Genre awareness in writing for testing purposes

Maria N. Melissourgou

A Thesis Submitted in Fulfilment of the Requirements
for the Degree of Doctor of Philosophy in
English Applied Linguistics

The University of the Aegean
School of Humanities
Department of Mediterranean Studies
2016

Η επίγνωση του κειμενικού γένους στο γραπτό λόγο ως μέρος εξέτασης

Μαρία Ν. Μελισσουργού

Διατριβή υποβληθείσα για την απόκτηση Διδακτορικού τίτλου στην Αγγλική Εφαρμοσμένη Γλωσσολογία

Πανεπιστήμιο Αιγαίου
Σχολή Ανθρωπιστικών Επιστημών
Τμήμα Μεσογειακών Σπουδών
2016

Advisory Committee

Associate Professor Katerina Frantzi (supervisor)
(Department of Mediterranean Studies, University of the Aegean)

Associate Professor Vasilia Kazoulis
(Department of Primary Education, University of the Aegean)

Assistant Professor Marianthi Georgalidou
(Department of Mediterranean Studies, University of the Aegean)

Examination Committee

Associate Professor Katerina Frantzi
(Department of Mediterranean Studies, University of the Aegean)

Associate Professor Vasilia Kazoulis
(Department of Primary Education, University of the Aegean)

Assistant Professor Marianthi Georgalidou
(Department of Mediterranean Studies, University of the Aegean)

Professor Georgios Mikros
(Department of Italian Language and Literature, University of Athens)

Associate Professor Efstathios Stamatatos
(Department of Information and Communication Systems Engineering,
University of the Aegean)

Assistant Professor Evangelia Vlachou
(Department of Mediterranean Studies, University of the Aegean)

Assistant Professor Ioannis Galantomos
(Department of Mediterranean Studies, University of the Aegean)

Date of Defence: 01/12/2016

Συμβουλευτική Επιτροπή

Αναπληρώτρια Καθηγήτρια Κατερίνα Φραντζή (επόπτρια)
(Τμήμα Μεσογειακών Σπουδών, Πανεπιστήμιο Αιγαίου)

Αναπληρώτρια Καθηγήτρια Βασιλεία Καζούλλη
(Παιδαγωγικό Τμήμα Δημοτικής Εκπαίδευσης, Πανεπιστήμιο Αιγαίου)

Επίκουρη Καθηγήτρια Μαριάνθη Γεωργαλλίδου
(Τμήμα Μεσογειακών Σπουδών, Πανεπιστήμιο Αιγαίου)

Εξεταστική Επιτροπή

Αναπληρώτρια Καθηγήτρια Κατερίνα Φραντζή
(Τμήμα Μεσογειακών Σπουδών, Πανεπιστήμιο Αιγαίου)

Αναπληρώτρια Καθηγήτρια Βασιλεία Καζούλλη
(Παιδαγωγικό Τμήμα Δημοτικής Εκπαίδευσης, Πανεπιστήμιο Αιγαίου)

Επίκουρη Καθηγήτρια Μαριάνθη Γεωργαλλίδου
(Τμήμα Μεσογειακών Σπουδών, Πανεπιστήμιο Αιγαίου)

Καθηγητής Γιώργος Μικρός
(Τμήμα Ιταλικής Γλώσσας και Φιλολογίας,
Εθνικό και Καποδιστριακό Πανεπιστήμιο Αθηνών)

Αναπληρωτής Καθηγητής Ευστάθιος Σταματάτος
(Τμήμα Μηχανικών Πληροφοριακών κι Επικοινωνιακών Συστημάτων,
Πανεπιστήμιο Αιγαίου)

Επίκουρη Καθηγήτρια Ευαγγελία Βλάχου
(Τμήμα Μεσογειακών Σπουδών, Πανεπιστήμιο Αιγαίου)

Επίκουρος καθηγητής Ιωάννης Γαλαντόμος
(Τμήμα Μεσογειακών Σπουδών, Πανεπιστήμιο Αιγαίου)

Ημερομηνία Υπεράσπισης και Κρίσης: 01/12/2016

4

To my family

ACKNOWLEDGEMENTS

ABSTRACT

This thesis aims to raise genre awareness in English language testing contexts. It first attempts to find out more about the learners' own views on writing and on their preparation for various exams. It then identifies the genres involved in text type categories and assesses both text type and genre representation in education material. From then on, it delves into genre analysis in order to present a thorough description of the most frequent genres in teaching material. It also explores similarity among these genres measuring the strength of relation between pairs of genres.

A specially designed questionnaire is used to gather information concerning learners' views and needs. To answer the rest of the research questions the study exploits the recent synergy between Corpus Linguistics and Genre Analysis. The creation of a pedagogical corpus consisting of writing model answers (WriMA corpus) from educational material becomes the basis for the assessment of representation of both text types and genres in teaching material, the extraction of the main features of individual genres as well as the measurement of the strength of the relation between genres. Genre analysis follows a Discourse Analysis-type approach for the exploration of individual genres as well as a quantitative-contrastive approach to explore genres from multiple perspectives.

Results form the questionnaire analysis show an increased preference for certification in level B2 followed by C2, the need for more emphasis on language demands than actual writing skills in B-level courses and the desire for more model answers in the classroom. Writing compared to the rest of skills tested is rated as the most difficult one by the majority of the learners of which young and male learners are the biggest part. The 'Review' and the 'Report' are seen as the most difficult text types and the less privileged ones in terms of preparation time.

Concerning educational material, findings reveal an uneven representation of text types and the overrepresentation of Essays. A great variety of genres is derived from these text types and several genres are found to be underrepresented in the material explored. The analysis has pointed to the distinctiveness of the 'Descriptive Essay' within the Essay category, the

similarity of the Personal Observation Report with genres that are not part of the Report text type categorisation and the weak relation between the 'Complaint' and the 'Advice Letter'. It has thus offered proof that text type categorization is too broad and can in some cases misguide teachers and learners as the linguistic variation between genres with the same text type label can in some cases be greater than the variation between genres of different text type categories. Genre categorisation is shown to be more helpful allowing for even the finest variation between texts to be revealed and explicitly taught assisting learners in the identification of generic requirements, a skill that is taken for granted in exams as the official writing guides reveal.

This study is one of very few that put learners at the center of interest and is the first to see various exams as a whole addressing this way practical needs caused by mixed groups of learners targeting various exams. The creation of a specialised corpus, not previously available, has offered insight in areas which were largely under-researched and remained vague. Findings provide precise information for the explicit teaching of genres, for a conscious and less subjective rating of answers and for clear and helpful feedback. Material designers can also benefit in order to address the weaknesses spotted in this research by using the detailed description of genres to provide specific guidance and by including more model answers of under-represented genres. They can also use the knowledge concerning the relations among genres to determine the sequence of the material. Guidance can this way be clear and based on evidence rather than intuition.

ΠΕΡΙΛΗΨΗ

Ο σκοπός αυτής της διατριβής είναι η ενημέρωση και ευαισθητοποίηση σχετικά με τη χρήση των κειμενικών γενών (genres) στο περιβάλλον των εξετάσεων πιστοποίησης της Αγγλικής γλώσσας. Αρχικά αναζητά τις απόψεις των ίδιων των μαθητών σχετικά με το γραπτό λόγο και την προετοιμασία τους για διάφορες εξετάσεις. Έπειτα ταυτοποιεί τα κειμενικά γένη που εμπεριέχονται στις κατηγορίες κειμενικών ειδών (text types) και αξιολογεί την εκπροσώπηση τόσο των κειμενικών ειδών όσο και των κειμενικών γενών στο εκπαιδευτικό υλικό. Το υπόλοιπο κομμάτι της έρευνας εστιάζει στην ανάλυση των κειμενικών γενών προκειμένου να παρουσιάσει μια ολοκληρωμένη περιγραφή των πιο συχνών από αυτά όπως εμφανίζονται στο υλικό διδασκαλίας. Επίσης εξερευνά την ομοιότητα μεταξύ αυτών των κειμενικών γενών υπολογίζοντας τη δύναμη της σχέσης ανάμεσα σε κάθε ζεύγος.

Για τη συλλογή πληροφορίας σχετικά με τις απόψεις των μαθητών χρησιμοποιείται ένα ειδικά σχεδιασμένο ερωτηματολόγιο. Για τα υπόλοιπα ερευνητικά ερωτήματα η μελέτη εκμεταλλεύεται την πρόσφατη συνέργεια μεταξύ της Γλωσσολογίας Σωμάτων Κειμένων (Corpus Linguistics) και της Ανάλυσης Κειμενικών Γενών (Genre Analysis). Η δημιουργία ενός παιδαγωγικού Σώματος Κειμένων που περιλαμβάνει μοντέλα απαντήσεων για το γραπτό λόγο από εκπαιδευτικό υλικό γίνεται η βάση για την αξιολόγηση της εκπροσώπησης τόσο των κειμενικών ειδών όσο και των γενών στο υλικό διδασκαλίας, για την άντληση των βασικών χαρακτηριστικών μεμονωμένων κειμενικών γενών όπως επίσης για τον υπολογισμό της δύναμης της σχέσης μεταξύ τους. Η ανάλυση κειμενικών γενών ακολουθεί μια προσέγγιση Ανάλυσης Λόγου (Discourse Analysis) για την εξερεύνηση μεμονωμένων κειμενικών γενών καθώς επίσης και μια ποσοτική-συγκριτική προσέγγιση για την εξερεύνησή τους από πολλές διαφορετικές σκοπιές.

Η ανάλυση των ερωτηματολογίων δείχνει αυξημένη προτίμηση για πιστοποίηση στο επίπεδο Β2, ακολουθούμενο από το Γ2, την ανάγκη για περισσότερη έμφαση στις γλωσσικές απαιτήσεις αντί για τις δεξιότητες στο γραπτό λόγο στα προγράμματα Β επιπέδου και την επιθυμία για περισσότερη πρόσβαση σε μοντέλα απαντήσεων στην τάξη. Ο γραπτός λόγος, συγκρινόμενος με τις υπόλοιπες δεξιότητες κατά την εξέταση, βαθμολογείται ως το πιο δύσκολο μέρος από την πλειονότητα των μαθητών, το μεγαλύτερο τμήμα των οποίων είναι νεότεροι και άρρενες μαθητές. Η Κριτική βιβλίου/ταινίας (Review) και η Αναφορά (Report)

θεωρούνται τα πιο δύσκολα κειμενικά είδη και τα πιο αδικημένα ως προς το χρόνο που αφιερώνεται στην προετοιμασία τους.

Σε σχέση με το εκπαιδευτικό υλικό τα ευρήματα αποκαλύπτουν μια άνιση εκπροσώπηση των διαφορετικών κειμενικών ειδών και την υπερ-εκπροσώπηση των Εκθέσεων-Δοκιμίων (Essays). Από αυτά τα κειμενικά είδη πηγάζει μεγάλη ποικιλία σε κειμενικά γένη αρκετά από τα οποία υπο-εκπροσωπούνται στο υλικό που ερευνήθηκε. Η ανάλυση καταδεικνύει τον ξεχωριστό χαρακτήρα του Περιγραφικού Δοκιμίου (Descriptive Essay) μέσα στην κατηγορία των Δοκιμίων, την ομοιότητα της Αναφοράς μετά από Προσωπική Επισκόπηση (Personal Observation Report) με γένη που δεν εντάσσονται στο κειμενικό είδος των Αναφορών και την αδύναμη σχέση ανάμεσα στην Επιστολή Παραπόνων (Complaint Letter) και τη Συμβουλευτική Επιστολή (Advice Letter). Έχει κατά συνέπεια προσφέρει αποδείξεις ότι η κατηγοριοποίηση σύμφωνα με το κειμενικό είδος είναι υπερβολικά γενικευμένη και μπορεί σε κάποιες περιπτώσεις να παραπλανήσει εκπαιδευτικούς και μαθητές αφού η διαφορετικότητα ανάμεσα σε κειμενικά γένη της ίδιας κατηγορίας κειμενικού είδους μπορεί σε κάποιες περιπτώσεις να είναι μεγαλύτερη από τη διαφορετικότητά τους με κειμενικά γένη που ανήκουν σε άλλες κατηγορίες κειμενικού είδους. Η κατηγοριοποίηση με βάση το κειμενικό γένος αποδεικνύεται πιο βοηθητική αφού επιτρέπει ακόμα και την πιο ισχνή διαφοροποίηση ανάμεσα στα κείμενα να φανεί και να διδαχθεί με σαφήνεια υποστηρίζοντας τους μαθητές στην αναζήτηση των απαιτήσεων κάθε κειμενικού γένους, μια ικανότητα που θεωρείται δεδομένη στις εξετάσεις όπως αποκαλύπτουν τα επίσημα εγχειρίδια για το γραπτό λόγο.

Η μελέτη αυτή είναι μια από τις λίγες που εστιάζουν το ενδιαφέρον τους στους μαθητές και η πρώτη που βλέπει διάφορες εξετάσεις συνολικά απαντώντας με αυτό τον τρόπο σε πρακτικά ερωτήματα όπως οι ανάγκες μικτών ομάδων μαθητών που προετοιμάζονται για διαφορετικές εξετάσεις. Η δημιουργία ενός ειδικευμένου Σώματος Κειμένων, μη διαθέσιμου στο παρελθόν, προσφέρει γνώση σε τομείς που δεν είχαν ερευνηθεί επαρκώς και παρέμεναν ασαφείς. Τα ευρήματα προσφέρουν πληροφόρηση για τη διδασκαλία των κειμενικών γενών με σαφήνεια, για μια συνειδητή και λιγότερο υποκειμενική βαθμολόγηση των απαντήσεων αλλά και για σαφή και υποστηρικτική ανατροφοδότηση στους μαθητές. Οι δημιουργοί εκπαιδευτικού υλικού μπορούν επίσης να επωφεληθούν ώστε να αντιμετωπίσουν τις αδυναμίες που επισημάνθηκαν σε αυτή την έρευνα χρησιμοποιώντας τη λεπτομερή περιγραφή των κειμενικών γενών για να προσφέρουν συγκεκριμένη καθοδήγηση και

προσφέροντας περισσότερα μοντέλα-απαντήσεις στα λιγότερο εκπροσωπούμενα κειμενικά γένη. Μπορούν επίσης να βασιστούν στη γνώση αναφορικά με τις σχέσεις ανάμεσα στα γένη για να καθορίσουν τη σειρά με την οποία αυτά παρουσιάζονται στο υλικό διδασκαλίας. Με αυτό τον τρόπο η καθοδήγηση μπορεί να είναι σαφής και βασισμένη σε αποδείξεις αντί στη διαίσθηση.

LIST OF ABBREVIATIONS

ACE   - the Australian Corpus of English

ALTE   - Association of Language Testers in Europe

BNC   - British National Corpus

BNC Web   - British National Corpus Web interface

CaMLA   - Cambridge Michigan Language Assessments

CEFR   - Common European Framework Reference for languages

CAE   - Cambridge English: Advanced

CARS   - Create A Research Space

CLC   - Cambridge Learner Corpus

CQPweb   - Corpus Query Processor Web interface

CRATER   - Corpus Resources And Terminology Extraction

CPE   - Cambridge English: Proficiency

EALTA   - Eurpopean Association for Language Testing and Assessment

EFL   - English as a Foreign Language

ECCE   - Examination for the Certificate of Competency in English

ECPE   - Examination for the Certificate of Proficiency in English

ELT   - English Language Teaching

ESL   - English as a Second Language

ESP   - English for Specific Purposes

ETS   - Educational Testing Service

FCE   - Cambridge English: First

GBWI   - Genre-Based-Writing-Instruction

ICE   - International Corpus of English

ICLE   - International Corpus of Learner English

ICT   - Information and Communications Technology

ID   - Introductory Data

IELTS   - International English Language Testing System

ILTA   - International Language Testing Association

KET   - Cambridge English: Key

ΚΠγ   - Κρατικό Πιστοποιητικό γλωσσομάθειας

L1   - first on the Left (collocates)

L2   - second Language

LB   - Lexical Bundles

LD   - Lexical Density

LLC   - Longman Learner Corpus

LOB   - Lancaster Oslo Bergen

MH   - Main Heading

MICASE   - Michigan Corpus of Academic Spoken English

OCR   - Optical Character Reader

PET   - Cambridge English: Preliminary

POS   - Part Of Speech

R1   - first on the Right (collocates)

RC   - Reference Corpus

RGS   - Rhetorical Genre Studies

RSLB   - Register Specific Lexical Bundles

SFL   - Systemic Functional Linguistics

SH   - Section Heading

STTR   - Standardised Type Token Ratio

TOEFL iBT - Test Of English as a Foreign Language - internet Based Test

TTR   - Type Token Ratio

WriMA   - Writing Model Answers

TABLE OF CONTENTS

LIST OF TABLES AND FIGURES

CHAPTER SIX

CHAPTER SEVEN

CHAPTER ONE

INTRODUCTION

1.1. The context

This chapter aims to offer a picture of the educational system and practice in Greece related to the teaching of English as a foreign language. Personal experience of the writer as a teacher in both state schools and private language centres makes it possible to present issues both at the official as well as the unofficial level regarding day-to-day practices and perceptions in both settings.

The following sections describe the educational settings separately as teaching in the state schools, private schools and language centres is based on different laws and objectives. Therefore, reference to their connection to international English language testing will be made. Even though private schools are supposed to follow the state's school curriculum and analytical programmes they are actively involved in language certification and they approach foreign language teaching in similar ways to language centres. For this reason, they are presented together with language centres. Higher education is not included in this presentation as it is believed that the majority of Greek students who opt for this kind of certification do so before they enter higher education (an assumption that will be examined in chapter 3).

1.1.1. Educational institutions in Greece, teaching writing practices and relation to English language testing

1.1.1.1. State secondary schools

English is taught as a foreign language in schools throughout Greece starting from an early age in primary school (usually at the 3rd grade although some piloting schools start from the first grade) with three teaching hours per week and two teaching hours for junior and senior high school. It is a compulsory subject until the very last year of secondary education but there is a choice of foreign languages in senior high school as French and German language courses are also offered. English is tested at the end of year examinations in secondary schools and these exams last for two hours as it happens with every other subject tested.

English is also tested in the secondary school context as a special subject at the end of the final year during the Pan-Hellenic university entrance exams but only for those wishing to enter faculties somehow related to languages (English Philology, European/International /Touristic/Maritime studies). It is necessary here to explain that in the Greek educational system, the secondary school has the responsibility of the students' selection and entrance to the tertiary education through national exams conducted at the same time all over the country. However, these English exams concern only a minority of students and this section has chosen to focus on the end-of-year exams during senior high school which are compulsory and involve the entire student population.

There is a provision in junior high schools for allocating students into two different levels for every grade, namely 'beginner' and 'advanced' learners only once as they enter high school. Teachers are free to decide what test to use and how to assess students. However, this does not allow for great flexibility since the groups have to be equal in the number of students attending, which in practice results in allocating the weakest and often indifferent students in the first group and the obedient, more aspiring students on the other. Personal experience prompts me to argue that this policy favours the good students but does not help the weak ones as it disregards the impact of peer-motivation in the beginners' groups. Besides, the beginners' classes have the same number of students as every other class despite the greater percentage of uninvolved or often undisciplined students, an imbalance that often reduces the actual teaching time. As this separation of beginner and advanced learners is only done once, students in the third grade continue in the same groups based on the assumption that all students move linearly and no great change is made due to personal effort and motivation.

Textbooks, supplied by the ministry of Education, are based on syllabi which assume a progression of proficiency levels (Παιδαγωγικό Ινστιτούτο, 2004-2011). There are different coursebooks and workbooks for each group in grades A and B but students of both groups use the same books in grade C. No audiovisual material is supplied even though coursebooks include listening activities. In the end-of-year exams teachers are obliged to include a 'dictation' part where they need to 'read aloud' a text offering students time to write. Pupils are allocated in various classrooms according to their year/grade, usually in alphabetical order, disregarding the 'beginner – advanced' classification which can result in classrooms with only few pupils of both groups. Teachers hurry in and out of classrooms trying to 'read

aloud' in such limited time distracting the rest of the students. Organisation problems combined with the shortage in audio material create a situation which is unfair to learners who try to cope with anxiety and need to concentrate.

In any case, the 'beginner – advanced' classification is not continued to senior high school where students are together again according to age and alphabetical allocation. The students' CEFR level (Common European Framework Reference for Languages, Council of Europe, 2001) is not a factor that influences allocation in groups and senior high school teachers often find themselves teaching students whose level may range from A2 to C2, in only two weekly sessions. Textbooks are still not supplied by the state as is the case with all other subjects, there is however a list of 'approved' books with reduced prices for school orders. This is an issue that can cause friction between teachers, students and sometimes even parents; teachers are left to their own devices as to the appropriate ways to handle students who do not bring textbooks in class or do not do their homework and who often use the lack of resources to buy a commercial book as an excuse.

Writing as a component in the end-of-year exams is compulsory only for senior high school students (15-18 years of age), and counts for 30 per cent in the final scoring. There is no special training for the assessment of writing or any specific criteria provided by the Ministry of Education apart from the end of year exam instructions mainly concerning the word limit and the gravity in scoring. Test format and assessment scales are strictly regulated by Ministry Decrees. The summative English language grade appearing on the school certificates indicates a pass score for the subject without any reference to CEFR levels or particular skills (e.g. reading, writing, listening or speaking). Papaefthymiou-Lytra (2012) comments on this gap between practices and actual needs:

> Therefore, language testing and assessment in state school has become a bureaucratic exercise of grade allocation in accordance with prescribed regulations rather than real assessment of learners' skills, abilities and knowledge. This practice has never meant much to stakeholders such as students and parents alike. (p.23)

Students rarely write in class. Dendrinos et al. (2013: 44) reporting on the 'European Survey on Language Competences' say that teachers state that they place little if any emphasis on writing in their language classes. Personal experience implies that most students are unmotivated towards writing in classrooms in the state schools resulting from the belief that they are going to be assessed favourably in any case and the view that 'writing' is difficult (the learners' view on writing as a skill will be examined in chapter 3).

Information and communications technology (ICT) although widely used in writing classes worldwide is rarely used in Greek schools apart from the classes where ICT is taught as a subject. A computer lab with basic equipment does exist in every school; however, it is not enough for 13-20 groups of students taking classes at the same time each hour. The lab is available only when the ICT teachers are not using it and even then, some planning is needed weeks before its use so that the lab is available. There is reluctance on both parts in using this space since teachers may be held responsible for any damage caused by students.

Spiris (2014: 357) verifies that according to teachers restricted access to the computer lab as well as the absence of computers in the classroom are major barriers in the incorporation of technology in the teaching process. Although teachers complain about the lack of training, experience has showed that most English as a Foreign Language (EFL henceforth) teachers at schools use a computer at home in order to search, learn and prepare student assignments and tests. Because of their fluency in English they have been the first ones, among their fellow teachers, to use the internet extensively for several reasons. Despite their favourable attitude towards technology EFL teachers do not feel that the Greek educational system supports the integration of technology in the classroom (Spiris, 2014:356, 359, 361).

Some classes have smart boards in them but they are rarely used as well. Most teachers do not know how to use them and even if they do the same reluctance exists since this is expensive equipment. Given that there is neither any incentive in Greek public schools for innovation in teaching methods nor any connection between learner satisfaction and teacher assessment, most EFL teachers do not use more than a cd-player during English classes. Even this has to be carried by the the teacher every hour in a different classroom since there is no foreign language lab in most schools.

Dendrinos et al. (2013:11, 16, 42) comment on the teaching hours offered in the Greek state school comparing them with the hours needed to reach B2 level, referring to this level as "the minimum goal for the children of the average Greek family" and conclude that these teaching hours are not sufficient. They also refer to "well qualified and experienced" English language teachers, "less than ideal teaching conditions" and the general view of foreign language courses as 'inferior' to other school courses.

Apart from assessment at the school's end-of-year examinations the only connection to external language testing is that of the 'ΚΠγ' (Κρατικό Πιστοποιητικό γλωσσομάθειας, KPG in English - State Certificate of Language Proficiency). 'External' in the sense that examinations are not organized by the schools themselves; they are however, organized by the state, in particular by the Greek Ministry of Education in cooperation with the foreign language departments of the National Kapodistrian University of Athens and the Aristotelion University of Thessaloniki (responsible for the English language).

This examination body certifies knowledge of several foreign languages according to CEFR levels. The exams take place all over the country at the same day and time for each language and level using the technological system also used by the Ministry for the Panhellenic university entrance exams, offering transparence and equal conditions for the candidates. This offers Greek students access to foreign language certification which is not possible at state school. There are two exam periods each year, one in November and the other in May. Candidates apply individually and the exam fees are less expensive than most external international exams. Although the state schools encourage participation to these examinations they do not formally prepare students or take any responsibility for their progress towards this goal. Greek is viewed as the common language of the candidates who are required to understand instructions and test rubrics in Greek or function in the role of mediator, transferring information from their native language into the target language. For example, the 'writing' Module 2 for levels B and C, is called 'Writing and Mediation'. This policy excludes any candidates who may reside in Greece but may not be fluent in Greek yet.

1.1.1.2 Foreign language centres and private schools

It is common for primary and high school students in Greece to attend extra private lessons outside school preparing themselves for language certification (Tsagari & Papageorgiou,

2012; Dendrinos et al., 2013; Sifakis & Fay, 2011). This is mainly due to the common spread attitude that work done at state schools may be insufficient (Scholfield & Gitsaki, 1996:118), but also due to the great need for foreign language certification and the access to employment this certification allows. The students' actual foreign language knowledge therefore is a result of joined forces, state and private, and is tested by various external examination bodies.

Private foreign language centres have managed to make their services 'necessary' and 'unavoidable', a view deeply rooted in the Greek mentality despite the financial consequences this view has. Dendrinos et al. (2013) describe the situation:

> Private language institutes are a burgeoning business in Greece, being almost exclusively oriented towards preparing students for language certification exams. A recent survey reports that there are 6,564 foreign language schools in Greece with 510,575 students, the vast majority of which (448,822) are preparing for English language certificates. (p. 16)

This could be attributed to the following characteristics associated with private language centres:

a. small class sizes,

b. groups according to the students' proficiency level, instead of age

c. three to five hourly teaching sessions per week, as opposed to the two 40-50- minute teaching sessions per week in state schools

c. easy access to multimedia equipment and a wide range of modern educational material

d. the freedom to teach towards specific external language certification exams and do the administrative work needed for participation in them.

In their study of vocabulary teaching practices comparing state schools and private institutes of foreign languages, Scholfield & Gitsaki (1996: 126), find the later to be more successful; This is not however "to be founded on overwhelmingly better teaching or learner training. It is perhaps the stricter environment with more class tests and greater discipline". They also refer to "the smaller number of students per classroom, and the greater number of teaching sessions" as a possible cause.

Even though the state does not offer preparation for these exams it fully recognizes the certificates as qualifications for access to jobs and promotion in the public sector. And even though the state school teachers need to be university degree holders, the great majority of language centre-teachers are C2-level certificate holders certified by the Greek Ministry of Education to teach (Sifakis, 2009:233). As Papaefthymiou-Lytra (2012: 23-24) notices "In this way, the State has equated mere language certificate holders with University degree holders, graduates of the Faculties of Foreign Languages and Literatures who are especially trained at a pre-service level to teach foreign languages!" Dendrinos et al. (2013: 78) report that foreign language teachers in Greek state schools not only hold a university degree but according to their study one third of those teachers have also completed postgraduate studies.

According to a study by the National and Kapodistrian University of Athens (Εθνικό και Καποδιστριακό Πανεπιστήμιο Αθηνών, 2007-2013), the yearly house expenditure related to foreign language education (materials and fees) rises up to 15,4% of the total yearly expenditure and is the third cost in terms of numbers affecting the family budget. "This financially expensive practice reflects parents' perennial lack of trust towards the quality of foreign language provision in Greek public schools and is a result of their deep-seated belief that foreign language instruction equals foreign language certification" (Dendrinos et al., 2013: 16, 17). As learners and parents are rarely aware of the state's share of responsibility in this situation, school teachers are often seen as the easy scapegoat for the extra financial cost each family has to go through. As Sifakis & Fay (2011: 293) explain, the close link between foreign language certification and private institutions has resulted in indifference towards state school TESOL classes and subsequently in teachers' low status.

In this strong testing-oriented context (Sifakis, 2009: 236), both language centres and private schools consider success in these exams as proof of quality for their services and invest a lot of time, work and funding toward this direction. *"In the private language schools' domain, professionalism and success depend largely on the number of candidates who pass a particular exam"* (Sifakis, 2009: 235). This may involve extra teaching time close to examination dates, participation in teacher seminars, book exhibitions, meeting with publishers and representatives of examination bodies. Language centres and private schools are free to choose from a wide range of international examinations organized in the country and advice stakeholders (e.g. students and parents) accordingly.

The KPG is still not embraced fully by these institutions who still prefer to promote international exams, possibly seeing them as their distinguishing asset from state schools. Even when these institutes decide to prepare learners for the KPG exams it is difficult to combine preparation for other language exams due to the written and oral "Mediation" factor, a skill not required by other examination bodies. It is definitely time and money consuming for the same group of learners to address such different examination requirements. Especially when mediation practice material is not available in the market the teacher will need time and personal effort to design such activities. (Δενδρινού και Καραβά, 2013: 149-150).

1.1.2. International English language exams: Writing task types, time, word-limits and CEFR levels.

This section describes basic features of well-known international examinations addressing CEFR levels B and C which attract large number of candidates in Greece and are recognized by the State. It is necessary information for the discussion of the research questions and results presented in the following chapters.

The selection of these exams is not based on data regarding participation as this information is not publically available but is instead based on years of professional experience in the field and discussions with students, colleagues and publishers. The rest of the data have been taken from documents supplied by the examination bodies, available in their official websites. (Cambridge English First, 2015, Cambridge English Advanced, 2015; Cambridge English Proficiency, 2015; ECCE: Sample Test Guide: 2012, ECCE 2015; ECPE, 2015; IELTS, 2016a; ETS TOEFL, 2011; ETS TOEFL iBT, 2015). Basic information concerning these exams is presented on table 1.1.

The CEFR levels related to all parts of this study range between B1 and C2, the most intensely examined levels for certification. Time limit ranges from thirty minutes for exams that only include one writing task (ECCE, ECPE- abbreviations explained in the table) to one hour and thirty minutes for two writing tasks (CAE and CPE). PET is the only exam here which tests 'Writing' together with 'Reading'.

In terms of time limitations and task requirements the most demanding exam seems to be the

TOEFL which allows only fifty minutes for two tasks followed by the IELTS exam with sixty minutes for two tasks. Some exams require a specific number of words for their writing tasks (FCE, CAE, CPE and IELTS 'general training' and 'academic' which specify the minimum required length) or advise candidates on approximate lengths of written items (PET, TOEFL iBT) whereas other exams do not pose a word limit (ECCE, ECPE). In cases, however, where there is a word limit "candidates are not penalized for over-length scripts per se and so exceeding the recommended word range is acceptable" (Spillet, 2012: 4). This advice however is provided to raters only and both teachers as well as candidates remain confused about the rigidness of raters on such issues. Confusion is justified as the exam bodies themselves revise and try to correct requirements or marking criteria that prove to be unfair. Lim (2012), for example, commenting on the process of revising the mark scheme of the Cambridge ESOL exams [PET, FCE, CAE, CPE- and Business Certificates (not included in table 1)] says:

> The revision also provided an opportunity to review how other issues (e.g. under and over length responses; varieties of English) should be dealt with. Where length of response is concerned, the writing tasks specify expected range of the output, and candidate responses that had not kept to the guidelines were generally met with automatic penalties. This however did not seem to be in line with a communicative construct, which would emphasise effectiveness of communication. In addition, it led to candidates and examiners spending significant time counting words, which did not seem to be a good use of their time. (p.8)

He notifies us of the change in the mark scheme which one could say still involves a great deal of subjectivity as any other issue connected to human rating: "The mark scheme has thus removed automatic length- related penalties, allowing for the effects of these to be dealt with under the four sub-scales."

Table 1 also offers information on writing tasks. A task "is defined precisely as a combination of rubric, input and response". The term 'rubric' refers to the "instructions given to a candidate on how to respond to a particular input". In some cases, there is 'input', that is, "material provided by the candidate for use in order to produce an appropriate response" (ALTE, 2009: 1).

Table 1.1 International English language exams related to this study: CEFR levels, task types, time and word limits

Examination	Developed by	CEFR level	Writing task types	time	Word limit
Cambridge English: Preliminary (PET)	Cambridge English Language Assessment	B1	1. Sentence transformation 2. Short communicative message 3. Continuous writing: one task to be selected from a choice of two: <u>an informal letter or a story</u> * Parts 1 and 2 carry 5 marks each - Part 3 carries 15 marks	1h 30 mins (R + W)	For part 3: Advised to write about 100 words
Cambridge English: First (FCE)	Cambridge English Language Assessment	B2	1. <u>Discursive Essay</u> (response to input text) 2. one task to be selected from a choice of three (text types: <u>article, informal letter or email, review, report</u>)	1h 20 mins	140-190 (for each task)
Cambridge English: Advanced (CAE)	Cambridge English Language Assessment	C1	1. <u>Essay</u> with a discursive focus (response to input text) 2. one task to be selected from a choice of three (text types: <u>letter/email, review, report, proposal</u>)	1h 30 mins	220-260 (for each task)
Cambridge English Proficiency (CPE)	Cambridge English Language Assessment	C2	1. <u>Essay</u> with a discursive focus (summarise and evaluate key ideas contained in two texts) 2. one task to be selected from a choice of five (either text types: <u>article, letter, report, review</u>, OR <u>article, essay, letter, report, review</u> based on a set book)	1h 30 mins	240-280 (part 1) 280-320 (part 2)
Examination for the Certificate of Competency in English (ECCE)	Cambridge Michigan Language Assessments (CaMLA)	B2	Test takers read a short excerpt from a newspaper article and then write a <u>letter or essay (choice)</u> giving their opinion on the situation or issue.	30 mins	No word limit (test takers advised to write about one page)
Examination for the Certificate of Proficiency in English (ECPE)	Cambridge Michigan language Assessments (CaMLA)	C2	<u>Essay</u> - One task to be selected from a choice of two.	30 mins	No word limit
International English Language Testing System (IELTS General training)	British Council –IDP: IELTS Australia – Cambridge English Language Assessment	B1 to C2 according to overall score (score scale 1-9)	1. Test takers are presented with a situation and required to write a personal response in the form of an informal, semi-formal or formal letter 2. a semi-formal/neutral discursive essay * Task 2 contributes twice as much to the final Writing band score as Task 1	60 mins (20 mins for task 1 + 40 mins for task 2)	At least 150 words for task 1
International English Language Testing System (IELTS Academic)	British Council –IDP: IELTS Australia – Cambridge English Language Assessment	B1 to C2 according to overall score (score scale 1-9)	1. Test takers are asked to describe visual information (graph/table/chart/diagram) and to present the description in their own words. 2. Test takers are given a topic to write about. Answers should be a discursive consideration of the relevant issues * Task 2 contributes twice as much to the final Writing band score as Task 1	60 mins (20 mins for task 1 + 40 mins for task 2)	At least 150 words for task 1 At least 250 words for task 2
Test of English as a Foreign Language TOEFL iBT	Educational Testing Service (ETS)	B1 to C1 or above according to overall score (score scale 0-120)	1. A task to write based on what is read and listened to 2. Essay- A task to support an opinion on a topic	50 mins (20 mins for task 1 + 30 mins for task 2)	No specific word limit (advised to write 150-225 words for task1 and a minimum of 300 words for task 2)

Most exam rubrics ask for a specific text type; in the exams presented in table 1, 'essays' seem to be the most common followed by 'letters' but one can also see 'articles', 'reports' and 'reviews'. One exam includes a 'proposal' (CAE) and another one includes 'stories' (PET) in the tasks where there is choice. Other exams do not specify the text type but describe instead what the candidate needs to do (IELTS academic and TOEFL iBT, task 1). The following is a sample task 1 in the IELTS academic exams (IELTS, 2016b) which does not name the required text type (although this is usually labelled as 'Report' in educational material).

[Sample task]
The chart below shows the number of men and women in further education in Britain in three periods and whether they were studying full-time or part-time. Summarise the information by selecting and reporting the main features, and make comparisons where relevant.

Exam writing tasks also differ in terms of the freedom available, that is, whether tasks are compulsory or there is room for opting for text types that candidates may be more familiar with. They may also differ in the 'weighting' of the tasks meaning that a task may count more or less in the total score. In table 1, tasks of the same exam carry equal weight unless otherwise stated.

1.2. Motivation and research questions

Researchers have diachronically commented on the complexity of the writing skill. Raimes (1994: 164), describes writing as "a difficult, anxiety-filled activity". Lines (2014: 83) explains the complexity:

> For any writing task, students need to draw on their knowledge of the topic and its purpose and audience, and make appropriate structural, presentational and linguistic choices that shape meaning across the whole text, as well as achieving specific rhetorical or aesthetic effects through manipulation of sentences and vocabulary. (p.83)

Writing is a cognitively challenging task activating several skills at the same time, similar to several musical instruments playing in the same orchestra. It is not surprising that we use the term 'composition' in both music and educational settings. Originating from the Latin 'componere' which means 'putting things together', a composition is the way a whole is made up by putting things together, or joining forces to make something beautiful or creative. Expressing thoughts in written form is "probably the most complex constructive act that most human beings are ever expected to perform" (Bereiter and Scardamalia, 1983: 20). Writing is a blend of different talents, and this complexity often results in student confusion. The phenomenon is more acute when the students have to write in class with time limitations and when the topic is rather general with no specific frames or previous discussion- preparation.

A large part of the research conducted on writing has been based on theories deriving from the native language writing area. It is only recently, that the difference of non-native students' writing has been discussed more lively. Writing in a second language however, adds extra obstacles to the student as there is clearly lack of linguistic resources and possibly significant cultural differences. Theories referring to native language writers focus more on rhetoric and organization strategies and less on language errors. Grabe (2003: 242-243), summarizes the complexities among L2 writers caused by cultural and language differences. He mentions "differing senses of audience and author, differing preferences for organizing texts, differing ways to use texts as learning resources, differing cultural socializations and belief systems and differing functional uses for writing". Examining seventy-two studies which compare research on first and second language writing, Silva (1993: 669) observed that writing in a second language is "strategically, rhetorically and linguistically different in important ways from L1 writing".

