, we can see that each essay in iteration 1 uses types that do not appear in any other essay. We can also see that each essay uses a different number of these types. Therefore, we can say that, taking this essay iteration as the starting point, each student has a different starting point because each student starts with a unique text as evidenced by the fact that each has words used by no other.

It must be noted that this is not the only area of uniqueness between each text in iteration 1, but for the purposes of examining whether the essays display the hallmarks of a chaotic system, this is perfectly sufficient as even the smallest difference is of significance. However, in Chapter 7, we saw that the distribution of types across the essay iterations, and hence across the texts within those iterations, is apparently random. Furthermore, in Chapter 5, we identified a set of features that seemed to stand out when frequency was correlated with distribution across texts. These features were the names of philosophers, words occurring in all texts such as *the* and the pronouns *I*, *you* and *we*. These three features were examined in further detail in Chapter 6, where it was argued that despite the similarity across texts in that they predominantly featured these words, each instantial use was unique. Despite the uniqueness of instantial uses, it was argued that these features are indicative in some way of a response to a task, genre and audience.

Although, as previously mentioned, there is certainly a uniqueness within each text, we can display that uniqueness using the features of the philosophers' name, *the* and the pronouns *I*, *you* and *we*. This is done in Figures 8.4, 8.5 and 8.6 below. The three figures are labelled *a*, *b* and *c* respectively as they represent three different views of the same 3-d plot.

Figure 8.4 17 texts in iteration 1 plotted according to *the*, *Socrates* and *I*, *you* and *we* (a)



In Figure 8.4, with *Socrates* on the bottom horizontal, it seems as if there are two clusters of texts. The first is towards the top of the figure and contains three texts. The second is closer to the middle of the figure and contains the other 14 texts.

Figure 8.5 17 texts in iteration 1 plotted according to *the*, *Socrates* and *I*, *you* and *we* (b)

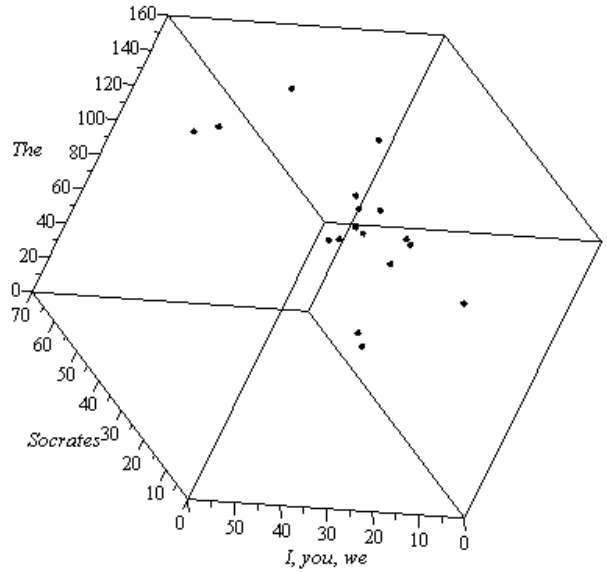
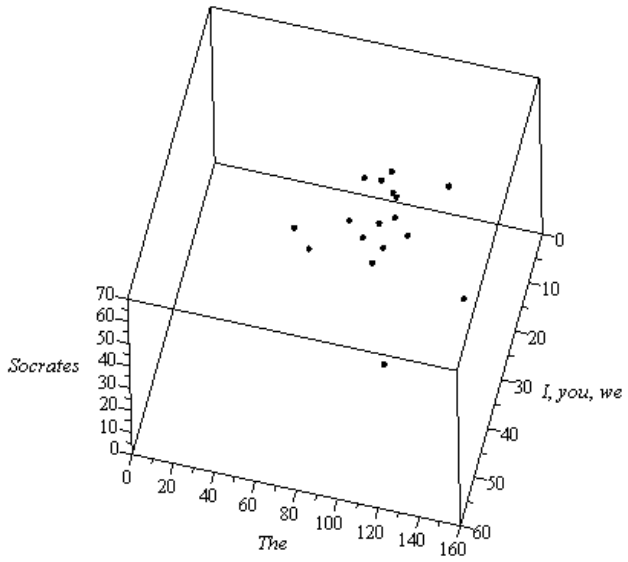


Figure 8.6 17 texts in iteration 1 plotted according to *the*, *Socrates* and *I*, *you* and *we* (c)



In figure 8.5, with *I*, *you* and *we* on the bottom horizontal, the clustering does not seem as separated as in Figure 8.4, yet the spread of texts seems to be greater. We must note here that this is based on appearance only as Figures 8.4 and 8.5 are in fact different perspectives of the same plot.

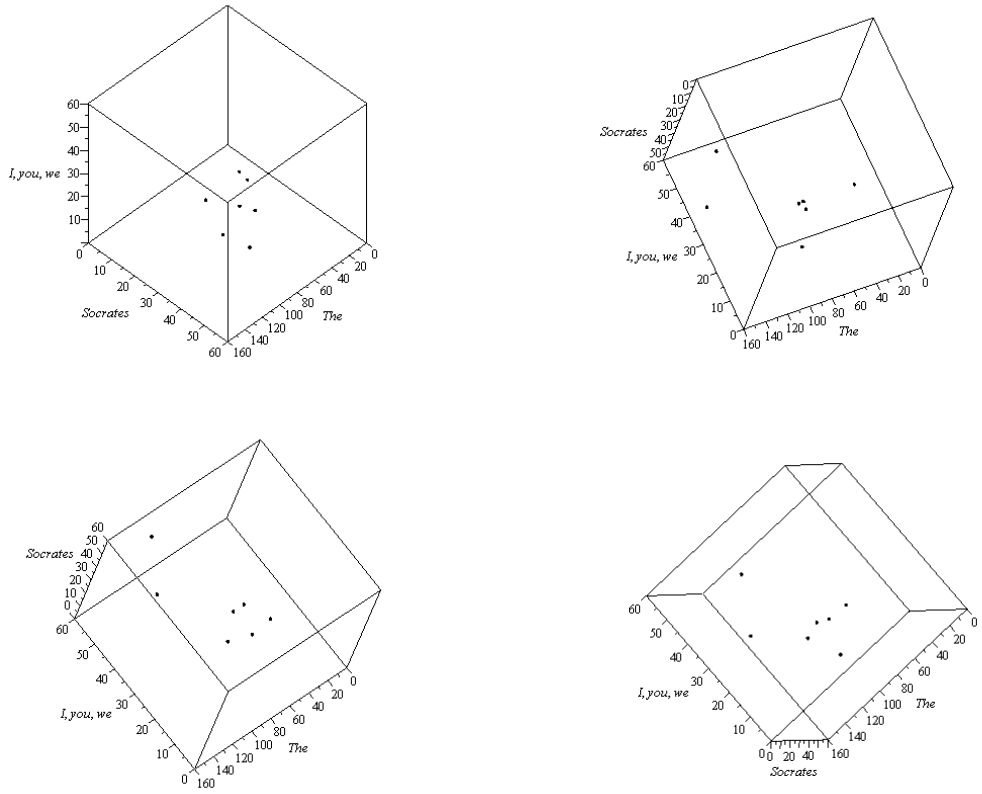
In Figure 8.6, with *the* on the bottom horizontal, there seems to be two clusters of texts. The first contains two texts and the second contains the other 15 texts.

In Figures 8.4, 8.5 8.6, which represent three different perspectives on the same plot, we can clearly see that each of the 17 texts has a different starting point.

Standardising for task

In Chapter 6, when looking at the distribution of the type *Socrates*, we looked at the seven essay texts that responded to the title *Socrates was a martyr for the truth. Discuss.* This was done to show that the apparent randomness in the use of types was not simply the result of differences in task selection. Here, we return to those seven essays to show that across the three features (name of philosopher, *the* and the pronouns *I*, *you* and *we*) there are differences in those texts. Figure 8.7 shows four different perspectives on a 3-d plot for those 17 texts responding to the same task plotted according to *Socrates*, *the* and *I*, *you* and *we*.

Figure 8.7 Seven essays responding to the same task in iteration 1



In Figure 8.7, we can clearly see that, despite the similarity in task, these seven texts are different and if we take iteration 1 as a starting point, each writer has a unique starting point.

At this point, we can see that, as evidenced through the empirical data of the texts themselves, in essay iteration 1, each student writer has a unique starting point. To show that change in the student writing is a chaotic dynamical system, the next step is to show that these unique starting points lead to dissimilar outcomes even under the same influencing factors. This will be done firstly at the level of the cohort (the sub-corpora) and secondly by looking at individual students.

8.3 Apparent randomness and non-linearity across sub-corpora

In section 8.2, we saw that each student has a unique starting point. The current section examines how the essay iterations change over time in relation to the cohort as a whole. To do so, we will concentrate on the indicators of genre, task and audience that were identified in Chapter 5 and investigated in further detail in Chapter 6. As stated previously, when examining change across the iterations as amalgamated sub-corpora, we will standardise to per 1,000 words, despite the fact that doing so negates the choices made by the writers.

Figure 8.8 below shows the changes in the frequencies of *I*, *you* and *we* over the course of the six essay iterations. These are then added together to create an indicator of interaction and shown together with *the* and the name of the philosopher in the task.

Figure 8.8 *I*, *you*, *we*, *the*, *philosopher* over 6 iterations

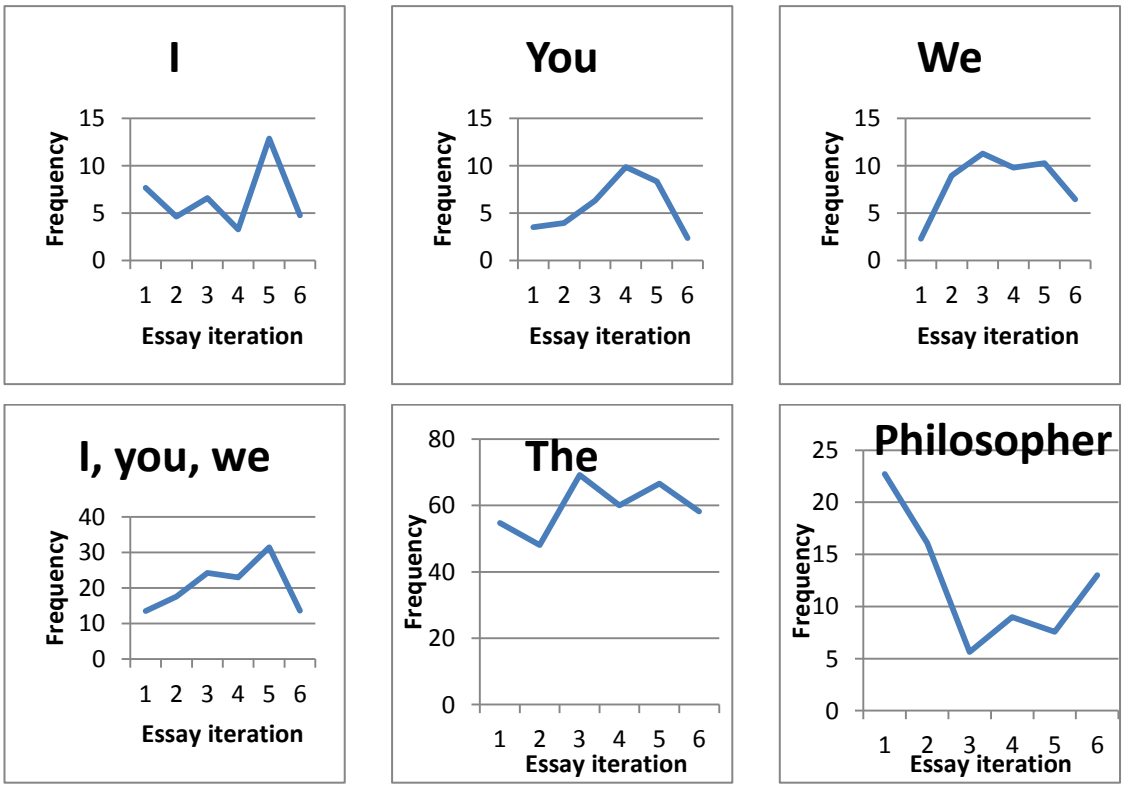


Figure 8.8 shows the changes over the course of the 6 essay iterations in some of the features already identified in the current research. It is striking that the frequency of each of these features does not change in a linear manner. Furthermore, in the main,

they are not similar to each other. For example, *I* peaks in essay iteration 5, *you* peaks in essay iteration 4 and *we* peaks in essay iteration 5. Each of the six frequency plots in Figure 8.8 has a different shape. It seems as if the similarity between the frequencies of these features is simply their non-linearity and lack of pattern.

Taking *I*, *you* and *we* combined, *the*, and the name of the task philosopher, we can plot the iterations in a 3-d plot using these features as the axes. This is done in Figures 8.9, 8.10, 8.11 and 8.12 which show four different perspectives of the same graph.

Figure 8.9 Change across iterations - perspective A

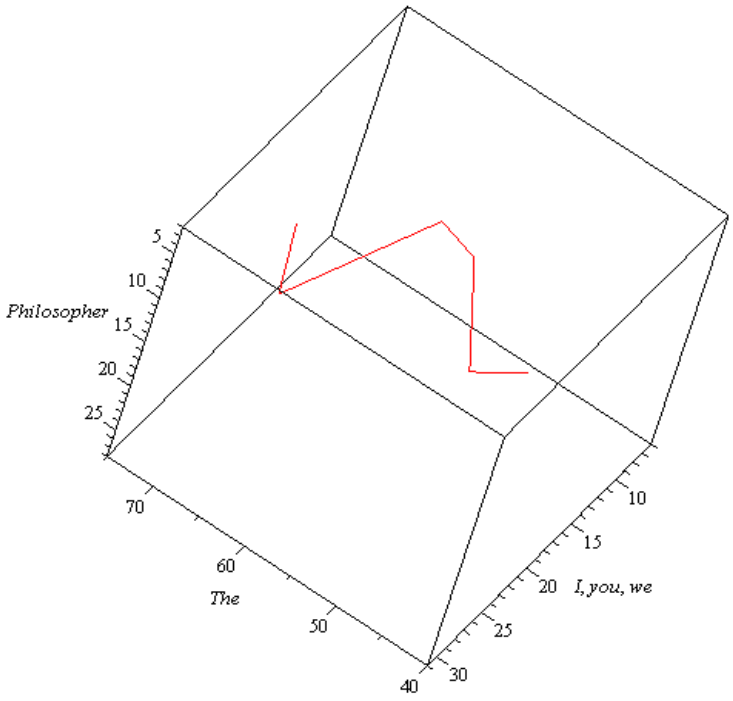


Figure 8.10 Change across iterations - perspective B

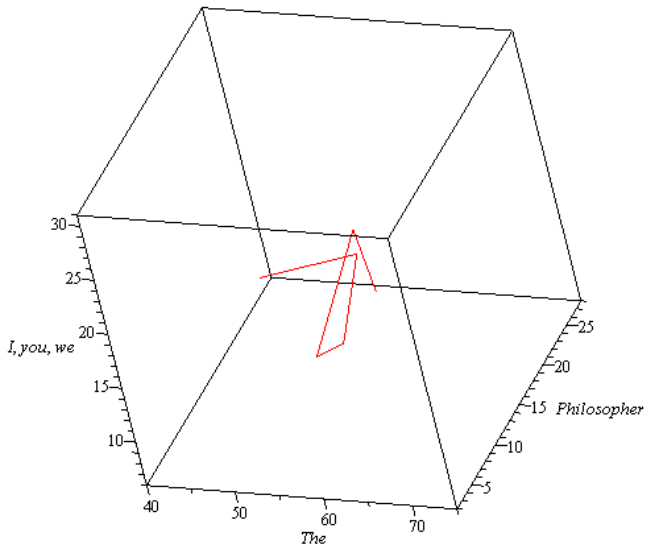


Figure 8.11 Change across iterations - perspective C

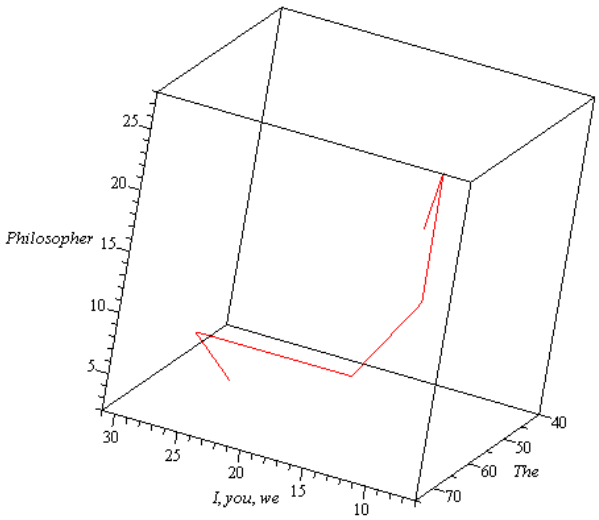
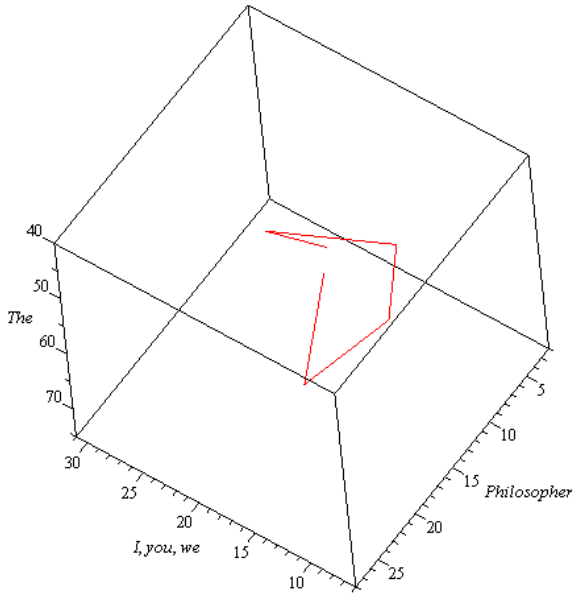


Figure 8.12 Change across iterations - perspective D



Figures 8.9 to 8.12 show that there is a non-linearity of change for the cohort over the course of essay iterations 1 to 6. From various perspectives, the change seems circular, spiral and disjointed. It is certainly non-linear and apparently random.

8.4 Apparent randomness and non-linearity across students

We have seen that there is an apparent ramdomness in the cohort in the changes over time. To allow looking at another level of magnification (attractors of chaotic systems are frequently 'self-similar', that is, display replication at different levels of magnitude) and to further investigate the individual instantial responses, the current section looks at the 94 texts as individual texts across the iterations, not amalgamated sub-corpora. The data in this section is not standardised to per 1,000 words for the reason, as previously stated, that standardisation negates the effect of some of the individual choices made by the writers. To begin with, we will examine change in the frequency of *I* for five students within essay iterations 2 to 5. Then, we will look at the cohort over the six essay iterations across *I*, *you* and *we* combined, *the* and the task philosopher.

Essay iterations 2 to 5 were submitted within 6 weeks of each other in the fourth semester of the degree programme. There are five students in MICUPE who composed essays on the same four tasks across those six weeks. These students are Students 6, 8, 10, 13 and 17. The first of the tasks, essay iteration 2, was *Explain Aquinas' account of being*. The frequency of *I* in the five essays is shown below in Figure 8.13.