Apart from investigating second language writing, this study differentiates writing in testing conditions from writing in classrooms or at home. Inferences drawn about test-takers' language abilities based on the scores from such assessments may result in high-stakes decisions such as university admission, graduation, citizenship and immigration or access to jobs and promotion in professional contexts. Viewing time and word constraints as well as anxiety levels as aspects seriously affecting response (Ferris, 2008; Gebril & Plakans, 2015; Hamp- Lyons, 1991) and influenced by the teacher-learner needs in the strongly testing-oriented context described above I decided to conduct research in the area of second-language writing with a focus on preparation for testing purposes.

Tasks posed in this kind of testing belong to genre categories that tend to have a repetitive character. Hamp – Lyons (2003: 174) notes that "even in so-called general test contexts, the types of tasks set have broad genre/ register patterns associated with them, and test-takers are often penalized if they do not write within those expectations." Reference to generic awareness is made in several parts of the official exam support materials without however mentioning the term 'genre'. In the Cambridge English First handbook (2015), for example, one sees the following: "Each writing task in Part 2 has a context, a purpose for writing and a target reader" and

> As with Part 1, candidates are expected to show that they are aware of the kind of writing required to accomplish a task, and must be able to demonstrate appropriate use of one or more of the following functions: describing, explaining, reporting, giving information, suggesting, recommending, persuading. (p. 28)

Candidates are supposed to be able to recognise basic generic elements such as the communicative purpose (using the term 'functions' in the handbook), to identify the target reader and the relation with the writer and to choose the appropriate register (referring to it with terms such as 'style' or 'tone'). For example, "candidates should consider carefully what level of formality is appropriate to the task" and "Your students need to think carefully about who the target reader is for each task and try to write in an appropriate style and tone."

Similarly, in the Cambridge English Advanced handbook (2015) they are told that they need the same knowledge and candidates get a warning for the possible treatment of different tasks based on topic similarity:

> It is important to have a balance between the functions required by the task and the relationship with the target reader. A pre-learned response on a similar topic is unlikely to meet the requirements of the specific task in the exam.
> (p. 31)

It is clear that the exam tasks theselves expect candidates to identify genres by reading the rubrics and subsequently adjust their response by using the appropriate register. In the

Cambridge English Proficiency handbook (2015) for example this is presented as a basic step:

> The question identifies the context, the writer's role and the target reader, which helps the candidate to choose the appropriate register. It is also very important that students learn to distinguish between the various task types required by the questions in Part 2. Even though a candidate may display an excellent command of the language, an answer will only achieve a high mark if all the above factors are taken into account. (p. 23)

However, the actual candidates' genre awareness is questionable. Personal professional experience suggests that the same is true for many teachers. Teaching material seems to lack important genre and register advice or provide information which is unclear and confusing. According to different coursebooks or teachers several characteristics should or should not be included in various text types. Rothery (1985) talks about contradictions in the educational system:

> It is a strange situation that we know what we want in written work and reward it when we receive it but we do not let our students know in advance exactly what we are looking for. In other words, we are operating within the framework of a hidden curriculum. (p. 80)

Even in cases where teachers are partly-informed on generic conventions and register variation the learners' exposure to these genres is limited due to time constraints in teaching, distance from authentic contexts and possibly due to lack of awareness on the part of the teachers of the importance of genre teaching. The sporadic sample answers in coursebooks are most of the time the candidates' way to deal with these mysterious requirements as well as with the diverse and vague feedback. Nesi & Gardner, (2012: 257) say that "pedagogic genres are the most occluded genres because they are written to be assessed and then discarded; published examples are rare." Even if someone has no experience of teaching English in this context, a visit on educational websites or blogs related to exam writing tasks can clearly show that:

a. candidates are constantly searching for sample or model answers

b. questions concerning genre/register issues are very common and the answers provided are often unclear or incosistent.

Young learners especially, who seem to be the main group of candidates in these exams in the Greek context are not well equipped to handle so many inconsistencies. Knowledge of genres seems to be implicit even though the exam tasks require sufficient knowledge and preparation in this area. As Reppen (2002: 321) explains, "for the L2 student, many writing conventions will remain a mystery unless teachers are able to bring these forms and patterns of language use to conscious awareness."

I choose to focus on writing as a highly demanding skill, often leading to failure. Prompted by the problems identified above I will try to find ways to make teaching more efficient and provide findings that could improve educational material. The aim is to clear up areas which remain vague and offer practical help to material writers, teachers and learners.

Research will first try to 'read' the situation by studying the learners' own views concerning writing for testing purposes. Then it will attempt to assess the text type/genre representation in educational material commonly used. The aim of the following part will be to identify genre and register features useful for candidates in the exams described above and provide information as to the relation between these genres. More specifically, this study will try to answer the following questions:

Research question 1: What are the learners' views on writing for testing purposes?

Although there have been several studies concerning teachers' or raters' attitudes towards writing (Cumming, 2003; Lines, 2014; Nesi & Gardner, 2006) and learners' performance (Dendrinos et al. 2013; Flowerdew, 2000; Henry & Roseberry, 2007), second language learners' views on Writing, especially for testing purposes, seem underrepresented in the literature. Hamps-Lyons (2003: 168) observes that "in writing assessment research the writer has too often been forgotten, probably because researchers are more distant from actual writing classrooms than they should be." She believes that the lack of student-focused research is "regrettable and problematic in all contexts" since that means that we cannot confidently advice teachers or education authorities and concludes that "the writer should never be perceived as a forgotten element".

The question does not aim to answer any specific issue related to 'writing' in detail but will instead try to 'touch' several issues connected with English language certification and the 'writing part'. Gebril & Plakans (2015: 1) say that "testing in general strikes a deep emotional chord in people" and use the terms 'ordeal' and 'daunting' to describe the experience. My interest derives from experience in preparing students for these exams. In full agreement with Gebril & Plakans, I wish to explore the learners' views on teaching practices and give them the chance to express their needs, difficulties and preferences.

Research question 2a: What is the representation of text types in model answers offered in English language teaching material?
2b: What would this representation be if texts were categorized according to genre instead of text type?

The aim is to assess the help provided in teaching material in terms of the various text types represented in model answers. Then, the same is going to be investigated for genres. Quantifying the representation of both text types and genres will make it possible to compare and check which text types and which genres are over or under-represented in educational material. Knowing how many and which genres are included in the same text type category can help teachers and learners see the range of possible writing tasks under the same label. It is a first step towards genre awareness. Measuring the representation of genres in educational material can show whether or not learners get the chance for adequate preparation on specific genres.

Research question 3: What are the most prevalent features of individual genres?

An analysis of the previously classified texts according to genre can bring typical generic elements to the surface aiming to make these elements available for explicit teaching. Model answers genre groups that have a sufficient representation in educational material will be examined in detail. The aim is to identify the lexicogrammatical features that stand out and explain how these features serve the main functions of the genre.

Research question 4: How are genres in this context related to each other? Does the similarity observed between genres under the same text type label justify the text type categorisation?

Research concerning the relation of each genre with the rest of the genres in this context can justify or provide evidence against the text type categorisation widely used in educational material. Genres belonging to the same category need to have greater similarity between them than with genres from other text type categories. If not, then text type categorisation could be misleading for learners.

This information can also assist the sequencing of various genres during the teaching process based on evidence rather than intuition. In the same way it is expected to be of help to educational material writers as they make decisions on structuring their material.

1.3. Organisation of the dissertation

There are altogether eight chapters in this dissertation. The first chapter introduces the context and the motivation for research. Four research questions are stated and the outline of the dissertation is presented.

Chapter Two reviews recent research on Genre theory, Genre-based-writing-instruction and Corpus Linguistics in relation to English language teaching. Previous genre and corpus analyses studies are presented showing what has been done and what still remains unanswered identifying this way the research gap and the significance of the current study.

Chapters Three, Four, Five and Six delineate the methodology applied for research questions one to four respectively and report the results of the current study.

Chapter Seven interprets the main findings with an in-depth discussion. It revisits the research questions addressed in Chapter One and explains how they have been answered.

The last chapter, Chapter Eight, ends the dissertation summarising the conclusions of the research, the contribution made as well as its limitations. Pedagogical implications and suggestions for future research are made based on the findings of this research.

CHAPTER TWO

LITERATURE REVIEW

This chapter explores two different fields, Corpus Linguistics and Genre Theory which tend to tend to come closer and synergise in some recent studies. In particular, the following literature review describes the way these two fields have contributed to language teaching. Section 2.1 reviews the use of Corpus Linguistics towards the improvement of teaching English both inside as well as outside the classroom. Section 2.2 explains the main principles of the Genre Theory and its applications in the teaching of writing. It also refers to the main purpose of Genre and/or Register analyses, the different approaches adopted as well as some of the issues that cause confusion due to the lack of consensus among researchers. The focus here is on the variation of perspectives and definitions given on the terms 'genre', 'text type' and 'register' in an attempt to both clear up terminology but also to justify the approach based on which corpus texts are going to be categorised later, in chapter four of this dissertation.

2.1. Corpus Linguistics and the teaching of English

'Corpus', means body in Latin and 'corpora' is the plural form although 'corpuses' is sometimes used in English. A corpus is "a collection of pieces of language text in electronic form, selected according to external criteria to represent, as far as possible, a language or language variety as a source of data for linguistic research" (Sinclair 2005: 16), or to put it more simply it is "a body of machine-readable text" (McEnery & Wilson, 2001: 197).

Corpus linguistics started in the 1960s with the Brown University creating what was later called the Brown corpus consisting of one million words of American English and Lancaster University starting the LOB corpus (Lancaster-Oslo-Bergen), (British English), in the 70s which was completed in Norway. The invention of the tape recorder enabled the collection of spoken data and led to the creation of the first electronic corpus of spoken language at the University of Edinburgh in the years 1963-5 (Tognini-Bonelli, 2010: 16). Its tipping point however, was probably in the mid 1990s, following the increasing power of computers. It became known among the ESL teachers mainly because of the creation of the *Collins COBUILD English Language Dictionary*, the first corpus-based dictionary for learners.

Cobuild corpus by Birmingham University was a 20-million-word corpus, a revolution compared to the previous 1 million word corpora. It was with the use of the optical character reader (OCR), an enormous machine at that time, that such a great deal of texts was recorded on computer (McCarthy & O'Keeffe, 2010: 5). Since then, more and more language materials are corpus-based. In addition, Corpus linguistics techniques started to spread to other disciplines outside linguistics such history, geography, sociology, politics education and media research.

Working with computers became "cheaper, faster" and ensured "total accuracy in text processing". (McEnery & Wilson, 2001:17). Computers process large quantities of linguistic data which would be impossible for humans to analyze on their own, they do not make mistakes and in contrast to human readers they are not subjective (Φραντζή, 2012: 25). Stubbs (1996: 231) emphasises the importance of this change by comparing it to great inventions such as the microscope and the telescope, "which suddenly allowed scientists to observe things that had never been seen before". As Sinclair (2004a: 10) explains this was "not only a qualitative change in the amount of language data available for study, but also a consequent qualitative change in the relation between data and hypothesis".

Corpus Linguistics is considered more as a powerful methodology rather than a new branch of linguistics but this methodology has the potential to change perspectives on language (Granger, 2002: 3). It studies natural language and provides real life data based on frequencies as opposed to content based on intuition. Many aspects of language can be investigated at once allowing the researcher to see how frequently words or patterns are used.

In the past, "linguistic descriptions relied very much on native-speaker intuition and introspection" (Tsui, 2004: 39). Now, it is possible to study how language is actually used rather than how it is perceived to be used. Patterns of use that are revealed in corpus analyses "often run counter to our expectations based on intuition" (Biber et al., 1994: 169), produce "only facts, but facts of previously unsuspected kinds" (Stubbs, 1996: 232) and can "shake our faith quite a bit in established models" (Sinclair, 2004a: 23).

Corpus linguistics shares common ground with Systemic Functional Linguistics as they are both concerned with naturally occurring language and with language as texts. They also take into account the context and are in favour of quantitative evidence. Corpus linguistics has the

ability to reveal tendencies in speakers' choices reminding that a strict adherence to grammatical rules can be unrealistic at times. It can discover functional units of meaning, patterns, that would be impossible to notice and categorise, as they do not correspond to traditional lexical or grammatical units. Sinclair (2004a: 165) sees the distinction between grammar and lexis as a very basic model of language and explains that the new evidence provided from corpora has prompted the reconsideration of this model. Corpus linguistics' respect to the actual use of the language goes hand in hand with the core viewpoint of the functional grammar which characterizes linguistic instances not in explicit formal terms but as "semantic patterns" relating them to their "non-linguistic universe of its situational and cultural environment" (Halliday, 1985, introduction xvii). Corpora provide a large empirical database of natural discourse ideal for functional analyses of language where the focus is the description of language as a communicative tool (Meyer, 2002: 3-4). For all these reasons it is quite clear why corpus linguistics research is a "thriving and productive area of applied linguistics" (Ferris, 2011: 187) and why Leech (1992: 106) describes this methodology as an "open sesame' to a new way of thinking about language".

Corpora rarely offer explanations on their own; their usefulness depends on the research question we pose and the way it is going to be utilized by the researcher. Studies with corpora can be either 'corpus- driven' or 'corpus- based'. The first one starts with no initial hypothesis and is theory-neutral. The data lead the researcher to notice any significant patterns which may provide new knowledge. The corpus-based study on the other hand moves within a specific theoretical frame. Applying a deductive approach, the corpus is used to test an initial hypothesis.

Depending on the research question the appropriate type of corpus needs to be selected. One of the distinctions between corpora relates to size. It may be a 'reference' or a 'monitor' corpus. The first is a fixed sized corpus (e.g., the British National Corpus - BNC) while the second is a corpus that keeps being expanded (e.g., the Bank of English, Cambridge Learner Corpus - CLC). 'Monitor' corpora are really useful for looking at rapid changes in language.

The other distinction has to do with content. The corpus may be 'general' or 'specialized'. A general corpus (e.g., Lancaster Oslo Bergen - LOB, BROWN) usually tries to reflect the general use of the English language whereas a specialized one (e.g., The Michigan Corpus of Academic Spoken English - MICASE) will focus on particular contexts and/or users. They will contain written or spoken language with the second type being rarer mainly due to the

difficulty and time one needs in order to transcribe oral speech. There are of course corpora which contain both written and spoken speech (e.g., BNC, The Australian Corpus of English - ACE)

Another type is the 'sampled' corpus; that is a finite collection of carefully selected texts. In addition, one may come across 'monolingual' or 'multilingual' (e.g. The International Corpus of English (ICE) with its various subsets of different languages) and 'parallel' (The Corpus Resources and Terminology Extraction - CRATER) or 'comparable' corpora. The ones composed of source text in one language and their translations in another language are called 'parallel corpora', while the term 'comparable corpora' typically refers to those which contain different components of different languages or varieties of the same language which are put together by using similar sampling techniques" (Baker et al., 2006: 126-127). Comparable corpora are similar in terms of balance and representativeness but texts included are not translations of each other. Parallel and comparative corpora are mainly used for translation and contrastive studies (McEnery & Xiao, 2007: 3,4).

There is also a distinction between 'synchronic' and 'diachronic' corpora. Synchronic corpora contain language as it is spoken or written at a particular point in time. Diachronic corpora allow us to look at language change over a long period of time (e.g. the Helsinki corpus covering the period from AD 750 to 1700) (Baker et al., 2006: 56).

Within the 'specialised' corpora family there are two types of corpora which are especially created and analysed for pedagogical purposes. The first and less often mentioned in the literature is the 'pedagogic' corpus (e.g. the corpus of TExtbook MAterial - TeMA) and the second, and more frequently used is the 'learner' corpus (e.g., the International Corpus of Learner English – ICLE, Longman Learner Corpus – LLC). Hunston (2002: 16) has defined the 'pedagogic corpus' as "a corpus consisting of all the language a learner has been exposed to". Meunier & Gouverneur (2009: 179-201), however, find this utopian as it is impossible to gather all the material a learner has used (e.g., coursebooks, readers, tapes) inside or outside the classroom and offers a more realistic definition: "a large enough and representative sample of the language, spoken and written, a learner has been or is likely to be exposed to via teaching material, either in the classroom or during self-study activities".

A 'learner corpus' is a computerized textual database of the language produced by foreign language learners (Leech 1998: xiv). They are often defined as systematic computerised collections of texts with the term 'systematic' meaning that texts have been selected according to a number of criteria such as the learners' age, proficiency level, and that the selection is representative of a certain learner group. Sylvaine Granger and her team were the first to compile a learner corpus (ICLE, Granger et al., 2002).

Corpora can be 'tagged' meaning that there may be special labels which offer linguistic information separately for each word or sentence. We then say that the corpus is 'annotated'. We may have 'syntactic/grammatical', 'semantic' or 'pragmatic' annotation depending on the type of question the researcher poses each time. Part-of-Speech (POS) tagging, is the most common way to add value to data. Corpora may also include non-linguistic information (e.g. the author, the date of publication). The information added in a text is called metadata. It is data about data.

Corpora can contribute in language teaching in a multitude of ways. One way to categorise their contribution to language teaching and learning is to see their use 'in the classroom' and 'out of classroom'. Broadly speaking the first use aims to provide extensive empirical data on language use and enlarge the students' exposure. The second refers to all the work done out of classrooms in order to support teaching. This work is usually aimed at creating corpus-based or corpus-informed material, informing teaching and learning practices, or assessment methods.

2.1.1. Working with corpora inside the classroom

Learner corpora can be studied in order for both the teacher and the students to notice mistakes in language use in a large sample of their work. Corpora of various types can be used as teaching material offering what might be called 'condensed language exposure' (Gabrielatos, 2005) or as a permanently available 'informant' allowing for the 'data-driven-learning' approach (Johns, 1991a, 1991b). This approach promotes autonomous learning based on the principles of 'learning by doing' and 'learning by discovery'.

Pedagogical corpora can be used to raise awareness by providing additional evidence on the

use of a word or phrase by referring back to examples from previous texts the students are familiar with (Baker et al. 2006: 128; Hunston, 2002: 16). General corpora can offer authentic texts and examples to compensate for the invented examples commonly seen in non-corpus-based material. As Stubbs (1996: 31) explains, these invented examples "have no independent authority or reason for their existence".

Using corpora in the classroom can be a challenge though. Clearly, learners need time to familiarise themselves with these new reference tools and acquire 'corpus literacy' (Chang, 2010, Mukherjee, 2004). For teachers however, working with corpora may seem like a daunting task. They need to be trained first to be able to understand the tremendous possibilities this activity has to offer. Apart from technical difficulties educators may find this authentic material difficult to handle. Sinclair (2004b) points out that corpus use can contradict well-known practices and existing views:

> From a classroom perspective the emergence of corpora may not seem to be good news—a large amount of new information to absorb, and an unsettling failure to confirm the consensus view of language that has been considered adequate for most classrooms for many years. (p. 271)

As this approach is 'learner-centered' as opposed to 'teacher-led', teachers usually regarded as experts may find it difficult to adjust to new roles where they would have to admit they do not know everything (Boulton & Tyne, 2015: 308, 309). McCarthy, M. and O'Keeffe, A. (2010) discuss this effect:

> By its nature, it turns the traditional order within the classroom upon its head. The corpus becomes the centre of knowledge, the students take on the role of questioner and the teacher is challenged to hand over control and facilitate learning. (p. 7)

Although there is development in this area, corpus research remains "largely invisible downstream to teachers and learners" (Boulton, 2010: 129) and the uptake of corpus pedagogy has been slow (Poole, 2015: 275). Teachers seem to use material they are more familiar with and benefit indirectly from the progress that is made in corpus-based or corpus-informed dictionaries, course books, grammars and software. Mukherjee, (2004: 242, 243) reporting on a survey in Germany, says that language teachers make use of corpus-based

material but they are not aware of the linguistic background of these products and that although the use of corpora have become mainstream in English linguistics only a tiny fraction of English language teachers are aware of the existence of corpus linguistics. McCarthy, M. and O'Keeffe, A. (2010: 9) stress the need for more wide dissemination of corpus linguistics findings as they can greatly help teachers and material designers.

2.1.2. Working with corpora outside the classroom

Insight from corpora can inform language teaching in a number of interesting ways. First, corpora can be used as 'item banks' by teachers and language testers. Authentic examples of language use can enhance activities focusing on specific language areas instead of made-up examples based on intuition. In the past one had to purchase a corpus and software in order to conduct analyses. Now, thanks to web-interfaces freely available (e.g., British National Corpus- BNCWeb, Corpus Query processor - CQPweb) anyone interested in real language use has access to word frequencies, collocates and patterns, the distribution of items in texts in various types of corpora as well as plenty of information and guidance in the form of tutorials. Access to appropriate software also means that teachers can build their own corpora. Even though these corpora would be small they may provide specialised information. Building a learner corpus, for example, can reveal their learners' frequent errors and major weaknesses. This information can then be used to assess progress and adjust teaching plans accordingly.

Apart from classroom-based assessment, corpora can improve assessment methods in high-stakes testing. The 'English Profile' programme, for example, aspires to create an inventory, a 'set of specifications of lexis and grammar', to be added to the functional characterizations of the proficiency levels in the CEFR (Hawkins & Filipović 2012; Tono, 2013).

The distribution of linguistic items in corpora can also be the basis for defining the goals and the content of the curriculum (Tsui, 2004: 41). A lot of materials produced for learners of language take this valuable information into account when deciding what and when to prioritise teaching to a particular learner. A language item for example which is frequent may be chosen to be presented earlier in the course than one that appears less. In textbooks that are not based on corpus studies it is very common to see this type of mistaken prioritisation.

The analysis of learner language provides insights into learner needs in different contexts, which then inform learner dictionaries and grammars (Gabrielatos, 2005). Formerly, publishing houses engaged in producing ELT (English Language Teaching) reference and course books were dependent primarily on the intuitions of highly skilled and experienced lexicographers to anticipate learners' difficulties with English.

These days, most of the dictionaries for the English language are being produced based on corpus data. A general corpus can offer attested examples and a learner corpus, seen as a database of errors and the contexts in which they commonly occur, can affect the way educational material present, prioritise or emphasise certain language items. Osborne (2004: 266) talks about top-down and bottom-up approaches to corpora in language teaching. In top-down approaches data from native speakers provide evidence of target usage and in bottom-up approaches the learners' own productions are the starting point for the enrichment of material. Evidence on the frequency of use has aided the ranking of certain words and phrases in dictionaries, the examples provided and the prominence of specific items of difficulty in learner dictionaries. (e.g., The Longman Essential Activator Dictionary).

Well-known publishing houses (e.g., Cambridge, Longman, MacMillan) have their own general and/or learner corpora providing access to their material writers. They use this feature as a 'selling point' adding special logos (e.g., Cambridge University Press) or phrases on their products to emphasise the authenticity of the language presented in them. (Meunier, Gouverneur, 2009).

The impact of corpus linguistics on grammar is manifested mainly in two reference grammars of English, the 'Longman Grammar of Spoken and Written English' (Biber et al. 1999) and the 'Cambridge Grammar of English' (Carter & McCarthy 2006). The perspective has been to reflect reality "moving the field beyond the dichotomous view of grammatical structures as acceptable versus unacceptable and accurate versus inaccurate" (Conrad, 2010: 238). Pedagogical grammars which are corpus informed are also available in the market nowadays. 'Exploring Grammar in Context' (Carter et al. 2000) and 'English Grammar Today' (Carter et al., 2011) are such examples.

Coursebooks such as the 'Touchstone' series (McCarthy, 2004; McCarthy et al. 2005), base their material on the North American English portion of the Cambridge International Corpus,

selecting the most frequent and typical uses of everyday words and grammar structures. Knowledge about spoken language from the corpus is presented in special 'In conversation' boxes.

Pedagogical corpora have been used to investigate the language that learners come across in textbooks in order to check its authenticity. Romer (2006: 239) insists that we need to compare the language that is used with the language we teach in order to improve language teaching materials. Koprowski (2005: 330), for example, examines the lexical syllabus in coursebooks and finds out that "designers, by and large, start with topic or theme and then intuit what they consider to be relevant or 'useful' lexical phrases. This selection process appears to be unscientific, largely grounded on the personal discretion and intuition of the writers." Romer (2004: 197), compares the use of modal verbs in spoken English and a German textbook series used in EFL calssrooms concluding that a number of frequent and important features of usage are underrepresented or even left out completely in textbooks whereas less important features are over-emphasised. She then investigates the use of progressive forms and finds out that the analysed EFL textbooks differ significantly from natural language use, and from each other (Romer, 2006: 239).

However, coursebooks in general have remained 'immune' to influence from corpus linguistics (Burton, 2012; Meunier and Gouverneur, 2009). Taking into consideration the points made about teachers' limited use of corpora in the classroom above and their lack of awareness of corpus linguistics methods and benefits it is natural for the demand for these products to be slow. Littlejohn (1992: 278) says that "potential commercial viability, in particular, appears to act as a 'filter' on innovation, leading materials to respond only very slowly to applied linguistic ideas". Sampson (2013: 257, 258) talks about a "time lag which applies to all new intellectual movements" and argues that "the unempirical research style which came into vogue in the 1960s" cannot survive.

2.2. The Genre Theory

According to Hymes (1972: 277) normal children acquire "knowledge of sentences, not only as grammatical, but also as appropriate". They understand how to talk in particular contexts.

Γεωργακοπούλου & Γούτσος (2011) explain how this appropriateness is acquired through experience:

> Speakers and listeners during speech production and comprehension strategically make use of models, which are based on specific social and cultural beliefs, knowledge, roles and relationships. These beliefs have cognitive representations organized in such a way that they can be recalled but also actively modified, according to the data of our experience. This way the text is directly dependent on its cognitive context. (p. 59, translated)

Texts are organized according to specific roles and the purpose they want to achieve in specific cultural and social contexts. These types of texts are called "genres". Hyland (2002: 114), describes genre as "socially recognized ways of using language". Swales (1990: 58) states that "a genre comprises a class of communicative events, the members of which share some set of communicative purposes. The purposes are recognised by the expert members of the parent discourse community, and thereby constitute the rationale for the genre." Christie (1984 :20) defines genre as "a purposeful, staged cultural activity in which human beings engage". Hyland (2003a), referring to Martin's (1992) similar definition, "a goal-oriented, staged social process", explains that

> genres are social processes because members of a culture interact to achieve them; they are goal-oriented because they have evolved to achieve things; and staged because meanings are made in steps and it usually takes writers more than one step to reach their goals (p. 19).

Casanave (2004: 84), talks about "the socially and politically situated contexts of writing and how these contexts influence both how writing gets done and the end products of writing". The term 'context' usually refers to the environment in which a chunk of discourse occurs. This includes not only the linguistic environment – the utterances which precede and follow the particular utterance – but also the ongoing situation in which the particular text is produced, and the wider culture. All these features work together to contribute meaning to utterances. According to Halliday & Hasan (1985), the term 'genre' is a short form for the more elaborate phrase 'genre-specific semantic potential'. The term 'context' itself reveals this interrelationship as it is made up of 'con' (with) and (text). The term 'text' (derived from

the Latin 'texere') originally meant something woven, a metaphor still transparent in that both a chunk of discourse and a piece of cloth can both be described as having 'texture'. (Sifianou, 2006: 52). Using Malinowski's (1923) terms, 'context of situation' meaning the environment of the text, and the 'context of culture', meaning the total cultural background, Halliday (1978: 5) stresses the importance of studying not only the language or text but also the "total environment in which a text unfolds" and Halliday and Hasan (1985: 5) remind us that "contexts precede texts" as "the situation is prior to the discourse that relates to it".

Every day we engage in conversations using mechanisms in a subconscious way in order to adjust. We try to match what is happening with a model of the context of situation in our minds. We assign it to a 'field', noting what is going on; we assign it to a 'tenor', noticing the persons and their relationships and we assign it to a 'mode' seeing what is being achieved by means of language. We "make predictions about the kinds of meaning that are likely to be foregrounded in that particular situation". We come with our "mind alert", in order to take part in this interaction. Every day "we are making inferences from the situation to the text and from the text to the situation" (Halliday &Hasan, 1985: 28, 37). We choose language which in our minds is typically associated with the situation at hand. When these discourses "become typified – that is, when the same events are carried out repeatedly through the same practices – they may be referred to as genres" (Tardy, 2009: 12).

Martin (2009), describing the 'Genre theory' explains that

> it tries to describe the ways in which we mobilize language – how out of all the things we might do with language, each culture chooses just a few, and enacts them over and over again – slowly adding to the repertoire as needs arise, slowly dropping things that are not much use. Genre theory is thus a theory of the borders of our social world, and our familiarity with what to expect. (p. 13)

The phrase 'as needs arise' and 'not much use' emphasizes the temporary element of what we teach and believe. Even though texts of the same genre share similar patterns, one should not forget that these patterns and general characteristics have not been stable over the years. Hyland (2002: 123) reminds us that genres are not "fixed, monolithic, discreet, and unchanging". Being part of the evolutionary system, genres can both be stable as well as flexible adapting to social change (Foley, 1990: 226; Bazerman, 2012: 230). This should be

stated and discussed in every classroom while teaching writing, referring to the e-mail genre for example, which has altered the conventional style of a letter or other changes brought by the digital evolution. Language in use demands constant study and observation, readjusting our knowledge, noticing the differences and stressing what is actually used at present. Writing is above all a social activity and it is bound to change form and style according to the changing context. Therefore, although genres are typified in the sense that they have a repeated structure making them recognisable, one needs to be aware that genre conventions should not be taught as rigid templates which could be applied eternally.

2.2.1. Generic competence and second language writers

McNamara & Roever (2006: 55), remind us of the difference between 'sociopragmatic' and 'pragmalinguistic' knowledge and the fact that both components are equally necessary. 'Sociopragmatic' knowledge helps the language user to adjust in the community's social rules, become familiar with appropriate norms and avoid unintentional offence or attitudes that disrespect common discourse practices. 'Pragmalinguistic' knowledge arms learners with the tools to interact, the necessary linguistic competence to convey messages. Because of the strong interrelation of these two competences "it is often difficult in practice to determine whether a given error was due to pragmalinguistic or sociopragmatic deficits". In the same line, Paltridge (2001: 7), considers generic competence as the combination of linguistic competence, that is, the mastery of the language code, and communicative competence based on pragmatic knowledge.

> We do not simply 'know' our mother tongue as an abstract system of vocal signals, or as if it was some sort of a grammar book with a dictionary attached. We know it in the sense of knowing how to use it; we know how to communicate with other people, how to choose forms of language that are appropriate to the type of situation we find ourselves in, and so on. All this can be expressed as a form of knowledge: we know how to behave linguistically.
>
> (Halliday, 1978: 13)

Even though in the development of the child as a social being this happens indirectly as Halliday (1978: 9) explains "through the accumulated experience of numerous small events",

second language learners need to acquire essential cultural competence through teaching and exposure to representative discourse samples of the new context. "Mere knowledge of meanings of words and the constructions in which they may occur does not guarantee successful interaction, since actual communication is located in particular socio-cultural contexts which contribute significant information" (Sifianou, 2006: 3). This means that especially second language teachers apart from teaching the language, have a responsibility to assist students in participating in discourses and become active members of social groups, communities and cultures they may be totally unfamiliar with. After all, "our ability to recognize the resemblance of any text to a genre prototype is [..] a consequence of exposure to these genres and our experience of using them in specific contexts" (Hyland 2002: 120). Learning a genre means learning how to "participate in the actions of a community". (Miller 1984: 165). The added difficulty on non-native speakers is mentioned in the New South Wales curriculum (Board of studies, 1998):

> All aspects of the English language, such as its sounds, ways of constructing meaning, its conventional patterns, as well as the appropriate language for a range of situations, are critical to the success of ESL learners. They have the double task of continuing to develop cognitively and of developing a new language at the same time. (p. 12)

In her article 'The Rhetoric/Syntax split', Kroll (1990: 41) also talks about this double task causing variation in performance especially in the case of ESL learners: "They must operate not only within a complex system of discourse and rhetorical rules that they have had limited exposure to but also according to an entire linguistic system (English) that may be but partially mastered."

Yasuda (2011: 113) says that "a heightened awareness of the relationship between the goals of a genre and the linguistic resources that realize them" could address both these areas of difficulty. The notion of genre-knowledge is important to L2 writing teachers because "it stresses that genres are specific to particular cultures and communities, reminding us that our students may not share this knowledge with us" (Hyland, 2004: 54), and because it urges us to "guide learners to participate effectively in the world outside the ESL (English as a Second Language) classroom" (Hyland, 2007: 148-149). The importance of teachers' genre

awareness is also stressed by Halliday & Hasan (1985: 69) who see it as an active ingredient of success.

Christie F. (1984: 20) argues that schools often fail to show pupils explicitly what the nature of each genre is, leaving children to work them out for themselves through their reading and through the few clues given by the teachers in their general instructions (e.g. 'remember to write a beginning, middle and end') and their evaluative comments (e.g. 'a good report, but where's your conclusion?). This may leave students worrying why they got better marks for some piece of writing than for others. She even doubts that teachers are always aware of the types of writing they are teaching. In the same line, Hyland (2003b: 151) criticises the lack of explicit guidance towards writing different types of texts in courses where instruction is not genre-based and where "learners are expected to acquire the genres they need from repeated writing experiences or the teacher's notes in the margins of their essays". While interviewing university tutors, Nesi & Gardner (2006), realised that although tutors appreciated argument, structure, clarity and originality in texts they could not be explicit on the ways these characteristics could be realised or recognised in text.

2.2.2. The 'genre-based-writing-instruction' (GBWI) approach

The emphasis on the accurate use of the language placed by many teachers during teaching does not mean that students do not have difficulty in writing. Several students 'struggle' for content the same way native language students do, while having extra lack of knowledge in social schemata and their corresponding form, process and content.

The literature on applied genre-based teaching suggests that learners even after a short period of instruction can benefit from genre-based teaching as they can learn faster (Amogne, 2013: 247), transform their genre knowledge from a receptive level to a productive level (Yasuda, 2011: 120) and improve preformance (Ellis et al., 1998: 153).

Supporters of teaching that focuses on genres view writing as an attempt to communicate with the reader. They place emphasis on the social and cultural context of genres, the real-world communicative practices and try to provide their students with generic competence. In a recent conference presentation, Tardy (2015) insists that students should have control over the texts they write rather than the texts controlling them.

This approach can also face problems related to the different competences of the same learner. Familiarity with a genre may enhance the learners' writing while they may prove incompetent in another genre. Students do not seem to move linearly from one developmental stage to another. "They travel back and forth depending on the complexity of the topics they write about, as well as the purpose, genre, and intended audience of their writing" (Fu, 2009: 75). Familiarity with a genre may result in quality writing whereas the same person may prove incompetent in another genre (Torrance, 1996a: 5).

Since genres are more or less typified and have a recognisable form this means that if students are aware of the genre, they can predict the general organization of the text and the stages needed to serve its purpose. (Derewianka, 1996: 8). "Genre knowledge demystifies writing, making writing very much like playing a part in an already exisitng scene with a script or map in hand" (Devit et al., 2004: 100).

After students understand the *social purpose* of a text (the 'why'), they are called to apply the register of the specific genre. A register is that variety of language that matches the specific context of situation. "A register is a semantic concept. It can be defined as a configuration of meanings that are typically associated with a particular situational configuration of field, mode, and tenor" (Halliday & Hasan, 1985: 38,39). These three variables influence the language used and changing even one variable may cause significant alteration in the register used. It is necessary therefore for learners to learn how to identify the 'field' (the 'what' - what is going on) the 'participants' (the 'who'- subjects involved) and the 'mode' (the 'how'- the role of the language). (Κονδύλη & Λύκου, 2009; Hyland, 2002).

Over the years, three Genre Schools have developed, the 'Rhetorical Genre Studies' (RGS, also called 'New Rhetoric'), 'English for Specific Purposes' (ESP) and 'the Sydney School', which is based on Systemic Functional Linguistics (Hyon, 1996). While all these Schools agree on the importance of genre awareness, they may have a different focus on their analysis. Flowerdew, (2002: 91,92) categorises them as linguistic (ESP and the Sydney School) and non-linguistic (RGS) approaches and explains their different starting points: "the linguistic approach looks to the situational context to interpret the linguistic and discourse structures, whereas the New Rhetoric may look to the text to interpret the situational context."

The RGS School also differentiates itself on explicit teaching, while the other two Schools, which "emerged out of a pedagogical imperative", favour explicit genre teaching based on text linguistic analysis (Smedegaard, 2015: 34). Johns, (2011: 21) based on a survey in the 2009 Second Language Writing Symposium states that the two 'linguistic' pedagogies (ESP and the Sydney School) have been most successful in reaching their goals in L2 contexts.

The Sydney School model, developed in the context of the Australian school system allows for a more systematic approach to teaching (Flowerdew, 2015: 6) and results from writing assessments demonstrate that students of all competency levels benefit from this approach (Knapp, 2002: 21). The ESP approach also supports explicit teaching of genres but its emphasis is on particular contexts (e.g. academic, professional) (Hyland, 2003a: 75).

On the other hand, Devit et al. (2004: 93,94) representing the RGS school, describe a four-stage pattern of teaching writing where the end purpose has been to see what rhetorical patterns are related to what situation. The stages are the following:
 1. Collect samples of the genre,
 2. Identify the scene and describe the situation in which the genre is used,
 3. Identify and describe patterns in the genre's features and
 4. Analyse what these patterns reveal about the situation.

The targeted populations also vary among the three schools. The RGS is best known in writing courses in North America (Devitt et al. 2004), the ESP mostly in university (Swales, 1990; Hyland, 2004) and professional settings (Bhatia, 1993) and the Sydney School has targeted mostly disadvantaged students in primary and secondary schools as well as L2 learners (Knapp, 1989; Martin, 1985) and only recently university students (Nesi & Gardner, 2012).