Figure 8.13 Frequency of *I* in iteration 2 across five texts

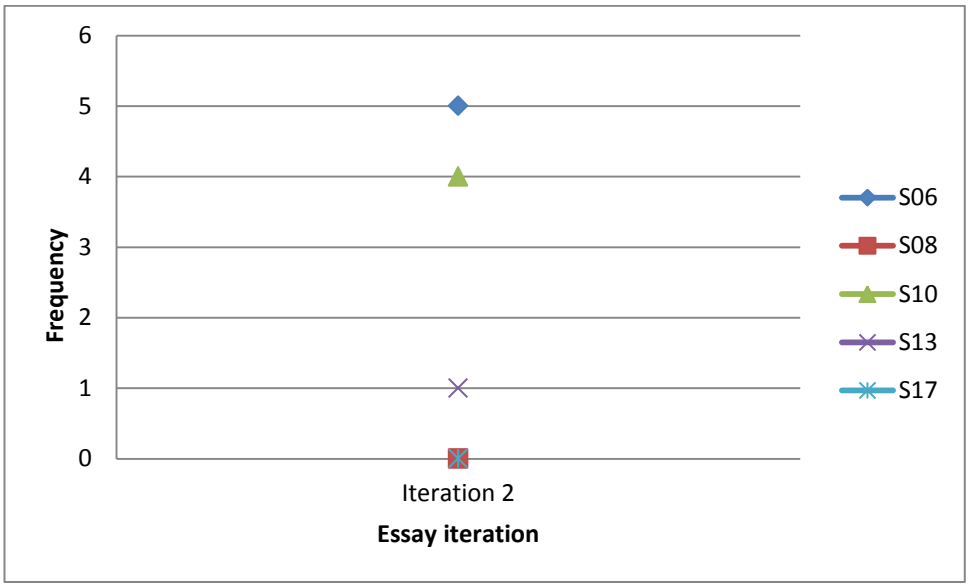
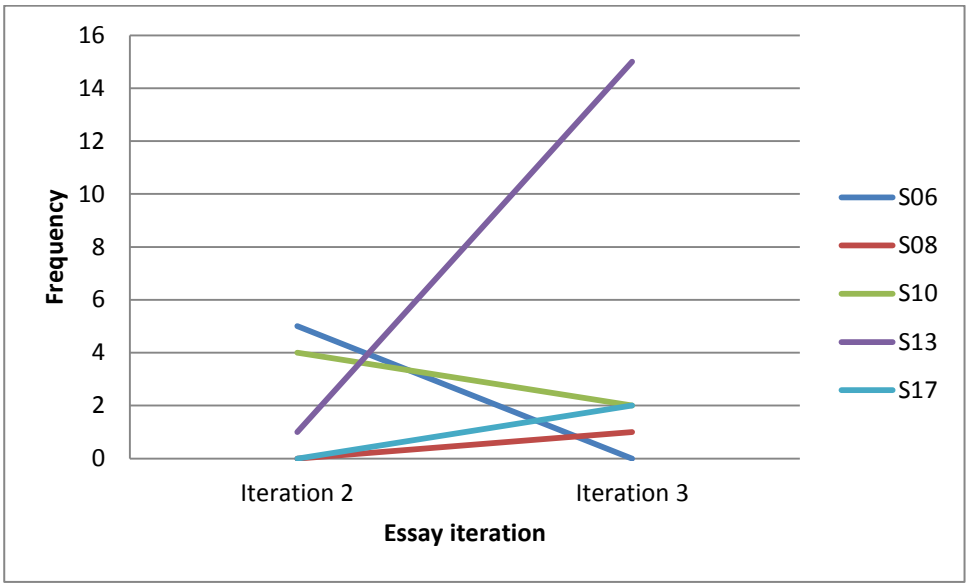


Figure 8.13 Shows that there is a range for the frequency of *I* in the five texts, despite consistency across task, audience and genre. Student 6 uses *I* 5 times in their text while Students 8 and 17 do not use *I* in their texts.

The essay title for the next task, essay iteration 3, was *Lonergan: the dynamic aspect of knowing*. This essay was submitted approximately two weeks after essay iteration 2. The frequency of *I* in those five texts, along with that of iteration 2, is shown below in Figure 8.14.

Figure 8.14 Frequency of *I* in iterations 2 and 3 across five texts

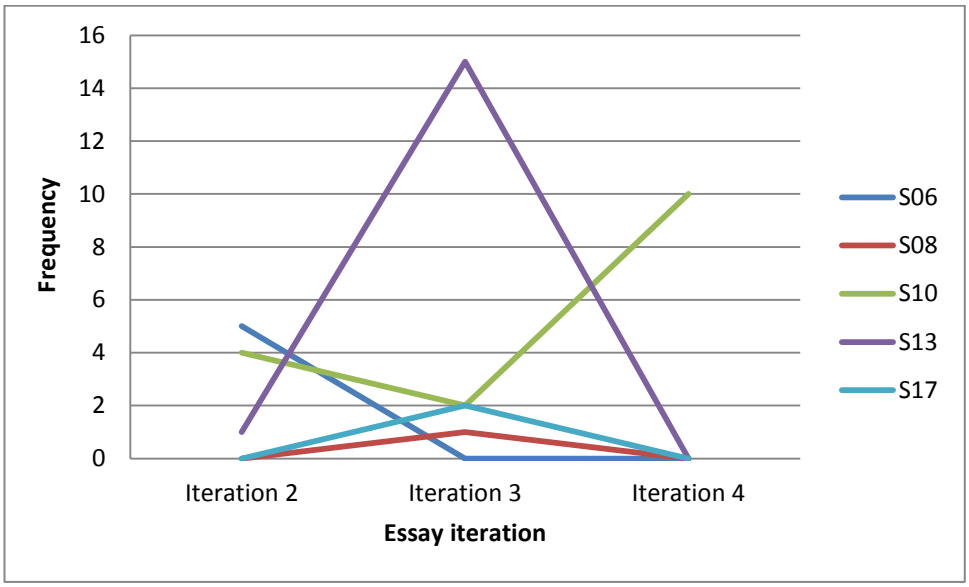


In Figure 8.14, we can see that the change in the occurrences of *I* is not linear from essay iteration 2 to essay iteration 3. Student 6, who had the most occurrences of *I* in iteration 2, now has the fewest occurrences. Student 13 has dramatically increased the frequency of *I* in their text in iteration 3 as compared to iteration 2. Students 10 and 17 have the same frequency of *I* in this iteration despite displaying a pattern differing from each other in the previous iteration. Taking the five students together, some students increase the frequency of their use of *I* from iteration 2 to iteration 3, while other students decrease their frequency of use of *I*.

The essay title for the next task, essay iteration 4, was *Dramatic Bias*. This essay was submitted approximately two weeks after essay iteration 3. The frequency of *I* in those five texts, along with that of iterations 2 and 3, is shown below in Figure 8.15.

In Figure 8.15, we can see a continuation of the non-linear change that was seen in Figure 8.14. Student 10 uses *I* a total of ten times in this essay. The other four students do not use *I* in this essay.

Figure 8.15 Frequency of *I* in iterations 2, 3 and 4 across five texts



The essay title for the next task, essay iteration 5, was *Distinguish the act of reflection that leads to judgement from the act of enquiry that leads to insight*. This essay was submitted approximately two weeks after essay iteration 4 and approximately 6 weeks after the first of these for essay iterations. The frequency of *I* in those five texts, along with that of iterations 2, 3 and 4, is shown below in Figure 8.16.

Figure 8.16 Frequency of *I* in iterations 2, 3, 4 and 5 across five texts

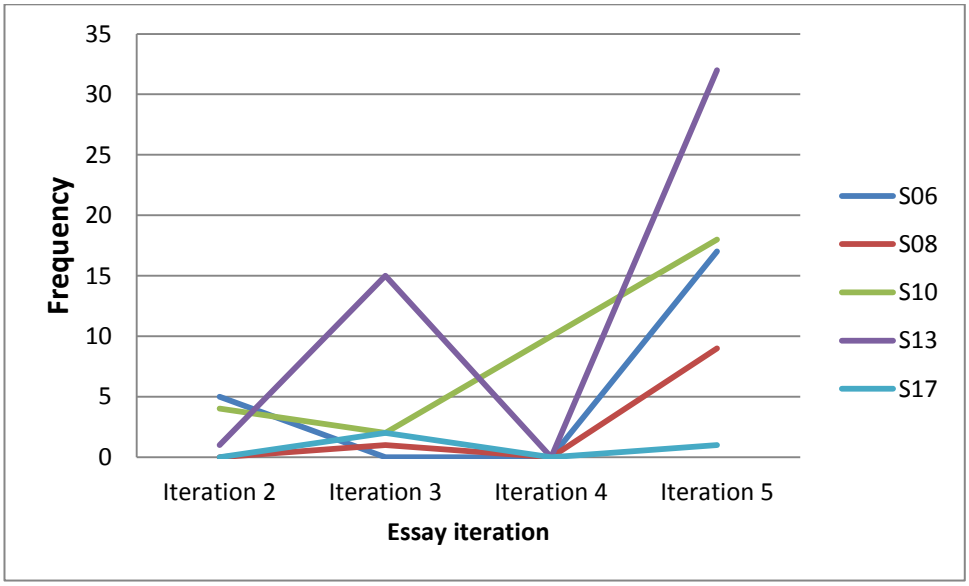


Figure 8.16 shows the frequency of *I* across essay iteration 2 to 5 for five students. Within each of the essay iterations, the task is the same for all five students. We can see a continuation of the non-linearity seen in Figures 8.14 and 8.15. Each student follows a different path in terms of frequency of *I* from iteration 2 to iteration 5. Students 6 and 10 have a similar frequency of *I* in iteration 2 (5 and 4 respectively) and iteration 5 (17 and 18 respectively), yet are dissimilar in iterations 3 and 4. Student 13 displays an erratic pattern of change, while Student 17 maintains a relatively stable pattern. Student 8 seems to have a stable of pattern of use for *I*, however this is not the case in relation to iteration 5.