For the ESP and the Sydney School genre teaching involves being explicit about how texts are grammatically patterned, but grammar is integrated into the exploration of texts and contexts rather than taught as a discreet component (Hyland, 2004: 89, 134; 2007:153; Frodesen & Holten, 2003: 141, 153). They argue that awareness of genres helps learners predict the organization, the stages and the linguistic elements characteristic of the genre, becoming competent writers. Therefore, they support explicit teaching (Derewianka, 1990,

1996; Halliday & Hasan, 1985; Hyland, 2003a, 2003b; 2007; Kress & Knapp, 1992; Nesi & Garner, 2012; Paltridge, 2001; Torrance, 1996a, 1996b) and disapprove of teachers who give unclear feedback to their students. The same disapproval is sometimes expressed towards English teaching material, especially because of the way grammar is presented and taught as a discrete component. (Hammond & Derewianka, 2001: 192; Hyland, 2004: 89).

It is common, in practice, for teachers to spend more time on language form rather than language use, especially in less advanced classes. Teaching writing to second language learners often means teaching grammar or/and vocabulary based on the actual mistakes each student makes. Feedback becomes a substitute for real teaching of writing and in a way explains the large number of articles on feedback and its effectiveness. (Cohen & Cavalcanti, 1990: 155; Fathman & Whalley, 1990: 178).

Another reason for this emphasis on grammar is the fact that these mistakes are much easier to correct and count (Casanave, 2004: 66) compared to other vague terms related to writing assessment such as 'flow', 'style', 'formality' or 'cohesiveness'. The common belief that students need to control core forms of language before they can write in English is also seen as a cause for neglecting the teaching of writing (Cummings' foreword in Fu, 2009: ix; Johns et al., 2006: 238). According to Reid (2008: 180), research has demonstrated that grammar exercises, particularly those that are not deeply rooted in the context of the assignment, do not transfer to future student writing (despite the high comfort level students and many teachers have with such exercises). Further on, he reminds us that by correcting and commenting on each and every grammatical mistake in students' writing, teachers "perpetuate the grammar myth". This obsession with grammatical mistakes during correction is perhaps justified considering the way these teachers were taught English themselves. The prevalent grammar theory when most teachers were young was Chomsky's (1957,1965) 'generative grammar' with an emphasis on the ideal language user producing grammatically correct sentences, a view that ignored the context in which language occurs.

In contrast to the formal grammars of this kind with little or no attention to meaning (semantics), context or language use (pragmatics), functional grammars seek "to explain why one linguistic form is more appropriate than another in satisfying a particular communicative purpose in a particular context" (Larsen-Freeman, 2001: 34). The shift which occurred during the 1970s moved from formal or structural to functional approaches in language education

and from an interest to 'grammatical competence' to an interest in 'communicative competence' (Hymes, 1972).

The 'Systemic Functional Linguistics' perspective which forms the basis of the Sydney School view of genre is based on Haliday's work (Halliday, 1985; Halliday, 1978; Halliday & Mathiessen, 2004). His systemic description of grammar sees grammar as "a network of interrelated meaningful choices" (Halliday & Mathiessen, 2004: 31). Language is considered to be a 'system' from which people make choices to convey meanings. It is called 'functional' because "everything in it can be explained ultimately, by reference to how language is used" (Halliday, 1985, introduction). As Thompson (2014: 21, 262) explains: "speakers do not go round producing de-contextualized grammatically correct sentences: they have reasons for saying something, and for saying it in the way they do". Seeing language from a Functional Grammar perspective means coming face-to-face with the "complexity and uncertainty that meaning and function inevitably introduce", accepting "the multidimensionality and fuzziness as an inherent and central figure of language" instead of "labelling isolated, decontextualized bits of language" with a focus on form. Although it often seems as a search in the dark the potential results are much more rewarding.

Genre-based-instruction makes sure that learners come to contact with genres that they will need in the future and that they understand the procedures and the abilities required while steadily the support is reduced as self-confidence is increased. This last tendency is influenced by Vygotsky (1978) and the notion of 'the Zone of Proximal Development', the area between what learners can do independently and what they can do with assistance. The teacher will 'scaffold' (Wood et al., 1976), the learner to develop through verbal interaction and task negotiation when and where it is needed. Scaffolding in this case may take the form of input and guided practice at first and then genre templates where learners need to fill some parts, until the teacher's guidance is finally withdrawn and the learner works alone.

```
                    ┌──────────────┐
                    │ Building the │
                    │   Context    │
                    └──────────────┘
        ┌──────────────┐        ┌──────────────────┐
        │   Linking    │        │  Modelling  and  │
        │Related Texts │        │  Deconstructing  │
        │              │        │      Texts       │
        └──────────────┘        └──────────────────┘
    ┌──────────────┐                ┌──────────────┐
    │ Independent  │                │    Joint     │
    │ Construction │                │ Construction │
    │   of Texts   │                │   of Texts   │
    └──────────────┘                └──────────────┘
```
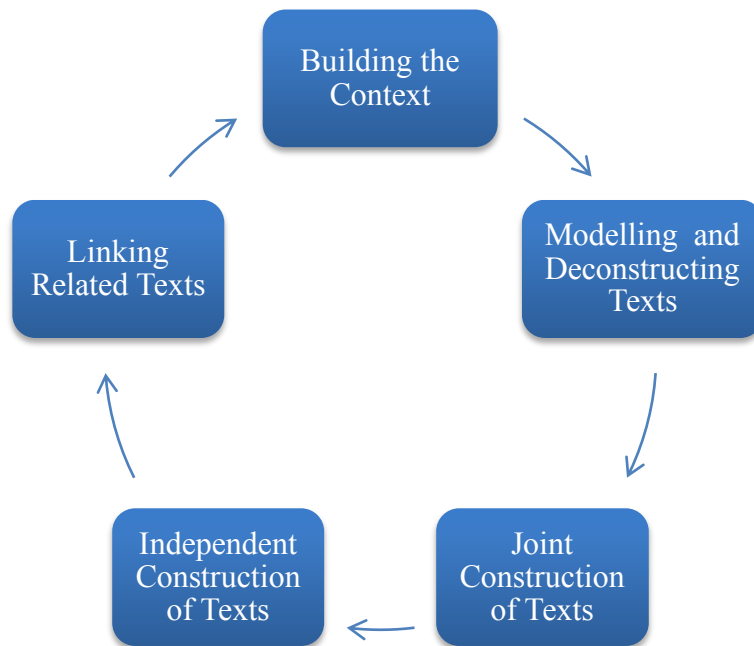
Figure 2.1 The cycle of teaching and learning (Feez, 2002: 65)

The genre-based 'Teaching and learning cycle', based on Vygotsky's (1978) notion of 'scaffolding', has been fundamental in the Sydney school approach. It consists of five key stages (figure 2.1): firstly, in 'setting the context' stage, the students are assisted in understanding the context and the communicative purpose of the genre. During the 'modeling and deconstructing' stage, they read representative texts of the chosen genre, trying to identify its key features and how it moves from one stage to the other. Later on, they collaborate to produce their own texts guided by the teacher ('joint construction' stage), before they attempt to write independently, monitored by the teacher ('independent construction' stage). Finally, at the 'comparing' stage, students link their work with other texts and comment on different genres and contexts (Feez, 1998; Hyland, 2004). An early application of this approach in the state of New South Wales (NSW) in Australia has been very popular both to teachers and learners. It "has mandated a genre-based pedagogy in the English K–6 Syllabus" which has largely benefited learners as demonstrated in the results of writing assessments (Knapp, 2002: 21).

The second stage of modelling and deconstructing texts in Feez's cycle is strongly influenced by Vygotsky's (1987: 211) belief that "what the child is able to do in collaboration today he will be able to do independently tomorrow" and the understanding that this 'scaffolding' may take the form of other resources such as texts or digital applications not necessarily involving

the support of another person. During deconstruction, learners will explore the cultural and situational context, they will understand the main communicative purpose, the topic and the relationships between the writer and the reader as well as the text's form and the channel of communication (e.g. written to be spoken versus written to be read). They will identify the basic rhetorical stages or moves needed to realize the genre's purpose and proceed to spot the lexicogrammatical features competent writers use to convey messages and adhere to the genre's conventions. Charney & Carlson, (1995: 90) attempt a definition of 'model texts':

> We will define a model as a text written by a specific writer in a specific situation that is subsequently reused to exemplify a genre that generalizes over writers in such situations. Such models are often used to supplement explicit guidelines or 'rules' (provided in a textbook or style guide) for spelling out some of the conventional features of the genre. (p. 90)

Several researchers support the use of model texts as an important stage in the learners' immersion in the genre (Charney & Carlson, 1995; Derewianka, 1990; Flowerdew, 1993; Hyland, 2004; Knapp, 1989; Tardy, 2006, 2009). Knapp (1989: 5) recognizes that modelling the genre may be time-consuming at first but insists that it will pay dividends when students write their own texts at the next stage and when teachers need to assess this writing. Tardy (2006: 94) observes that learners often seek out models when they are not provided.

Genre-based teaching provides explicit criteria for both learners and teachers as to what is being assessed and what needs to be improved. It is common for teachers to feel unable to justify why a text seems weak resulting in poor feedback. Genre awareness means that teachers have a conscious knowledge of what is wrong and how it should be improved avoiding unambiguous feedback. A scoring rubric based on genres such as the 'asTTle Writing Assessment Rubrics' (Glasswell et al. 2001: 17-24) or Beck & Jeffery's (2007: 66) include specific criteria for the fulfillment of a genre instead of the too general or fuzzy descriptions seen in most assessment scales which leave room for subjectivity and variance among raters. They are more 'teacher-friendly' as they define what the teachers are marking and move them away from "relying on gut feeling or professional intuition" (Glasswell et al. 2001: 12).

Genre explicit teaching has been criticised for its reliance on forms that may restrict students' creativity as well as for its 'product perspective'. If applied carelessly, it could lead to 'decontextualisation' and overgeneralization. (Charney & Carlson, 1995; Freedman, 1993). Knapp (2002) responds to criticism:

> The fears of a genre-based pedagogy producing uniform, robotic writers have been unfounded and if anything the opposite is being demonstrated. Competent writers are able to demonstrate an ability and enthusiasm to use generic structures and forms creatively and to great effect. Less competent writers use the boundaries and scaffolds provided by generic forms to write texts that fulfil the demands of the writing task. Without such structures these writers often struggle to know where to start. (p.21)

Johns (2011: 65) reporting on a survey of literacy instructors conducted at the 2009 Second Language Writing Symposium, says that most L2 instructors in EFL contexts argued for the teaching of fixed text structures. These responses together with learners' assignment results show that being 'productive' is not seen as a problem; on the contrary all interested parties involved in the educational teach-learn-assess cycle are looking for goal-oriented approaches to address real-world demands, even if teaching itself may take different forms.

It is true that overuse or over-dependence on one and only technique may result in pre-packaged products lacking character. "Genre teaching can indeed be formulaic and constraining, if genres are taught as forms without social or cultural meaning" (Devitt, 2009: 337). Hyland (2004: 19) admits that "inexperienced or unimaginative teachers fail to acknowledge variation and choice" which makes "students see genres as a how-to-do list". Charney & Carlson, (1995: 89) raise concerns about students who may "treat generic conventions as a Procrustean bed distorting their material to fit the outline rather than bending the rules" and Johns (2011: 13) commenting on the Sydney School curriculum warns that "if inappropriately presented, the eight key genres could be memorized as rigid formats, rather than as problem-spaces or open to critique and change".

It would be unfair, however, to deprive learners of such explicit step-by-step guidance just because some teachers may use it inappropriately. Learners need to see model texts as prototypes that may have differences among them not as fixed formulas to be used step-by

step and teachers need to be careful when choosing model texts for classroom use. Although in principle model texts do help learners the choice of the model text can sometimes be misfortunate. "Texts that are not clear in their purpose or which shift almost aimlessly between genres can provide poor models for student writing." (Knapp & Watkins 1994: 26).

Paltridge (2001: 69, 123) sees a challenge for teachers there; the ability to guide and direct learners while fostering independence and learning at the same time. According to him learners need to understand to what extent they might need to imitate certain genre patterns and on what occasions they might need to resist them. Johns (2015: 116) differentiates 'genre acquisition' pedagogies from 'genre awareness'. For the first "the focus is upon students' ability to reproduce preconceived text types that are organized, or 'staged,' in a predictable way" while the second is "designed to assist students in developing the rhetorical flexibility necessary for adapting their previously held socio-cognitive genre knowledge ('schemas') to ever-evolving contexts".

Most concerns related to conformity and prescriptivism refer to fiction or poetry obviously concerning first-language, rather than second-language writing. It is natural for first-language writing to be more demanding in creativity, keeping in mind that the linguistic obstacle is not that great. The kinds of topics that occur in international English language exams however, do not tend to have a creative character to such an extent and word-limits do not leave much freedom for extensive self-expression.

Hyland (2003a: 9) explains that teachers supporting the 'creative' or 'personal' approach "try to avoid imposing their views, offering models, or suggesting responses to topics beforehand". What is valued in this approach is the presence of an authorial voice in contrast to the genre-based view which emphasizes the fulfillment of a social purpose. Beck & Jeffery (2007: 74) considering the two approaches say that the genre-based view is more appropriate for academic success. Johns (1995) supporting the same view explains:

> This movement's emphases on developing students as authors when they are not yet ready to be second language writers, in developing student voice while ignoring issues of register and careful argumentation, and in promoting the author's purposes while minimizing understandings of role, audience, and community have put our diverse students at a distinct disadvantage as they face

academic literacy tasks in college classrooms where reader and writer roles,
context, topic, and task must be carefully considered and balanced. (p. 181)

Besides, creativity and innovation requires a previous knowledge of the conventions of the genre or in other words, to recognize or attempt innovation one needs to know what has been the standard, the commonly accepted product, the prototype. Writers cannot modify too many essential features fo the genre as it will not be recognizable by its audience. Genre awareness involves a sense of genre boundaries and an understanding of the extent to which these boundaries can be bent, leaving room for both convention and innovation (Devitt et al. 2004: 149; Hyland, 2004: 64; Tardy, 2015).

Even though some genre-based instruction supporters recognize the risks of teaching genres at schools and inevitably moving them from authentic contexts, they still believe that it will help learners. Devitt (2009: 341, 343) recognizes that generic traits can never be articulated even by experienced users but supports the practice of giving learners access to particular genres even when articulation is less than full. Hyland (2004: 63,64) warns that even though "genres allow a great deal of individual choice these choices are not unlimited". He remains in favour of explicit teaching of genres and stresses that second language writers are not in the best position to initiate changes and manipulate conventions. Besides, criticism based on the decontextualisation of genres fails to suggest a solution for writers from non-English speaking backgrounds who are considerably disadvantaged and deprived of natural, situated acquisition (Hyland, 2004: 17,18; Rothery, 1985: 76).

2.2.3. The suitability of the GBWI approach for this context

Apart from the reasons why genre-based-writing-instruction is suitable for second language learners explained in detail above, there are some extra features of the local context that make this approach even more appropriate for the Greek EFL classroom.

The first feature has to do with the identity of the learners. Even though they are considered second language learners this is a broad group categorization used mainly to distinguish them from native speakers. As this is the case in most parts of the world, in countries where English is the dominant language and students are usually visitors (students, immigrants) from other countries (e.g. the USA and the UK) and in countries where English is the official

language but local languages are also used (e.g. India and Kenya) the term applies to the majority of non-native English speakers/learners. Research therefore and discussion about this group has been intense and the term second language learners is often considered as including English as a Foreign Language (EFL) learners.

It is however important to distinguish these groups. For ESL learners in the initial and narrower sense the context provides sufficient input for immersion. Learners have access to the language and the culture constantly. In EFL contexts, however, learners are in their home culture attempting to learn a language "imbued with foreign cultural connotations" (Brown, 2007: 133). These students will not use English outside the classroom, they do not have the chance to listen to native speakers and their need for authentic communicative situations is obviously greater but still opportunities to participate in this kind of authentic discourses are rare. The fact that linguistic competence on its own is not enough has been discussed by numerous researchers. As Hyland puts it (2007: 151), L2 writers are often at a considerable disadvantage in such unfamiliar naturalistic settings and genre-based writing teaching can short-cut the long processes of situated acquisition. The word shortcut is also used by Johns (2003: 196), connecting students' familiarity with common genres to the successful processing and production of written texts.

The learners' young age especially as exam candidates is another factor that differentiates the Greek context. A reason behind this premature urge to participate in foreign language proficiency exams is the pressure of the PanHellenic exams (university entrance exams) for which students usually go through a three-year, senior high school intensive preparation, often combined with evening extra tuition leaving no time for foreign languages. Papaefthymiou - Lytra (2012) states:

> As a matter of fact, the practice promoted among school-age students and their parents has been for students to acquire a B2 level certificate in their lower secondary school years in order to secure a language certificate for life and 'get done with foreign language learning for good' as the popular saying goes. As a result, students as young as 12 years old may sit a B2 level exam in English, in particular, which is the compulsory foreign language in the Greek primary school system. Thus, by lowering the age entry level for such adult certificates,

students can sit for the C1 or even C2 level English certificate as young as 14 or 15 years old. (p. 24)

These exams however, especially B2 level and beyond are designed to address adult learners (Papaefthymiou - Lytra, 2012: 25). As a result, writing tasks sometimes have no relevance to young students' lives. 'Applying for a job', or 'asking for a refund' for example is something they have not done before even in their native language. These students lack background information in order to be able to build arguments about 'Drugs' for example and lack the life experience needed to describe 'the psychology of a pensioner'. Assessment in the above mentioned exams which is designed for international use will not take this particularity into account and will grade for poor content although the students may be competent in topics more familiar to them. Training young learners, especially in the writing part, to apply their judgement and perspective in topics designed for adults is a difficult activity demanding time not necessarily connected with language issues. Asking for the adjustment of this content in specific text types (e.g. letter, article), in which the writer needs to infer particular communicative purposes, relation to readers, appropriate register simply by reading the rubric is probably too much to ask from this age group. Unfortunately, studies with a focus on young ages remain rare as most research refers to tertiary education (Harklau & Pinnow, 2009: 126).

Offering 'general English' courses in this context means that the rest of the skills tested in these exams (reading, listening, speaking) need to be taught in the same course. General English mainstream coursebooks are used, combined with material that address a specific exam closer to the exam dates. Limited time and non-specialised writing courses and material for writing means that students are in need of a goal-oriented approach. When the pressure of time in this context is combined with competitiveness and effectiveness interpreted as success rates in these exams, then teaching writing becomes a tremendously demanding role and the need for explicit teaching of genres is even more urgent. Applying teaching approaches without taking in mind specific contextual factors and the particular learners' specific-purpose objectives can be unfair to the learners.

2.2.4. Genre/register analysis

Stubbs (1996: 48) talks about the necessary fiction of 'the English language' which has to be maintained in preparing dictionaries and grammars and the use of long texts as evidence of the language as a whole. Even though he supports corpus analysis he notices the conflict between theory emphasizing variability and practice where differences within individual long texts and across text types are averaged away. He therefore concludes that there is large scope both for studies of genre variation and for analyses of the internal organization of whole texts. According to Hyland (2004), 'Genre analysis' is

> a branch of discourse analysis that explores specific uses of language. It is driven by a desire to understand the communicative character of discourse by looking at how individuals use language to engage in particular communicative situations. It then seeks to employ this knowledge to support language education (p. 195).

Flowerdew (2011a, abstract) also shows the connection of genre analysis to pedagogy explaining that "good genre descriptions can feed into pedagogy in the form of syllabus and materials design".

According to the three Genre Schools (described in section 2.1.2) the focus of genre analysis may be on contextual factors (Rhetorical Genre Studies) or linguistic features (English for Special Purposes and the Sydney School). Examining the text's contextual characteristics, the RGS analysts look outside the text at factors that influenced the way it was written. They often focus on the discourse structure of genres (also mentioned as generic or schematic structure) and provide the basic 'moves' in the texts, that is, particular stages or steps considered to be necessary for the achievement of the communicative purpose of the specific genre. Knowledge of this structure in a variety of genres can assist learners' generic competence. Swales' seminal work (1990) has been the foundation for this kind of analyses. His 'Create a Research Space' (CARS) model describes the rhetorical moves typical in introductions to research articles. Sometimes these moves are divided in obligatory and optional ones. Henry & Roseberry, (2001), for example, through an analysis of the genre 'Letter of Application', find eleven moves, of which five, are thought to be obligatory while six moves, are optional. The obligatory moves are: 'Opening', 'Polite Ending', 'Signing Off',

'Offering Candidature' and 'Promoting the Candidate'. In order to see how these moves are realised they identify the lexicogrammatical features of the strategies used.

Some corpus-based studies 'tag' texts to indicate move structures while other studies use concordancing software to uncover phraseologies which relate to specific sub-sections of the text (Flowerdew, 2005: 325, 326). This tagging is mainly manually and therefore is inappropriate for large-scale corpora. However, software tools for coding move structures are becoming more sophisticated (e.g. 'AntMover'- Anthony, 2003).

This focus on the schematic structure usually excludes or gives little information about linguistic features. Flowerdew (2011a: 148) notices that even though Swales (1990) and Bhatia (1993) - coming from the ESP School - stress that the lexico-grammatical realization of the genre is an important part of the analysis, they place much more emphasis on move structure than on lexicogrammatical features. He argues that ESP analysts should balance out the more extensive work on move structure by showing more interest in lexico-grammar and should combine these approaches, "the sum of the two approaches, used in combination, equaling more than the two parts taken separately" (Flowerdew, 2011a: 149; Flowerdew, 2011b, abstract). More often though ESP genre analysts combine move analysis with a detailed analysis of the linguistic features involved (e.g., Hyland, 2000), showing particular interest in situational contexts, professional, academic or cultural. Sometimes this is combined with exploring genre practices observing genres in use and analyzing users' perceptions of what they are doing (Hyland, 2004: 209).

Analysts of the Sydney School following mostly Systemic Functional principles also refer to these stages but are more interested in breaking down texts into segments which makes it easier to associate the writer's purposes to the linguistic features used to achieve them (e.g., Glasswell et al. 2001). Researchers often join consecutive stages of a genre with the carat sign '^'. For example, according to Nesi & Gardner (2012: 100) investigating texts in university settings, the Discussion Essay has the following schematic structure: Issue ^ Alternative Arguments ^ Final Position. Both ESP and the Sydney School explore linguistic features. However, linguistists that belong to the first School usually discuss genre-specific language in grammatical terms (e.g. verbs, nouns) while the those belonging to the second School tend to use functional terms (e.g., verbs of action, attributive adjectives) for their description (Paltridge, 2001: 13).

When the purpose is to investigate variation among genres a more quantified approach is usually chosen in order to make precise comparisons. Researchers compare the use of core linguistic features in terms of frequency and see how genres are similar or different. Biber (2014: 14) insists that in register analysis we need to identify the features that are "typical" that is, features, which are "especially frequent and pervasive in some text varieties in contrast to other varieties". Biber's influential work (1988) and his powerful method known as 'Multi-Dimensional Analysis' uses statistical procedures to compare core linguistic features in texts of heterogeneous varieties of spoken and written discourse. He then associates these measurements to specific textual properties called 'dimensions' showing how various registers are placed between the measured poles in a positive-negative style (e.g., Abstract vs. Non-Abstract Information). His work more recently though (Biber, 2006), incorporates descriptive, qualitative analyses with a focus on stance, lexical bundles and vocabulary use.

Researchers may also choose to study a range of genres according to one or a small set of variables and focus on specific features. This type of analysis may refer to one genre (e.g., Hyland, 2001 - investigating self-mention in research articles; Hyland, 2009 - investigating 'engagement' in academic reports) or contrast several genres looking at how various linguistic features vary systematically according to genre (e.g., Gardner & Holmes, 2009- investigating section headings in 13 genre families of student writing; Huang, 2013- investigating lexical bundles in legal genres).

The advantage of the first type of work (analyses of individual genres) is that one can go into depth in each individual genre/register and provide a detailed description with frequent examples from concordance tables. Studies which do not involve corpora for genre analysis though have always been qualitative, based on one or a small number of text samples. (e.g. Derewianka, 1996; Devit et al. 2004; Knapp & Watkins, 1994; Martin, 1985). Their observations and explanations have been based on both knowledge and rich experience in teaching and research. Although corpus linguistics studies have offered much to genre analysis the contribution of non-corpus-based analyses has been great and has often been the foundation for later analyses.

The advantage of the second type of work (contrastive analyses) is clearly the contrastive

character of the description offering information on a range of genres/registers measured against the same metrics and methods each time. Findings from such analyses are very useful as we know little about the ways that genres form 'constellations' with neighbouring genres (Swales, 2004). The first type of analysis is more prone to subjective evaluations as to what is high and low in terms of frequency of occurrence unless the results are contrasted to a general corpus. The contrastive quantitative type of analysis on the other hand, may leave much to be interpreted by the reader who is not an expert in genre/register variation metrics. Although the numbers presented may be objective, this type of presentation has the additional disadvantage that a reader interested in one genre is deprived of the rich interpretation discourse analyses can offer.

Generally, studies that explore the rhetorical moves or stages in a genre are of the first type with more emphasis on qualitative data and an interest in individual genres, whereas studies that explore the lexicogrammatical features of genres are either largely qualitative, descriptive analyses, especially in the case of individual genres, or largely quantitative, contrastive ones. The combination of quantitative and computational techniques with qualitative interpretations however, has been rare (Biber et al. 1998: 157). It is often the case that researchers choose one or the other perspective. Stubbs (1996) encourages the combination of perspectives in genre analysis:

> However, Biber's analysis is across representative samples of genres and sub-genres, with no analysis of the discourse structure of individual instances of the genres. The most powerful interpretation emerges if comparisons of texts across corpora are combined with the analysis of the organization of individual texts. (p. 34)

Biber (2004: 53-54), comments on the distinction between corpus-based studies which investigate the linguistic characteristics of texts and those studies based on a small number of texts with a focus on the internal structure of texts from a single register. He makes that is, a distinction between studies with a different focus and a different quantity of texts on which researchers base their findings. He also observes that the combination of perspectives is rare, despite its obvious advantages. In fact, the range of methods and perspectives in genre analysis is so wide that Ferencik (2004), says that "elaboration of a fully exhaustive and

universally applicable method of text typology remains one of the most challenging tasks of text linguistics, stylistics and rhetoric".

The first and most important step in genre analysis is the identification of genres. If done carelessly it can jeopardise the results of the analysis. If texts which are not good prototypes of the genre are chosen for classroom genre analysis the results can be misguiding and difficult to interpret. The same applies when one analyses corpora. Texts under the same genre label which fail to represent the genre can produce unreliable results. Various researchers state that identifying categories of texts during the corpus compilation process is an important consideration. They also see the need for informing future users of the criteria used for this categorization. (Biber et al., 1998; Biber, 2010; Lee, 2001; Sharoff, 2015). Using corpora that are not categorized according to genres can cause problems due to the lack of homogeneity and may produce misleading findings regarding features especially those associated with style (Biber, 2006; Stamatatos et al., 2001). Despite the pressing need to classify texts in corpora in terms of genres, Sharoff (2015: 306) observes that "getting a suitable set of genre labels is surprisingly difficult. The major corpora disagree with respect to their genre inventories".

I can see three main reasons for this inconsistency. Firstly, the fact that there is still no consensus in the literature on what the terms 'genre', 'text type' and 'register' actually represent and secondly, the fact that there has been no systematic and widely accepted method of categorizing large groups of texts yet (Lee, 2001; Passonneau et al. 2014; Sharoff, 2015; Stubbs, 1996). The third reason has to do with size. The advantage of a large set of texts in analysis as opposed to one or two texts has been one of the strongest arguments in favour of corpus linguistics for years. There is a belief that the bigger the corpus the better. This is because when there is more data the researchers can be more confident about their findings. Especially statistic results are more reliable when based on a large corpus.

It has been shown, however, that specialised corpora can be much smaller and that foreign language researchers tend to use smaller corpora, which are easier to compile and analyse but designed according to strict criteria and created for specific research (Flowerdew, 2005; Pravec, 2002; Henry & Roseberry, 1996; Ooi, 2001; Tribble, 2001; O'Keeffe et al. 2007). Flowerdew (2004), says there is no ideal size for a corpus; it all depends on what the corpus contains and what is being investigated. When characterizing corpora in terms of size, she notes that there is general agreement that small corpora contain up to 250,000 words. Biber

(1990), also supports smaller corpora representing the full range of variation as opposed to larger general corpora when the focus of analysis is text variation. Besides, general corpora do not always contain complete texts but excerpts, which makes it impossible to examine the position of certain words in the text and their role in the overall discourse structure (Hanford, 2010). Adding tags to enrich genre analysis, a manual procedure feasible only on small corpora, (Flowerdew 1998) and the fact that in these corpora the analyst is "probably also the compiler and does have familiarity with the wider socio-cultural dimension in which the discourse was created" (Flowerdew, 2004: 16) are additional advantages of specialized corpora for genre analysis stated by Hanford (2010: 259).

Closely related to the issue of 'size' is the issue of 'representativeness'. It "refers to the extent to which a sample includes the full range of variability in population" (Biber, 1993). Representativeness is a critical factor of a quality corpus (Biber, 1993; O'Keefe et al., 2007; McEnery & Wilson, 2001) and especially in the case of specialised corpora, size becomes a secondary issue (Lee, 2010: 114). Sardinha & Pinto (2014: xix), complain about some of the researchers' choice to disregard representativeness saying that "Biber's (1993) recommendations have largely fallen on deaf ears".

So, the reasons why corpus compilers do not classify texts according to genres have to do with the lack of consensus in theory, practical concerns when the case is big general corpora, and ignorance about the importance of size when the case is small specialized corpora.

Lee (2001: 37) argues that "genre is the level of text categorisation which is theoretically and pedagogically most useful and most practical to work with". Classifying texts according to text type though can allow for a much greater number of texts to be included under the same label while using genre classification would create many small sub-corpora. This could be understood if for example one starts thinking how many generic categories would result under the text type label 'Letters'. This category may include different genres such as 'Complaint Letters', 'Letters of Application', 'Letters of Advice' to name just a few.

Another way to bypass these complications is to classify texts according to domain, that is, according to big and broad categories related to topic. This kind of classification though permits such broad and varied categories of texts to be included in a corpus or sub-corpus

that confines the potential types of analyses later on and is certainly unsuitable for genre analysis.

The impact of genre has gained such attention recently that researchers coming from different scientific areas work actively towards genre recognition. Although in corpus-based studies with a pedagogical focus the genre recognition process is usually a part of genre analysis there is some work that has focused on genre recognition. McCarthy et al. (2009) for example, investigate reading ability and its relation to genre recognition. An array of scientists not necessarily involved in education is actively engaged in automatic genre identification/ recognition. The idea is to train computers in order to recognize and classify texts based on statistical procedures and pre-determined genre features (e.g. Passonneau et al. 2014; Stamatatos et al. 2000, Stamatatos et al. 2001). One of the most common applications of this type of studies has to do with web genres. The need has sprung mainly out of the vast amount of information the Web has offered us and the subsequent demand for quick and reliable results when web users perform searches (e.g., Santini, 2006) or the need to collect and annotate corpora form the web based on generic categories (e.g. Sharoff et al. 2010).

Despite the difference in the purpose of these studies, progress in genre identification/recognition can only improve genre analysis and the results from these seemingly varied fields are, in fact, mutually dependent and largely interrelated.

2.2.5. 'Genre', 'text type' and 'register': clearing up terminology

In this section I try to clarify the confusion among the terms 'genre', 'text type' and 'register' reviewing their use in the literature. There is naturally the danger of overgeneralizing in an attempt to describe convergence and divergence but this interpretation may help clear the fuzziness in this area.

2.2.5.1 'Genre' versus 'text type'

Most researchers referring to genre do not mention the term text type (e.g. Halliday 1978; Halliday and Matthiessen 2004; Nesi & Gardner 2012) or use the two terms interchangeably (Stubbs 1996). Those on the other hand who do use both terms and make a distinction do not share the same basis for the differentiation of the terms.

Biber (1988, 1989) sees genre as defined and distinguished on the basis of systematic non-linguistic criteria and text types on the basis of strictly linguistic criteria, that is, similarities in the use of co-occurring linguistic features. For him, text types are groupings of texts that share linguistic features irrespective of genre. Based on this conceptual framework he has found that the same genre can differ greatly in its linguistic characteristics and that different genres can be quite similar linguistically. He has thus, come to the conclusion that "genre distinctions do not adequately represent different text types". Lee (2001: 40), commenting on Biber's multi-dimensional approach says that this classification "is at the level of individual texts, not groups such as 'genres', so texts which nominally 'belong together' in a 'genre' (in terms of external criteria) may land up in different text types because of differing linguistic characteristics" and that "these typological categories should be taken as indicative rather than final".

Paltridge (2001: 63, 123) defines the term text type as patterns of discourse organization that occur across different genres such as 'description', 'narrative', 'instruction', 'explanation' but later on he refers to a 'letter', a 'story' and an 'advertisement' as genres too.

Knapp & Watkins (1994) link the term genre to language processes such as 'describing', 'explaining', 'arguing' and the term text type to texts seen as products or things such as 'reports', 'expositions' and 'stories'. They encourage teaching genres as processes, rather than products as the generic features remain consistent and can be applicable to all text types written by students. From this perspective commonly used text types often deploy several genres. They support that such an approach has no problem with multigeneric texts.

For Glasswel et al. (2001), genre is driven by functional purpose whereas text type is affected by mode (text form). They point out that the purpose is able to change even if the type of the text remains the same and uses 'letters' to explain:

> Letters may be written to make complaints, to argue a point, to recount an event, to make an explanation, to tell an anecdote, or to advertise a product. In short, letters may have different purposes and, thus, the structuring of these texts and their lexicogrammatical resources will differ significantly, regardless

of the fact that each will still be considered a letter in terms of layout and transmission. (p. 2-3)

Even though this view differentiates text type from genre it gives prevalence to the term genre (seen as functional purpose) instead of text type (seen as text form). It is therefore quite different from Biber's distinction and his preference in studying text types irrespective of genre.

Cummings (2003: 194), sees text types as components of genre. He labels 'narrative', 'description', 'exposition', 'dialogue' and 'monologue' as genre categories and 'novel', 'travel brochure', 'article', 'conversation' and 'oration' as text types.

Huang (2013: 147-154) refers to the term register as the greater grouping name of text varieties used by other researchers (e.g. conversations, academic writing, academic spoken language) encouraging future research on lexical bundles between specific genres for "a more fine-grained picture" of lexical bundles. Although the study is interested in language varieties rather than genre or register variation it appears that she considers register a broader category than genre. However, in other parts of her thesis she uses the terms interchangeably. The following extract is indicative of the additional terminological confusion between genre/mode: "It compares multi-word combinations across different <u>genres</u> of English dialogues. The investigation aims to describe the linguistic characteristics of lexical bundles in two <u>modes of spoken data</u>: <u>private dialogues and public dialogues</u>". (p. 4)

What is obvious here is that there is no consensus on the term text type especially regarding its relation to genre. Stubbs (1996: 12), looking back at categorisations that have been proposed based on text types, says that "none is comprehensive or generally accepted". Paltridge (1996: 237) notices that "the terms 'genre' and 'text type' seem to have been conflated with the term 'genre' being used to include both of these notions." Lee (2001: 41), says that the term text type remains an "elusive concept" and that it seems redundant to have two terms which cover the same ground.

2.2.5.2. 'Genre' versus 'register'

For Halliday (1978: 111, 31, 32, 35) "a register can be defined as the configuration of semantic resources that the member of a culture typically associates with a situation type. It is the meaning potential that is accessible in a given social context". A situation type is characterized by three factors: what is happening, who is taking part and what part the language is playing. These three variables are called 'field', indicating the type of social action, 'tenor' referring to role relationships and 'mode', denoting the symbolic organization. According to him these three variables, taken together, determine the 'register', that is, "the range within which meanings are selected and the forms which are used for their expression". Exploring register means attempting "to understand what situational factors determine what linguistic features". Later on, he defines register as "a syndrome of lexicogrammatical probabilities" (Halliday, 1992: 68).

Derewianka (1996: 47) also sees register as the configuration of field, tenor and mode and associates genre with purpose. She sees the notions of genre and register as inseparable and considers an awareness of the genre as the basis for the prediction of the overall organization of the text (stages) and an awareness of the register as the basis for the prediction of the language features that generally characterise such a text.

Martin (1993: 156) sees genre as a layer above register and as encompassing register. He says that "genre is a way in; it works to raise awareness, and it works in a way which register analysis alone had not been able to work before". Thompson (2014), sees genre as register plus communicative purpose and gives an image to show the difference between register and genre. He suggests that we see register as cloth and genre as garment:

> the garment is made of an appropriate type of cloth or cloths, cut and shaped in conventional ways to suit particular purposes. Similarly, a genre deploys the resources of a register (or more than one register) in particular patterns to achieve certain communicative goals. (p. 52)

Biber (2006: 11), explains that he uses the term register as "a general cover term, with no implied theoretical distinction to genre". In a detailed description of the term register, Conrad & Biber (2001: 3) say that "register distinctions are defined in non-linguistic terms, including the speaker's purpose in communication, the topic, the relationship between speaker and hearer, and the production circumstances". This way of identification sounds very similar to the Systemic Functional Linguistics approach which sees register as the configuration of field, tenor and mode but it includes the notion of purpose which in the SFL perspective is associated with genre. So the basic distinction between the Hallidayan approach and that of Biber's is the consideration of the text's communicative purpose in the first approach as the basic criterion for defining genre whereas in the second approach purpose is one of the criteria determining register. For SFL linguists, genre is a greater notion that encompasses register whereas in Biber's perspective no theoretical distinction is made between the terms and the term register is often preferred.

In practice, however, Biber's view of text classification based on register does not necessarily involve communicative purpose as a criterion. Conrad & Biber (2001: 3) distinguish between a specialized register "corresponding to the extent to which the register is specified situationally" and a general register in which texts "tend to exhibit a wide range of linguistic variation". This is a way of admitting that in the second case, register is not specified situationally. And this is very clear as for example in register categories such as 'conversation' or 'newspaper language' there are texts with all sorts of different topics, relations between speaker and hearer as well as purposes which will certainly cause great linguistic variation within the text category. Such a broad grouping of texts can give us the basic differences between conversation and newspaper language but it hides important differences between the 'news article', 'the advertisement' or the 'obituary' for example. The absence of communicative purpose as a criterion for grouping texts in such broad register categories can make the interpretation very difficult later on as researchers will need to define the contextual factors causing variation among texts.

Having seen the use of these terms by prominent researchers I conclude that the basic difference between researchers who usually investigate broad register categories with those who analyse genres is not on the criteria that define genre or register. It is rather a different ordering and sense of priority in the procedures. Those who refer to genres place more emphasis on defining the precise nature of genre and its sub-genres by studying closely the

contextual factors (communicative purpose, field, tenor, mode) and grouping texts in much more consistent categories which in corpus linguistics studies may result in smaller corpora but with far greater chance of bringing the typical features of the specific genre into light. When only purpose and structural organization are explored then the term genre is commonly used. When linguistic features are also investigated or are the main focus of the study both the terms genre and register are used in the literature.

Researchers who use the term register throughout their study usually start in a less preoccupied manner concerning the initial grouping of the texts, compiling easily much bigger corpora, which then bring into light the predominant features of each broad category. This can easily highlight the differences between oral and written speech for example but can also raise a lot of questions about the variation within each of the categories. The features of subordinate categories, what Conrad & Biber (2001) call 'specialized registers' or what others prefer to call 'sub-genres' or 'sub-registers' are not presented in these cases. They are studies with a different scope, broader and more general in nature.

In line with Martin (1993) and Thompson (2014), I see genre as encompassing register where the communicative purpose together with field, tenor and mode determine the overall structure of the text, what is going to be written or discussed, affected by the relations between the writer/reader or the speakers, organised in a suitable form in order to achieve its purpose. I see the notion of text type as related to mode, denoting text form, in line with Glasswell et al. (2001), and thus as one of the variables determining register. Essays or Letters are text types in this sense, not genres.

In this view, genre and register are inseparable (Derewianka, 1996; Finegan & Biber, 1994) but genre can be studied on its own if the case is investigations of purpose and structure. Register studies on the other hand have to consider genre (in its traditional sense, associated with purpose) in the choice of texts to be explored. In agreement with Thompson's example previously mentioned it is the garment (genre) that will determine which type of cloth is suitable (register).

CHAPTER THREE

**TEACHING AND TESTING SECOND LANGUAGE WRITING: THE LEARNERS'
VIEWPOINT**

3.1. Methodology

To find out more about learners' views on writing for testing purposes I designed a short questionnaire. It included mostly closed answers and was brief and anonymous in order to be easily accepted by teachers and students. Because of the students' young age keywords in the questions were written in bold and phrases of guidance on how to answer the question were underlined to avoid confusion. It was handed out to students of five secondary schools in Rhodes, Greece, during the first months of 2015. Both the questionnaire and the analysis refer generally to language certificates and testing bodies, avoiding specific names and comparisons between specific certificates. Teaching practices, learners' needs, difficulties and preferences are explored and linked to learners' age, proficiency level or gender. The following seven questions were posed to learners:

1. In English as a foreign language exams or during your preparation for them **which part is the most difficult for you**? Rate with 1-4 (where 1 is the most difficult)

a. Reading Comprehension _____

b. Listening Comprehension _____

c. Writing _____

d. Speaking _____

2. During your English classes (in total) **how much time was devoted (approximately) to preparation for 'Writing'**? Circle only one answer:

a. 1/3 of the total time, b. 1/4 of the total time, c. 1/5 of the total time, d. 1/6 to 1/10 of the total time, e. almost no time at all

3. Out of the following text types commonly asked in English language exams **which is the hardest for you**? Rate with 1-6 starting with the hardest (1). You need to use all numbers.

a. Formal letter, b. Informal letter, c. Essay, d. Short story, e. Report, f. Review

4. How much time was spent for your preparation for each of the following **text types**?

Use the letters A-D where A= Enough, B= Some, C= Little, D=No preparation

a. Formal letter, b. Informal letter, c. Essay, d. Short story, e. Report, f. Review

5. Which of the following is **the hardest for you** when you are being tested in **'Writing'** in the English language? Circle only one.

a. Vocabulary and appropriate phrases

b. Grammar/ syntax

c. Content (ideas)

d. The word-limit

e. Understanding the question

f. None of the above

6. When you write (exams or classroom) there is a time limit. Does this **stress** you and **in what way**? Circle only one answer.

a. not at all, b. slightly, c. moderately, d. substantially (it affects the quality of my writing negatively), e. seriously (it affects me so much that I do not manage to complete the task)

7. During the **teaching** of writing **I would like**:

 Circle as many as you wish.

a. more time in general

b. to know the assessment criteria beforehand

c. to write more often for practice purposes

d. to be given 'model answers'

e. feedback on my assignments to be more detailed

f. some discussion on the topic of the task before writing

g. Something else: ………………………………….

Learners were also asked to fill in some personal data such as their age group, gender, CEFR level and whether they already had a language certificate stating the CEFR level of the highest one in case they had more than one. Finally, they were informed in writing that the questionnaire was part of ongoing research at the local university. The sample included 600 students, 268 were male and 332 female. 389 of them belonged to the first age group (12-17), 186 to the second (18-24) and only 25 were older than that. Data was stored and analysed using Microsoft Excel. Results are presented in tables and figures to facilitate comprehension.

3.2. Results

Sixty-five percent of the students stated that they had already been certified, seventeen per cent of them stated that they had not yet gained a certificate in English and eighteen per cent of them preferred not to answer the question. Figure 3.1 shows the levels of language certification acquired by the participants. Almost half of the students had a B2 level certificate, nearly a quarter of them had a C2 level certificate, one eighth of them had either a C1 or a B1 level certificate and only a small minority stated they had a certificate of a level lower than B1.
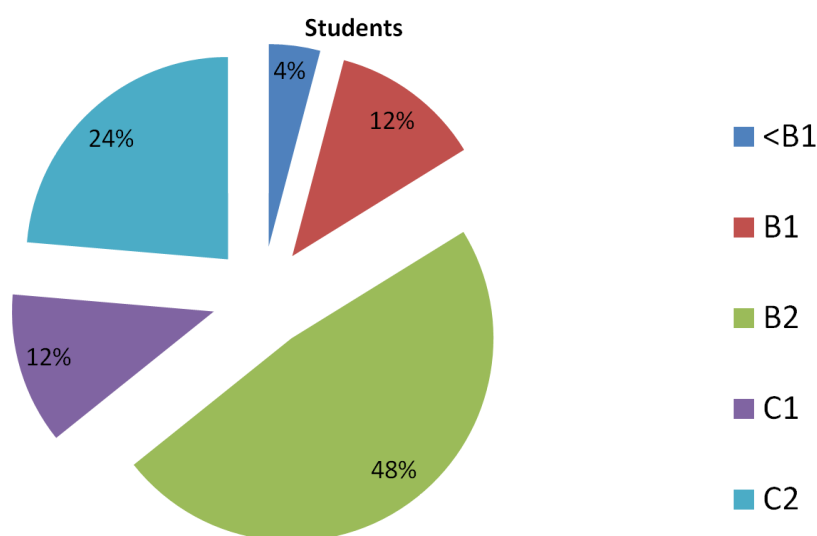
Figure 3.1 The CEFR levels of the certified students

As the majority of the students belonged to one of the first two age groups their answers were contrasted to these age groups in order to see if there were any indications of preferences concerning age and proficiency levels. Table 3.1 shows that B2 and C2 levels have attracted the majority of the students in both groups. B2 is the first choice and C2 is the second in both groups. C1 and B1 certificates as well as those that are lower than B1 level have had fewer participants in the second age group. The data show that there is an upward trend for certification at these levels in younger students.

Table 3.1 The CEFR levels of certified students in two age groups

_____

CEFR level	Age group A (12-17)	Age group B (18-24)
<B1	4%	1%
B1	19%	8%
B2	43%	59%
C1	13%	9%
C2	21%	23%

_____

In the same way, the CEFR level of the certified students was seen in relation to gender but results show that there is not much difference in any level between male and female students.

In the first question the students were asked to rate the four skills (Reading Comprehension, Listening Comprehension, Writing, and Speaking) in terms of difficulty. Writing was the most difficult skill for almost 42% of the participants. This percentage was far bigger than any other percentage.

In table 3.2, one can see the profile of the students who rated Writing first. Regarding gender, male students were more than female. The fact that the initial sample was not balanced in respect to gender and that female students were more than male ones can perhaps raise the gravity of this finding. Overall, almost half of the students who rated 'writing' first were male students while thirty-seven per cent of them were female. The rest refers to answers were information about gender was not provided. In the first age group nearly half of the students rated 'writing' as the most difficult part while in the other two age groups the percentage was smaller (32% for each category). Then each age group was broken down in gender categories showing that male students outnumbered female ones in every age group. It was fifty per cent for males compared to forty-three for females in the first age-group, forty-four per cent for males compared to twenty-nine for females in the second age group and forty-two per cent for males compared to thirty-eight for females in the third group.

Table 3.2 Students who rated Writing first in terms of difficulty

Gender		Age-group		
M	F	A (12-17)	B (18-24)	C (25-)
49%	37%	46,5% [M: 50%, F: 43%]	32% [M: 44%, F: 29%]	32 % [M: 42%, F: 38%]

'Writing' therefore is considered the most difficult skill by the majority of the students and this difficulty occurs mostly between twelve and seventeen years of age rather than in the older age-groups. In addition, this difficulty preoccupies mostly male students in every age-group. Although gender differences were not evident previously in relation to candidates' certification in the various levels, it seems that looking at 'writing' in particular, male students have more difficulties than female ones. This must mean that boys supplement this weakness with higher achievements in other parts or that they work harder during preparation and eventually minimise this difference. This leaves room for further investigation on gender and writing competence.

Concerning classroom preparation time, the majority of the students (37%) stated that they had spent 1/3rd of the total classroom time on writing preparation. Almost thirty per cent of them chose 1/4th of the time and nearly twenty per cent chose 1/5th of the time. Of course this has been a rough estimation by students as they have taken courses in several contexts (schools, private language centres, one to one lessons) but it shows that according to them time has been allocated fairly compared to the rest of the skills. However, there was an eight per cent that marked the answer 1/6 to 1/10 of the total class time and six per cent for which almost no time in the classroom was spent on writing preparation. Figure 3.2 shows the answers concerning time spent on writing preparation.
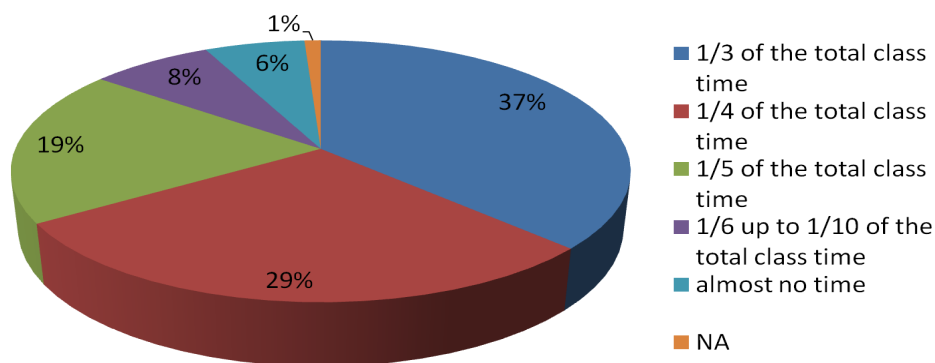
Figure 3.2 Classroom time spent on writing preparation

The next two questions referred to the text types that candidates are usually asked to write in these exams. Students were asked to rate these text types in terms of difficulty. According to their answers, the Review is the most difficult text type, followed by the Formal Letter. The Report is also considered difficult as it gathers most answers at the second level of difficulty. The Short Story has most answers at the sixth level which means that most students do not find it difficult. The Informal Letter has most answers at levels five and six. It is interesting how the two Letter types are seen as completely different in terms of difficulty. Concerning the Essay, the answers are spread across levels of difficulty showing that students have various views. The biggest number however, is at level four which shows that most students see the Essay as a text type of medium difficulty. This could also mean that Essays are a big group name for a range of texts with varied levels of difficulty. Table 3.3, shows the dispersion of answers for each text type. The biggest value for each text type is in **bold** and the second biggest value is in *italics*.

Students were also asked to estimate the time spent on preparation for each text type in the classroom. As shown in table 3.4, the majority of the learners consider the time spent on the first 4 text types – Formal and Informal letter, Essay, Story - to be enough. They seem to have spent some time in preparation for Reports and little time on Reviews. It is interesting that answers for Reviews are dispersed across the four choices and it is the only text type that got a high number of answers for No preparation. This coincides with answers in the previous question where most learners found Reviews to be the most difficult genre to write. It is also very interesting that according to the learners enough time has been spent on preparation for

formal Letters but it is still seen as a difficult text type. In table 3.4 the biggest value is marked in **bold** and the second biggest value is in *italics*.

Table 3.3 Rating text types in terms of difficulty
*(1 stands for the most difficult one)

Text types	1*	2	3	4	5	6	Total answers
Formal Letter	**122**	*103*	83	80	76	51	
Informal Letter	24	31	69	100	**148**	*143*	
Essay	82	*90*	88	**105**	81	69	515
Story	47	61	84	79	*89*	**155**	
Report	67	**141**	*122*	87	67	31	
Review	**174**	*88*	68	62	54	69	

Table 3.4 Classroom preparation time for each text type

Text types	A=Enough	B=Some	C=Little	D=No	Total
Formal Letter	**266**	*198*	77	41	
Informal Letter	**302**	*168*	77	35	
Essay	**251**	*177*	97	57	582
Story	**201**	144	*145*	92	
Report	*155*	**184**	145	98	
Review	129	139	**168**	*153*	

Figure 3.3 shows learners' main difficulty when tested in 'writing'. Grammar/syntax is the first problematic area (31%) with vocabulary/appropriate phrases being close (28%). They were given the choice to answer 'none of the above' and this was chosen by sixteen percent of the students. Content came fourth in their choices (11%) and word-limit seemed to be a problem only for nine per cent of the students. Only five per cent stated that understanding the question was their main difficulty.
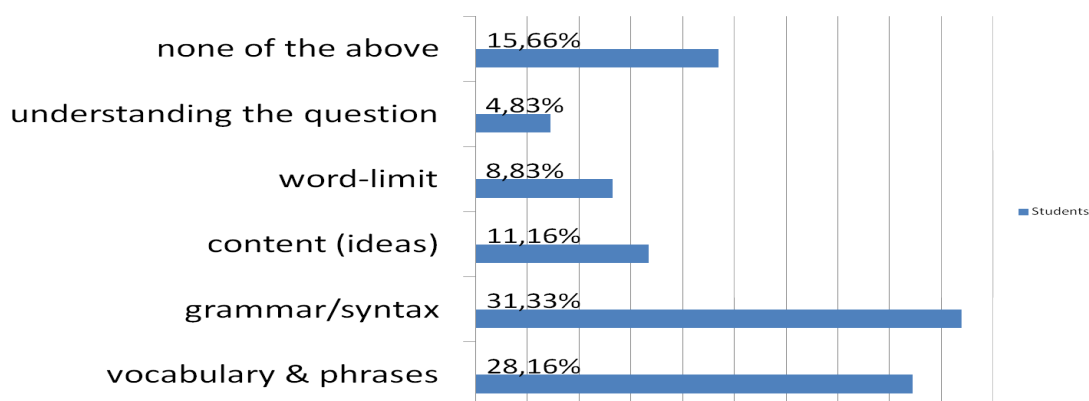
Figure 3.3 Learners' main difficulty when tested in Writing

Then the main difficulties in levels C1 and C2 were studied more closely in order to check if they remained the same at higher levels. It is interesting here that none of the above was the first answer (29,5%), which shows that problems are not so acute at this level and that grammar/syntax (20%) was not higher than vocabulary and phrases (26%) as was the case for the whole sample.

The answers to the question concerning the time limit when writing either in the classroom or the exams show that almost thirty-five per cent of the students are affected moderately and thirty-three of them are affected slightly. Students were specifically asked whether time affected the quality of their writing, or worse, if they had problems completing the task because they did not have enough time. Seeing the numbers (figure 3.4), this is not true for the majority of candidates. For almost 14% of them, time affects the quality of their writing but only six per cent state they have to hand in incomplete texts.

Finally, the students were given a set of choices in order to show what they would like to have more of in the writing classroom. Here they could choose more than one answers so the percentages refer to the total positive answers for each option. As we can see in the following pie (figure 3.5), the need for model answers is first (22%) while the need to write more often together and the need to participate in discussions concerning the specific topic before writing come second with 19% for each option. More time in general is the next preference with 17% followed by the need for more detailed feedback on their writing with 14% of total answers. It is interesting that knowing the assessment criteria beforehand is not really

important for students or perhaps we could assume they do not understand what this means. We can also assume that this is so because most of them have not discussed the assessment procedure in classrooms. In this question, students had an open choice in case they wanted to add anything else. Only 8 students chose to write something there. This is what they asked: no word-limit (3) - more thematic vocabulary (1) - more help with syntax (1) - clear topic questions (1) - help with structure (1) - some first ideas for development (1).
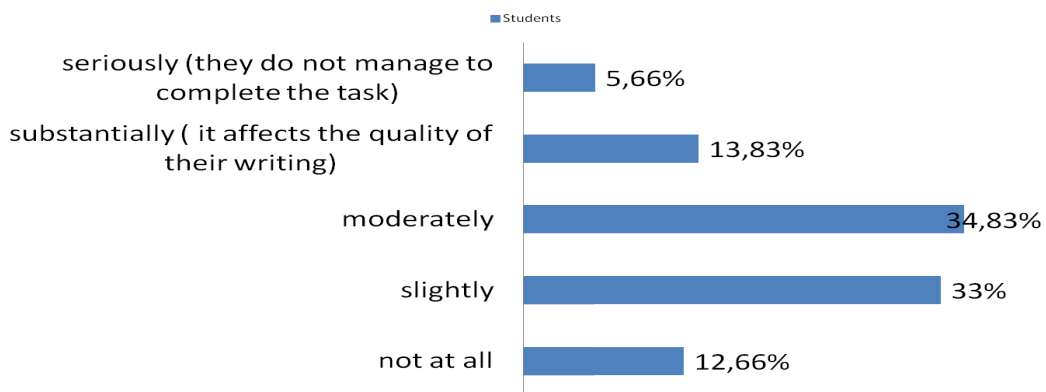
Figure 3.4 Time limit affecting writing in language testing
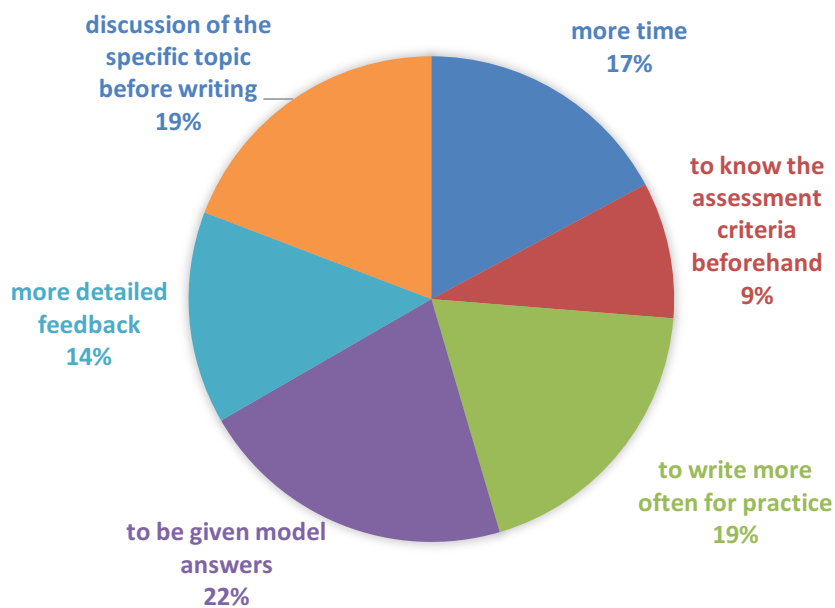
Figure 3.5 Learners' needs in the writing classroom

CHAPTER FOUR

THE REPRESENTATION OF TEXT TYPES AND GENRES IN ENGLISH LANGUAGE MATERIAL

Corpus linguistics combined with Genre analysis is the methodology chosen for questions two, three and four. To investigate genres in writing tasks of well-known exams a large database of texts representing this context is necessary. In order to understand the contextual factors of each text the database should include the rubrics of the writing tasks as well. This will help determine the generic category for each text and see how text types are interpreted in genre categories (question 2) but will also form the basis for textual analysis in order to study the typical features of each genre (question 3) as well as the relations among them (question 4).

4.1. Methodology

As the aim of this chapter is to find out what the representation of text types as well as genres is in English language teaching material a corpus consisting of such texts is built and then assessed for its representativeness. First, the stages of corpus building are described in detail, explaining the criteria on which the collection of data has been based and the way it has been annotated (a stage needed for genre analysis in chapters 5 and 6). Then, the process of genre identificiation adopted in this study is described explicitly (drawing on theoretical frameworks reviewed in chapter 2). The range of genres identified through this process are presented as parts of the initial text type categories in order to evaluate the number of genres included in each category as well as their representation in the corpus as a whole.

4.1.1 The WriMA corpus building and processing

Stimulated by the impact of the 'modeling and deconstructing' stage of the Sydney School, I decided to use samples of model writings which are actually used in EFL classrooms. A pedagogical corpus consisting of writing model writings from published course books and educational websites specifically targeting international English exams was considered ideal for studying genre issues in a specific context based on large-scale data. Since such a specialized corpus was not available I had to build one especially for this study.

Rizzo (2010: 21), reporting on experience from the compilation of a specialised corpus, notices that there are some gaps in the literature concerning the guidelines for the compilation of such corpora. It is true that specialised corpora have all sorts of different end-purposes and it is natural for their creators to follow their own steps but it is because of this uniqueness that the design criteria of each corpus have to be stated clearly in every study. After all, the kind of data and the way it has been collected is the basis of every corpus-based study and has great impact on the validity of the results. For this reason, detailed information is given here concerning the criteria set for the inclusion of texts as well as the steps followed for the creation of the WriMA corpus (Writing Model Answers).

4.1.2 Criteria for data collection

4.1.2.1 Representativeness

Christie (1993) uses the term 'curriculum genres' and Smedegaard (2015) talks about 'examination genres'. Although some of these may be similar, the range of the genres investigated here are not part of a state school curriculum. They are chosen by international examination bodies and affect teaching in various contexts such as schools, language centres, one-to-one tutoring. Therefore, a better naming of the genres explored in this study would be 'EFL examination genres' as a sub-category of the larger group of 'pedagogic genres' which includes the previous terms.

The WriMA corpus was created because the texts needed for this research were specialised. For this reason, any material not strictly connected to the exams studied here was excluded. This material may have been model answers of different institutions and examinations, such as secondary or tertiary curriculum related work. In these different contexts, for example, text length limitations would be completely different and CEFR level categorisation non-existent. In this corpus, texts are marked for their relation to the Common European Framework for Languages (CEFR) levels. I have included texts targeting levels B1 up to C2, at least as far as their publishers claim so, as these are the most intensively tested levels and because work on the writing skill is not very common in earlier stages. This information is included in the metadata of the corpus.

Representativeness was also a criterion for the type of texts to be included. A text type category, for example, needed at least twenty sample texts to be representative. As some text types were easily found and other types were rare in this material this cut off point was set in order to keep looking for more sources in case a category was not sufficiently represented. Model writing answers were collected from widely used English language textbooks and educational websites addressing international English language examinations. As representativeness is a critical factor of a quality corpus (Biber, 1993; O'Keefe et al., 2007; McEnery & Wilson, 2001; Sardinha & Pinto, 2014), I tried to collect writing answers from as many sources as possible. This prevents the data set from being affected by the idiosyncrasies of specific writers. In total, 93 different sources have provided the content, with 56% coming from the Web and 44% from printed books.

4.1.2.2 Model answers

Educational websites are a useful resource especially for learners who desperately need 'free' help. However, because of the constantly growing easiness connected to the creation of a website today one needs to be careful on the quality of advice and the expertise of the creators. Right at the start of this procedure I noticed that a few of the websites name 'model writing answers' texts that have been sent by various learners which of course is a good source for learner performance but cannot be called a 'model'. This type of data was disregarded. These texts are usually called 'sample answers' and one has to be careful of this distinction. Texts from the web included in this corpus have been written or modified by teachers, material writers or exam raters, according to the information given in the website. The testing institutions' official sites were given priority but it seems that as a general policy these sources give away samples of learner writings in order to explain the assessment process. In those cases, I only included texts with top scores in the assessment scale.

Gathering texts from books was easier from that perspective as they only provided 'model' answers. It is interesting though that none of the textbooks used, provided any learner material for feedback on errors or any advice on assessment criteria related to specific samples the same way as some websites do. Even though this kind of information was not needed for this study, this observation is interesting as it gives the web as a source for pedagogic material the advantage of more practical help and more up-to-date information based on recent applied linguistic findings on the exploitation of learner material.

4.1.2.3 Content and size

Size is not the priority in specialized corpora where representativeness is a more crucial issue (as explained in chapter 2.2.4). According to Biber (1990) at least 1000 words spread across at least five different samples per genre is regarded adequate for detecting basic linguistic properties in a genre. This corpus exceeds this size by far consisting of 1151 texts in total with no less than 24 texts per category and texts ranging between 80-300 words approximately depending on the category. However, since the following analysis would involve new categories after the genre-based classification the same criterion was set for the new text categories. The sub-corpora investigated in detail in this corpus (after the naming and classification of texts- chapter 6.2), are within these limits with the smallest consisting of 10.000 words (across 61 different text samples) and the largest consisting of 47.153 words (across 176 text samples). The final corpus has 253.025 tokens broken down in nine sub-corpora (initial text type categories). As there is general agreement that 'small corpora' generally contain up to 250,000 words (Flowerdew, 2004) this corpus is at the border and could be called a large corpus of the 'small corpora' category. The corpus content as initially classified is presented in table 4.1. At this stage the interest is in the number of texts in each category in order to measure representation in the material. Furthermore, the initial categories (based on text types) will break into new sub-categories (based on genres), so the number of tokens is not presented here but later on when the sub-corpora based on genres are analysed.

Table 4.1 The WriMA corpus content (initial classification according to text types)

Text categories	Number of texts	Text categories	Number of texts
Essays	415	Reports	176
Formal Letters	171	Informal Letters	105
Articles	88	Stories	78
Informal Emails	55	Reviews	39
Formal Emails	24	Total texts: 1151	

4.1.3 Corpus compilation and annotation

According to Leech (1998: xvii), "The compilation of a corpus (with proper attention to quality, design criteria and so on) always takes twice as long as one thought, and sometimes ten times as much effort". Apart from the strict criteria set for data collection, the compilation has been time-consuming because of the state of the material. When the texts are already in electronic form, for example material from the web, it is far easier to compile a corpus. In this case, almost half of the texts were collected from books and texts were either keyboarded or scanned using OCR (Optical Character Reader) software (MP navigator EX, included in Canon MP550 scanner). All texts were then converted to plain text (.txt) format.

Manual annotation for the purposes of the specific study was employed before storing the texts. This concerned marking:
a. text and paragraph borders
b. headings and sub-headings
c. Greetings (in letters)
d. proper names (human names, location names)
e. nationalities
The first three elements were chosen for their contribution to the special structure and layout of specific genres. By grouping proper names (substituting proper names with <name>), it is possible to count the occurrence of these categories as a group and notice significant use or lack of use in a genre without specific names coming up as key vocabulary in a genre. Although this manual tagging has been time-consuming, it can significantly contribute towards a detailed description of a genre.

Each text was given a four-digit number and the initial letters of the category; ModRep0058, for instance, is the 58th text in the Model_Report sub-corpus. Then, the corpus was POS (Part-Of-Speech) tagged with TagAnt 1.1.2 (Anthony, 2014), built on Tree Tagger developed by Schmid (1994), relating words to their grammatical class (for a sample POS tagged text see appendix 3; the tag set used is provided in appendix 4). The added information facilitates analysis both at word-level as well as at text-level. At the word-level the researcher can differentiate words used for example, both as nouns and as verbs (e.g., look, study, increase). At text-level this tagging offers the ability to look at grammatical categories as a whole (e.g. the extent of noun/adjective/pronoun use) or calculate ratios (e.g. Noun/Verb ratios) and

compare them to other genres. It also enables the researcher to obtain derivative statistics such as the lexical density of the texts (Ure, 1971).

Metadata considered important for external contextual information was stored in Microsoft Excel spreadsheets. This corpus metadata included information on the CEFR levels, the rubric, the targeted examination and the source (for a sample see appendix 2).

4.1.4 Genre identification, classification and labelling

I follow the Systemic Functional perspective in defining genre as expressed by Martin (1993) and Thompson (2014), where genre encompasses register and where register refers to the linguistic choices with respect to context. The first precondition for the view to be described is that changing one contextual parameter can indeed affect register. Therefore, we cannot describe register disregarding genre and its contextual factors. The second precondition is that linguistic choices need to be justified not just presented quantitatively and to achieve that we need to have previously understood the context of use.

I base genre identification mainly on functional purpose (Martin, 1985) taking into consideration the main purpose in case of multi-generic texts. Swales (1990: 10) reminds us of the importance of communicative purpose: "It is communicative purpose that drives the language activities of the discourse community; it is communicative purpose that is the prototypical criterion for genre identity, and it is communicative purpose that operates as the primary determinant of task". The importance of 'purpose' in the identification of genres is stated clearly also in recent and enlightening studies with a pedagogical scope where genre classification has been implemented for further analysis of the language (Nesi & Gardener, 2012):

> Whilst reading and re-reading the assignments, we looked for statements of assignment purpose which might be found in abstracts, or in introductions and conclusions; headings and subheadings were useful in extracting assignment skeletons or macrostructures (Gardner and Holmes, 2010) to inform the grouping process. The first and last sentences of each section and paragraph gave a rough idea of the content of that section / paragraph and could be quickly skim-read, and reading and re-reading the texts enabled us to determine

the purpose and stages of what had been written, and to recognize it in other texts. (p. 33-34)

It is implied that this procedure requires time, knowledge and experience. They however, extract this information from texts rather than the prompts, which in this case has only been applied only where the prompts did not offer all the clues needed.

Then I investigate the 'ideational', 'interpersonal' and 'textual' metafunctions to understand the context (Halliday & Hasan, 1985). As the corpus metadata includes the prompts all this information is retrieved without the need for reading the text itself. The prompt also defines the text type, asking for example, specifically for an 'essay' or a 'letter'. There is however, need for reading individual texts where the prompt leaves choice as to the development of the texts. This may happen for example in some argumentative essay prompts where the writer may be free to choose between Exposition or Discussion. Going into the text itself is also needed in cases the prompt is not clear about main purpose or field but presents these as a series of necessary elements to be included. Countinho & Miranda (2009: 42) call function mechanisms for identifying genres 'markers': "the marker is a semiotic mechanism (of any sort) that functions like any clue or indication of the updating of a generic parameter with distinctive value". They identify two big classes of genre markers: the 'self-referential' and the 'inferential'. Examples of the first type of markers are the labels used in the prompts (e.g. article, essay, letter). In this sense, most information for identifying genres in this study has come out of self-referential markers. What is not evident or clear is inferred from clues in the body of text, (inferential markers). Phrases for example such as 'on the other hand', 'some people believe', 'while others', help the experienced reader activate genre knowledge and distinguish a 'discursive' from an 'expository' essay.

In one category the procedure was more complicated. Several textbooks included the letter/email text types in the same prompt as if the same task could be written as a letter or an email. The model answer provided under the prompt was not defined as a text type leaving the impression that it could be used in either case, letter or email. That is, for a lot of textbook writers the two text types as labels were used for the same text. The same was noticed in official examination guides when describing the text types needed (Cambridge English First, 2015):

AN EMAIL/A LETTER is written in response to the situation outlined in the question. Letters and emails in the Cambridge English: First Writing paper will require a response which is consistently appropriate in register and tone for the specified target reader. Candidates can expect to be asked to write letters or emails to, for example, an English-speaking friend or colleague, a potential employer, a college principal or a magazine editor. (p. 30).

There was a confusion on text type in this case which did not occur in other text types. These model answers were carefully studied looking for inferential markers but no obvious difference was evident. There has been a preference for the term email in lower proficiency levels but this seems to be related to the demand for a shorter text (word-limit). As the proficiency levels increase there seemed to be no outstanding clues in the body of the texts (inferential markers) whatsoever. For that reason, my criteria for grouping texts have remained consistent concerning main purpose, ideational and interpersonal metafunctions but have not distinguished between letters and emails unless there was a genre category that all prompts asked for a letter or email only. In the results section the category may include for example the 'Application Letter/email' meaning that this genre was presented in material under a double text type labelling but it may also include the 'Reference Letter' meaning that this genre was found only under the 'letter' labelling.

An appropriate name is then given which best illustrates the basic features and requirements of the genre for less experienced writers such as students. This naming is only a suggestion for improving the students' understanding of the requirements in a few words. I have based this 'naming' on purpose and mode. Terms for 'purpose' have been chosen because I share the view that purpose is the "prototypical criterion for genre identity" (Swales, 1990). Terms for 'mode', in the sense of text types (Glasswell et al. 2001), are widely known and can trigger some subconscious knowledge of the genre at hand. Both terms function as signposts for writers with no expert knowledge of genres. Figure 4.1 illustrates the main procedure followed:

Task prompts   **Body of text**

a) Main purpose + b) Register variables + (if needed) 'inferential' markers

field   tenor   mode

why?

(What?) (Who to whom?) (Text form)
(relations)

Inferential markers

main purpose

register variables
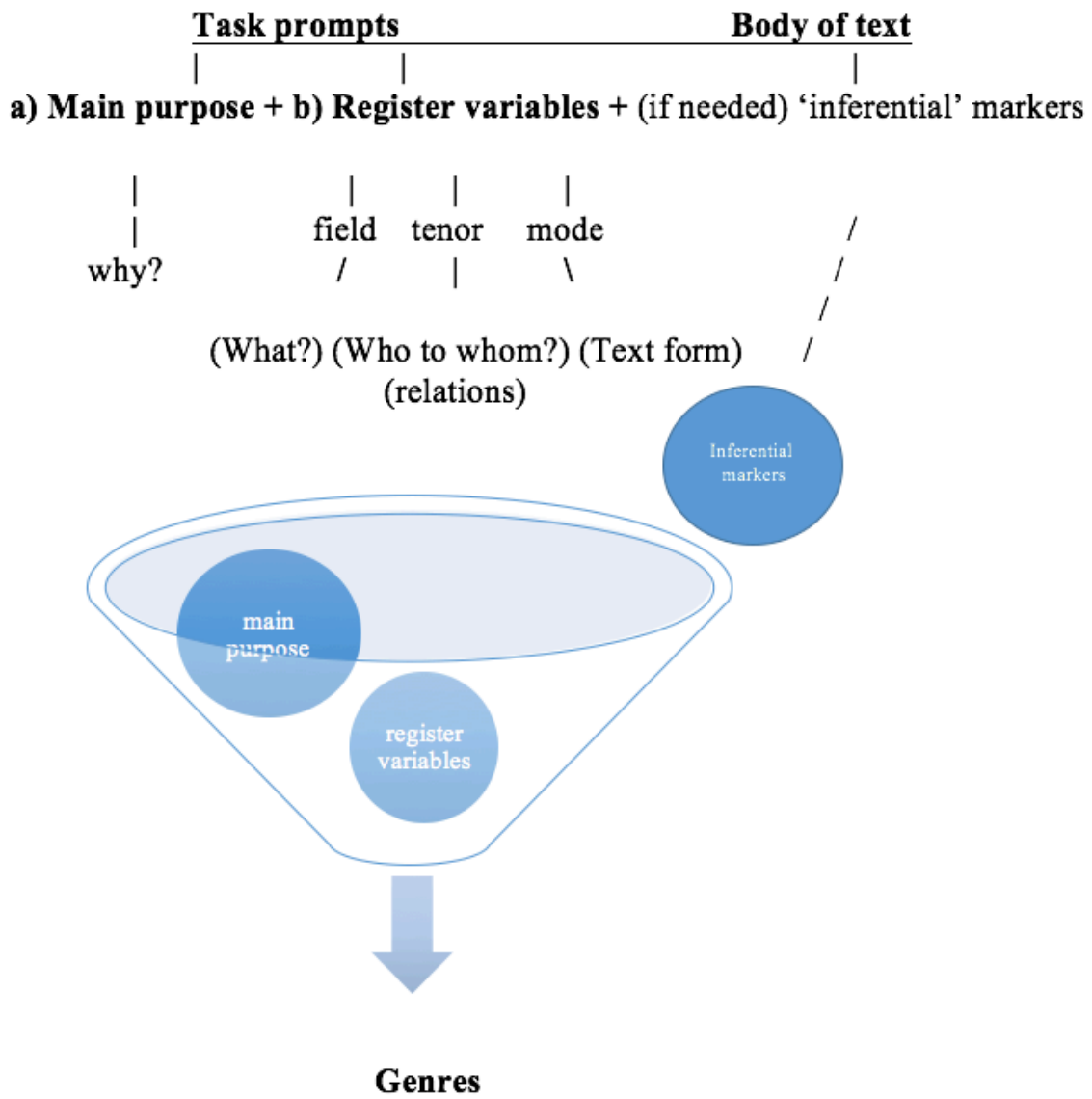
Genres

Figure 4.1 Method for identifying genres from task prompts

4.2. Results

Four main functions are targeted in these prompts: argumentation, description, narration and communication. Clearly there are other secondary functions associated with some of these text categories resulting in what some researchers call 'embedded genres' (Paltridge, 2001) but it is the main purpose that determines the genre 'naming' and classification here.

Although two texts may share the same purpose the metafunctions may be different. The purpose may have been for example 'to offer solutions to a problem' but the text type asked may have been an essay, an article or a letter which have different targeted readers (tenor) and different text formats (mode). In this case three different genre categories have been created: 'Solutions to a problem Essay', 'Solutions to a problem Article' and 'Solutions to a community problem Letter'.