Figures 8.13 to 8.16 demonstrate a non-linear change in the use of *I* across individual students over the course of four essay iteration submitted within six weeks of each other. This change is apparently random and is not standardised according to task, genre or audience as these factors remain constant. When combined with the instantiality of meaning dependent on context of use explored in Chapter 6, this points to change within the corpus as a chaotic dynamical system. To further investigate this we will look at the essay texts within MICUPE across *I*, *you* and *we* combined, *the* and the task philosopher.

Figures 8.17, 8.18, 8.19 and 8.20 below outline four different perspectives as each text is plotted according to *I*, *you* and *we* combined, *the* and the task philosopher. Essay iteration 1 is represented by red, iteration 2 by blue, iteration 3 by green, iteration 4 by black, iteration 5 coral and iteration 6 by orange.

Figure 8.17 94 texts according to essay iteration - perspective A

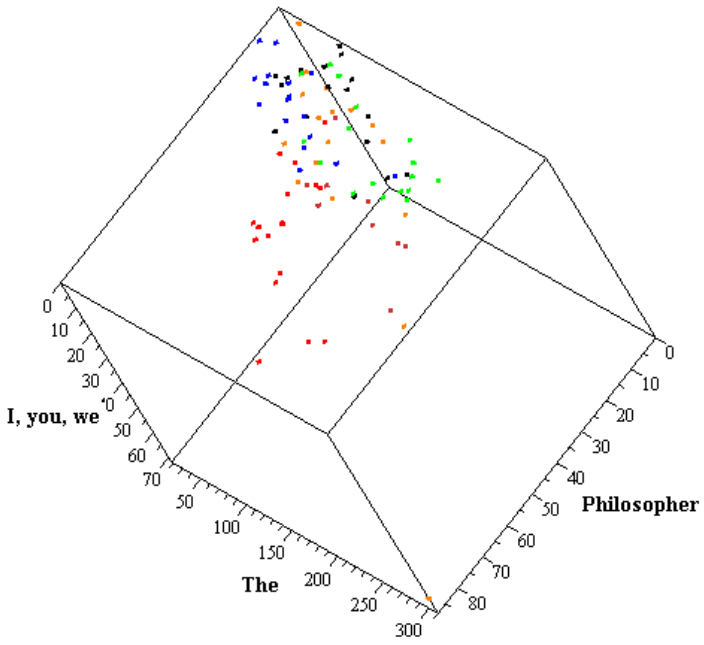


Figure 8.18 94 texts according to essay iteration - perspective B

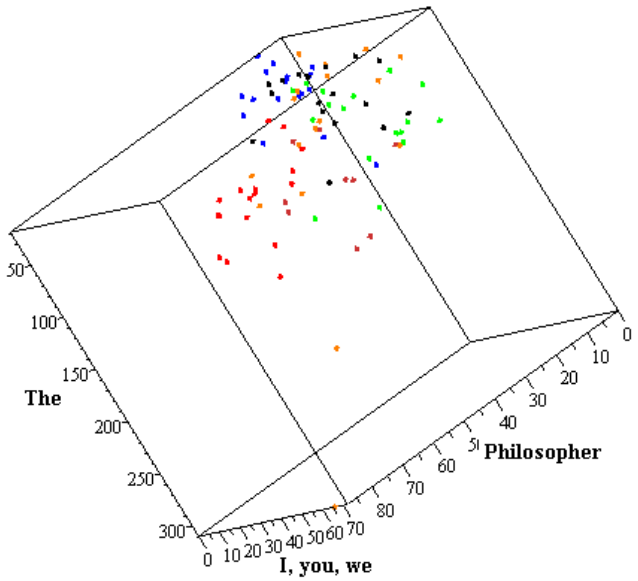


Figure 8.19 94 texts according to essay iteration - perspective C

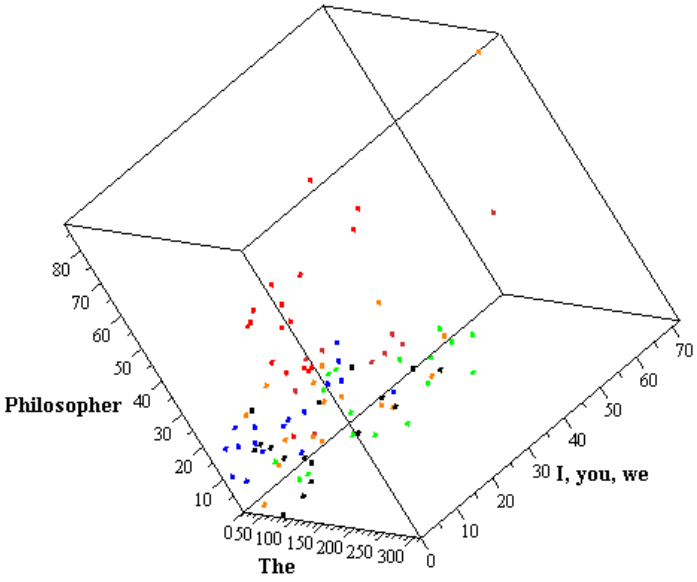
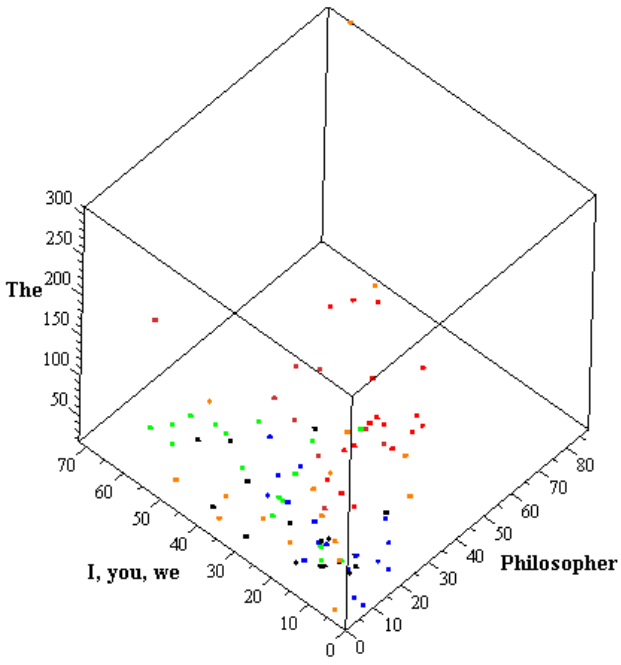


Figure 8.20 94 texts according to essay iteration - perspective D



Figures 8.17 to 8.20 show that the individual texts within the iterations are not grouping according to iteration, showing many overlaps. If we look at the same data set but with the points for each student joined together, we can see how each individual student changes on the three parameters under investigation over the course of the degree programme. This is represented in figures 8.21, 8.22, 8.23 and 8.24 where each individual is represented by a different colour.

Figure 8.21 Change over the 6 iterations by student - perspective A

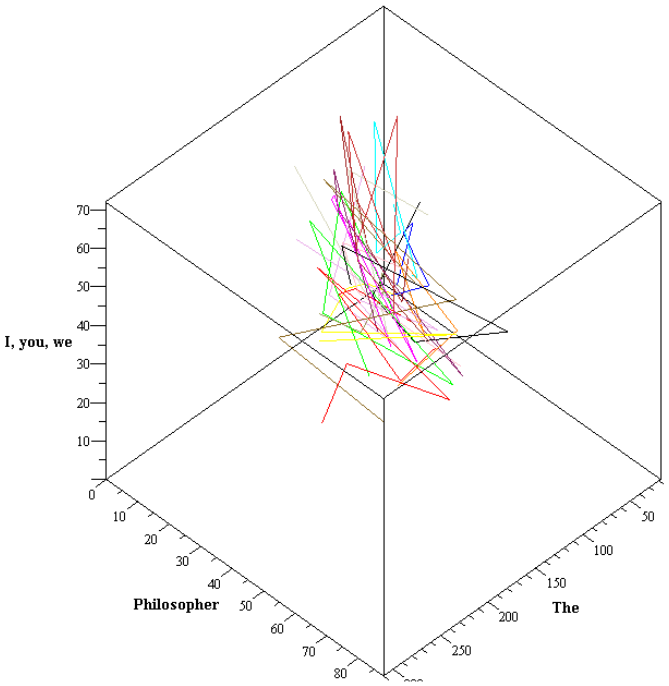


Figure 8.22 Change over the 6 iterations by student - perspective B

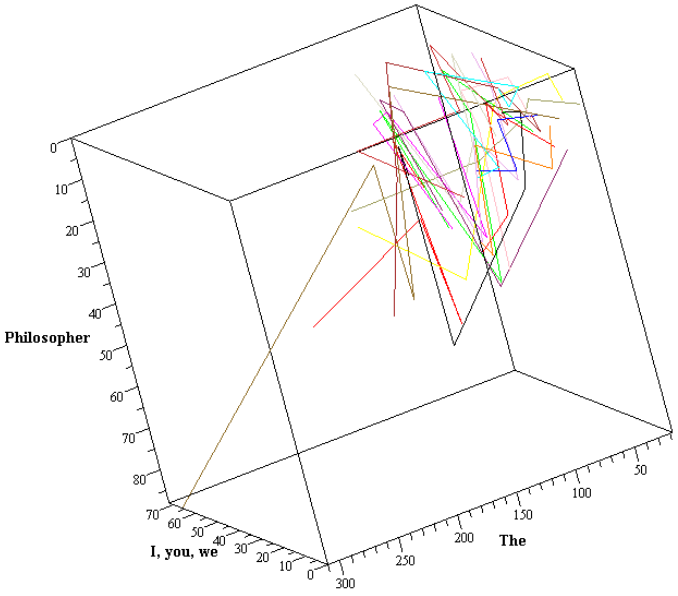


Figure 8.23 Change over the 6 iterations by student - perspective C

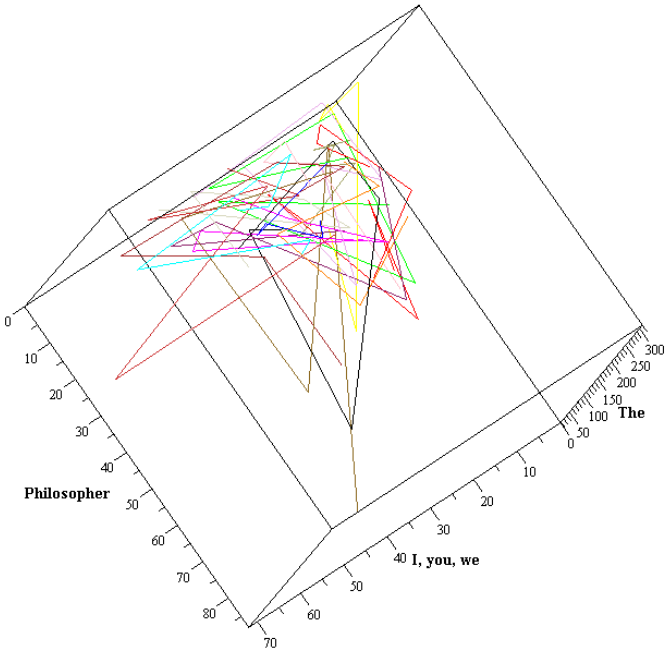
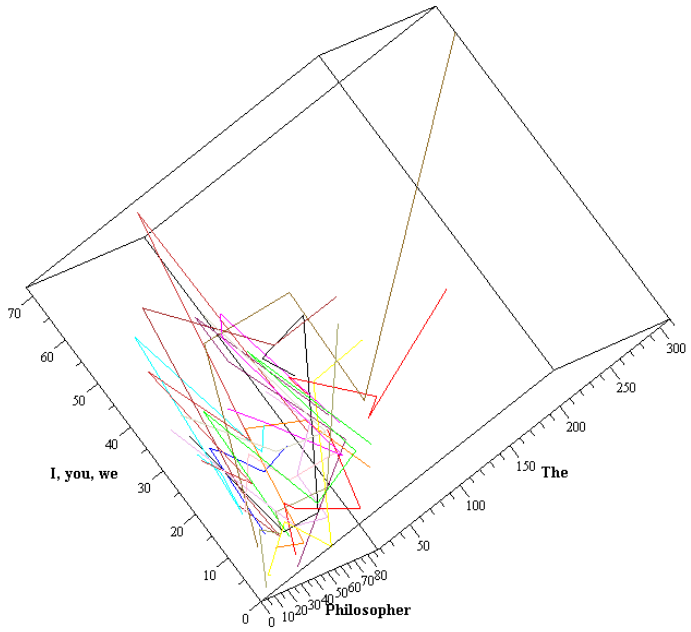


Figure 8.24 Change over the 6 iterations by student - perspective D



Figures 8.21 to 8.24 show the changes over time for each individual student. A number of points become apparent:

- Each student follows a unique path;
- This path is not linear;
- Similar contexts can generate different outcomes.

Similar essay titles

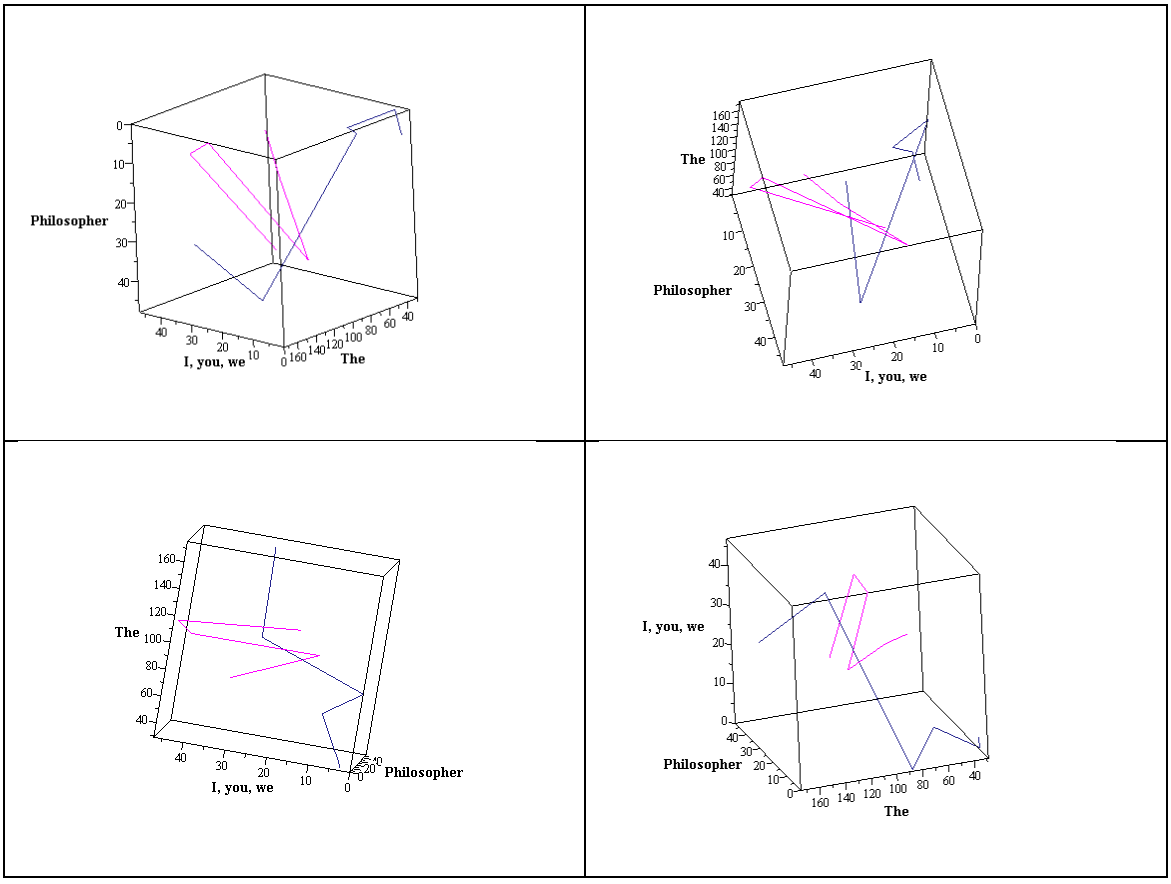
In each essay iteration, there are various titles that the students write to. Student 6 and Student 17 write to the same essay titles over the course of the 6 essay iterations. Those titles are shown in Table 8.1 below.

Table 8.1 Essay titles for Student 6 and Student 17

Iteration	Title
Essay 1	Socrates committed suicide. Discuss.
Essay 2	Explain Aquinas' account of being.
Essay 3	Lonergan: the dynamic aspect of knowing.
Essay 4	Dramatic bias.
Essay 5	Distinguish the act of reflection that leads to judgement from the act of enquiry that leads to insight.
Essay 6	Evaluate the argument in the first three of the five ways of Aquinas.

If we take these two students as an example, we can compare how two students with unique starting points can have dissimilar outcomes despite the apparent similarity in determining factors. In Figure 8.25 below, the task is standardised.

Figure 8.25 Student 6 and Student 17 across 6 essay iterations



In Figure 8.25, we can see that although task, genre and audience are standardised, the response to these factors by the individual students is far from standardised and their respective paths of change are highly individualised.

8.5 Conclusion

One of the criteria of a complex dynamical system is that there is a sensitivity to initial conditions. The current chapter has shown that there is unique starting point for each student as evidenced by the uniqueness inherent in each text in essay iteration 1.

Another criterion of a complex dynamical system is that there is an apparent randomness and a disproportionate effect despite what seems to be a similar cause. The current chapter has shown that for the factors previously identified that indicate genre, task and audience, there is an apparent randomness and a disproportionate effect in the writing of the students in the study.

The third criterion of a complex dynamical system is that there is a non-linearity in the rates of change. The current chapter has shown this non-linearity in the frequencies of *I*, *you* and *we*, *the* and the name of the philosopher dictated by the task.

The fourth criterion of a complex dynamical system is that the system displays self-similarity on various scales and levels. The current chapter has shown this self-similarity in relation to both the cohort and the individual students who make up the cohort changing over time. In both cases, non-linearity, sensitivity to initial conditions, apparent randomness and disproportionate effects from what seems a similar cause were demonstrated.

Taking these four criteria into account, we can posit that the change in the student writing over the course of their degree programme follows that of a chaotic dynamical system in an empirical manner, not just in a metaphorical sense as has hitherto been used in Applied Linguistics.

The research sub-questions posed in Chapter 1 of the current research are as follows:

- Is each text unique?
- How do writers respond to genre, task and audience?
- Is the response to genre, task and audience standardised across writer and/or situation?

From the four criteria outlined above, in relation to the research sub-questions, we can say that:

- each text is unique;
- writers respond to genre, task and audience in a complex, instantial manner;
- this response is unique to that particular point in time and is not standardised across writer and/or situation.

Having answered these research sub-questions, we can now address the main research question. This chapter has demonstrated that there is not a linear change in the student writing over time. That change, however, is patterned in the way that a chaotic dynamical system is patterned. This means that the pattern is determined but does not mean that the pattern is predictable.

The following chapter, Chapter 9, will summarise the current research, re-address the research questions and examine the implications arising from this study.

Chapter 9 Conclusion

9.1 Summary

The research questions of the current research, as outlined in Chapter 1, are as follows:

- Main research question:

  Is there a patterned, linear change in the student writing over time?