In other cases the mode may be the same, asking the students to write an Essay for example, but there may be varying purposes resulting in various genres such as the 'Discursive Essay' (to discuss two opposing views), the 'Expository Essay' (to state and justify personal view), the 'Factorial Essay' (to argue about the causes of a problem/situation), the 'Consequential Essay (to argue about the consequences of a problem/situation), the 'Solutions to a problem Essay' or the 'Descriptive Essay' (to describe) under the same text type label. It is evident that educational material writers, obviously being based on task prompts and the labelling used by testing bodies, use to label text categories based on mode and that there is an emphasis on text form rather than purpose, that is, a classification based on text types rather than genres.

Another case of same mode but different purpose is the Letter to the Editor often named a genre in the literature. This has been a broad typification characterized by tenor, placing emphasis on a specific addressor-addressee relationship underestimating the importance of purpose and the variation it can cause in the register. A letter of this broad group can be written to inform about new facilities in the area, to praise the editor about a well-written article or to complain about a change that affects the public for example. As purpose has been the main criterion for classification in this study these letters have been allocated to different genres.

Reports were divided in two categories as the basis for reporting is completely different for the two tasks. The first type (named 'Data Report') asks students to report and summarise based on data provided in the rubric, presented in graphs, while the second type (named 'Personal Observation Report') asks them to report based on personal experience. This difference in both field and purpose was considered important having the potential to alter basic features of the language used and led to the creation of two separate categories for Reports.

While the label 'formal letter' is informative I show that this broad labelling can result in twelve different genre categories which can directly aid the writer as to the purpose of the task and can pave the way for the choice of different lexicogrammatical features. The terms 'formal' and 'informal' were retained only in cases where the purpose was the same but tenor was different (e.g. 'Formal Apology Letter' versus 'Informal Apology Letter'). The extent to which these genre categories result in significant register variation is going to be explored later on conducting a detailed quantitative and qualitative analysis.

The new 'genre categories' (table 4.2) with a more student-friendly naming, offer a more accurate view of what is required in these tasks. Both the initial as well as the final classification in the table offer information about the representation in the educational material used for this corpus. For each category the number of texts found is given together with the percentage for the coverage in the material.

Out of seven initial text types (nine if letters and emails are seen as different categories) this process revealed thirty-three different genres. Seven genres for the Essay tasks, two for the Reports, one for the Reviews and the Stories, four for the Articles, twelve for Formal Letters/Emails and six for Informal Letters/Emails. Even though I do not see formality as a variable affecting text type, I have kept the initial 'formal/informal' distinction in letters/emails in order to assess the representation of text types in the material. Looking at the representation of the initial categories in the ninety-three educational material sources used for the corpus, the results show that Essays are overrepresented while Articles, Stories and Reviews are underrepresented.

Table 4.2 The representation of text types and genres in English language educational material

Initial corpus categories	# of texts	Initial representation in corpus	Genre categories	# of texts	Genre representation in corpus
Essay	415	36%	Discursive Essay	176	15.3%
			Expository Essay	85	7.4%
			Factorial	31	2.7%
			Consequential	27	2.3%
			Factorial and Consequential Essay	15	1.3%
			Descriptive Essay	58	5%
			Solutions to a problem Essay	23	2%
Report	176	15.3%	Data Report	100	8.7%
			Personal Observation Report	76	6.6%
Article	88	7.6%	Descriptive Article	40	3.5%
			Expository Article	31	2.7%
			Informational Article	9	0.8%
			Solutions to a problem Article	8	0.7%
Review	39	3.4%	Book/Film Review	39	3.4%
Story	78	6.8%	Short Story	78	6.8%
Formal Letter Formal Email	171 24	14.9% 2.1%	Complaint Letter/email	47	4.1%
			Formal Informational Letter/email	38	3.3%
			Application Letter/email	30	2.6%
			Opinion as a response letter/email	28	2.4%
			Reference Letter	12	1%
			Formal invitation Letter/email	8	0.7%
			Making suggestions formal Letter/email	8	0.7%
			Formal Apology Letter/email	6	0.5%
			Formal Letter of Request	6	0.5%
			Solutions to a community problem Letter	5	0.4%
			Nomination Letter	4	0.3%
			Resignation Letter	3	0.3%
Informal Letter Informal Email	105 55	9.1% 4.8%	Advice Letter/email	61	5.3%
			Personal news Letter /email	33	2.9%
			Informal Informational Letter/email	29	2.5%
			Informal invitation Letter/email	27	2.3%
			Informal apology letter	7	0.6%
			Informal letter of request	3	0.3%
Total	1151	100%	Total	1151	100%

When one does the same for the genre categories it is clear that there are plenty of 'Discursive Essays' in the material compared to other essay genres. For Formal Letters/Emails a great variety of derived genres is evident combined with uneven distribution in material. Clearly, educational material fails to offer adequate guidance in a lot of genres especially those derived from the Formal Letter/Email category.

Overall, the classification of texts according to genre following the above method has shown that:

a. The present labelling of categories based on mode conceals important information. Classification of texts according to genre provides more detailed categories allowing even the finer variation to be identified. A more informative labelling of these categories based on genre and text type can be a shortcut to the candidates' understanding of the task requirements.

b. Learners may have very limited exposure to specific genres.

c. Some genres are overrepresented compared to other genres within the same text type category (e.g. the 'Discursive Essay' in the 'Essay' category). When this happens learners run the risk of assuming that the requirements of the task set in examinations will be similar to the one they have been extensively taught under the same label (e.g., Essay).

d. The derived generic categories cause serious doubt as to the similarity of the language used under the broad initial categories. Corpus compilation based on text types could conceal considerable internal linguistic variation.

CHAPTER FIVE

A DESCRIPTION OF THE MOST PROMINENT GENRES IN ENGLISH LANGUAGE MATERIAL

This chapter investigates the most frequent genres in the teaching material (identified in the previous chapter) in order to find and describe their typical features. It explains the approach chosen for the analysis and proceeds to the description of each individual genre in order to contribute to a more explicit and evidence-based teaching of written genres in the particular context.

5.1 Methodology

This section investigates in detail the genres that were shown to be prevalent in educational material and therefore have been largely represented in the corpus (sub-corpora with more texts). Table 5.1 shows the sub-corpora the number of tokens and texts included as well as the CEFR levels of the texts in each sub-corpus. WordSmith Tools v. 6 software (Scott, 2015) is used for text analysis.

Table 5.1 Sub-corpora used for genre analysis (research questions 3 and 4)

Sub-corpus	# of tokens	# of texts	CEFR levels
Expository Essay	24.347	85	B1, B2, C1, C2
Discursive Essay	47.153	176	B1, B2, C1, C2
Descriptive Essay	15.448	58	B2, C1, C2
Personal Observation Report	17.702	76	B1, B2, C1, C2
Data Report	16.828	100	C2
Short Story	15.558	78	B1, B2
Complaint Letter	8.770	47	B2, C1, C2
Advice Letter	9.938	61	B1, B2, C1, C2

For the study of individual genres, a discourse analytical approach is chosen. Several linguistic features are analysed quantitatively comparing frequency and range showing which features are pervasive and typical of the genre. Qualitative analysis of these features then aims to describe the ways writers manipulate them to realise the main functions, the rhetorical moves and the stylistic conventions of the genre. In this view, the schematic structure is linked to the main purpose but also linked to the specific register.

Specific linguistic features are explored because of their key-role and because they may reveal useful information about the construction of the text as a whole. Particular language features explored in all categories are top frequent common words and keywords. Frequency refers to the number of times a token (a single linguistic unit) appears in the corpus and range shows the number of texts in which this token has appeared. As each genre is explored individually in this part, I present absolute frequencies.

All the words with high frequency can offer information about the corpus content and the prevalent information in a corpus of texts. However, in genre analysis studies, keywords have additional properties and are largely used as the basis for the analysis. In corpus studies, keywords are those whose frequency is statistically significant, when compared to a reference corpus. (Baker, 2004; Scott & Tribble, 2006). This means that compared to a much larger corpus, usually a general one, positive keywords are significantly more frequent in the corpus we are interested in. Using Dunning's (1993), Log Likelihood statistical procedures this is a robust basis for analysis compared to an individual's intuition about the keywords in any text, especially in the case of multiple comparisons of keywords in a large number of texts. If the corpus happens to be tightly designed in terms of genre as this one is, then we can have words that are typical of the genre. These keywords also guide me in finding relevant recurrent patterns in multi-word sequences.

Another advantage of using keywords for the analysis instead of common wordlists is that grammatical or functional words such as 'the' or 'because', for example, occurring frequently in many genres will crop up in certain genres (sub-corpora in this case) as keywords. These words would not usually be identified by the human reader as key, but they may be indicators of style and their appearance in the keyword list can prompt the researcher to go back to the concordances to search the reason for this.

I use the whole WriMA corpus as a reference corpus for the extraction of keywords. This helps me contrast each genre to the rest of the genres in the same context, a process that allows even the slightest difference between similar genres of this context to be revealed.

Another aspect that is taken into consideration during the analysis is the extent to which the words with high frequencies recur consistently in a range of texts. This function, termed 'range' by Nation (2006), comes automatically with any word list in WordSmith Tools software. It may happen, for a word to occur frequently but to appear in two or three texts in the corpus. This means that it has been topic-dependent, strictly linked to the subject discussed, but cannot be regarded as characteristic of the genre. Therefore, frequency is important but if not seen in relation to range it may be misleading. Keyword lists provided in this study, include information about the frequency of each word in the specific sub-corpus as opposed to its frequency in the whole corpus. This contrast leads to the 'keyness score' which shows the strength of the relation between specific words and genres. The number of texts each word occurred in is also provided. Negative keywords are presented with their own 'keyness scores' as they indicate what is rare or unusual in the genre at hand.

The 'Concordance' tool, incorporated in WordSmith Tools software, provides all of the instances of the search word showing them in context. This way the researcher can have a more comprehensive view instead of decontextualized words. Apart from viewing concordance lines, one can further investigate collocation, that is, words in the neighbourhood of the search word (e.g. first collocates on the left for 'education' may be the words 'rounded' or 'higher' and first ones on the right may be 'system' and 'should'). When this involves grammatical items it is often called colligation. I refer to these collocates as L1 or R1 for left and right positions showing the immediate neighbouring words.

In fact, the corpus linguist has the ability to search within a much larger collocate horizon and the further on they move from the search word the more difficult it becomes to observe patterns. By examining 'friend' words the researcher is able to retrieve common lexical and grammatical patterns of co-occurrence. As Firth (1957:11) puts it, "You shall know a word by the company it keeps". This way, I am able to present multi-word clusters where these include keywords and are frequent. These may be continuous or discontinuous clusters. For example: [Most people (believe/think/feel) that…] is a discontinuous pattern with a slot for choice in between. But [in terms of] is a continuous or uninterrupted cluster.

Using the 'plot dispersion value', another tool incorporated in WordSmith Tools, it is possible to know how the occurrences of a specific word are distributed in one or more texts. This way, it is possible to see where in the text (beginning, middle, end) the keyword is mostly used which can be related to the function of the word in particular rhetorical stages. Although I do not perform a detailed 'moves analysis', searching for the context of keywords and frequent common words helps me identify basic stages/moves realised by particular linguistic features.

Seeing prevalent linguistic features in relation to neighbouring words and within sentences in concordances it is possible to interpret their functional role in the specific context. Words and phrases are chosen because "they are particularly well-suited to the purposes and situational context of the register" (Biber, 2012: 192). These functions may be the realization of necessary rhetorical moves, cohesion, the writer's expression of personal stance or the engagement of the reader. 'Personal stance' refers to feelings, attitudes, opinions or judgements that the writer allows to reveal. 'Reader engagement' refers to techniques used by the writer to address readers "pulling them along through their argument" or "including them as discourse participants" (Hyland, 2012:417). Interpretation may also refer to stylistic functions adding for example a formal or informal tone where the writer may be distant or involved.

The overall structure of a text is a fundamental factor in its readability and potential to communicate and can be one of the characteristic conventions of generic types. Headings and subheadings in reports for example, can help organise thematic sections in order to help readers follow the writer's point of view. In the same way, greetings are an essential part of letters and are always written in separate lines in the opening and closing of the text. Wherever such features are recurring, their function is investigated.

Therefore, I explore the existence of characteristic structural elements of the texts and the use of recurrent words, keywords and patterns associating them with basic functions. This process leads to the identification of the most prominent, the typical features of the genre explored each time.

5.2 Results

5.2.1 The Expository Essay

In the Expository Essay writers are called to put forward a viewpoint, provide arguments in defense of or as objections to the proposition made. They need to justify their position and reach a conclusion. A common structure of the Expository Essay, observed also in this context, is the following:

^ Introduction of the issue ^ Thesis statement ^ Arguments (2-3) ^ Conclusion

In most texts in this corpus the first two to three sentences introduce the issue and the thesis statement is expressed in one sentence right after the introduction. Two to three arguments are put forward, developed and illustrated by examples in most cases, followed by the conclusion in one or two sentences. In terms of paragraphs, the first two moves are usually in the same paragraph, arguments are developed in separate paragraphs and the conclusion is always in the last paragraph. There is a heading in only 4.7% of the texts in this sub-corpus.

As can be seen from the range of nouns in key words (table 5.2), the focus is on social issues, human concerns that can be debatable and seen from different perspectives ('life', 'lives', 'learning', 'money', 'education', 'opportunities'). These are issues of public interest which tend to transcend local or national boundaries ('world', 'society', 'human').

Reference to subjects is general and non-specific. The subjects of concern are general groups of people ('children', 'people', 'parents', 'students') or individuals but seen broadly as representatives of larger groups not as entities ('child', 'individual'). Reference to particular persons is highly uncommon either in the form of pronouns or in specific names. In fact, proper names and pronouns referring to specific subjects ('I', 'me', 'my', 'you', 'your', 'he', 'she') are in the negative key word list. The pronouns 'they' and 'their', however, are positive keywords showing a preference for general reference.

Table 5.2 Keywords in the 'Expository Essay' sub-corpus

Positive					Negative	
Keyword	Freq. %	Texts	RC freq. %	Keyness score	Keyword	Keyness score
their	0.88	68	0.42	81.80	<Name>	-299.35
child	0.19	14	0.04	63.35	#	-229.17
learning	0.16	15	0.04	46.40	you	-206.08
children	0.41	27	0.18	45.52	I	-174.75
should	0.47	52	0.24	37.47	was	-142.79
they	0.76	72	0.48	31.53	<location>	-81.00
money	0.25	23	0.11	29.32	your	-77.33
life	0.29	36	0.14	28.83	me	-58.71
world	0.29	39	0.14	27.32	he	-58.34
education	0.18	18	0.07	27.23	had	-43.75
need	0.27	41	0.13	25.57	am	-33.15
lives	0.15	27	0.05	24.27	she	-29.95
not	0.72	63	0.47	24.15	day	-24.78
such	0.30	44	0.15	23.76	my	-23.98
society	0.16	21	0.07	22.17	very	-23.51
skills	0.13	12	0.04	21.86	year	-21.48
human	0.09	17	0.03	21.34	the	-17.65
people	0.78	67	0.54	21.19	were	-15.79
learn	0.14	18	0.05	19.53	at	-15.33
parents	0.20	24	0.09	18.29		
provide	0.12	23	0.05	18.15		
developing	0.06	11	0.01	17.61		
or	0.55	63	0.37	17.42		
help	0.21	26	0.11	16.93		
young	0.23	24	0.12	16.06		
of	2.89	85	2.46	15.92		
opportunities	0.07	12	0.02	15.65		
students	0.30	19	0.17	15.63		
are	1.12	80	0.87	15.63		
important	0.20	34	0.10	15.56		
individual	0.07	14	0.02	15.38		
cannot	0.10	21	0.04	15.33		
in	2.43	84	2.05	15.17		

Present verb tense gives the text a diachronic perspective in the sense that what is said applies generally and reference to past is avoided ('was', 'were' in the negative keyword list). Generalising is in fact a common technique in argumentation used to give the impression that all or the majority of the people make the same choices and share the same perspective with the writer.

E.g. 1. *To my mind, tourism is only eminent nowadays because people are more concerned with the money it generates than the environment it destroys.*
E.g. 2. *When people move to a new country and culture, they naturally want and need to adapt to and become a part of it.*

Writers need to present the issue and adopt a viewpoint which will then be supported by arguments. To do so, they may present a general 'need' to which their suggestion will later respond

E.g. *Growing up, children need to know there is someone there for them emotionally* (thesis: against 'working mothers')

They may also highlight the issue to attract interest or to differentiate between the degree of importance between two different views using the word 'important':

E.g. 1. *Education is an important factor/ teachers play and will continue to play an important role in the classroom.*
E.g. 2. *There is no doubt that regular exercise is good for your mental and physical health, and it is especially important for young people.*

The issue is often presented as a dilemma using the [whether… or not] pattern. The following example is using the pattern to introduce the thesis statement at the second sentence of the text:

E.g. *Whether it is better to do household chores or not, the advantages of having children help with household jobs are more than not doing such.*

Then writers often oppose an idea or habit ('not', 'cannot'):

E.g. 1. *The high school curriculum <u>cannot</u> give students a clear enough picture of academic fields like law or engineering.*

E.g. 2. *Although family time and individual learning are important, the range of expertise and resources that schools offer <u>cannot</u> be matched in the home.*

And finally, writers suggest and advise ('need', 'should'):

E.g. 1. *What governments <u>need</u> to do to make this happen is to ensure there is a global programme to educate people of all ages about the environmental consequences to their actions.*

E.g. 2. *Criminals <u>need</u> to feel that their violent crimes will be punishable by death, making them think twice about committing a deadly offense.*

Adding connectors are used when arguments are added on the same side while contrastive ones, link sentences of opposing views. Since writers are mainly supporting their own arguments in this genre, adding connectors are more common than contrastive ones (table 5.3). Causal/consequential connectors are also frequent as writers try to show how particular attitudes lead to particular problems or solutions to problems.

Table 5.3 Connectors in the 'Expository Essay' sub-corpus

Adding connectors	**frequency**	**Contrastive connectors**	**frequency**
and	829	but	75
also	67	however	59
in addition	20	although	39
furthermore	13	while	25
moreover	6		

The use of nominalisation allows writers to create long sentences which are 'packed' with meaning.

E.g. *However, in my view, success is a reflection of internal human qualities, such as compassion, kindness, honesty, understanding and more.*

Writers position themselves in the beginning of the text as seen earlier, but keywords show that self-reference is less often observed than in other genres of the same context ('I' in the negative keyword list). Looking at the dispersion plot writers refer to themselves mostly in the first paragraph to express their view and the last paragraph to draw conclusions after the argumentation. The pronoun 'they' is preferred however, occurring in nearly 85 per cent of the texts referring to general participants (table 5.4). Their expression of stance is more often implicit rather than explicit even though patterns such as [in my opinion], [in my view], [I believe that], [I think that] are used (figure 5.1). Modals in frequent lexical bundles such as [it would be], [that we should], [we need to], are employed for the discreet expression of personal judgements. Writers need to convince the reader through argumentation and reasoning rather than merely stating their preference. Strong arguments smooth the path leading to the conclusion chosen by the writer. Table 5.4 presents the frequency of personal pronouns in the sub-corpus and figure 5.1 shows the most frequent collocates of the pronoun 'I'.

Table 5.4 Pronouns in the 'Expository Essay' sub-corpus

Personal pronouns	frequency	Texts %
I	153	64.71
You	41	15.29
He	6	5.88
She	12	4.71
It	246	87.06
We	154	43.53
they	185	84.71
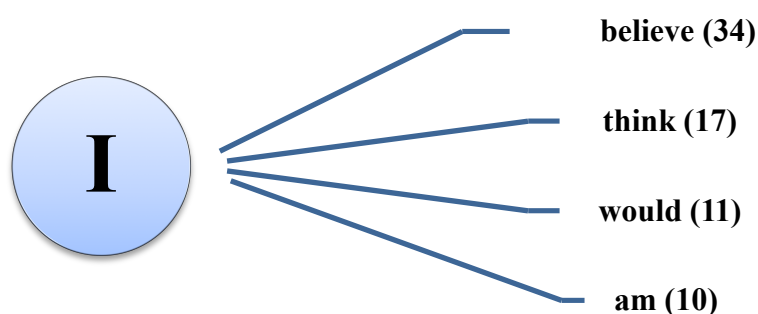
Figure 5.1 R1 collocates of 'I' in the 'Expository Essay' sub-corpus

Writers try to include themselves when mentioning people's beliefs and attitudes rather than address them directly by using 'you'. By using the plural first person pronoun in almost half of the texts they make the reader feel part of the group, creating a circle of shared habits, beliefs and attitudes relevant to both the reader and the writer. Inclusive engagement (we) is preferred over straightforward reader engagement (you).

E.g. *First of all, instead of spending time we spend money for the ones we love*

The pronoun 'we' is often followed by 'should' and 'can' in suggestions, and 'are' in generalisations.

E.g. 1. I think we should all remember that money is only paper.
  E.g. 2. If we spend our money sensibly, we can enjoy our lives and help others at the same time.
  E.g. 3. As social beings, we are naturally influenced by those we meet in the course of our lives

5.2.2 The Discursive Essay

These tasks usually ask the writer to discuss two opposing views presenting arguments for and against and then form an opinion based on these arguments. During corpus analysis the following structure has been observed:

^ Introduction of the issue ^Argument in favour of one side ^ Argument in favour of the other side ^ Summary of pros and cons + Conclusion in favour of one side.

Each move is a separate paragraph resulting in a 4-paragraph essay most of the times. In a few cases the writers choose to state where they stand regarding the issue early in the text, that is, before the presentation of any argument. This is an additional sentence after the 'Introduction of the issue' move, but apart from this deviation these texts also follow the structure shown above. There is a main heading in only ten percent of the texts in the sub-corpus.

The issue to be discussed is troubling people, it can be seen from more than one perspective and is worthy of attention as it affects societies, countries or large groups of people. What is discussed has an impact to the 'world', 'society' or 'country' (table 5.5). Writers are challenged to look outside their micro-world and personal lives.

E.g. As the <u>world</u> becomes more integrated, the need for communication is becoming more pressing.

Apart from being socially significant issues they are also related to youth interests and lifestyle. 'Nowadays' or 'these days', emphasise the present and are often used in the first sentence to introduce a modern tendency or a development that has changed our lives.

E.g. *<u>Nowadays</u>, an increasing number of students are turning to the Internet as a primary source of research material.*

Subjects are mostly general groups rather than individuals. Large groups of people are manipulated by writers to achieve their purposes. First, to show how important the issue is, affecting a lot of people; second, to add strength to the arguments made by the writer based on the assumption that what is said/thought/done by the majority of people is probably true or sensible. Later on, referring to parts of the group the writers build their arguments dividing people as they wish. There is a tendency to divide people based on perception. Sometimes the writer admits this is a personal judgement as in the following example, but this is not always the case.

E.g. *<u>It is probably true to say that</u> most people believe that a university degree is the only way to get a good job.*

Table 5.5 Keywords in the 'Discursive Essay' sub-corpus

Positive					Negative	
Keyword	Freq. %	Texts	RC freq. %	Keyness score	Keyword	Keyness score
is	2.27	171	1.58	105.02	I	-480.28
people	0.95	150	0.54	98.41	<name>	-447.86
that	1.93	166	1.32	97.75	#	-413.37
hand	0.23	82	0.07	79.46	was	-279.90
are	1.28	166	0.87	68.82	you	-161.80
they	0.79	126	0.48	67.46	<location>	-116.80
more	0.70	124	0.41	64.99	had	-108.90
advantages	0.12	45	0.03	59.06	me	-102.61
many	0.42	105	0.22	58.43	am	-100.49
their	0.69	123	0.42	56.66	he	-95.08
other	0.43	128	0.23	54.98	my	-92.02
disadvantages	0.09	34	0.02	51.39	your	-67.30
argument	0.09	29	0.02	49.55	were	-51.62
benefits	0.13	40	0.04	49.27	the	-36.29
can	0.74	143	0.48	46.09	know	-32.28
not	0.72	137	0.47	44.89	at	-30.18
believe	0.24	84	0.11	41.45	year	-25.68
society	0.17	43	0.07	40.84	like	-23.32
internet	0.15	24	0.06	40.29	<nationality>	-18.94
arguments	0.07	27	0.02	39.93	we	-18.47
may	0.26	66	0.13	38.73	last	-18.11
however	0.30	110	0.16	36.65	with	-18.06
such	0.28	86	0.15	36.08		
example	0.18	64	0.08	35.20		
argue	0.09	36	0.02	34.84		
or	0.56	120	0.37	33.92		
country	0.17	35	0.07	33.85		
it	1.40	162	1.08	32.55		
should	0.40	89	0.24	32.28		
education	0.16	39	0.07	31.99		
be	1.18	157	0.90	29.48		
these	0.31	81	0.18	27.95		
some	0.40	118	0.25	27.25		
often	0.17	58	0.08	26.36		
world	0.24	67	0.14	25.29		
others	0.12	52	0.05	24.54		
both	0.20	68	0.11	22.70		
argued	0.04	21	0.01	22.19		
there	0.61	127	0.44	21.47		
issue	0.07	30	0.03	20.56		
important	0.18	52	0.10	20.35		

of	2.81	175	2.46	19.91		
against	0.06	25	0.02	19.19		
better	0.16	57	0.09	18.98		
on	0.89	159	0.70	18.23		
countries	0.15	37	0.08	18.10		
cannot	0.08	34	0.04	17.47		
case	0.08	30	0.03	17.16		
does	0.09	31	0.04	16.57		
children	0.28	51	0.18	16.40		
this	0.87	147	0.69	16.38		
means	0.10	37	0.05	16.29		
opinion	0.10	42	0.05	16.07		
view	0.09	33	0.04	15.95		
those	0.15	40	0.09	15.92		
benefit	0.07	25	0.03	15.87		
major	0.06	23	0.02	15.72		
development	0.08	22	0.03	15.61		
social	0.12	30	0.07	15.54		
who	0.30	87	0.20	15.53		
own	0.14	44	0.08	15.20		
has	0.39	97	0.28	15.15		

(table 5.5 continued)

Of course the writer can also use this arbitrary quantification to persuade readers by adding people to the argument of their preference. [Some people believe], but [more people believe] for example. This abstractness and generalization is quite different from precise numbers and percentages one can find in reports. The word 'people' is found in 85% of the essays in the sub-corpus. In figure 5.2 one can see the adjectives preceding the word *people* and their frequencies. The variety of expressions filling the slot in the frequent pattern [quantifying phrase + of + people] is shown in table 5.6.
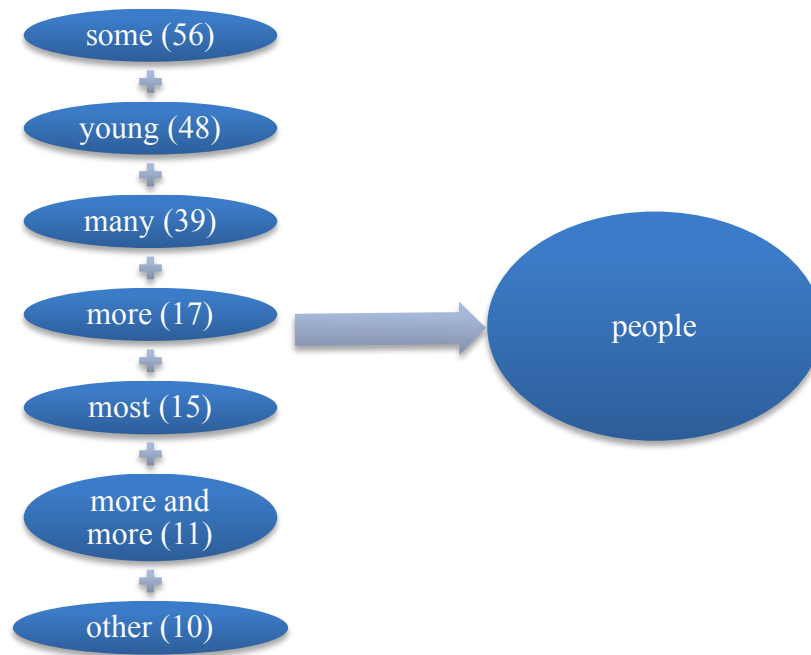
Figure 5.2 'People' L1 collocates in the 'Discursive Essay' sub-corpus

Table 5.6 [Quantifying phrase + of + people] pattern in the 'Discursive Essay' sub-corpus

millions	a growing number	
thousands	a minority	
a group	billions	
large numbers	a small proportion	**of people**
the majority	a large majority	
two types	an increasing number	
a lot of	a large number	
a small number	a wide variety	

Writers present peoples' beliefs, thoughts, feelings, using the pattern [people + mental/saying verb + that] as shown in figure 5.3. 'Believe' is the most frequent verb occurring in 48% of the texts. A similar pattern [others + mental/saying verb + that], occurs when the writer presents the second argument moving from one general group to another. Presenting these two groups in one sentence is also very common in the first or second sentence during the 'Introduction of the issue' stage.

E.g. <u>Some people claim</u> that it is cruel to keep wild animals in cages while <u>others believe that</u> zoos are the only way we have to come into contact with some rare species of the animal kingdom.
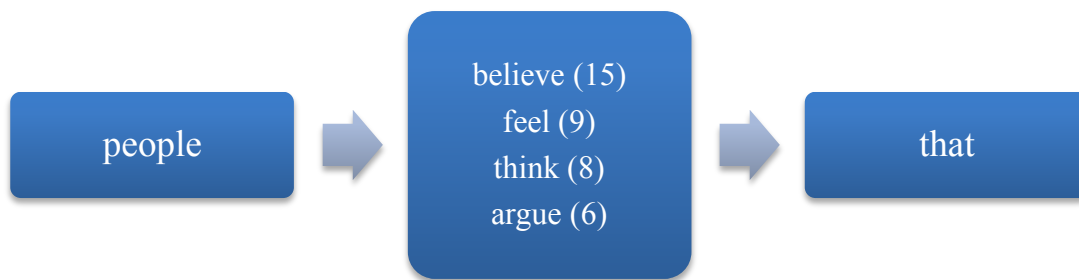
Figure 5.3 [People + verb + that] pattern in the 'Discursive Essay' sub-corpus

The writer's view comes as a result of the examination of evidence and appears in the last paragraph. Even though 'they' and 'it' are the most frequent pronouns in these essays, the frequency of 'I' shows that writers express personal opinions especially near the end of the essay. The clusters [in my opinion] and [in my (own/personal) view] are frequent at the summary/conclusion stage. Figure 5.4 shows the percentage of texts including each personal pronoun in the sub-corpus:
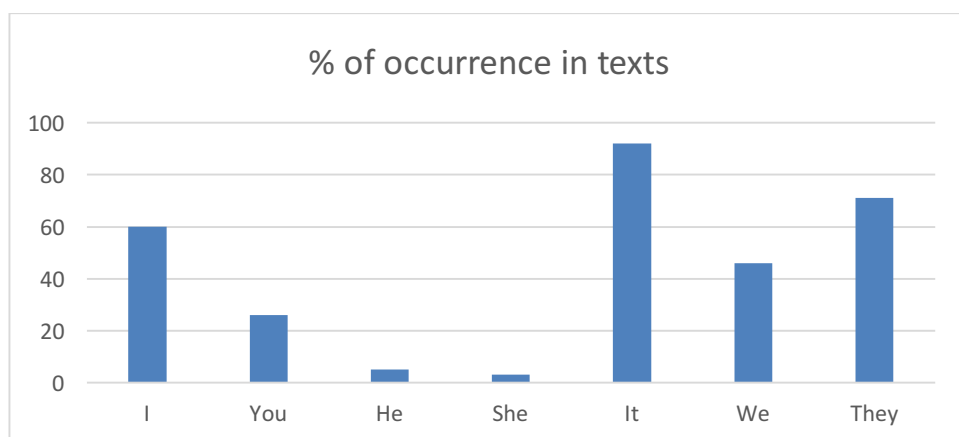
Figure 5.4 Personal pronouns in the 'Discursive Essay' sub-corpus (occurrence in texts)

The pronoun 'we' is sometimes used to generalise instead of 'people' or 'they'. Using the pattern [if + we + verb], writers sometimes attempt to draw the reader closer, to show that the positive or negative consequences of an action or choice are bound to affect all people including themselves. The pronoun 'you' is also used to generalize although much less.

This is very interesting as the use of pronouns in essays is a confusing issue mostly caused by teachers who insist that students should avoid 'I', 'you' and 'we' and that style should be impersonal. The analysis shows that 'I', 'you', 'he' and 'we' are negative keywords which means they are used rarely compared to other genres. This does not mean however, that these essays are totally impersonal. Writers try to sound objective presenting arguments but they need to take stance and there is some involvement even if only at the last stage. Therefore, generalized teacher guidance concerning the strict avoidance of these pronouns would be inaccurate.

Presenting the issue to be discussed in the first sentence means placing the reader at the centre of important choices or dilemmas in life. These choices are often introduced with 'or'.

E.g. When they finish school, teenagers face the dilemmas of whether to get a job <u>or</u> continue their education.

Extended use of 'that-clauses' are used to show that some trait or pattern of behavior is characteristic of all or the majority of humans. Arguments supporting one side often take the form of facts, presented as universally accepted propositions especially using the phrase [the fact that] and the somewhat milder expression [it seems that], mostly used at the final stage leading to a conclusion.

E.g. 1. One of the main advantages of being self employed is <u>the fact that</u> you are completely self-reliant and can make decisions on your own.
E.g. 2. All things considered, <u>it seems that</u> although air travel does have its bad points, it has one overwhelming positive aspect – <u>the fact that</u> it is far quicker than any other means of transport.

Certain nouns signal the presentation of opposing sides as shown in table 5.7. Reference to positive points is preferred. 'Advantages' occur in more texts than 'disadvantages' and the

same is true for 'benefits' compared to 'drawbacks'. Keywords are marked as 'Kw'; the other words are presented here mainly for comparison purposes. Percentages refer to the texts including the word in the corpus. The same tendency appears with evaluative adjectives. 'Positive' is more frequent than 'negative' although only slightly.

Table 5.7 Nouns supporting opposing sides in the 'Discursive Essay' sub-corpus

	Occurrence in texts		Occurrence in texts
Advantages (Kw)	25%	Drawbacks	9%
Advantage	9%	Drawback	0.5%
Disadvantages (Kw)	19%	Benefits (as a noun) (Kw)	21%
Disadvantage	5%	Benefit (as a noun)	3%

Consequences of certain actions or choices are often presented as possible rather than certain to happen. Modals 'can', 'may' and 'would', serve this function.

E.g. In addition, trying another cuisine <u>can</u> expand our knowledge about food and we <u>can</u> discover new and enjoyable tastes.

Some writers prefer to make mild statements using modals even at the conclusion stage.

E.g. In conclusion, while continuous assessment <u>may</u> be fairer in some contexts, there are still times when traditional exams <u>may</u> be more appropriate.

There are two main categories of verbs in these essays. The first are mental verbs (believe, think) and the second are relational verbs in present tense. The second category occurs frequently in formulaic expressions. The next table shows the most frequent 3-word clusters in the corpus, including the verb 'be'. It is clear that existential 'there' and 'that-clauses' are often used to present the situation.

Table 5.8 Three-word clusters including the verb 'be' in the 'Discursive Essay' sub-
        corpus

That it is	It is true	It is also	It is a
That there is	What is more	It can be	Seems to me
There is a	It is not	Is that there	It is often
However there are	In conclusion it	This is not	is true that
It would be	Is that the	To be a	Is not always
Be able to	That this is	Be argued that	There are some
It is the	It is important	It seems that	There are also
Is that it	View is that	It seems to	Need to be

Adding arguments to reinforce the expressed proposition is mainly done using 'also', 'furthermore', 'what is more', 'in addition', and the pattern [not only… but also].

E.g. A convincing argument can be made about globalization <u>not only</u> playing a pivotal role in the development of technology and economy, <u>but also</u> promoting the cultural exchange between different countries.

It is common for writers to offer examples after each argument. The phrase [for example], is very common, with most instances at sentence initial position. The pattern [such as + noun + and + noun] is also used to exemplify.

E.g. It can also be argued that continuous assessment is a more effective way of testing some subjects <u>such as design and technology</u>, which are more creative and less academic.

As stated earlier, the main body of the essay is carefully structured to include two paragraphs of arguments concerning the views on the issues discussed. This usually takes the form of two separate paragraphs, one for each side. These paragraphs often open with the phrase [on the one hand] for the first paragraph and [on the other hand] for the second. The second phrase always occurs at the middle of the text and is far more frequent (74 instances of 'on the other hand' compared to 24 for the phrase 'on the one hand'). The opposing view is also introduced with conjunctions such as 'however', also at the middle of the essays in opening sentence position. The same is achieved with 'while', often in the pattern [while some people + verb… others + verb].

Negation with the verb be ('is not', 'are not') is also a way to contradict previous arguments.

E.g. One significant counter-argument is that the purpose of education <u>is not</u> just to prepare children for later careers, but also to develop their all round "culture".

As the final paragraph is a summary of the basic points made leading to a conclusion, the same negation is sometimes used at this stage.

E.g. To sum up, although some people argue killing animals for research and food is ethical, I would argue there is sufficient evidence to demonstrate that this <u>is not</u> the case, and therefore, steps must be taken to improve the rights of animals.

The modal 'should', occurs in half of the essays in the corpus and is often used in recommendations at the final paragraph.

E.g. 1. In conclusion, it is clear that it <u>should</u> be made illegal to smoke in public places.
E.g. 2. As long as a person is competent and has the will to contribute, then age <u>should</u> not be a barrier.

This analysis suggests that the 'conclusion' is probably the most challenging stage from a linguistic point of view. Multiple functions are served in a sentence, summarising the points made previously, expressing personal stance as a result of specific evidence and perhaps making recommendations. The following sentence is a good example of clever manipulation of lexicogrammatical features achieving multiple functions in only one sentence. The writer manages to summarise the main positive and negative effects, include himself/herself in those affected ('we', 'our') and show which side he/she supports ('but', 'extremely negative').