- Sub-questions:
  1. Is each text unique?
  2. How do writers respond to genre, task and audience?
  3. Is the response to genre, task and audience standardised across writer and/or situation?

Chapter 2 outlined the concepts of viewing writing as process-based and writing as genre-based. The process approach is primarily concerned with the steps a writer takes in the production of a text. These steps are non-linear and recursive, and their continual realisation involves the self-actualisation of the writer and, hence, the production of a text. Genre-based approaches are focused on the finished product, the text, and the linguistic features contained within the text and sees those features as a direct result of the social space in which the text was created in tandem with the communicative purpose of that text.

Chapter 3 outlined the limitations of process-based and genre-based approaches to academic writing. A theoretical framework for the current research was then developed, taking into consideration the limitations of both approaches, based around the notions of writer, context of situation and text. It was argued that each text must be unique and differ from every other text because each individual writer is engaged in a process of

change and this change is non-linear. The writers' instantial subjective interpretation of the context of situation, including their interpretation of genre, task and audience, together with the instantial linguistic choices they make must result in unique texts.

Chapter 4 outlined the data gathering and methodology used in the current research as designed in order to answer the research questions as outlined in Chapter 1. The corpus, MICUPE, was created with 94 texts written by 17 students at 6 submission points over the course of a degree programme for one assessor. From this, an analytical framework for the analysis chapters, Chapters 5-8, was developed.

Chapter 5 examined various sections of the frequency list, comparisons of these sections with another academic writing corpus (BAWE) and comparisons of frequency lists between the six sub-corpora that constitution MICUPE. Chapter 5 also correlated raw frequency with number of texts. Outliers within MICUPE in terms of frequency versus distribution were identified and included function words such as *the, in, of,* and *for* that appear in each of the 94 essay texts, the pronouns *I, we* and *you* and the names of the authors of the primary texts for each essay such as *Socrates*, *Aquinas* and *Lonergan*. This chapter answered one of the research sub-questions, namely the question concerning the uniqueness of each text. In terms of use of types, an analysis of the empirical data has shown each text to be unique.

Chapter 6 examined some of the outliers as identified in Chapter 5, including some of the 14 words that appear in all 94 essay texts, the use of philosophers' names and the three pronouns *I*, *you* and *we*. Taken together, the analysis of these three features suggested that the choices that the writers make are based on influences such as task, the genre and audience. However, it was also apparent that these influences have differing effects on each writer at particular points in time. In relation to one of the research sub-questions (How do writers respond to genre, task and audience?), it was demonstrated that all students respond uniquely to the concerns of genre, task and audience. The complete text is a response to these concerns and not simply certain words, and by focusing on certain indicators of the writer response, we see that each text has a different realisation of these indicators. No essay iteration elicits the same response from two or more students and each student changes how they respond over time.

Chapter 7 showed that writers can change the number of types they choose to use to complete a set task. Furthermore, individual writers completing the same set task can choose to use a variety of types, both in frequency and in meaning. It was argued that it is a set of instantial decisions that give rise to these uses. Furthermore, there was an apparent randomness in the sharing of types across essay iterations. In relation to the research sub-questions, this chapter has demonstrated that uniqueness goes the whole way through the corpus and that responses to genre, task and audience are not standardised across writer or situation.

Chapter 8 investigated whether the change in student writing could be considered a chaotic dynamical system. Due to sensitivity to initial conditions, apparent randomness and a disproportionate effect despite what seems to be a similar cause, a non-linearity in the rates of change and self-similarity on various scales and levels, it was concluded that the change in student writing over time could be considered such a system. This chapter addressed the main research question by demonstrating that there is not a linear change in the student writing over time. That change, however, is patterned in the way that a chaotic dynamical system is patterned. This means that the pattern is determined but it is not predictable.

Throughout the analysis contained in Chapters 5 to 8, a by-product of answering the research questions was a questioning of the nature of corpora, corpus linguistic techniques and what the techniques applied to a corpus can tell us.

9.2 Implications of the research

Academic writing

In relation to the teaching of academic writing, learning and change take place naturally. Somebody involved in the teaching of academic writing must be aware of this and also aware that natural change is not linear. Every individual student changes and progresses at their own unique rate. Items can be learned and then unlearned and there is uniqueness in each essay due to the uniqueness of each individual and their unique way of interpreting the context and responding to it. For this reason, the best way of

enhancing the students' learning is in individualised sessions that are specifically tailored to the needs and place of the student at that particular point in time.

With regard to learning materials, prescriptiveness must be avoided. This study has shown that the range and variety of instantial meanings that each writer will communicate to the reader is not through a narrow set of options. Although materials may be based on empirical data, the prescriptive nature of suggesting that one set of choices is preferable to another set for every situation is at best questionable.

Corpus linguistics

The implications of the current study for corpus linguistics are interesting. A corpus is generally synchronic, meaning that it provides a snapshot in time. Anybody working with a corpus must be aware of the limitations inherent in analysing one point in time and must acknowledge that should the language producers represented in that corpus take part in similar events at another time, the corpus would be different, giving rise to another set of data for analysis.

A corpus is constructed from a series of texts, and at times those texts can also form sub-corpora. The texts themselves exist due to a series of choices made by the writer (or speaker in some corpora). The language producers create the texts from which a corpus is constructed, resulting in a certain frequency of a word or phrase and a different frequency of another word or phrase in the same corpus. Furthermore, those words or phrases, despite similarities, have instantial meanings, but are an attempt to communicate with the reader (or listener) in relation to the task in a manner deemed appropriate. When we look at a corpus, we are looking at the relic of those choices, but those choices were not made before the text was created. They are instantial at the time of text creation.

Furthermore, this research has shown that influencing factors have individualised, instantial effects on language. These effects are a complex combination of those factors as realised at any given time. For this reason, researchers involved in corpus linguistics need to be extremely careful about attributing simple, linear explanations for the

appearance and meaning of a word or phrase, even if their empirical data suggest that a majority within the particular corpus adhere to one particular pattern.

Theory of language

The current research has shown that in the particular context of academic writing, change is apparently random and non-linear. However, we know that the appearance of words is not random. We can say this is because the unique and unpredictable, although at times patterned, combinations of those words communicate a meaning to the reader. If the appearance of words were truly random, meaning would not be communicated. Our understanding of language, hence our analysis of language, is better enhanced by conceptualising language in use as a chaotic dynamical system.

There are, however, drawbacks to conceptualising language as a chaotic dynamical system. The first is based on the fact that this notion is a mathematical concept. Although the data in the current research seem to adhere to such a system, the transfer of a mathematical concept to our understanding of language is questionable. For this reason, further investigation is needed and a framework and theory for understanding language needs to be developed based around the empirical data of the language itself. It is probable that such a theory would include some of the elements of chaos theory, but yet it still needs to be tailored to language analysis and understanding.

In mathematics, chaos theory is the result of a deterministic system. This means that all elements and outputs are predetermined. As a researcher who values the individuality and creativity involved in language production, it is disheartening to suggest that all language production is predetermined, despite the difference between predictability and predetermination. This is also reminiscent of an observation by Lorenz regarding the predictability of the weather:

> When our results concerning the instability of non-periodic flow are applied to the atmosphere, which is ostensibly non-periodic, they indicate that prediction of the sufficiently distant future is impossible by any method, unless the present conditions are known exactly. In view of the inevitable inaccuracy and incompleteness of weather observations, precise very-long-range forecasting would seem to be non-existant.
>
> (Lorenz, 1963: 141)

The reason language is not predictable is because there is an implied acceptance that we cannot know everything about everything and therefore cannot predict language because the smallest differences in initial conditions can lead non-similarity of outcomes. However, conceptually, the notion that we can predict all language use if it were possible to know everything is problematic.

Be that as it may, this research has shown quite clearly that any analysis, and hence theory, on language needs to take into account all empirical instances evident in the data that is being used. If we base our assumptions on the majority, even if that majority is 99%, our assumptions are still lacking.

9.3 Limitations of the current study

This research has reached a tentative conclusion that the change evident in the cohort of undergraduate philosophy students under investigation can be viewed a chaotic dynamical system. At present, we can hypothesise that such may be the case. Any certainty or generalisability is constrained by the three following limitations:

1. The current research is based on one cohort of students in one site.
    - In a similar manner to the individuality and uniqueness of text as evidenced in the current research, the cohort partaking in the current research and their interaction with the institution must also be seen as unique. This has an adverse effect on the generalisability on any findings.

2. The current research is limited to academic writing;
    - The current research tentatively claims that the change in the academic writing of the individuals comprising the cohort in the study resembles that of a chaotic dynamical system. In addition to the limitation previously mentioned, namely the data contains one cohort, the data is also limited in specificity to academic writing.

3. The current research is limited to six iterations.

- The tentative claim that change in the current data resembles a chaotic dynamical system is limited by the fact that there are only six iterations in the data. Further iterations could either enhance the tentative claim or, on the other hand, negate its validity.

Additional limitations to this study include its 'time-boundness' – taking any point in time generates issues of replicablity. A point in time cannot be replicated. To properly replicate this study, one would need to find a similar sized dataset from the same period in time, in the same institution, etc., albeit in different disciplinary domain. A contemporary replication would have to take into account that many conditions have changed both locally and globally since these data were collected. The use of the internet is pervasive, far more so than when this cohort wrote their essays. Therefore, a similar dataset from 2013 – 2016, for example, would potentially be open to more influences from internet sources than this cohort. Additionally, all essays are typed and redrafted using word processing software. At the time of data collection, some still wrote their essays by hand. The additional processing that technology affords may have an altering effect on the data as a whole.