E.g. In conclusion, it may save money in the short term if we allow minority languages to disappear, but in the long term this would have an extremely negative impact on our cultural heritage.

5.2.3 The Descriptive Essay

It is necessary for candidates to know some basic things about Description as it is often embedded in other genres. This means that a text belonging to a letter genre, for example, could have descriptive elements if the task required so or if writers chose to expand more on this function. The writer's main purpose in a letter may be, for example, to tell the reader what they can do together when he/she visits but this may require some description of the location or the events they are going to attend. Another example is the 'Story' which usually includes a detailed description of a central figure or the setting. These descriptive parts however, when seen in other genres, are short and do not hide the main purpose of the genre. When description is embedded in another main genre, it is easy for the writer to lose focus while writing, as I am going to show during the analysis, and mishandle the proportion of length for each function.

The 'Descriptive Essay' in this context is about a person/location or event that has impressed the writer. Apart from the detailed description, writers are often asked to explain why this person/place or event has a special place in their heart.

[sample prompts]
E.g. 1. Describe a person who has influenced your life and explain why you admire him/her.
E.g. 2. Describe an annual event that you attended and explain why you enjoyed it.

Therefore, stages usually follow this pattern:
^ Introduction of the subject of description ^ Extended description ^ Explanation of the writers' feelings

The fact that proper names were manually tagged in the WriMA corpus, makes it possible to see names as an entity and find out that both human and location names are keywords in this genre (table 5.9). Specific reference is made to people and places here, which become the focus of attention.

Table 5.9 Keywords in the 'Descriptive Essay' sub-corpus

Positive					Negative	
Key word	**Freq. %**	**Texts**	**RC freq. %**	**Keyness score**	**Key word**	**Keyness score**
her	0.66	14	0.14	139.97	you	-85.49
she	0.69	11	0.18	119.16	your	-65.67
he	0.68	15	0.20	98.98	# (number)	-49.44
his	0.44	14	0.12	71.06	should	-32.81
was	1.13	39	0.55	68.17	that	-31.24
<location>	0.83	32	0.38	57.17	this	-30.11
him	0.26	11	0.06	49.55	can	-22.40
beautiful	0.15	18	0.03	39.69	be	-22.36
admire	0.09	8	0.01	31.35	do	-21.42
hair	0.07	11	<0.01	27.40	to	-20.44
<name>	1.19	44	0.79	25.40	are	-19.92
warm	0.09	10	0.01	25.03	am	-19.01
street	0.10	9	0.02	23.87	may	-18.96
when	0.45	39	0.23	23.49	believe	-17.84
appearance	0.07	9	<0.01	22.82	will	-17.58
always	0.21	16	0.07	22.15	more	-15.85
streets	0.09	11	0.02	22.05		
person	0.17	13	0.05	21.58		
a	2.87	58	2.28	20.35		
were	0.40	27	0.21	19.10		
city	0.19	11	0.07	18.78		
had	0.44	26	0.24	18.78		
ago	0.09	14	0.02	18.00		
eyes	0.08	11	0.02	17.17		
event	0.09	7	0.02	16.96		
seemed	0.06	8	<0.01	16.72		
never	0.14	18	0.05	16.68		

The description refers to something or someone that played a significant role in the writer's life. Sometimes the prompt itself implies that this is going to be a description of a pleasant experience using words such as 'admire'*, '*like'*, '*exciting'*, '*enjoyed'*,* and 'favourite'*,* among others. Writers, however, opt for the description of a positive character, event or visit, even in the limited cases where this is not clearly requested. Therefore, adjectives with positive connotations such as 'warm', 'beautiful', 'fascinating', 'friendly', 'good', 'caring', 'famous' and 'best' are frequent. They are often preceded by 'very'. Sometimes a chain of adjectives is used to describe people or objects.

E.g. 1. <Name> is a <u>warm, fun-loving, intelligent</u> person.
E.g. 2. I soon learnt, however, that her true character was <u>calm, warm and kind</u>, yet unbendingly <u>professional</u>.

Human behaviour traits or physical appearance features seem to last through time adding credibility to the writer, justifying their good impression. Using the adverb 'always'*,* and the pattern [would + infinitive], the writer manages to show that these qualities were habitual and characterised the person described over long periods of time.

E.g. 1. Not only was she a good listener, but the advice she offered was <u>always</u> sound as well.
E.g. 2. As we lived near a major port, he <u>would visit</u> us whenever his ship came in.

What is described has a strong connection with the writer's life, past or present. It may refer to a relative, friend, classmate, teacher, location visited or event attended, but the writer is not at the centre of this description; it is the impact of the person, the visit to a place or the attendance to an event that matters. First person-pronouns are frequent; the description is closely related to the writer (A person I admire/ The house of my dreams/ A wedding I'll never forget). It is, however, third person-pronouns ('he', 'she', 'his', 'her', 'him', 'himself') that are keywords in this genre.

Moving from the person described to the writer is common in this genre and requires some skill. It is possible for the writer to get carried away writing about himself/herself and devote more space than one should when describing the impact. Considering the short length of these essays these risks should be explained during teaching. The following is an example of skilled writing regarding this changing of persons.

E.g. I had forgotten to take an umbrella, had missed the bus and was now ready for a good scolding. It was then that I met <name>, <surname>, the Managing Director, who was to have an enormous influence on my life and career. My first impression of her was one of politeness verging on stiffness. She was…

Reference to the writer himself/herself is made in the beginning and/or at the end of the text. The central and biggest part of the essay is about the subject of description. The size of the descriptive part therefore, is the main difference between the descriptive genre and other genres that include some description.

Writers in descriptive texts often use language that departs from literal meaning to help the readers form images in their minds and to evoke imagination. In the texts studied here, there is some use of similes, metaphors and idioms, however not extensively used probably due to the fact that these are model texts for second-language writers and figurative language can be a challenging task for writers with limited awareness of these expressions. If, however, this feature could be taught it would be very useful in descriptive texts. The following examples show some of the uses of figurative language in the sub-corpus:

E.g. 1. *Surrounded by high, medieval stone walls, the old part of the town is like a journey back in time.*
E.g. 2. *His hair is as crazy as his clothes, going from yellow at the front, to red, to blue to green at the back, and it sticks up as if he's styled it with the aid of a powerful electric current.*
E.g. 3. *She emerged, graceful as a swan, in her romantic white dress.*

Another aspect of difficulty for second-language writers is the transition from one verb tense to another and the great range of tenses needed in this genre. There is a constant connection of past experiences, actions or habits to current feelings and memories. Therefore, present simple, past, simple and continuous, present and past perfect, may all be used in the same text. In fact, the present perfect tense, often confusing learners for its relation both to present and past can be nicely illustrated using the last sentence of these essays that usually refers to the impact of the experience on the writer's life.

E.g. 1. *The combination of ancient and modern held such a fascination for me during that first visit that I have been back many times since, each time discovering something new.* (End of text)

E.g. 2. *I met him in a summer camp four years ago and we've been best friends ever after.* (Beginning of text)

5.2.4 The Personal Observation Report

Research on the report genre as a whole has mostly focused on Research Reports (Nesi & Gardner, 2012; Paltridge, 1997) and Workplace (or Professional) Reports (e.g. Bondi & Danni, 2015; Flowerdew & Wan, 2010) or both (Devit et al., 2004). This is understandable as the first type is necessary for every student or scholar in higher education and the second is necessary to many people working in companies, organisations or other professional contexts. Research has also involved school-based reports, naming them Informational Reports (Board of studies NSW, 1998; Cope & Kalantzis, 1993; Derewianka, 1990; Knapp & Watkins, 1994; Martin, 1985).

However, this type of report resembles neither 'Research' nor 'Informational Reports'. 'The Personal Observation Report' is not similar to Research Reports, as reporting is not based on raw data, facts and figures. The themes are not generalisable as in the case of school-based Information Reports (e.g. 'Frogs', 'The life cycle of a grasshopper', 'Our solar system'). The subject is often very specific and the writer is asked to assess and draw personal conclusions based on subjective views and the proximity to a place or personal experience, which the recipients do not share. This is obvious by looking at the prompt:

[Sample prompt]

 You work for a local magazine. A new take away restaurant has opened in your area. The editor has asked you to visit it and write a report saying whether you recommend it or not.

The official guidelines in the Cambridge English proficiency handbook for teachers (2015: 24), state: "The content of a report is mainly factual and draws on the prompt material, but there will be scope for candidates to make use of their own ideas and experience".

The task often asks the writer to recommend a place explaining the reasons for this recommendation. Thus the purpose is not simply reporting but in most cases it encompasses recommendation and explanation.

[Sample prompt]
A group of American students is coming to visit your school in a few months. They have never been to your town before so their coordinator, Mr. <Surname>, has asked you to write a report about interesting places worth visiting in the area. Write a report describing the places and explain why you think they may be of interest to the group of visiting students. Write your report to Mr. <Surname>.

Most of these tasks seem to share more similarities with workplace reports since they are more product-oriented, that is, the ability to conduct research is not a prerequisite as in the case of academic Research Reports. In the previous prompt, the use of pronouns (e.g. whether *you* recommend it or not) shows that personal involvement is not only allowed but also required.

In the Personal Observation Report specific recommendations are often asked for instead of general conclusions. Here there is real-like need for immediate sharing of personal knowledge or views based on personal experience for practical reasons, as is the forthcoming assessment of an employee or a group of people visiting the writer's town soon. There is a practical need. As Nesi and Gardner (2012: 172) explain, in workplace genres the methodology employed by the writer may be of little concern to the reader. Clients are unlikely to care about the replicability of results; the focus is on the practical outcome.

We could say then, that the Personal Observation Report is a workplace-like report. Since candidates for these exams are often quite young, these tasks are often adjusted to their interests and experience as mentioned before. The writer for instance, works in a children's camp and needs to assess an employee or works for a teenage magazine or needs to recommend places to visit for a visiting group of students. Workplace simulation is understandable, as the candidates of these exams may need to use the language they are being tested on in workplace environments. However, seeing this genre in high-stakes language exams, one cannot help thinking that the certificates obtained may also be used for university entrance and this kind of tasks does not prepare them for the Research Reports they are going

to use in that case. In fact, the need for reading and interpreting graphs in a scientific way, as in the case of academic IELTS (International English Language Testing Services) exams reports, is not present here. Few rubrics mention a previously conducted survey and include its results. These may be presented in the form of diary-notes or main points, hypothetically written by the writers themselves. Mainly, the writers are asked to report based on their own views.

[Sample prompt]
You went to <place> for five days to attend your sister's wedding. Look at the extracts from the diary you kept during your trip there. Then write your report on the trip.

Between these two reports in language testing, this has seemed to me the most prevalent difference and has been the reason for coining the names 'Personal Observation Report' and 'Data Report' respectively.

In general, these reports have the following structure:

^ Introducing the subject and stating the purpose of writing ^ Description of key features (in different sections) ^ Conclusion (± Recommendation + Justification of recommendation).

The presence of headings and subheadings is a characteristic feature of this genre. Sometimes there is a main heading and other times there is some introductory data before the text is divided into smaller parts. In table 5.10, we can see that in 22,3% of the total texts there is use of section headings only. A main heading is used in 28,9% of these reports and it is combined with section headings in 26,3 % of the texts. Nearly half of these reports (44,6%) include introductory data in a steady pattern of [To (recipient/reader) + From (writer) + Date +Subject]*,* showing the emphasis for specificity and factual information in this genre right from the start. This is combined with section headings in 38,1% of the texts. Finally, in 4,2 % of the reports there is neither use of main headings, section headings nor data and the layout resembles any type of essay that is, the text is simply divided by paragraphs.

Table 5.10 Headings in the 'Personal Observation Report' sub-corpus

Section heading only	Main Heading		Introductory data		Nothing
22.3%	28.9%		44.6%		4.2%
	M.H*. + S. H.*	only M.H	I.D.* + S. H	only I.D.	
	26.3%	2.6%	38.1%	6.5%	

*(M.H.= Main heading, S.H.= Section heading, I.D.= Introductory data)

In cases where there are sub-headings their role is to divide the texts in thematic sections and classify interesting parts of the object/place/person described. Sometimes these sections are defined by the question itself:

[Sample prompt]
You are a secretary for a primary school. The headmistress has asked you to assess a newly established local wildlife park to see if it is suitable to use for field trips. Write your report, describing <u>what there is to see</u> and <u>do</u> there and what the <u>facilities</u> and <u>prices</u> are like.

However, most of the times the writers have to decide on their own and prioritise on the qualities they are going to focus on. This is an important step as it will define the outline of the report and information included in these sections will be used later on to justify recommendations. Since there is a word limit in these tests careful planning at this stage seems necessary. The word 'report' is often used in main headings and nominalisation [e.g. 'Introduction', 'Conclusion', 'Recommendation'] helps writers create short sub-headings and 'pack' more meaning in sentences. Thematically speaking there are two categories which appear frequently using varied combinations of words. One of them has to do with 'cost' and the other with 'strengths and weaknesses' as shown in table 5.11.

Table 5.11 Common thematic categories in section-headings of the 'Personal   Observation Report' sub-corpus

'Strength/weaknesses' headings	'Cost' headings
Weaknesses	Prices
Strengths	Cost
Positive features	The food and the prices
Positive/ Negative points	Projected costs
Positive/ Negative sides	Prices and service
Positive/ Negative aspects	Availability and cost
Problems	Room prices
Problem areas	Quality and price
	Running costs
	Cost and service
	Facilities and prices
	Prices young people can afford

Going back to the prompts we notice that the writer may be working for a magazine or may be a student representing the college students, a company employee, a pupil privileged to be asked by teacher, a part-time employee in a place where young people spend time or a member of a committee. In other words, the writer has been chosen to write a report because of particular abilities or roles in the community. The addressee is usually of higher status; it can be the editor of a magazine, the principal, the company manager, the coordinator, the chairman of the local board of school governors, the visiting group-leader, the teacher or head teacher. The writer is trusted to be critical and objective based on hypothetical previous evidence related to his/her performance. In some cases, the ability of the writer to combine studies with a part-time job may also be used as a trick to add extra competence on his/her profile and real-life experience gained from the working sector. However, most of the times the interest remains local and the writer knows the hypothetical reader of the report. There is some status on the part of the writer and he/she usually reports to someone with even more status. This is why formality is non questionable and why specific, factual language is needed.

As seen in table 5.11, it is common in these reports to mention prices. This is done by using specific numbers and percentages and referring to specific currencies. Numbers are also used

for dates and ages. However, quantification may also involve grouping sizes in a more general sense (e.g. 'a wide range of', 'a number of', 'many', 'several', 'few').

Humans are identified by name and surname; specific locations are also named ('restaurant', 'school', 'sports club', 'street', 'shopping centre'). When humans are not named they are grouped together in distant and general categories ('students', 'people', 'teenagers', 'children', 'teachers', 'visitors', 'staff', 'group'). Whether it is for people, places or facilities the adjective 'local' is a keyword (table 5.12) which shows the proximity as a common cause for the person chosen to report. Figure 5.5 shows the most frequent collocates of 'local' on the right.
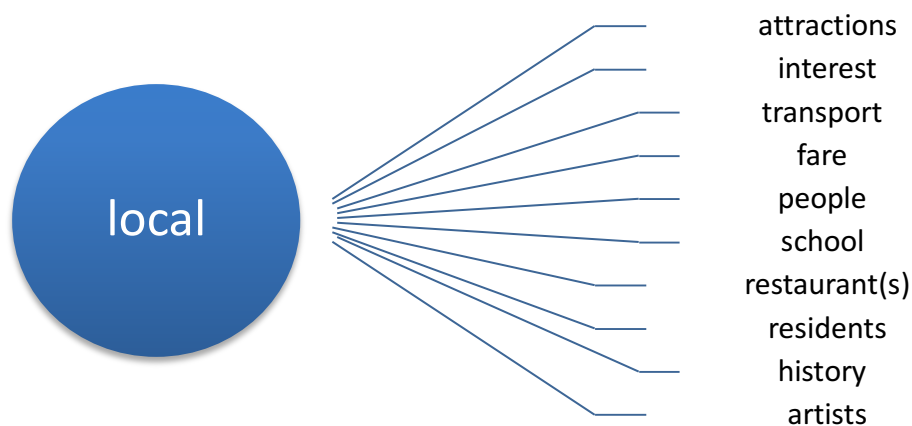
Figure 5.5 R1 collocates of 'local' in the 'Personal Observation Report'

The fact that the definite article which is normally found in any text of any genre is a keyword here shows the degree of specificity in this genre (table 5.12).

Table 5.12 Keywords in the 'Personal Observation Report' sub-corpus

	Positive				Negative	
Key word	**Freq. %**	**Texts**	**RC freq. %**	**Keyness score**	**Key word**	**Keyness score**
report	0.41	59	0.03	192.15	you	-97.45
introduction	0.26	46	0.02	115.90	I	-71.45
purpose	0.16	28	0.02	65.97	me	-37.77
subject	0.18	31	0.02	63.71	my	-33.87
students	0.49	26	0.17	63.44	your	-26.30
recommendations	0.13	18	0.01	62.86	life	-25.11
prices	0.19	20	0.03	62.75	world	-19.06
<name>	1.39	54	0.79	60.07	if	-18.73
conclusion	0.30	45	0.08	57.71	just	-18.26
recommend	0.17	25	0.03	47.34	know	-16.80
range	0.15	17	0.02	49.43	have	-16.38
facilities	0.17	18	0.03	47.34	so	-15.90
date	0.13	24	0.02	46.37		
the	6.16	76	4.99	45.02		
restaurant	0.17	13	0.03	44.99		
centre	0.21	18	0.05	43.06		
menu	0.10	9	<0.01	41.24		
service	0.17	15	0.04	40.43		
survey	0.09	9	0.01	33.33		
aim	0.08	14	<0.01	33.05		
offers	0.11	17	0.02	32.05		
restaurants	0.10	9	0.02	29.21		
<surname>	0.45	38	0.22	28.74		
recommendation	0.06	11	<0.01	28.62		
town	0.18	15	0.06	27.78		
local	0.24	22	0.09	26.18		
food	0.24	16	0.09	26.18		
assess	0.06	10	<0.01	25.82		
wide	0.10	14	0.02	25.44		
suggestions	0.08	10	0.01	25.34		
excellent	0.12	17	0.03	24.41		
meals	0.08	9	0.01	23.54		
staff	0.10	12	0.02	21.42		
reasonable	0.06	9	<0.01	18.37		
visitors	0.10	12	0.03	16.90		
atmosphere	0.09	11	0.02	16.79		

Recommendation is a key stage in these reports and often specifically expressed as a section-heading. Then writers may choose to justify their recommendation more clearly. Depending on the type of the question, justification may appear close to the end of the text if the purpose has been to present/describe and then recommend and justify or it may be present throughout the text when the question has asked mainly for recommendations. The words 'recommendation(s)', 'recommend' and 'suggestions' are keywords in this genre (table 5.12).

Adjectives are essential during the recommendation stage. They are used to describe existing positive qualities ('excellent', 'reasonable') and then recommend the place/person because of them, or to suggest ways of improvement so that the place/person will acquire these qualities. 'Well + past participle' is frequently used to evaluate (e.g. 'a well-stocked bookshop', 'the complex is well-managed', 'a large well-equipped gym'). Positive connotation adjectives outnumber those with negative connotations.

E.g. *New computers with Internet connections would be an <u>excellent</u> resource for students.*

Modals are also used to suggest improvements, mainly 'could' and 'should' *(e.g. although improvements <u>could</u> certainly be made, students <u>should</u> be encouraged to).* Conjunctions such as 'because', 'as', 'so'*,* are used to justify the recommendations made:

E.g. 1. *There are two attractions that may be taken into consideration, <u>as</u> they seem suitable.*
E.g. 2. *My suggestion is to visit the <location> <u>because</u> it is special.*

Out of the texts used in this sub-corpus 64% include the pronoun 'I' and 33% the pronoun 'we'. In figure 5.6 the most frequent collocates of 'I' show some participation in visiting the places mentioned as well as instances where personal beliefs, opinions and suggestions are expressed.
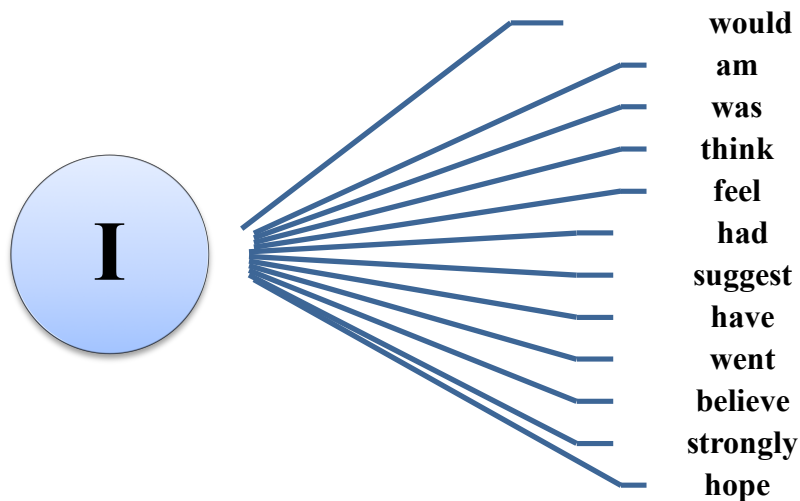
Figure 5.6 R1 collocates of 'I' in the 'Personal Observation Report' sub-corpus

Even though writers are allowed to express personal views and in some cases are specifically asked to do so, it seems that this is done mostly by describing and presenting features rather than imposing personal opinion and being openly involved. Through description and evaluation, the writers lead the reader to the conclusion that what is recommended is in fact the best option available. Modals often make these suggestions sound mild (e.g. 'I would recommend', 'may be taken into consideration'). While there is some subjectivity in these reports, this is discreet and controlled.

5.2.5. The Data Report

Tasks of this kind usually ask the writers to 'summarise' the information provided selecting and reporting only the key points. Writers also need to understand where a comparison is possible and useful in presenting this information.

[Sample prompt]
The table shows the Proportions of Pupils Attending Four Secondary School Types Between Between 2000 and 2009. Summarize the information by selecting and reporting the main features and make comparisons where relevant.

The stages in general adhere to the following pattern:

^ Introducing the subject and stating the purpose of writing ^ Description and comparison of key points (from overall trend to specific features or simply main features) ^ Conclusion (summary or overall trend).

Where the second stage begins stating the overall trend and then going to smaller features this is not repeated at the conclusion. When the writer starts presenting the main points the text usually ends with a presentation of the overall trend.

Topics refer to social issues and tendencies of general interest which are measurable. Money spent and income, travelling habits and employment issues, electricity or water consumption, crime and punishment are some examples. The information provided requires a basic worldview on the part of the reader and some knowledge on reading tables and graphs and therefore it is not appropriate for very young people. However, there seems no requirement of pre-existing specialised or scientific knowledge for the reader to be able to comprehend the information offered.

Concerning the type of data, the prompts show a preference for bar charts and diagrams. There are also a lot of tasks in which the data is combined, the candidate may need to read and understand two or three same or different types of data visualisation. Figure 5.7 shows the type of data provided in the reports of this sub-corpus. Sometimes there are more than two variables to be compared raising the complexity level of the task. For example, presenting 'the electricity consumption of two countries in a year', is less complicated than 'the hours worked and the stress levels of workers in eight different countries'. Considering the limited size of the text asked, the candidates need to carefully select and group factors according to similarities. This should be done for example, when the prompt presents 'the Olympic medals in 12 different countries' or 'the production of fruit and vegetables in 27 countries of the European union'.
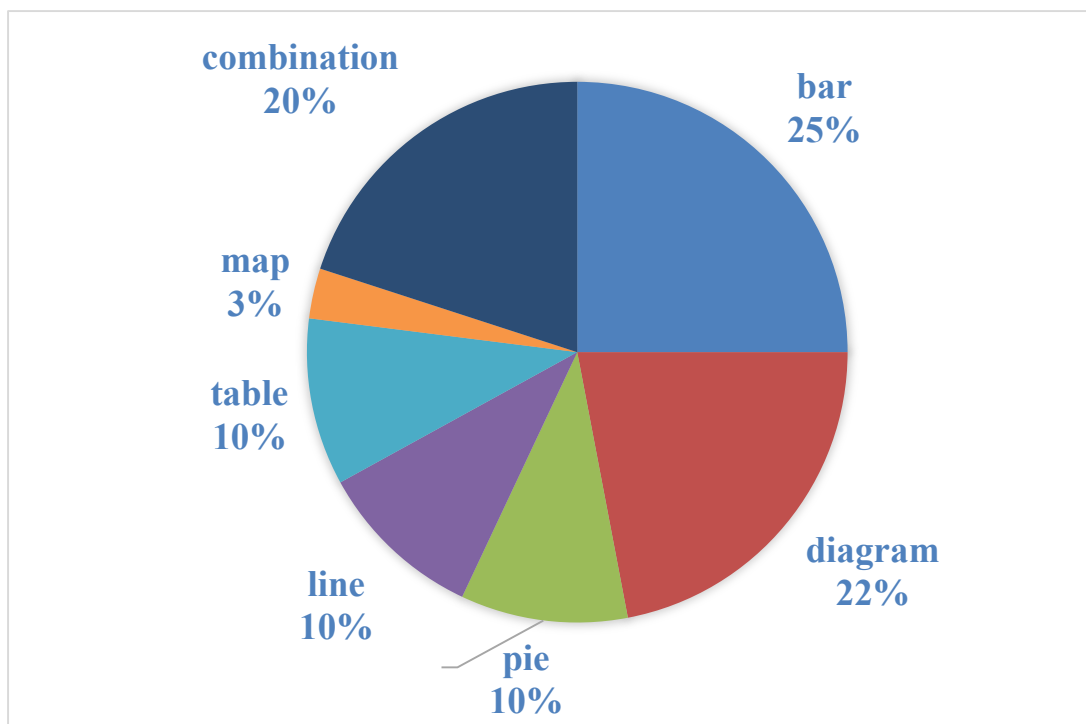
Figure 5.7 Type of data provided in the 'Data Report' sub-corpus

This report is always based on specific data. Therefore, factual language and objective interpretation of the facts is required. Words referring to the type of data such as 'chart', 'graph', 'bar', 'line', 'pie' and 'diagram' are keywords (table 5.13). Description usually begins with the [type of data + shows] (51 instances in 43 texts), *or* [type of data + illustrates] (16 instances in 16 texts).

Information refers to specific facts and the definite article is a keyword (table 5.13). This means that it is a statistically significant word, typical of the report genre. Its keyness score though is a lot higher than the one seen in the 'Personal Observation Report' (table 5.12, 5.13). 'The' occurs in several formulaic expressions. I present the most frequent collocates of 'the' and the extent of its patterning with specific words in figure 5.8 and table 5.14.

Table 5.13 Keywords in the 'Data Report' sub-corpus

Positive					Negative	
Key word	Freq. %	Texts	RC freq. %	Keyness score	Key word	Keyness score
# (number)	6.36	89	0.75	2,494.94	be	-107.29
the	8.29	100	4.99	320.20	we	-101.95
over	0.69	54	0.12	204.16	have	-73.74
period	0.37	35	0.04	161.74	not	-69.03
in	3.43	99	2.05	132.04	if	-62.78
chart	0.26	35	0.02	129.10	to	-49.03
number	0.48	31	0.09	124.17	so	-48.33
shows	0.28	43	0.03	110.84	many	-45.26
of	3.74	98	2.46	99.06	they	-41.94
respectively	0.19	30	0.01	92.96	some	-39.07
approximately	0.19	23	0.02	87.30	will	-34.19
overall	0.21	33	0.02	86.70	that	-32.17
year	0.39	33	0.09	85.20	do	-29.05
graph	0.16	25	0.01	81.37	can	-25.66
process	0.23	16	0.03	79.09	them	-24.55
than	0.53	48	0.17	75.34	it	-23.35
around	0.35	40	0.08	74.63	world	-19.24
figures	0.17	19	0.02	73.93	make	-18.81
increased	0.21	21	0.03	73.63	when	-17.90
percentage	0.16	15	0.01	69.56	well	-16.68
proportion	0.15	22	0.01	68.85	are	-16.25
bar	0.18	29	0.02	68.77		
total	0.16	17	0.02	63.87		
between	0.26	32	0.06	60.90		
from	0.87	68	0.42	59.16		
half	0.17	19	0.02	58.23		
slightly	0.13	21	0.01	55.64		
line	0.15	18	0.02	53.97		
stages	0.11	13	<0.01	53.30		
per	0.17	16	0.03	52.39		
pie	0.11	17	<0.01	51.47		
contrast	0.14	25	0.02	50.65		
highest	0.11	14	<0.01	47.63		
similar	0.12	18	0.02	45.27		
rose	0.09	14	<0.01	44.88		
charts	0.09	15	<0.01	44.88		
figure	0.10	17	<0.01	44.85		
by	0.68	62	0.34	44.08		
three	0.21	29	0.05	43.90		
illustrates	0.09	16	<0.01	42.50		
trend	0.12	16	0.02	40.37		

amount	0.16	16	0.03	40.12		
increase	0.14	20	0.03	39.37		
four	0.13	22	0.02	39.36		
were	0.47	40	0.21	39.10		
higher	0.15	16	0.03	38.30		
years	0.28	29	0.09	38.17		
whereas	0.11	19	0.01	37.39		
under	0.16	21	0.03	36.98		
diagram	0.07	12	<0.01	36.46		
at	0.84	67	0.48	36.46		
just	0.30	34	0.11	35.39		
almost	0.16	22	0.04	32.84		
remained	0.08	12	<0.01	32.83		
compare	0.07	13	<0.01	30.10		
only	0.40	42	0.18	29.93		
next	0.21	28	0.07	29.37		
main	0.17	16	0.05	28.56		
while	0.29	42	0.12	27.60		
significantly	0.07	12	<0.01	27.55		
second	0.13	18	0.03	26.55		
rise	0.09	14	0.02	25.30		
less	0.19	19	0.07	24.34		
popular	0.14	15	0.04	24.33		
and	3.45	99	2.80	23.95		
nearly	0.08	12	0.01	21.79		
five	0.10	12	0.02	21.08		
countries	0.19	21	0.08	20.08		
end	0.12	14	0.04	19.83		
clear	0.11	19	0.03	18.35		
both	0.24	34	0.11	18.26		
used	0.16	14	0.06	17.67		
two	0.22	32	0.10	16.59		
different	0.19	28	0.09	16.44		
given	0.11	15	0.03	15.80		
most	0.33	41	0.18	15.49		
show	0.09	13	0.03	15.22		

(table 5.13 continued)

in (134)	<location name> (44)
the (1490)	
to (57)	number (63)
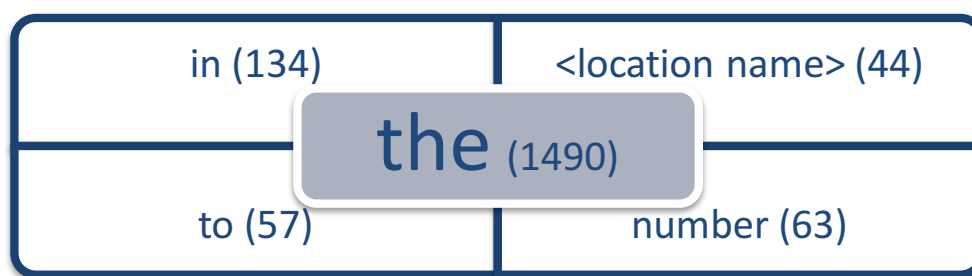
Figure 5.8 Collocates of 'the' in the 'Data Report' sub-corpus

Table 5.14 Three-word clusters including 'the' in the 'Data Report' sub-corpus

The number of	Of the population	Beginning with the
The bar chart	The percentage of	In the process
End of the	The line graph	In the number
Of the period	Over the next	Half of the
Over the period	East of the	Stage in the
In the <location name>	The pie chart	Than the other
The end of	The figures for	The pie charts
The proportion of	At the end	The internet in
Part of the	Most of the	<location name> and the
The most popular	Of the village	

Humans are either identified specifically by name or are grouped together as 'people', 'population', 'visitors', 'men' 'men and women'. The type of data itself categorises people in groups according to nationality or age (e.g. People aged 60 or over). There is a large proportion of named locations and time is presented in specific years, periods or hours using information from the prompt. The preposition 'in' is used in recurrent patterns of [in + specific years/months].

There is extensive reference to numbers related to time, age and humans. Actual numbers are preferred over numbers in words. Trying to see if there is a systematic preference for the use of numericals for some categories I noticed that they are preferred for years (2010), decimals (22.5), and numbers in tens (60), hundreds (600) or thousands (6000), probably in an effort to cut down on the number of words used. For very big numbers the combination of [numerical + million/billion] is systematically used. Words are often used for small numbers (one to six). The word 'number' is everywhere especially in the cluster [the number of]. Other expressions

of measurement are seen frequently in 3-word clusters or n-grams such as [the amount of], [the figures for], [in the number], [average number of].

Sometimes approximation is preferred over specific numbers possibly for variation in the language of the text or in order to show the ability to interpret and explain a graph without copying the exact information provided. The prepositions 'over' and 'under' are modified with 'just' to present round numbers or percentages (70% instead of 69,1% for example) as seen in figure 5.9. The numbers show the instances where each word precedes or follows the central one (two-word combinations). 'To' shows the change of the numbers whereas 'at' shows the end of the change. In these cases, 'at' is preceded by 'finish', 'remain' or 'stood'. The patterns [At+ adverb ('approximately', 'about', 'around')] and [in almost] are also used to express approximation.

E.g. 1. *Minutes spent on local calls fluctuated over the time period, with just over 70 billion minutes in 1990, peaking at approximately 90 billion in 1994 and then steadily decreasing <u>to just over</u> 70 billion minutes in 2000.*
E.g. 2. *It started lower <u>at about</u> $350 per month, falling in the following month, and then increasing significantly to finish <u>at just under</u> $600 in April].*
E.g. 3. *Turning to the bar chart, poverty rates were highest amongst children, and the rates were roughly equal for males and females, <u>at around</u> 21% for under 5s and 15% for 5-17 year olds.*
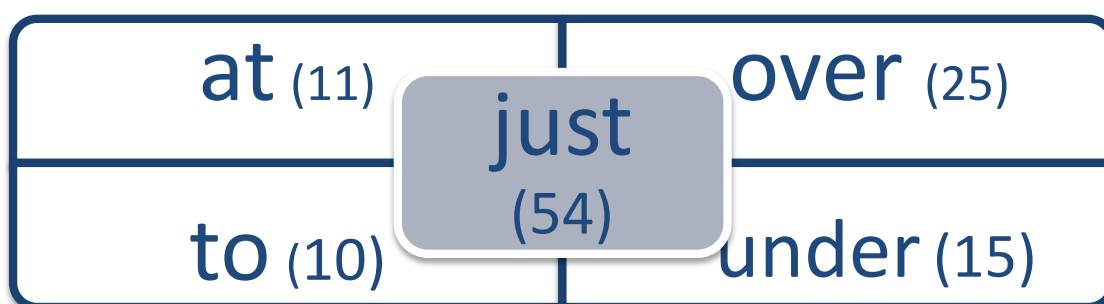
Figure 5.9 Collocates of 'just' in the 'Data Report' sub-corpus

Another way of showing complete understanding of the data and to avoid a text that looks like an accountant's spreadsheet is to refer to numbers as fractions. Several clusters (table 5.15) show part of the total entity measured.

Table 5.15 'Of' clusters showing a fraction in the 'Data Report' sub-corpus

The proportion of	Part of the	Of total
Of <nationality>	Most of the	A quarter of
The percentage of	Half of the	

Comparison is a key element in this genre. The pattern [the + (type of data in plural) + compare(s)] often initiates comparison (e.g. <u>The graphs compare</u> the proportion of various nutritional compounds contained in two different foods; macaroni and medium baked potato). Clusters with the word 'contrast' such as [in contrast], [in contrast to], [by contrast], are also frequent. The nouns 'trend' and 'pattern' and the phrases [similar pattern], [opposite pattern/ trend], [different pattern] or [upward trend] show the direction of change. 'While' and 'whereas' connect clauses expressing contrast.

Verbs indicating movement are mostly in the past simple and express upward trends ('increased', 'rose'), downward trends ('decreased', 'dropped'), both ('fluctuated') or stability through different periods of time ('remained'). When this change is significant adverbs are employed to illustrate the degree or speed of this change [increased dramatically/ significantly/ steadily]. Time is specified with the patterns [between (year) and (year)], [over the period], [over the next + year(s)] and [from (year) to (year)]. The example below contains all three elements:

E.g. The facts show that the average number of cars passed on <name of road> <u>increased significantly from 1993 to 1999</u>, the same year in which methods to slow down traffic was introduced'.

Comparative and superlative forms of adjectives ('more', 'less', 'higher', 'highest', 'largest', 'least', 'lower', 'lowest') also show the degree of change. 'More' and 'higher' are much more frequent than 'less' and 'lower', showing a preference for the description of upward trends (table 5.16). Comparatives are sometimes premodified by adverbs as in the pattern [slightly +

comparative adjective + than], and 'only' is sometimes used to emphasise small numbers in trends:

E.g. Almost half of the commuters take less than 15 minutes to get to their place of work, while <u>only around 3 per cent</u> spend over an hour commuting].

Table 5.16 Comparatives and superlatives showing the degree of change in the 'Data Report' sub-corpus

Comparatives and superlatives	Freq. # / # of texts
more	68/38
less	39/19
higher	27/16
highest , least	19/14
Largest, lower	10/10
lowest	10/8

These tasks have a limited text length. Words need to be carefully selected to present the information provided, highlighting the main points. The word 'overall', a keyword of the genre, often occurs in the second sentence (first sentence includes the type of data and the topic of the report). It is used early enough in order to give the reader the main trend, the conclusion, before the writer goes further to report on change over time, subjects and location. This satisfies a practical need simulating perhaps the workplace report where professionals may wish to find out the outcome and decide if they want to go into detail:

E.g. *The bar chart depicts the monthly expenditure on food, gas and clothing of a family living in the <location> in 2010. <u>Overall,</u> it can be seen that levels of expenditure fluctuated over the period.*

Participles (e.g. 'the given bar graph', 'the information shown', 'the minutes spent') and nominalisation (e.g. 'expenditure', 'consumption', 'production'), are used by writers in order to create dense and meaningful sentences. The use of agentless passive voice adds a distant tone in the text (e.g. The greatest increase can be observed in the total number of people

coming from <location>). Writers are expected to report retaining from comments and any expression of view. The rarity of personal pronouns and other stance markers such as modals or attitude adverbs make this genre the most impersonal of all the genres studied in this context.

It is necessary to say that although the previous findings are well spread across the corpus I found a small number of elements that were constantly emerging in reports that referred to a procedure. These task types illustrate a process or a sequence of stages (e.g. 'the carbon cycle', 'how solar panels provide electricity', 'waste paper recycling') and the language used can be slightly different. Temporal connectors such as 'next' and 'then' which occur in the Data Report are a lot more frequent in this case:

E.g. 1. Next, the mixture is frozen and then passed once again through the grinder
E.g. 2. *These beans are then dried, roasted, and cooled before being put in a grinding machine, which turns the beans into coffee granules*

The nouns 'stage(s)' and 'process' are seen only in procedures:

E.g. *It is clear that there are six distinct stages in this process, from the initial collection of waste paper to the eventual production of usable paper.*

Numbers, past simple verbs or the comparison element are not key features here and may not be seen at all throughout the texts. This is a kind of less typical reports mostly based on diagrams, flow charts or pictures, where ordering and sequencing is necessary and there is more need for explanation based on detailed description rather than quantification. Clearly, in teaching contexts this differentiation needs to be shown and exemplified.

5.2.6 The Short Story

Basic stages in a story include the 'Orientation' where the reader learns about the setting (where, when and who), followed by the 'Complication' stage where the lives of the characters are complicated and problems arise. Then the complication may be resolved for better or worse at the Resolution stage (Board of studies NSW, 1998: 37; Derewianka, 1990:42). Abbs & Richardson (1990: 107) name the first part 'Exposition' and add the 'Climax' stage between the Complication and the Resolution stage. Writers need to include all the previously mentioned stages in these short texts in order to build a good story. Unskilled writers can easily get carried away while creating imaginary settings and interesting plots. Therefore, it is a challenging task to control text, paragraph and sentence lengths taking care to include all necessary stages at the same time.

Nearly a quarter of the stories in the corpus have a two to three-word title. Almost half of these cases are responses to prompts asking for a specific title (e.g. write a story entitled…). We can say that in cases where the tasks do not ask for a title, the majority of writers in this corpus do not add one in their texts.

One of the most common elements is the emphasis on people as agents or people as recipients of what is happening. Characters are often named and there is great involvement of relatives and friends (table 5.17). Surnames are less frequent as the people involved usually belong to the writer's immediate context. Names were found in 40 texts while only 6 texts included surnames. In fact, <surname> is a negative keyword. Location can also be named but not as often as humans. Stories take place in everyday places (home', 'school'). It seems that for the majority of the stories studied here, scenes taken from everyday life are sufficient and there is no need for imaginary or unfamiliar settings.

Table 5.17 Keywords in the 'Short Story' sub-corpus

Keyword	Freq. %	Texts	RC freq. %	Keyness score	Keyword	Freq. %	Texts	RC freq. %	Keyness score
Positive key words									
was	3.01	77	0.55	747.41	later	0.13	21	0.03	29.44
he	1.42	43	0.20	426.38	climbed	0.06	9	<0.01	29.31
had	1.51	64	0.24	417.18	after	0.27	30	0.10	28.67
I	3.95	65	1.59	363.02	lost	0.11	13	0.02	28.50
my	1.67	53	0.55	212.30	into	0.29	32	0.11	28.47
his	0.60	27	0.12	137.69	happened	0.09	13	0.01	28.41
she	0.65	32	0.18	106.31	nervous	0.08	12	<0.01	28.33
suddenly	0.24	36	0.02	105.60	<name>	1.21	40	0.79	27.69
we	1.26	43	0.55	99.17	front	0.10	13	0.02	27.45
felt	0.26	29	0.03	89.29	quickly	0.10	16	0.02	26.94
said	0.31	29	0.05	84.49	bed	0.07	10	<0.01	26.41
me	0.71	44	0.25	80.95	took	0.14	19	0.03	26.24
saw	0.26	29	0.04	77.30	soon	0.21	27	0.07	26.03
got	0.27	33	0.05	63.07	arrived	0.10	15	0.02	25.90
him	0.29	24	0.06	62.38	surprise	0.08	10	<0.01	25.50
walked	0.13	17	<0.01	60.52	smiled	0.06	9	<0.01	25.14
her	0.44	22	0.14	59.12	relieved	0.06	9	<0.01	25.14
back	0.26	31	0.05	57.75	at	0.79	62	0.48	24.97
looked	0.18	24	0.02	57.50	window	0.06	9	<0.01	24.32
were	0.56	46	0.21	55.76	couldn't	0.06	9	<0.01	24.32
went	0.25	30	0.05	55.04	fell	0.08	10	0.01	23.54
door	0.17	18	0.02	54.58	down	0.15	22	0.04	23.51
morning	0.18	23	0.03	53.73	friends	0.24	20	0.09	22.90
day	0.40	40	0.13	53.32	thought	0.12	15	0.03	22.71
didn(t)	0.13	15	0.01	53.24	trip	0.12	13	0.03	22.56
decided	0.17	22	0.02	52.83	reached	0.09	14	0.02	21.90
started	0.17	24	0.02	52.22	no	0.27	26	0.12	21.27
told	0.17	17	0.02	50.71	lucky	0.06	10	<0.01	21.25
asked	0.17	23	0.03	49.88	tried	0.08	11	0.01	21.24
realised	0.13	18	0.01	49.51	off	0.14	18	0.04	20.93
out	0.40	44	0.13	49.44	slowly	0.06	9	<0.01	20.59
then	0.33	38	0.10	48.16	mother	0.11	10	0.03	20.35
when	0.55	53	0.23	45.60	stopped	0.07	9	0.01	20.25
came	0.15	17	0.02	44.98	going	0.18	21	0.06	19.95
ran	0.10	14	<0.01	43.38	us	0.27	17	0.12	19.56
but	0.70	60	0.34	43.25	car	0.17	11	0.06	18.59

eyes	0.13	13	0.02	42.51	minutes	0.10	14	0.02	17.66
turned	0.13	19	0.02	41.16	up	0.33	38	0.17	17.65
heard	0.13	15	0.02	40.75	man	0.09	9	0.02	17.50
knew	0.12	17	0.01	38.25	teacher	0.10	10	0.03	17.28
dark	0.10	13	0.01	37.75	strange	0.06	9	<0.01	17.20
opened	0.12	14	0.01	37.57	excited	0.06	9	0.01	17.09
shouted	0.08	11	<0.01	37.29	towards	0.08	12	0.02	17.05
so	0.60	53	0.29	36.54	been	0.34	32	0.18	16.87
wanted	0.13	17	0.02	35.23	left	0.10	12	0.02	16.71
noticed	0.09	14	<0.01	33.54	seemed	0.06	9	<0.01	16.62
began	0.09	14	<0.01	33.54	never	0.14	20	0.05	16.50
woke	0.07	9	<0.01	32.36	walking	0.07	11	0.01	16.32
friend	0.14	12	0.03	30.59	set	0.10	13	0.03	15.58
explained	0.07	11	<0.01	30.14					

Negative keywords

Keyword	Keyness score	Keyword	Keyness score	Keyword	Keyness score
is	-349.60	this	-52.89	may	-22.62
are	-173.42	more	-49.70	would	-21.79
be	-100.26	which	-49.33	need	-17.94
of	-81.27	or	-48.86	much	-17.94
will	-80.59	also	-47.30	such	-17.79
people	-79.75	should	-43.50	<surname>	-17.19
you	-76.05	in	-42.15	make	-16.30
can	-72.67	your	-30.40	however	-15.72
have	-63.92	many	-30.24	from	-15.65
has	-63.09	for	-25.77	most	-15.40
their	-58.03	not	-24.29		
# (number)	-55.69	if	-23.62		

(table 5.17 continued)

The pronouns 'I' and 'my' are keywords in this genre (table 5.17) and they are a lot more frequent than the rest of the pronouns. Table (5.18) shows the frequencies for personal and possessive pronouns and the number of texts they occurr in.

Table 5.18 Pronouns in the 'Short Story' sub-corpus

Personal pronouns	freq./ # of texts	Possessive pronouns	freq./# of texts
I	617/65	My	260/53
It	222/71	His	94/27
He	222/43	Her	69/22
We	197/43	Our	48/25
She	102/32	Your	18/13
they	61/30	Their	14/11
you	49/27	its	2/2

The narrator is often actively engaged in the story and shares own real or imaginary experiences. When a story is told through first person narration "the storyteller adopts the voice of autobiography and the listener or, reader, imagines that he is listening to an intimate first-hand account of experience" (Abbs & Richardson, 1990: 127).

Studying the questions posed we are reminded that tasks often constrain the writer in the type of narration to use. For example, 'The most exciting experience of my life' leaves no choice other than first-type narration and 'Write a story that ends with the words: She knew the events of the day would change her life forever', calls for third-type narration. In third type narration, "the author presents the story but remains invisible. The characters are 'out there'. We are introduced to them and learn about them as the narrative unfolds" (Abbs & Richardson, 1990:164). I measured the instances where based on the question, the writer has a choice on the type of narration and found that only 41% of the questions in this corpus really offers this choice (table 5.19). First-person narration is explicitly asked in 52% of the prompts and only 7% ask for third-person narration. Where choice is offered, 62% of the answers are written as first-person narration. This can be an important finding for teachers preparing students for story writing, as there is little need for practicing 3rd person narration.

Table 5.19 Prompts and type of narration in the 'Short Story' sub-corpus

Prompts and type of narration	% of prompts in the corpus
1st person narration	52
Free choice	41 (of which 62% written as 1st person narration)
3rd person narration	7

A typical feature of this genre is the fast-paced action; therefore, verbs are crucial. The ratio for nouns to verbs in the corpus is 1:1. In a study on university registers, Biber (2006: 46) finds written registers using nouns to a much greater extent than any other word class and spoken registers using nouns and verbs to about the same extent. From this perspective, the language used in Short Stories seems to be closer to spoken than written registers.

As it is normally expected in storytelling, most action takes place in the past. However, when quantified, the extent of this phenomenon is astonishing. I measured the frequency of lexical verbs (not including auxiliaries) used to form four active voice tenses: Present simple, Past simple, Present Perfect simple and Past Perfect simple. Past simple verbs far outweigh other verb tenses as figure (5.10) shows.
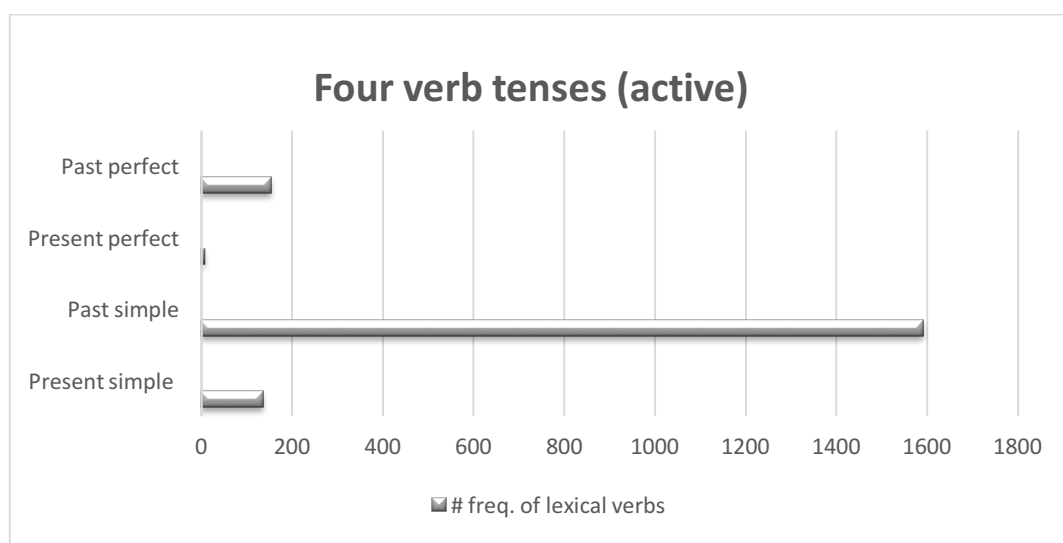
Figure 5.10 Four verb tenses (active) in the 'Short Story' sub-corpus

Action verbs predominate but there are also verbs referring to what humans 'felt' or 'said'. Table (5.20) shows the most frequent verbs related to their function. Verbs in bold are also positive keywords, meaning that they are typical of this genre and statistically they have less chance of appearing in any of the other genres involved in this study.

Table 5.20 Frequent verbs and their function in the 'Short Story' sub-corpus

Functional category	Verbs	# freq./ # of texts	Verbs	# freq. / # of texts
Material (action)	**got**	42/33	**left**	15/12
	went	39/30	**ran**	15/14
	go	32/24	**began**	14/14
	going	28/21	**reached**	14/14
	started	27/24	find	13/12
	get	26/20	found	13/12
	came	23/17	help	13/13
	do	22/21	**fell**	12/10
	took	22/19	**tried**	12/11
	walked	21/17	**stopped**	11/9
	turned	20/19	**walking**	11/11
	opened	18/14	**woke**	11/9
	lost	17/13	**climbed**	10/9
	made	16/15	come	10/9
	set	16/13	getting	10/10
	arrived	15/15	walk	10/9
Mental & Behavioural	**felt**	41/29	**thought**	19/15
	saw	40/29	**knew**	18/17
	looked	28/24	looking	17/15
	decided	27/22	know	15/11
	realised	25/23	**noticed**	14/14
	heard	21/15	like	12/10
	see	21/16	remember	11/10
	wanted	20/17		
Verbal (saying)	**said**	49/29	**shouted**	13/11
	asked	27/23	called	11/10
	told	26/17	**explained**	11/11
Relational (being, having)	**was**	469/77	**been**	53/32
	were	87/46	have	28/20
	is	19/15	are	14/12

Narratives construct a pattern of events with a problematic and/or unexpected outcome that entertains and instructs the reader or listener (Board of studies NSW, 1998: 37). Reflecting on experience or imaginary situations, they aim to hold the reader's interest with the unexpected development of events. Writers manipulate language to create a feeling of suspense in their stories. Nouns such as 'surprise' and adverbs such as 'back', 'quickly' and 'suddenly' add to the sense of fast-paced action and change in the pattern of events. 'Back' is often used as the desired end-point, symbolising safety and calm, during or after a negative experience. The following sentences are examples of this use:

E.g. 1. *He knew he had to make it <u>back</u> to the boat…*
E.g. 2. *… as we shakily made our way <u>back</u> to the camp.*

Time related connectors ('then', 'when', 'later', 'after', 'soon', 'as soon as') consistently connect sentences in the middle paragraphs where the action is dense and the reader needs help to follow the narration of the events. 'Suddenly' is a key connector showing that this unexpected change in the pattern of events is not necessary in the other genres included in this corpus. The dispersion plot reveals that 'suddenly' is commonly used to introduce the Complication stage. Certain adjectives used to set the scene are carefully chosen to prepare readers for the 'Complication' stage. 'Dark', for example, is easily associated with possible danger. It is often found close to 'cool' or 'cold'.

E.g. 1. *While we were walking along the beach, we saw a cave, so we went inside. It was <u>cool and dark</u>. I turned on the torch and my friends followed me.*
E.g. 2. *Soon it got <u>dark and cold</u> and I had nowhere to go.*

A rather unexpected genre-specific word is 'door'. It is frequently associated with a person's entrance to the scene as well as with strong feelings.

E.g. 1. *Then the front <u>door</u> opened and an old lady came out.*
E.g. 2. *I was so excited that I slammed the front <u>door</u> behind me and ran away.*

Olson (2012: 607) talks about the unexpected, exceptional event at the heart of the story giving a sentence which includes the word 'door' and a person entering as an example of this characteristic in stories. He reminds us that "in a story the unexpected occurs. The more unexpected, the better the narrative". It seems that in the stories studied here a way to introduce the unexpected is someone not previously mentioned entering a room. The most frequent collocate of 'door' on the left is the word 'front' and the most frequent one on the right is the word 'opened'. Looking at the plot dispersion, 'door' is mainly used in the middle of the texts. 'Window' is also a keyword used in a similar way but is less frequent (table 5.17).

Finally, the use of contractions (e.g. 'didn't', 'wasn't', 'couldn't'), certain punctuation marks such as exclamation and question marks and dialogue (used in 58% of the texts), support the general liveliness of the texts.

The emphasis on feelings is also a typical feature of this genre. Humans' psychological state is constantly described. In table 5.20, there are a lot of verbs related to feelings and mental state. The dispersion plot of the verb 'felt' indicates that it is mainly used in the middle of the texts but also towards the end. 'Nervous', 'relieved', 'excited', 'proud', 'happy', 'terrible' and 'embarrassed' usually follow the verb 'felt'. In particular, the protagonists usually feel 'nervous' at the 'Complication' stage and 'relieved' or 'not nervous' at the 'Resolution' stage as in the following examples:

E.g. 1. *As the rescuer climbed down carefully and pulled me out, I <u>felt relieved</u>.* (Resolution stage)

E.g. 2. *"You're <u>not nervous</u> about sailing anymore, then," he smiled.*

(Resolution stage)

The word 'eyes', is also key in this genre. As the following examples show it has been used to describe both humans (e.g. 1a, 1b) as well as feelings (e.g. 2a, 2b).

E.g. 1a. … with dark affectionate <u>eyes</u>

E.g. 1b. … with orange skin and huge blue <u>eyes</u>

E.g. 2a. He closed his <u>eyes</u> for a moment, praying…

E.g. 2b. He drove so fast that I kept my <u>eyes</u> tightly shut all the way.

5.2.7. The Complaint Letter

Out of the genres investigated here, the 'letter genres' are perhaps the ones that are more connected to real life, the type of writing that students will most certainly need to engage with during the course of their lives. The 'Complaint Letter' in particular, is connected to some problem the writer has come across, causing trouble and inconvenience, which needs to be resolved as soon as possible in order to alleviate frustration and the feeling of injustice. These are everyday problems: faulty product or delivery, poor facilities or unsatisfactory service. The writer is looking forward to some response, a full or partial refund, a compensatory voucher or even a non-materialistic response such as a simple apology coming from a person of high status in a company and the recognition of their fault. Letters of complaint in this sub-corpus have the following structure:

^ Purpose of writing ^ Reasons for complaining (description of problem) ^ Expectations

A typical feature of this genre is the presence of conventionalised greetings in opening and closing positions. This is usually the most revealing characteristic for the reader to realise that this is a formal letter even if its main purpose is not yet recognised. Two main patterns of opening and closing greetings are found in 66% of the texts in the sub-corpus:

1) Dear Sir/Madam   => Yours faithfully <name> <surname>

   (or Dear Sir or Madam)   (or Yours faithfully Mr (or Mrs) <surname>)
2) Dear Mr (or Mrs) <surname> => Yours sincerely <name> <surname>

The first pattern occurs three times more often than the second one. It is largely accepted therefore, that if the writer addresses the recipients with their surname, this will be signed off with the expression 'Yours sincerely' and if the writer chooses to address readers without naming them (Sir or Madam) the signing off will be 'Yours faithfully'. The rest of the texts use slight variations such as 'Dear Sir(s)' and because three letters addressing newspaper editors have been included I also came across 'Dear editor' or 'To the editor'. But the pattern of opening with an unnamed addressee and closing with the phrase 'Yours faithfully', is the same in these cases too.

Apart from greetings there are a number of keywords and formulaic expressions in these letters which are positioned consistently in certain parts of the text related to certain moves/stages. The first sentence for example is always about the reason for writing and the pattern [I am writing to complain] occurs frequently:

E.g. *I am writing to complain about the dreadful service at the <location> fast-food restaurant in <location> street.*

Then one always sees words related to the problem itself. Approximately one fourth of the letters in the sub-corpus is about bad 'service' and another fourth mentions a dishonest 'advertisement' with 'misleading' claims. Talking about the problem often involves detailed information about times, dates and location. The words 'matter', 'refund' and 'apology' always occur just before the closing at the stage where the writer explains what he/she expects the reader to do (table 5.21).

Temporal connectors are used to sequence events and consequential connectors introduce the justification of the inconvenience caused, the feeling of dissatisfaction and the request for a refund or an apology. Contrastive connectors are employed in order to show the difference between the initial expectation and what was actually offered. The use of these connectors is also consistent with specific moves and positions in the texts. For example, events are set in time in the second stage ('Reasons for complaining') which is the longest part of the letter.

E.g. 1. *After a further long delay, our food finally arrived, but as we began to eat, we realised that it was cold.* (temporal)
E.g. 2. *I was therefore shocked to discover that the same tickets were being sold for 10 pounds each at the door on the day of the concert.* (consequential)
E.g. 3. *According to your advertisement, the place is perfect for holding private conversations in a relaxing atmosphere. However, it turned out that the music was so loud that we could hardly hear each other.* (contrasting)

The writer is directly involved in the 'Complaint Letter' referring to a personal unpleasant experience. The pronouns 'I' and 'my' are keywords in this genre. The most frequent collocates of 'I' are shown in figure 5.11. Words of general reference such as 'people' or 'they' are negative keywords (table 5.21). Writers address the reader directly but using

147

surnames and titles ('Mr', 'Madam', 'sir') in this case. The pronoun 'you' is also a keyword here.

Table 5.21 Keywords in the 'Complaint Letter' sub-corpus

Positive					Negative	
Key word	**Freq. %**	**Texts**	**RC freq. %**	**Keyness score**	**Keyword**	**Keyness score**
I	4.31	47	1.59	271.57	are	-61.03
your	1.56	43	0.34	192.33	people	-49.51
am	0.87	41	0.18	113.57	is	-40.88
yours	0.51	42	0.07	95.25	they	-37.51
sir	0.36	32	0.03	88.91	and	-37.05
madam	0.34	30	0.03	84.58	their	-36.36
writing	0.48	40	0.07	82.35	can	-35.46
was	1.42	37	0.55	80.76	she	-15.19
faithfully	0.35	31	0.03	80.16		
complain	0.23	20	<0.01	78.07		
refund	0.20	16	<0.01	72.34		
dear	0.52	46	0.12	60.75		
<surname>	0.68	45	0.22	50.60		
service	0.26	13	0.04	45.62		
matter	0.20	18	0.02	40.78		
disappointed	0.12	11	<0.01	34.22		
my	1.08	34	0.55	33.93		
forward	0.26	23	0.05	33.62		
advertisement	0.16	9	0.02	33.53		
were	0.55	23	0.21	32.26		
reply	0.12	11	0.01	30.24		
Mr	0.19	14	0.04	24.37		
had	0.54	27	0.24	23.55		
express	0.11	9	0.01	23.55		
holiday	0.17	7	0.03	23.42		
misleading	0.07	6	<0.01	23.29		
dissatisfaction	0.07	6	<0.01	22.19		
attention	0.14	11	0.02	21.97		
sincerely	0.16	14	0.03	21.70		
advertised	0.09	6	<0.01	20.00		
two	0.29	20	0.10	19.39		
look	0.23	20	0.07	18.69		
hour	0.11	8	0.02	18.52		
company	0.14	12	0.03	18.02		
request	0.08	6	<0.01	18.02		
did	0.17	11	0.04	17.86		
unfortunately	0.11	8	0.02	17.28		
furthermore	0.17	14	0.04	16.94		
received	0.09	8	0.01	16.88		
you	1.35	40	0.90	16.81		
matters	0.07	6	<0.01	16.66		
room	0.14	6	0.03	15.73		
feel	0.25	17	0.09	15.60		
trust	0.07	6	<0.01	15.52		
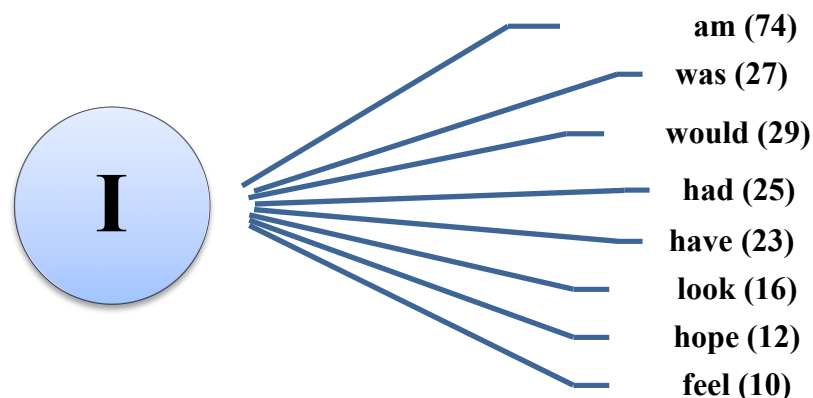
Figure 5.11 R1 collocates of 'I' in the 'Complaint Letter' sub-corpus

Keywords in this genre show the negative feelings spread throughout these texts ('disappointed', 'dissatisfaction', 'unfortunately'). Even though the writer is clearly frustrated the tone remains polite and formal. Writers often trust the recipient to look into the problem and ask for a response within a reasonable time. The words 'trust', 'matter', 'reply' and 'attention' are key words in these letters (table 5.21). The phrase [look(ing) forward to] is very frequent as it signals the transactional character of the letter.

E.g. I trust that you will give this matter immediate attention and look forward to receiving your reply as soon as possible.

Less often these letters include some warning for further action if the writer receives no reply or if the response is not considered satisfactory. In these cases, writers may become more aggressive, it is not however the norm.

E.g. *If an official apology is not forthcoming and I am not compensated in some way with a shopping voucher for example – then be warned that I will take further steps to ensure that I get justice.*

Optimism on the part of the writer that the issue will be resolved is preferred in order to encourage the recipient to act.

5.2.8. The Advice Letter

The 'Advice Letter' refers to the offer of advice to a friend who has previously asked for help on everyday issues or problems. There is a range of problems from vital and pressing ones such as being bullied at school to less serious ones such as visiting a friend not knowing what clothes to pack and what there is to do there. The 'Advice Letters' investigated here, consistently follow these stages:

^ Reference to previous communication stating the problem ^ Offering advice and justification ^ Expression of hope for resolution ^ Request for further communication and updating

Writers open these letters with either 'Dear <Name>' or a simple 'Hi <Name>'. There is no preferred choice characterizing CEFR levels or emails compared to letters. The greetings at the end of the text however, are far less consistent. There is a great range of expressions used to sign off such as [Love/ Best wishes/ Yours/ Lots of love/ All the best + <name>].

The 'Advice Letter' is part of a series of communication exchanges, previous letters and letters expected to follow between the same participants/friends. This continuity is embedded in the stages of the letter. First, by expressing feelings concerning previous communication immediately after the opening greetings with the pattern [I was + (glad/happy/pleased/sorry/ shocked/excited) + to hear] or [it was + (great/ good/nice) + to hear], and secondly at the end of the letter where the writer expresses sincere interest, hopes that the advice offered will be of help and asks to be further informed. In fact, 'hope' and 'soon' are keywords (table 5.22) and the most frequent 3-word lexical bundle is [Let me know].

E.g. *I hope that this advice will be of some help. Write back soon. Let me know how things are.*

Both first singular and second person pronouns are very frequent and are keywords in the 'Advice Letter'. Compared to the 'Complaint Letter' however, it is the other way round. 'You'*,* is far more frequent in this genre; it is the person having the problem who becomes the central figure and 'I' is the person offering help. The word 'here' is often used to introduce the 'Offering advice stage' and the phrase [I think] is often used in this general

introduction before moving into specific advice sounding more confident with phrases such as [I'm sure] and [I know].

E.g. *Anyway, I have looked into the whole accommodation issue for you and <u>here's what I think</u>.*

Writers use the modal verbs 'can' and 'should' to suggest changes in the behavior or the lifestyle of the reader in order to deal with problems. Writers address the readers directly and express their opinion.

E.g. *1. Finally, you <u>should</u> make sure that the place you work makes you feel positive and comfortable.*
E.g. *2. Another thing you <u>can</u> try is joining a gym.*

Although the imperative is present in some letters, it is more frequent for the writer to make suggestions rather than direct the reader even though he/she has been called to help. The writer sometimes puts himself in the reader's shoes using the phrase [if I were you], to make suggestions that would work in his case.

E.g. <u>*If I were you*</u> *I would put my favourite song to wake me up instead of the alarm.*

The writers are usually able to help either because they more experienced in general or because they have had similar problems in the past. The phrase [I know] is very frequent (e.g. <u>I know</u> how terrible it is to be bullied). Due to the close relationship of the writer with the reader, the writer often shows knowledge of certain behavior traits or details of the reader's life.

E.g. *1) <u>I know</u> pretty well that you had a busy year and you were under stress.*
   *2) <u>I know</u> you have a sweet tooth, but maybe you could cut down on sweets and
      eat more fruit and vegetables.*

Table 5.22 Keywords in the 'Advice Letter' sub-corpus

	Positive				Negative	
Keyword	**Freq. %**	**Texts**	**RC freq. %**	**Keyness score**	**Keyword**	**Keyness score**
you	5.21	61	0.90	911.70	the	-85.45
your	1.73	56	0.34	262.55	is	-56.07
(')m	0.45	27	0.03	151.55	of	-47.34
(')ll	0.40	25	0.02	134.48	in	-42.23
(')t	0.60	35	0.09	117.39	#	-39.51
(')s	0.73	34	0.14	117.29	their	-31.78
know	0.61	47	0.12	95.76	people	-29.00
don(t)	0.34	28	0.03	94.97	our	-19.16
I	2.93	59	1.59	87.07	who	-18.68
(')re	0.28	17	0.02	83.38	however	-15.99
(')ve	0.20	13	0.01	66.24		
<name>	1.60	60	0.79	59.75		
dear	0.48	48	0.12	57.72		
let	0.29	27	0.05	53.36		
me	0.70	48	0.25	51.57		
sure	0.28	20	0.05	47.70		
hear	0.22	20	0.03	44.55		
get	0.42	30	0.12	41.69		
lots	0.19	14	0.02	40.73		
(')d	0.15	12	0.01	40.67		
advice	0.16	13	0.02	40.46		
tips	0.10	8	<0.01	37.39		
try	0.21	17	0.04	36.01		
hope	0.29	27	0.07	36.00		
love	0.26	21	0.06	34.72		
if	0.66	39	0.29	33.88		
how	0.41	28	0.14	33.43		
hi	0.14	14	0.02	31.13		
great	0.32	27	0.09	31.00		
summer	0.18	14	0.03	29.18		
thanks	0.16	15	0.03	28.06		
can	0.90	42	0.48	27.89		
soon	0.25	22	0.07	27.53		
coming	0.14	11	0.02	27.24		
here	0.19	15	0.04	27.03		
glad	0.10	10	<0.01	26.89		
write	0.15	14	0.03	26.54		
really	0.27	23	0.08	24.84		
think	0.33	22	0.12	23.84		
best	0.27	22	0.09	23.57		
some	0.54	35	0.25	23.38		

way	0.37	22	0.15	22.29		
so	0.57	32	0.29	20.75		
tell	0.15	12	0.03	19.67		
about	0.53	35	0.27	19.26		
idea	0.16	16	0.04	18.93		
email	0.09	9	0.01	18.78		
what	0.38	28	0.17	18.39		
things	0.21	15	0.07	18.19		
good	0.33	28	0.14	18.09		
sorry	0.09	9	0.01	17.94		
letter	0.14	14	0.03	17.69		
going	0.20	15	0.06	17.45		
visit	0.19	13	0.06	16.09		
do	0.46	30	0.24	16.05		
<nationality>	0.35	11	0.16	15.33		

(table 5.22 continued)

Writers often suggest solutions, which they know are not easy to apply, and this is acknowledged by using the verb 'try'. They also sound optimistic about the resolution using the pattern [I am sure], and show confidence in the reader's ability to handle the situation boosting the reader's self-confidence.

E.g. 1. *I understand why you are nervous but I'm sure you'll do well and make friends at your new school.*

E.g. 2. *I know you're great at organizing, so I'm sure you'll have no trouble planning an enjoyable trip for your school.*

They explain how their advice is going to help by using the phrase [this/that way].

E.g. *Another good idea is to join a school club. This way you can make new friends and build your confidence.*

The issues discussed involve the immediate context and close friends or relatives so a casual friendly tone is expected. A number of features distinguish this register as an informal one. First, it is the opening and closing greetings, analysed before, which signal the tenor, the relationship between the writer and the reader, early enough without even reading the rest of the letter. Second, it is the presence of punctuation that is closer to spoken genres. Direct

questions (e.g. how are you?) and exclamation marks (e.g. What exciting news!) are very frequent. Third, it is the preference of contractions over the full forms of auxiliary and modal verbs. As these contractions are rare in the rest of the genres they have been marked as keywords in the 'Advice Letter' (table 5.22). Words such as 'thanks' and 'really' that are commonly used in everyday speech are keywords in this genre. The use of 'really' is interesting here; it is used to emphasise the sincere interest, the fact that the writer is not saying something out of politeness.

E.g. 1. *I'm really excited to hear that.*
E.g. 2. *I'm really looking forward to seeing you.*

There is also frequent use of phrasal verbs, especially with the verb get [get + (a job/a break/in touch/tired/better/ready/back/here/on)].

E.g. *Well, I have to get back to work now. Please let me know how you get on.*

CHAPTER SIX

GENRE RELATIONS

This chapter is focused on quantitative data that can be contrasted in order to identify typical characteristics of each genre as well as the shared features among the eight genres studied in detail.

6.1 Methodology

This analysis is divided in two parts. The first contrasts genres based on shared vocabulary and looks into the relation of each genre to the others. Conducting a 'Detailed Consistency Analysis', a function incorporated in Wordsmith tools v.6, I compare the eight sub-corpora to see which words are shared among which genres. This analysis is chosen because it can reveal the lexical overlap between each pair of genres. Barker (2008), used this type of analysis to report on lexical overlap between levels regarding texts from the Reading texts in question papers at the five different Main Suite levels of Cambridge ESOL exams (KET, PET, FCE, CAE and CPE).

The cut-off point for the number of texts each word occurs in has been five, maintaining Biber's (1990) minimum number of text samples needed to assess central generic tendencies. Therefore, words that occur in less than five texts are not presented no matter how frequent they may be. Then the relation of each corpus to the other is measured statistically using 'Dice coefficient' which is based on the joint frequency and the word counts of the two texts or corpora. A Dice coefficient ranges between 0 and 1 and can be thought of like a percentage. If, for example, the consistency relation is 0.597 there's about a 60% overlap between the vocabularies of the two texts or corpora. Wordsmith tools (Scott, 2012, 2015) uses the following formula to measure the relation:

Relation strength= (J times 2) divided by (F1 + F2)

[where: J = joint frequency, F1 = frequency of word 1 (or corpus 1 word count), F2 = frequency of word 2 (or corpus 2 word count)].

Although these words are seen from a quantitative point of view in this study they could be further explored for their use in various genres and taught in writing classrooms. For this reason, in Appendix Six, I provide lists of the 400 most frequent shared words stating the genres they occurred in.

The second part is also looking for similarity and difference but seen from a number of linguistic features that have been associated to specific genres through various genre/register studies. The selection of sixteen features as the basis for this analysis has been based on their prominent role in distinguishing genres observed during the discourse analysis of individual genres conducted earlier (chapter 5) and/or their use in genre analyses in previous studies. These are selected features from grammatical categories (e.g. pronouns, modals), derivative statistics (e.g. lexical density, Standardised Type/token ratios) and statistics related to text structure (e.g. mean word length, words per sentence).

When the purpose of the analysis is to identify features that are "especially frequent and pervasive in some text varieties in contrast to other varieties" (Biber, 2012: 191), there is a need for quantitative data that can be compared with each other and a basis on which one can say if a value is high or low. Using the POS tagged sub-corpora I have extracted frequencies of specific grammatical categories. Values have been normalised in order to be comparable. This means that when sub-corpora are of different sizes it would be wrong to compare frequencies of one feature in a sub-corpus with the occurrences of the same phenomenon in a longer or shorter sub-corpus. I present normalised grammatical features per thousand words across sub-corpora using the following formula:

$$\text{Normed frequency} = \frac{\text{Observed frequency}}{\text{\# tokens (corpus size)}} \times 1000$$

Pronouns and modals are important genre features as they are associated to involvement, reader engagement and the expression of stance (Biber, 1988; Hyland, 2005). Using first person pronouns writers become involved and may adopt a particular stance. Research has shown that presenting a discoursal self is more common in humanities and social sciences as writers are not afraid to identify themselves with a particular argument and indicate their perspective. In the sciences however, writers use to downplay their personal role suggesting that research outcomes would be the same irrespective of the individual conducting it (Hyland, 2005: 181). Sometimes writers choose to address readers as participants by using

the second person pronoun, to pull them into the discourse at critical points and guide them to particular interpretations. Modals being part of a larger group called 'hedges' are often a way to present information as an opinion rather than acredited fact (Hyland, 2005: 178) and according to Biber's (Biber, 2006: 95) study they are by far the most common grammatical device used to mark stance in university registers.

For verb tense and aspect I have manually counted the frequencies using the concordance tables to separate present and past simple verbs and active and passive voice as the tag set and software used does not provide such detailed information.