An obvious limitation is the size of the dataset. Though it is speculated that doubling the size would not change the results this has yet to be proven. Be this as it may, regardless of whether the dataset were increased in size, the essays in the current corpus would still exist and would still differ from each other. It is however challenging for a researcher to gain access to one cohort longitudinally in an academic context because of the possibilities of students exiting, students failing and having to repeat, and so on, in addition to the possibility of essays being sent to external examiners and not being returned.

9.4 Directions for further research

With regard to further research, based on the current study, the following directions arise:

- Further research with diachronic corpora through the lens of chaos theory is needed. This should be both with written and spoken corpora. This is to establish

whether the chaotic patterns evidenced in MICUPE are limited to the context of situation within MICUPE or whether they are also evident in other situations. Such research should incorporate greater numbers of iteration as compared to the current research.

- Further research with regard to the nature of instantial choices made by language producers in needed. This should be through the production of texts, not based on informant data. This would allow us to see how a text is produced, the order it takes and the instantial decisions made, some of which will supersede previous decisions (editing). Such research should not be limited to academic writing.
- Further research in relation to the distribution of types across texts is needed. This can be done in various contexts, not just in relation to academic writing. From this, the omission of types from a small number of texts, the uniqueness of types within and across texts, the nature of sharing of types across texts and the co-occurrence, or lack of, of these categories of types can be investigated.

9.5 Conclusion

The current research has examined change in student writing within the subject of philosophy over the course of a degree programme. The texts under investigation were produced over three years for the same assessor by 17 students from the same cohort. Based on an analysis of a corpus of these texts, it was posited that the change evident in the writing of these students can be viewed as a chaotic dynamical system.

List of references

Abercrombie, D., 1969. Voice qualities. In N. N. Markel, ed. *Psycholinguistics: an Introduction to the Study of Speech and Personality*. London: The Dorsey Press, pp. 109–127.

Allison, D., 1999. Key concepts in ELT. Genre. *ELT Journal*, 53(2), p.144.

Askehave, I. & Swales, J.M., 2001. Genre Identication and Communicative Purpose : A Problem and a Possible Solution. *Applied Linguistics*, 22(2), pp.195–212.

Barras, R., 1995. *Students must write: a guide to better writing in coursework and examinations*, London: Routledge.

Bartholomae, D., 1986. Inventing the University. *Journal of Basic Writing*, 5(1), pp.4–23.

BAWE, BAWE (British Academic Written English) and BAWE Plus Collections. Available at: http://www2.warwick.ac.uk/fac/soc/al/research/collect/bawe/ [Accessed July 5, 2013].

Bazerman, C., 1981. What written knowledge does: Three examples of academic discourse. *Philosophy of the Social Sciences*, 11(3), pp.361–87.

Beaufort, A., 1997. Operationalizing the concept of discourse community: A case study of one institutional site of composing. *Research in the Teaching of English*, 31, pp.486–529.

Beckner, C. et al., 2009. Language Is a Complex Adaptive System: Position paper. *Langauge Learning*, 59(Suppl. 1), pp.1–26.

Bhatia, V., 1993. *Analysing Genre: language use in professional settings*, London: Longman.

Bhatia, V., 2004. *Worlds of Written Discourse: A Genre-Based View*, London: Continuum.

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

Biber, D. et al., 1999. *Longman grammar of spoken and written English*, London: Longman.

Biber, D., 2006. *University language: A corpus-based study of spoken and written registers*, Amsterdam: John Benjamins.

Biber, D., Conrad, S. & Reppen, R., 1998. *Corpus linguistics: Investigating language structure and use*, Cambridge: Cambridge University Press.

Biggs, J., 1987. *Student Approaches to Learning and Studying. Research Monograph.*, Hawthorn: Australian Council for Educational Research.

Bizzell, P., 1982. Cognition, convention, and certainty: What we need to know about writing. *Pre/text*, 3(3), pp.213–43.

Bizzell, P., 1986. What happens when basic writers come to college? *College Composition and Communication*, 37(3), pp.294–301.

Bizzell, P., 1987. Literacy in culture and cognition. In T. Enos, ed. *A sourcebook for basic writing teachers*. New York: Random House, pp. 125–37.

Bloom, B.S., 1956. *Taxonomy of educational objectives: The classification of education goals by a committee of college and university examiners*, New York: David McKay.

De Bot, K. et al., 2013. Dynamic Systems Theory as a comprehensive theory of second language development. In M. M. A. María del Pilar García Mayo, María Junkal Gutierrez Mangado, ed. *Contemporary Approaches to Second Language Acquisition*. Amsterdam: John Benjamins Publishing Company, pp. 199–220.

Cadman, K., 1997. Thesis writing for international students: A question of identity? *English for Specific Purposes*, 16(1), pp.3–14.

Carter, R. & McCarthy, M., 1988. Word lists and learning words: some foundations. In R. Carter & M. McCarthy, eds. *Vocabulary and Language Teaching*. London: Lon, pp. 86–112.

Carter, R. & McCarthy, M., 1997. *Exploring Spoken English*, Cambridge: Cambridge University Press.

Charles, M., 2003. "This mystery": a corpus-based study of the use of nouns to construct stance in theses from two contrasting disciplines. *Journal of English for Academic Purposes*, 2(4), pp.313–26.

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), pp.310–31.

Clyne, M., 1987. Cultural differences in the organization of academic texts: English and German. *Journal of Pragmatics*, 11(2), pp.211–41.

Collins, J., 1995. Basic Writing and the Process Paradigm. *Journal of Basic Writing*, 14(2), pp.3–18.

Corbett, E. and Connors, R., 1999. *Classic Rhetoric for the Modern Student*, Oxford: Oxford University Press.

Coulthard, M., 2004. Author Identification, Idiolect, and Linguistic Uniqueness. *Applied Linguistics*, 25(4), pp.431–447.

Coxhead, A., 2002. The academic word list: A corpus-based word list for academic purposes. *Language and Computers*, 42(1), pp.73–89.

Department of Education and Skills, 2011. *National strategy for higher education to 2030*, Dublin: Department of Education and Skills.

Eggins, Suzanne and Martin, J.R., 1997. Genres and registers of discourse. In T. Van Dijk, ed. *Discourse as Structure and Process*. London: Sage, pp. 230–256.

Ellis, N., 2008. The dynamics of second language emergence: Cycles of language use, language change, and language acquisition. *The Modern Language Journal*, 92(2), pp.232–49.

Ellis, N. & Larsen-Freeman, D., 2006. Language emergence: Implications for applied linguistics—Introduction to the special issue. *Applied Linguistics*, 27(4), pp.558–89.

Ellis, N., O'Donnell, M.B. & Römer, U., 2013. Usage-Based Language: Investigating the Latent Structures That Underpin Acquisition. *Language Learning*, 63(Supp. 1), pp.25–51.

Emig, J., 1971. *The composing processes of twelfth graders*, Urbana: NCTE.

Entwistle, N. & Ramsden, P., 1982. *Understanding student learning*, Beckenham: Croom Helm.

Eronen, S., Nurmi, J. & Salmela-Aro, K., 1998. Optimistic, defensive-pessimistic, impulsive and self-handicapping strategies in university environments. *Learning and Instruction*, 8(2), pp.159–177.

Evison, J., 2010. What are the basics of analysing a corpus? In M. O'Keeffe, A. and McCarthy, ed. *The Routledge Handbook of Corpus Linguistics*. London: Routledge, pp. 122–35.

Expert Group on Future Skills Needs, 2007. *Tomorrow 's Skills: Towards a National Skills Strategy*, Dublin: Government Publications.

Faigley, L., 1985. Non-academic writing: The social perspective. In D. Odell, L. and Goswami, ed. *Writing in non-academic settings*. New York: Guilford, pp. 231–48.

Ferguson, C.A., 1996. Genre and Register: One path to Discourse Analysis. In T. Heubner, ed. *Sociolinguistic Perspectives: Papers on Language in Society, 1959-1994*. Oxford: Oxford University Press, pp. 167–176.

Firth, J.R., 1935. The Technique of Semantics. *Transactions of the Philological Society*, 34(1), pp.36–73.

Flower, L. & Hayes, J.R., 1981. A cognitive process theory of writing. *College Composition and Communication*, 32(4), pp.365–87.

Freedman, A., 1987. Learning to write again: Discipline-specific writing at university. *Carleton papers in applied language studies*, 4, pp.95–115.

García, O. and Kleifgen, J.A., 2010. *Educating emergent bilinguals*, New York: Teachers College Press.

Ghadessy (Ed.), M., 1993. *Register analysis: Theory and practice*, London: Pinter.

Gleick, J., 1997. *Chaos - The Amazing Science of the Unpredictable:*, London: Vintage.

Grabe, W. & Kaplan, R.B., 1996. *Theory & Practice of Writing: an applied linguistic perspective*, London: Longman.

Gregory, M. & Carroll, S., 1978. *Language and situation: Language varieties and their social contexts*, London: Routledge and Kegan Paul.

Gumperz (Ed.), J.J., 1982. *Language and social identity*, Cambridge: Cambridge University Press.

Halliday, M.A.K., 1978. *Language as social semiotic: The social interpretation of language and meaning*, London: Edward Arnold.

Halliday, M.A.K., 1985. *Introduction to Functional Grammar*, London: Edward Arnold.

Halliday, M.A.K. & Hasan, R., 1989. *Language, context, and text: Aspects of language in a social-semiotic perspective*, Oxford: Oxford University Press.

Halliday, M.A.K., McIntosh, A. & Strevens, P., 1964. *Linguistic Sciences & Language Teaching*, Indiana: Indiana University Press.

Harrington, S. & Adler-Kassner, L., 1998. "The Dilemma that Still Counts": Basic Writing at a Political Crossroads. *Journal of Basic Writing*, 17(2), pp.3–24.

Harris, J., 1989. The idea of community in the study of writing. *College Composition and Communication*, 40(1), pp.11–22.

Hasan, R., 1978. Text in the Systemic-Functional Model. In W. Dressler, ed. *Current Trends in Textlinguistics*. Berlin: Walter de Gruyter, pp. 228–246.

Heikkilä, A. & Lonka, K., 2006. Studying in higher education: students' approaches to learning, self-regulation, and cognitive strategies. *Studies in Higher Education*, 31(1), pp.99–117.

Herrington, A., 1985. Writing in academic settings: A study of the contexts for writing in two college chemical engineering courses. *Research in the Teaching of English*, 19(4), pp.331–61.

Hirvela, A. & Belcher, D., 2001. Coming back to voice: The multiple voices and identities of mature multilingual writers. *Journal of Second Language Writing*, 10, pp.83–106.

Hunston, S., 2002. *Corpora in applied linguistics*, Cambridge: Cambridge University Press.

Hyland, K., 2001. Humble servants of the discipline? Self-mention in research articles. *English for Specific Purposes*, 20(3), pp.207–226.

Hyland, K., 2002. Authority and invisibility authorial identity in academic writing. *Journal of Pragmatics*, 34(8), pp.1091–1112.

Hyland, K., 2003. Genre-based pedagogies: A social response to process. *Journal of Second Language Writing*, 12(1), pp.17–29.

Hyland, K., 2006. Representing readers in writing: Student and expert practices. *Linguistics and Education*, 16(4), pp.363–77.

Hymes, D., 1974. *Foundations in sociolinguistics: An ethnographic approach*, Pennsylvania: University of Pennsylvania Press.

Hyon, S., 1996. Genre in Three Traditions : Implications for ESL. *TESOL Quarterly*, 30(4), pp.693–722.

Ivanič, R., 1998. *Writing and identity: The discoursal construction of identity in academic writing*, Amsterdam: John Benjamins.

Ivanic, R and Camps, D., 2001. I am how I sound: Voice as self-representation in L2 writing. *Journal of Second Language Writing*, 10(1-2), pp.3–33.

Johns, A., 2002. *Genre in the classroom: Multiple perspectives*, New Jersey: Lawrence Erlbaum.

Kay, H. & Dudley-Evans, T., 1998. Genre: What teachers think. *ELT journal*, 52(4), pp.308–14.

Kellert, S.H., 1993. *In the Wake of Chaos: Unpredictable Order in Dynamical Systems*, Chicago: University of Chicago Press.

Kilgarriff, A., Kovář, V. & Krek, S., 2010. A quantitative evaluation of word sketches. In *Proceedings of the 14th EURALEX International Congress, Leeuwarden, The Netherlands*. pp. 372–79.

Killingsworth, M., 1992. Discourse Communities. Local and Global. *Rhetoric Review*, 11(1), pp.110–22.

Kress, G., 1993. Genre as social process. In B. Cope & M. Kalantzis, eds. *The powers of literacy: A genre approach to teaching …*. Bristol: Falmer Press, pp. 22–37.

Larsen-Freeman, D., 1997. Chaos/Complexity Science and Second Language Acquisition. *Applied Linguistics*, 18(2), pp.141–165.

Larsen-Freeman, D., 2013. Complexity theory: a new way to think. *Revista Brasileira de Linguística Aplicada*, 13(2), pp.369–73.

Larsen-Freeman, D. & Cameron, L., 2008. *Complex systems and applied linguistics*, Oxford: Oxford University Press.

Lorenz, E., 1963. Deterministic Nonperiodic Flow. *Journal of the Atmospheric Sciences*, 20(2), pp.130–141.

Lunsford, A., 1979. Cognitive development and the basic writer. *College English*, 41(1), pp.38–46.

MacDonald, S., 1987. Problem definition in academic writing. *College English*, 49(3), pp.315–31.

Maimon, E., 1983. Maps and genres: Exploring connections in the arts and sciences. In W. B. Homer, ed. *Composition and literature: Bridging the gap*. Chicago: University of Chicago Press, pp. 110–25.

Malinowski, B., 1923. The problem of meaning in primitive languages. In C. Ogden & I. A. Richards, eds. *The Meaning of Meaning*. London: Routledge and Kegan Paul, pp. 296–336.

Malinowski, B., 1935. *Coral Gardens and Their Magic, Vol. 2: The Language of Magic and Gardening*, New York: American Book Company.

Martin, J.R., 1985. Process and text: two aspects of human semiosis. In J. D. Benson & W. S. Greaves, eds. *Systemic perspectives on discourse*. Norwood: Ablex, pp. 248–74.

Martin, J.R., 1992. Genre and Literacy-Modeling Context in Educational Linguistics. *Annual Review of Applied Linguistics*, 13(-1), pp.141–172.

Martin, J.R., 1993. Genre and literacy-modelling context in educational linguistics. *Annual Review of Applied Linguistics*, 13(1), pp.141–72.

Martin, J.R. & Rothery, J., 1993. Grammar: Making meaning in writing. In B. Cope & M. Kalantzis, eds. *The powers of literacy: A genre approach to teaching writing*. Bristol: Falmer Press, pp. 137–53.

Marton, F. & Säljö, R., 1976. On qualitative differences in learning: I-outcome and process. *British Journal of Educational Psychology*, 46(1), pp.4–11.

Matsuda, P.K., 2001. Voice in Japanese written discourse: Implications for second language writing. , 10(1), pp.35–53.

McCarthy, M., 1998. *Spoken language and applied linguistics*, Cambridge: Cambridge University Press.

McEnery, T., Xiao, R. & Tono, Y., 2006. *Corpus-based language studies: An advanced resource book*, London: Routledge.

Miller, C., 1984. Genre as social action. *Quarterly Journal of Speech*, 70(2), pp.151–167.

Mitchell, T.F., 1957. The language of buying and selling in Cyrenaica. *Hesperis*, 44, pp.31–71.

Nation, P. & Meara, P., 2002. Vocabulary. In N. Schmitt, ed. *An introduction to applied linguistics*. London: Arnold, pp. 35–54.

O'Keeffe, A., 2000. Undergraduate academic writing: an analysis of errors and weaknesses in syntax, lexis, style and structure. In G. Shiel, U. Ni Dhalaigh, & E. Kennedy, eds. *Language and Literacy for the New Millenium*. Dublin: Reading Association of Ireland, pp. 167–86.

Paltridge, B., 1994. Genre analysis and the identification of textual boundaries. *Applied linguistics*, 15(3), pp.288–99.

Paltridge, B., 1996. Genre, text type, and the language learning classroom. *ELT journal*, 50(3), pp.237–43.

Pennycook, A., 1994. The politics of pronouns. *ELT Journal*, 48(2), pp.173–178.

Perl, S., 1979. The composing processes of unskilled college writers. *Research in the Teaching of English*, 13(4), pp.317–36.

Perl, S., 1980. Understanding Composing. *College Composition and Communication*, 31(4), pp.363–369.

Pogner, K., 2003. Writing and interacting in the discourse community of engineering. *Journal of Pragmatics*, 35(6), pp.855–67.

Reither, J., 1985. Writing and knowing: Toward redefining the writing process. *College English*, 47(6), pp.620–28.

Scott, M., 2010. What can corpus software do? In A. O'Keeffe & M. McCarthy, eds. *The Routledge Handbook of Corpus Linguistics*. London: Routledge, pp. 136–51.

Seedhouse, P., 2010. Locusts, snowflakes and recasts: complexity theory and spoken interaction. *Classroom Discourse*, 1(1), pp.4–24.

Shaughnessy, M., 1977. Some needed research on writing. *College Composition and Communication*, 28(4), pp.98–103.

Silver, M., 2003. The stance of stance: a critical look at ways stance is expressed and modeled in academic discourse. *Journal of English for Academic Purposes*, 2(4), pp.359–74.

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

Sommers, N., 1980. Revision strategies of student writers and experienced adult writers. *College Composition and Communication*, 31(4), pp.378–88.

Stapleton, P., 2002. Critiquing voice as a viable pedagogical tool in L2 writing: Returning the spotlight to ideas. *Journal of Second Language Writing*, 11(3), pp.177–90.

Swales, J., 1988. Discourse communities, genres and English as an international language. *World Englishes*, 7(2), pp.211–20.

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

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, pp.S23–S39.

Taylor, G., 1988. *Literacy by Degrees*, Milton Keynes: Society for Research into Higher Education & Open University Press.

Tognini-Bonelli, E., 2010. Theoretical overview of the evolution of corpus linguistics. In A. O'Keeffe & M. McCarthy, eds. *Routledge Handbook of Corpus Linguistics*. London: Routledge, pp. 14–28.

Tucker, P., 2003. Evaluation in the art-historical research article. *Journal of English for Academic Purposes*, 2(4), pp.291–312.

Ventola, E., 1984. Orientation to Social Semiotics in Foreign Language Teaching. *Applied Linguistics*, 5(3), pp.275–86.

Verspoor, M., 2013. Dynamic Systems Theory as a comprehensive theory of second language development. *Contemporary Approaches to Second Language …*.

Watkins, D., 2001. Correlates of approaches to learning: A cross-cultural meta-analysis. In R. J. Sternberg & L. Zhang, eds. *Perspectives on thinking, learning, and cognitive styles*. New Jersey: Lawrence Erlbaum Associates Publishers, pp. 165–95.

Wegener, P., 1885. *Untersuchungen über die Grundfragen des Sprachlebens*, Amsterdam: John Benjamins.

Woodward-Kron, R., 2002. Critical analysis versus description? Examining the relationship in successful student writing. *Journal of English for academic purposes*, 1(2), pp.121–43.

Woodward-Kron, R., 2004. "Discourse communities" and "writing apprenticeship": an investigation of these concepts in undergraduate Education students' writing. *Journal of English for Academic Purposes*, 3(2), pp.139–61.

Zipf, G. K. (1935). *The psycho-biology of language: an introduction to dynamic philology.* Michegan: MIT Press.