I have also used the tagged version to calculate 'Lexical Density' which shows how dense or informative the text is. This is the ratio of the content or lexical words to the number of tokens in the corpus. The higher the value the more content words in the corpus. Ure (1971) has found written texts to have a lexical density of over 40 per cent and for spoken texts to be under 40 per cent. Content or lexical words belong to the major word classes of nouns, main verbs, adjectives and certain adverbs (lexical adverbs such as 'honestly' or 'beautifully' as oppposed to grammatical adverbs such as 'when' or 'where') and are the words that usually carry more meaning than function words. They are also called 'open classes' as new members can be added to these categories. Function or grammatical words tend to occur frequently in any genre and do not tell us much about the content of the text. Stubbs (1996: 71) calls them 'minor', 'empty' and 'structural'. To extract lexical verbs, I have manually separated auxiliary verbs. Phrasal verbs have been counted as two separate items, one as lexical (verb) and one as grammatical (preposition). I have used Ure's method (1971) to measure lexical density:

$$LD = \frac{\text{\# Lexical or Content words}}{\text{\# Total words (tokens)}} \times 100$$

I have used the untagged sub-corpora for the extraction of the rest of the features. I provide the frequencies of connectors as individual analysis of genres at the previous stage corroborated findings on the prime role of connectors in accomplishing the purpose of each genre (So, 2005; Glasswell et al. 2001). I have grouped connectors according to their

functional role in four main groups (a few connectors have been omitted because of their multiple functions):

a) Temporal: after, during, finally, later, next, soon, suddenly, then, when.

b) Adding: also, and, furthermore, in addition, what is more.

c) Contrasting: although, but, despite, however, moreover, on the one hand, on the other hand, whereas, while.

d) Causal/consequential: because, consequently, therefore.

Nominalisation is a technique that writers use to add more information and create longer sentences. It is the use of nouns for meanings that are more typically expressed in a verb, adjective or whole clause (Martin, 1985, 1991). For example, 'introduction' and 'popularity' are nominalisations derived from 'introduce' and 'popular'. When writers use nouns instead of verbs the style may become more impersonal and sound more objective as there are less pronouns. Meaning is more condensed as this allows other verbs to be included in the sentences. "Nominalization can turn actions into static things and therefore attribute objective reality to states of affairs" (Stubbs, 1996: 226). It has been observed that nominalisation is more frequent in academic prose than in fiction or spoken registers (Biber et al., 1998: 62). In this study the value for nominalisation is the sum of the frequencies of the words with the following derivational endings: -tion, -sion, -ness, -ment, -ity, -ship and –ism (filtering out manually nouns that are not instances of nominalisation such as 'station' or 'city').

Several terms have been used for specific groups of words that tend to appear frequently such as multi-word units (Moon, 1997), lexical bundles (Biber et. al., 1999) lexical phrases (Lewis, 1993), formulaic sequences (Wray & Perkins, 2000), clusters (Scott, 1997) or n-grams (Kanaris & Stamatatos, 2007). There are however, minor differences on what researchers choose to study in this category. I am using the term lexical bundles the way Biber et al. (1999: 989) see them, as an umbrella term to include formulaic phrases, idioms and recurring "bundles of words that show a statistical tendency to co-occur". These multi-word units can give us some idea of the degree of repetitiveness and standardisation of the language used in a genre. Kopaczyk (2012: 6) studying legal texts talks about 'prefabricated, standardised formulas and phrases, expected by the participants in a given communicative situation' and suggests that 'deviations from the norm may result in failing to recognise the text as the one intended'. Huang (2013: 32, 35), in her thesis on lexical bundles observes that the research scope of contrastive studies based on bundles has been limited to registers in a

general sense (for example, 'academic written discourse' and 'academic spoken discourse', or 'native academic writing' and 'learner academic writing') and that "studies of lexical bundles in other genres are scarce". She cites Gries et al., (2011: 13) who encourage researchers to "shift their focus more on the finer divisions of corpora than the fewer, more convenient coarser divisions", an issue I also discuss in chapter three.

My focus here is on 3-word uninterrupted sequences. I have chosen this size as 2-word lexical bundles are usually included in 3-word ones and because 4-word lexical bundles which are less frequent in this corpus, are often an extension of 3-word bundles (Huang, 2013). According to Biber et al. (1999), lexical bundles must spread across at least five different texts in a register in order to exclude individual user idiosyncrasies so I have included any 3-word lexical bundle that occurs in at least five texts in the sub-corpora and I have named this feature '3-word Register-Specific-Lexical Bundles'. I treat each bundle as a single lexical item and calculate the relative frequence in each sub-corpus in the following way:

3-word RSLB = <u>Sum of the frequency of 3-word bundles (occurring in at least 5 texts)</u>   X100
Tokens

[where: RSLB = Register Specific Lexical Bundles]

Although the interest here is on the percentage of lexical bundles from a quantitative perspective, these lists of lexical bundles can be studied in detail for their functions in discourse and can be of practical value to anyone teaching or learning how to write these genre-specific texts. For this reason, lists of the most frequent 3-word lexical bundles per genre are provided in Appendix Seven.

The type/token ratio (TTR) is also a value which is considered important in this analysis as it is usually seen as an indicator of the variety of the vocabulary used. The TTR is the ratio of the types (number of individual words that appear in the corpus) to the tokens (total number of words in the corpus). A low TTR value means that many of the same words are used repeatedly. High values show that texts include a variety of words and that less words are used repeatedly. The problem with the type/token ratio is that it is highly dependent on text length or corpus size. As the texts become longer, words tend to be repeated so there are relatively fewer word types. (Biber, 2006; Μικρός, 2003; Scott, 2012; Tweedie & Baayen,

1998). So, the conventional TTR is informative if dealing with corpora comprising lots of equal-sized texts. If the case is corpora with texts of various lengths WordSmith Tools (Scott, 2015), offers the Standardised Type/ Token Ratio (STTR) measurement as a more reliable solution. The researcher has the choice to compute TTR for every n words and get an average based on consecutive chunks of n words. Texts with less than n words (whatever n is set to) will get a standardised type/token ratio of 0. As the texts in these sub-corpora are short, n is set to 100 words so that the biggest number of texts are included in calculating the average.

The two final features investigated are 'word-length' and the 'number of words per sentence'. These features, often called 'corpus token-level properties', 'shallow discourse features', or 'basic text properties', have been used in various genre/register studies (Crossley et al. 2014; McCarthy et al. 2009; Nesi & Gardner 2012; Stamatatos et al. 2001). In studies with an educational interest, the length of words (in letters) and the number of words per sentence may reveal a more advanced use of vocabulary and an ability to cope with complex sentences. The means for these two features are provided by WordSmith Tools in Statistics.

After the initial presentation of the values for these sixteen features, I associate them with text properties based on previous studies making this way pairs of linguistic features and overall stylistic choices. This helps us understand how some vague and implicit qualities of texts often mentioned in assessment rubrics and teachers' feedback can be attributed to specific linguistic features. Looking at these text properties in various genres adds to the knowledge about individual genres gained form the previous discourse analysis. Moreover, the contrast of these values makes it easier to notice which genres are closer to each other or more distant in terms of significant text characteristics.

More specifically as can be seen in table 6.1, sixteen features (left) have been chosen as best indices of specific text properties (right). High values of the linguistic feature measured in a genre show a prevalence of the textual property linked to it. The right side column shows associations made in previous studies. Biber's associations (1988, 1995) refer to linguistic features from the factors in his Multi-Dimensional analysis. For a few linguistic features, a text property name has been coined as a better alternative to existing ones or existing names have been slightly altered. The selection, matching and naming of text properties has been based on previous research on academic genres and this study's discourse analysis (chapter 5). The text property names for the associations used in this study are written in bold.

Table 6.1 Features and associated text properties
(name for properties used in this study **in bold**)

Features	Association with text properties
1. Passive Voice verbs (4 tenses)	- Abstractness/ **Objectivity and Formality** (Glasswell et al., 2001: 6,8) - Abstract style (Biber, 1995: 141-163), - Formality (Michos et al., 1996: 194) - Detached writer (Czernieska, 1992)
2. Lexical density	**- Informational density** (Fang et al., 2006: 259; Nagy &Townsend, 2012:94; Yates, 1996: 37) - Lexical denseness (Ventola, 1996: 162)
3. Nominalisation	- Condensed meaning / Objectivity and Formality (Glasswell et al., 2001: 6) **- Objectivity** (Beck & Jeffery, 2007: 66) **- Elaborated** (Biber, 1995: 141-163), - Information density (Fang et al., 2006: 254) **(Objectivity and Elaboration** in this study)
4. Standardised type/token ratios	**- Lexical variety** (Fialho et al., 2012: 65; Meara & Bell, 2001: 6; Viana et al., 2008: 276) - Informational production (Biber, 1988)
5. Word length	**- Lexical complexity** (Štajner et al., 2015: 387) - Informational production (Biber, 1988)
6. Words per sentence	- Elaboration (sentence level), Formality and Syntactic complexity (Michos et al., 1996: 193, 194) **- Syntactic complexity** (Štajner et al., 2015: 387)
7. 3-word register-specific lexical bundles	Standardisation (Kopaczyk, 2012: 4,6) **Standardised language** (in this study)
8. Active s. present verbs	- Timeless present (Glasswel et al., 2001: 6,7) - Involved (Biber, 1988) **- Reference to timeless present** (this study)
9. Active past simple verbs	- Narrative discourse (Biber, 1995), Imaginative narrative (Glasswell et al., 2001: 8), Narrative genre (Czerniewska, 1992) **Reference to past** (this study)
10. Connectors- temporal	**- Events set in time** (Glasswel et al., 2001: 8)
11. Connectors- Adding	- Additive relations (Glasswel et al., 2001: 6,8) **- Addition** (this study)
12. Connectors- Contrasting	**Contrast** (So, 2005: 74)
13. Connectors- Causal/ consequential	- Causal relationships (So, 2005: 72; Glasswel et al., 2001: 6) **- Causality** (this study)
14. Modals	**- Authorial stance** (included in 'hedges') (Hyland, 2005: 177) - The degree of certainty/commitment of the writer (Beck & Jeffery, 2007: 66) - Overt expression of argumentation (Biber, 1988) - The degree of obligation or certainty involved in the argument (Glasswell et al., 2001: 6) - Engagement (modals of obligation), (Hyland, 2005: 184)
15. 1st person sing. pronoun	**- Involvement** (Biber, 1988), - Stance (Hyland, 2005: 177)
16. 2nd person pronoun	**- Reader engagement** (Hyland, 2005: 177) - Involved (Biber, 1988)

Next, the values for the features in each genre which are presented in various forms (normalised POS categories, ratios, percentages) are converted to z-scores. This offers a point of reference as to the mean; the value for each feature seen in each genre is contrasted to the values for the same feature in other genres so that it is clear which genre is close to the mean and exactly how close. The second advantage of z-scores is that values of different scales are translated to a single scale and can be comparable (Ζαμπετάκης, 2013). For the conversion the IBM SPSS v. 22 statistics software has been used. A z-score is a standardized variable with a mean of zero and a standard deviation of 1. Values that are less than the mean produce negative scores whereas values that are larger than the mean produce positive scores. Means, standard deviation as well as minimum and maximum values used for the calculation of z-scores are provided in the Results section. The formula for z scores is:

$$Z = \frac{(X - \bar{X})}{S}$$

[where X is the variable score, $\bar{X}$ is the (sample) mean and S is the (sample) standard deviation (Ζαμπετάκης, 2013)].

Results are shown in graphs from two different perspectives. The first presentation refers to each text property as this relates to different genres. This view is more helpful to those who study particular text properties and would like to know how these vary across genres. Then each genre is presented on its own with its positive (over the mean) and negative (below the mean) text properties. This offers a view of each genre to those interested on specific genres and their features.

Then I measure distance for all possible pairs of genres in this study based on each text property. When this is done for all text properties and all pairs of genres I calculate the mean distance for each pair of genres. This gives us a value that shows the degree of similarity between each pair of genres based on all 16 text properties measured in this study. Each genre is presented as it relates to the rest 7 genres ranked according to strongest relation (smallest distance) so that it is easy to see which genre should be taught or presented in teaching material closer to the one already taught/presented in order to minimise the time and effort needed by learners to take control of the texts they need to write. I also find out where

exactly two different genres converge or diverge in order to place emphasis on particular text properties (and related lexicogrammatical features) during teaching. The table provided at the end of chapter 6 shows which genres in the WriMA corpus are closer to each other and which are more distant. At the same time, it shows whether this similarity is based on text properties, vocabulary or both presenting the results of both types of analysis.

6.2 Results

The extent to which each individual genre relates to the rest of genres studied here is investigated from two different perspectives. The results in the first part show the similarity based on shared vocabulary between each pair of genres while the results in the next part show similarity based on sixteen linguistic features and associated text properties.

6.2.1 Similarity based on shared vocabulary

Results show that similarity based on shared vocabulary among the eight genres ranges from 0.3 to nearly 0.6 or in other words, there is a lexical overlap of 30% up to 60% between various pairs of genres (the strength of relation is measured on a scale of zero to one) (table 6.2).

Within the first positions one sees the Expository - Discursive Essay pair as well as the Personal Observation Report with all Essay genres. It is interesting that the Descriptive Essay paired with the other two Essay genres occurs in the seventh and eight position and that there is greater similarity with the Personal Observation Report and the Short Story.

Another interesting finding is the relation between the two Reports. Even though there is a 40% overlap between them, the Personal Observation Report is strongly related to all other genres before its match with the Data Report in the 11th position.

The relation between the two Letter genres is ranked 9th with 42% overlap but it is not the first involving the Letters. The Advice Letter is related to the Personal Observation Report at the same position and the Complaint Letter is matched with the same Report in the 6th position.

In general, the Personal Observation Report has strong relations with all genres as it not seen below the 11th position while at the same time the Data Report is shown to be the least related to all other genres as its first relation occurs in the 11th position.

Table 6.2 Similarity across genres based on shared words (Consistency analysis)

Ranked according to strength of relation	Pairs of genres	Strength of relation based on shared words
1.	Expository Essay - Discursive Essay	.583
2.	Descriptive Essay - Personal Observation Report	.476
3.	Discursive Essay - Personal Observation Report	.474
4.	Exposition Essay - Personal Observation Report	.460
5.	Descriptive Essay - Short Story	.459
6.	Personal Observation Report - Complaint Letter	.446
7.	Discursive Essay - Descriptive Essay	.443
8.	Expository Essay - Descriptive Essay & Short Story - Advice Letter	.437
9.	Personal Observation Report - Advice Letter & Complaint Letter - Advice Letter	.424
10.	Descriptive Essay - Advice Letter & Personal Observation Report - Short Story	.408
11.	Personal Observation Report - Data Report	.403
12.	Short Story - Complaint Letter	.389
13.	Expository Essay - Data Report	.387
14.	Descriptive Essay - Complaint Letter	.384
15.	Expository Essay - Advice Letter	.380
16.	Discursive Essay - Data Report	.379
17.	Expository Essay - Complaint Letter	.364
18.	Discursive Essay - Advice Letter	.362
19.	Descriptive Essay - Data Report	.360
20.	Data Report - Complaint Letter	.350
21.	Discursive Essay - Complaint Letter	.348
22.	Discursive Essay - Short Story	.343
23.	Expository Essay - Short Story	.337
24.	Data Report - Advice Letter	.333
25.	Data Report - Short Story	.318

6.2.2 Similarity based on linguistic features and text properties

Table 6.3 illustrates the sixteen linguistic features measured for each genre. These include values in various forms (normalized values, ratios, means, percentages) and are the starting point for the analysis of text properties in various genres.

Table 6.3 Sixteen features in eight genres

[where (N)= normalised values, (M)= mean]

Linguistic features/Genres	Expository Essay	Discursive Essay	Descriptive Essay	Data Report	Personal Observation Report	Short Story	Complaint Letter	Advice Letter
1. **Passive voice verbs** (4 tenses-sum), (N)	7.1	6.3	7.6	12.9	10.4	5.9	11.4	1.9
2. **LD** (Lexical density), (%)	51.8	51.1	49.6	53.4	52.9	45.3	45.2	47
3. **Nominalisation** (N)	27.1	25.7	12.3	18.6	22.9	4.4	18.3	6.3
4. **STTR** (Standardised type/token ratio on a 100word basis)	71.8	71.1	70.9	57.7	68.3	68.3	70.9	71.1
5. **Word length** (M)	4.62	4.65	4.26	4.50	4.75	4	4.33	3.87
6. **Words per sentence** (M)	18.92	17.90	20.43	19.50	17.60	12.95	17.33	13.45
7. **3-word register-specific lexical bundles** (%)	1.5	4.9	.6	2.7	2.6	2.1	3.7	2.6
8. **Simple present verbs** (N)	57.2	63.3	42	32.1	44.4	8.7	25.6	56.4
9. **Simple past verbs** (N)	4.4	3.6	37.6	28.4	18.4	102.2	41.7	14.8
10. **Temporal connectors** (N)	3.9	4.2	9.3	8.9	5	19.8	10.9	7.9
11. **Adding connectors** (N)	32.6	32.3	33.9	37.8	36.2	29	23	26.3
12. **Contrasting connectors** (N)	7.7	11.9	8.2	9.3	6.8	9.2	9.2	6.6
13. **Causal/ consequential connectors** (N)	2.6	3	2.1	.29	1.5	1.5	1.9	1.6
14. **Modals** (possibility, prediction, necessity- sum) (N)	22.9	25	10.6	4.1	17.4	6.6	13.6	27.9
15. **1st person sing. pronoun** (N)	6.3	4.5	19.2	.3	8.6	40	43.4	29.4
16. **2nd person pronoun** (N)	1.7	3.8	2.8	.05	2.9	3.1	13.6	52.2

Associating these features with text properties provides a more holistic view of the genres and clarifies vague and subjective notions characterising whole texts. Each text property is presented in bar graphs showing all genres to facilitate comparison. This presentation is mainly of interest to anyone seeking detailed information about genre variation on specific text properties while the presentation at the next stage sees each genre as a whole with its typical properties. Both views, however, raise awareness of genre variation.
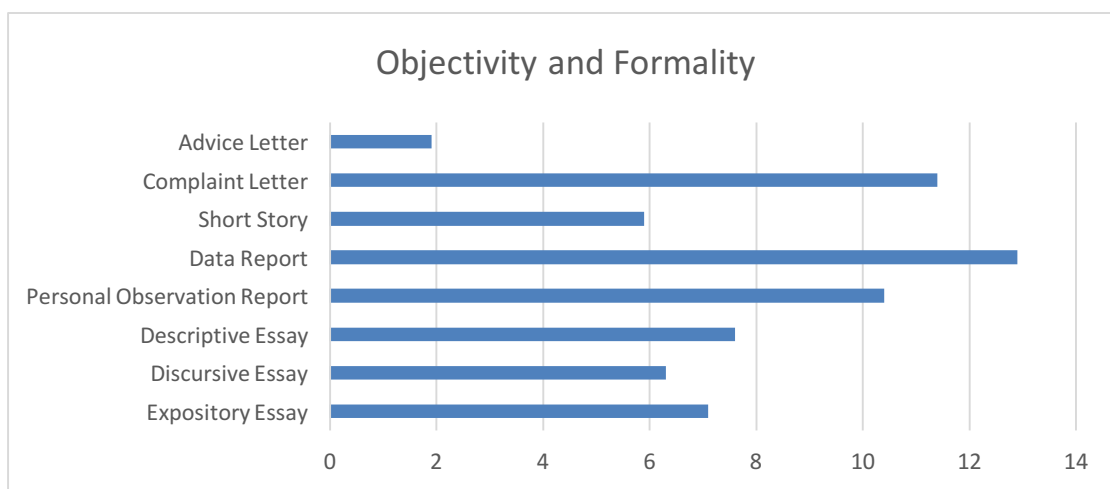
Figure 6.1 Genres and text property 1: 'Objectivity and Formality'
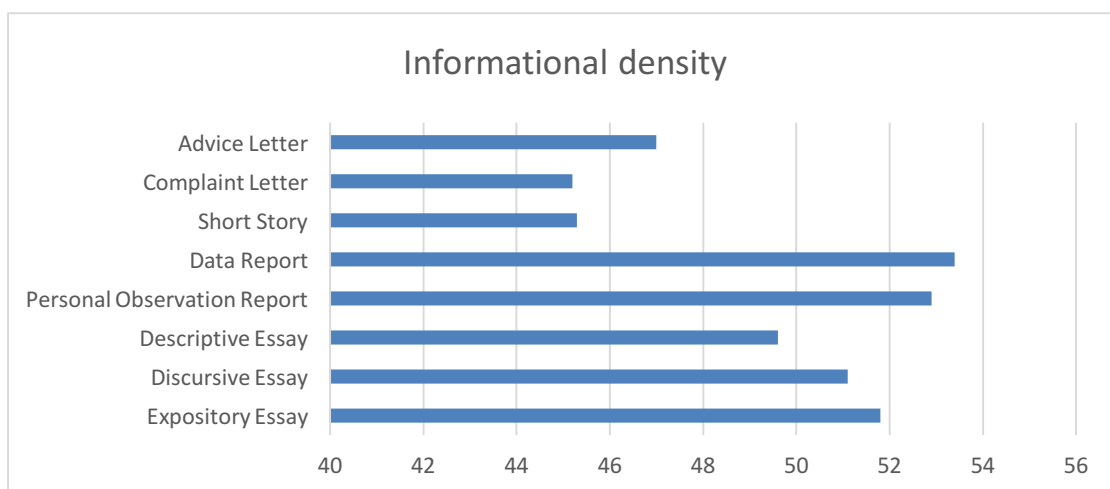
Figure 6.2 Genres and text property 2: 'Informational density'
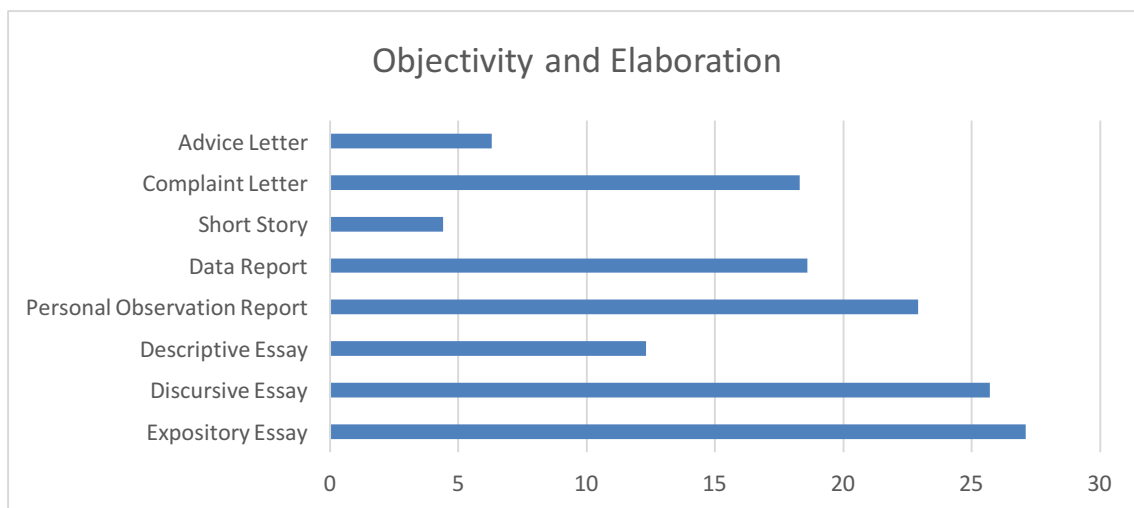
Figure 6.3 Genres and text property 3: 'Objectivity and Elaboration'

The Data Report is the most 'Objective and Formal' genre followed by the Complaint Letter and the Personal Observation Report (figure 6.1). The Advice Letter has the lowest values on this text property compared to other genres. The two Reports are dense in information followed closely by the Discursive Essay (figure 6.2). This means that more meaning is compressed in the text in the form of content words. The two Letter genres and the Short Story score low in this text property.

At the same time, the Expository and the Discursive Essay are both 'Objective and Elaborated', that is, there is a preference for nominalisation which allows for condensed meaning in an impersonal style. The Personal Observation Report follows and scores higher than the Data Report (figure 6.3). The least 'Objective and Elaborated', genres in this sample are the Short Story and the Advice Letter.

Looking at 'Lexical variety' (figure 6.4), there are not great differences apart from the Data Report in which many words are repeated and less new words are used. Sentences however, are syntactically more complex than those in the Personal Observation Report (figure 6.6). The Descriptive Essay has the highest score on 'Syntactic complexity' while the Advice Letter and the Short Story are the least syntactically complex genres with less words per sentence. These two genres also have the shortest words (figure 6.5). Matching the short length of sentences in Short Stories and Friendly Advice Letters together with the shorter words it is possible to detect a fast paced rhythm with less complex vocabulary than all the

167

other genres, resembling the style of spoken genres. The Personal Observation Report scores high on lexical complexity followed by the Expository and the Discursive Essay (figure 6.5).
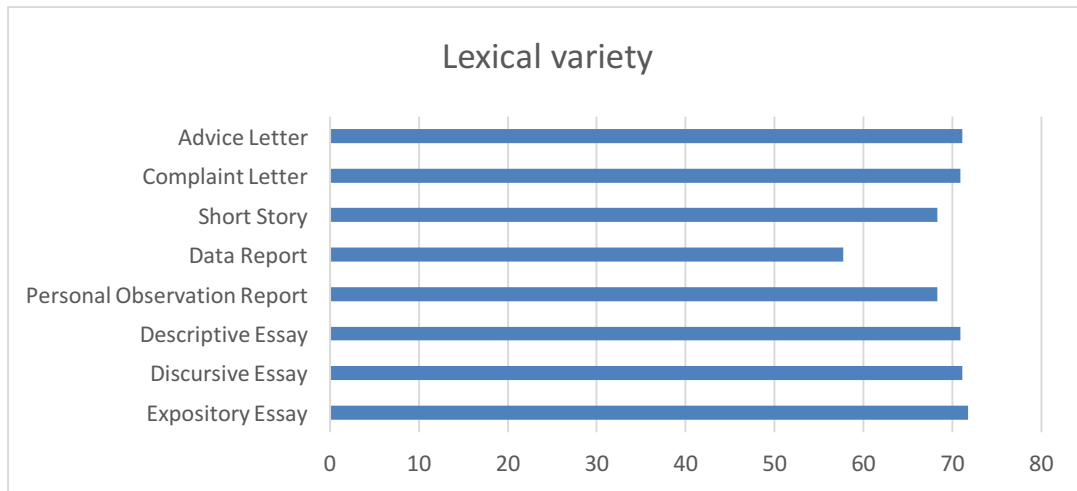
Figure 6.4 Genres and text property 4: 'Lexical variety'
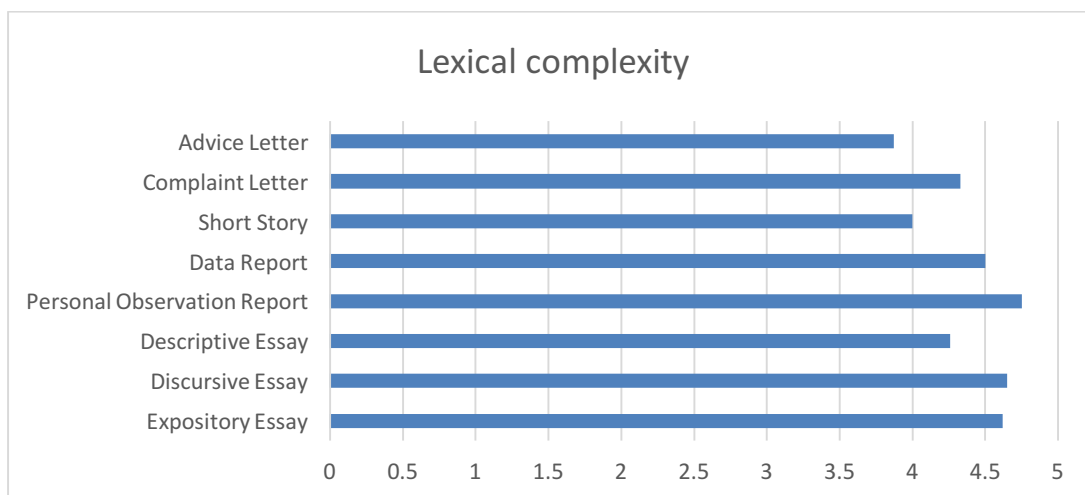
Figure 6.5 Genres and text property 5: 'Lexical complexity'
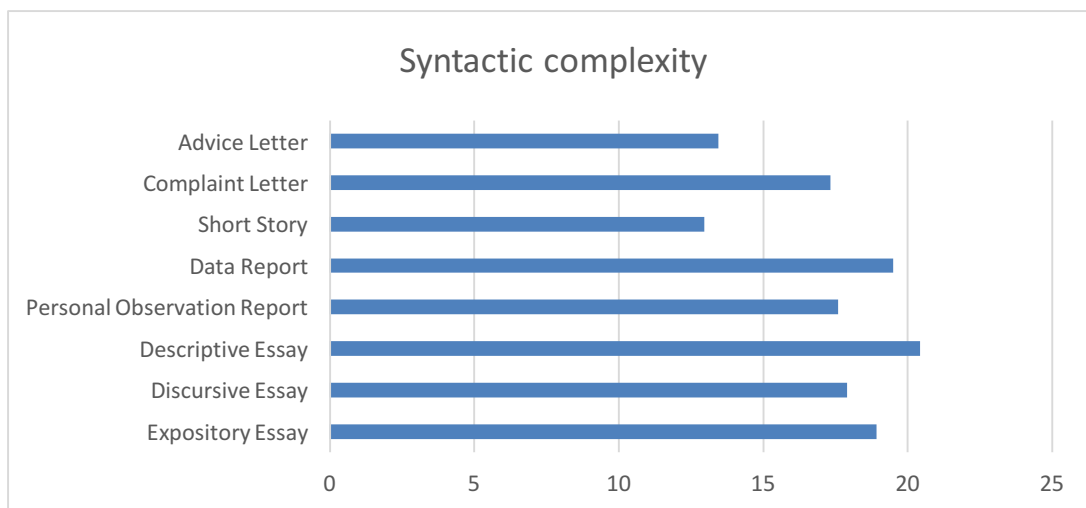
Figure 6.6 Genres and text property 6: 'Syntactic complexity'
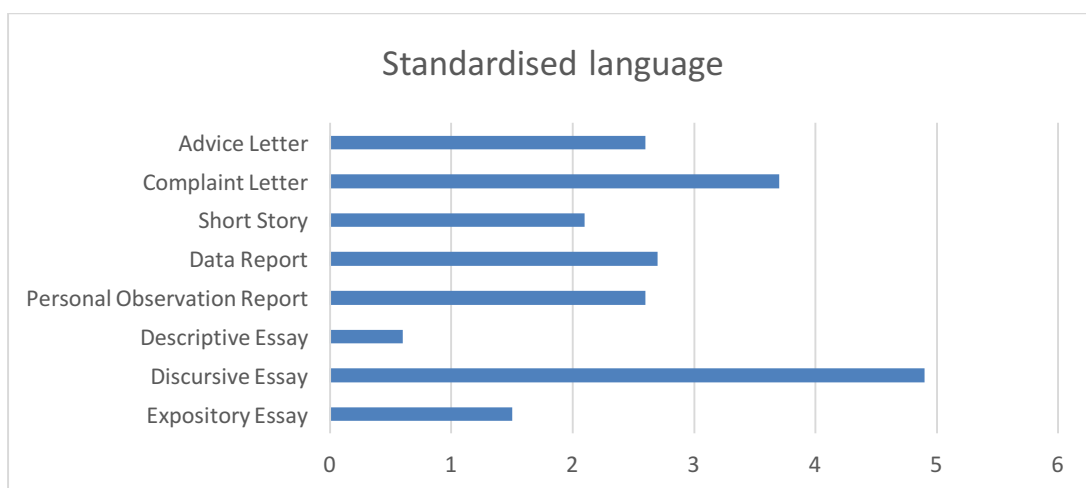
Figure 6.7 Genres and text property 7: 'Standardised language'

The use of lexical bundles is very high in the Discursive Essay and there is a sharp contrast with the other genres (figure 6.7). This Essay genre and the Complaint Letter have a more 'Standardised language' while the Descriptive Essay is the least standardised. It is interesting that between the two argumentative Essays there is a significant difference on the use of lexical bundles making the Discursive Essay a lot easier to identify as a genre as well as to teach than the Expository Essay.

Concerning time, the Short Story is very different from other genres in its 'Reference to past' (figure 6.8). The Complaint Letter is the only other genre where past reference exceeds the 'Reference to timeless present'; the use of past verbs however, is a lot more frequent in the Short Story. There is past reference in the Descriptive Essay and the Data Report but it does

not exceed the reference to timeless present. The Discursive and the Expository Essay are the genres in which past reference is almost non existent.

Events are set in time in the Short Story where temporal connectors are more frequently used than in any other genre (figure 6.9). This tendency is also seen in the Complaint Letter, the Descriptive Essay, the Data Report and the Advice Letter but to a lesser extent.
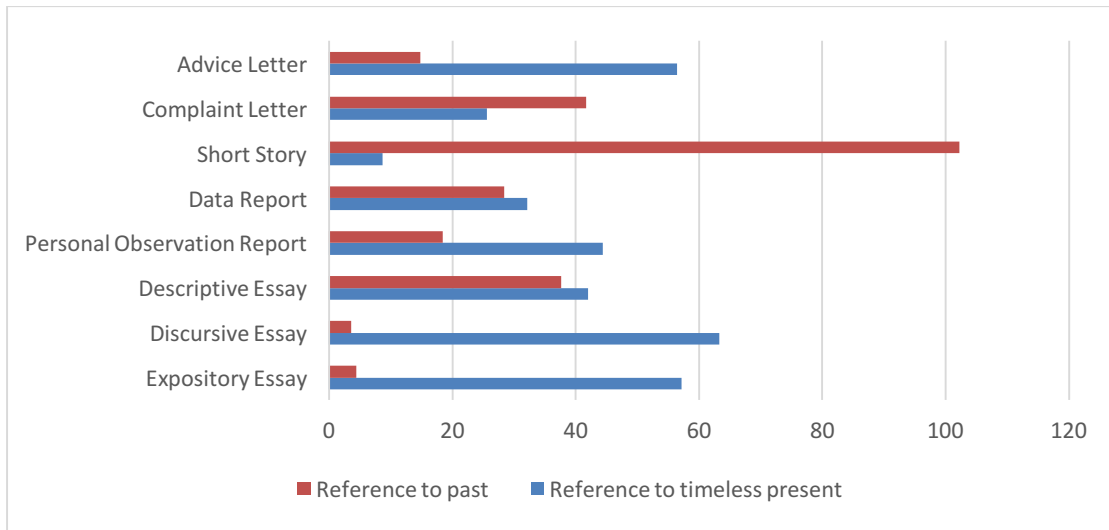
Figure 6.8 Genres and text properties 8 & 9: 'Reference to timeless present' and 'Reference to past'
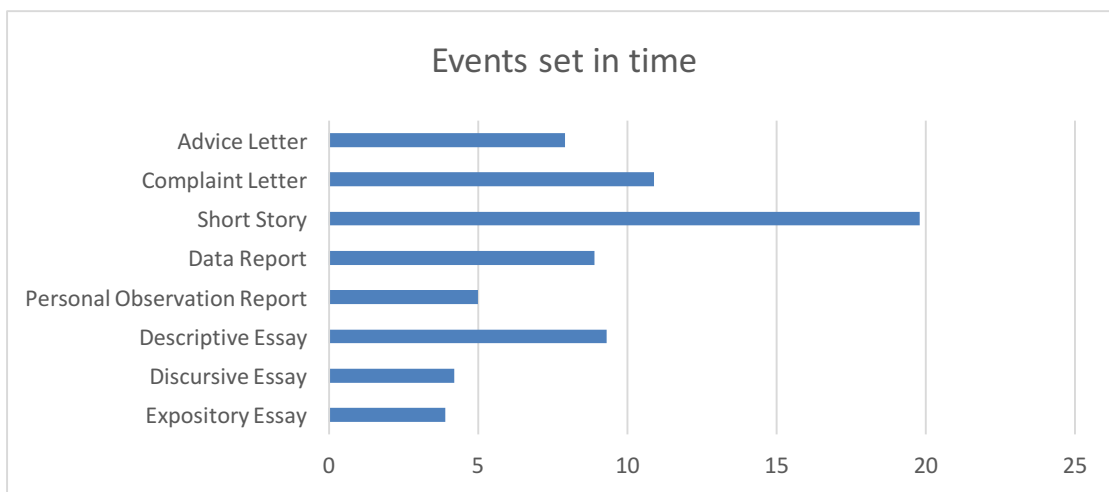
Figure 6.9 Genres and text property 10: 'Events set in time'
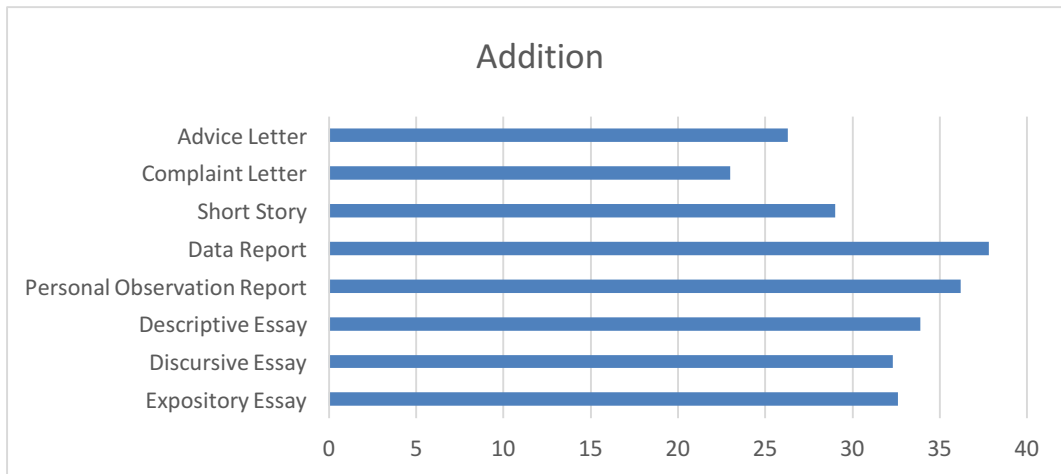
Figure 6.10 Genres and text property 11: 'Addition'

The two Reports and the three Essay genres have the highest scores in the use of adding connectors (figure 6.10) while the Discursive Essay is also full of 'Contrast'. Compared to the Expository Essay this seems a major difference (figure 6.11). There is also some difference in this text property between the two Reports. There is much more contrast in the Data Report.

The Discursive Essay also comes first in 'Causality' (figure 6.12). There is a sharp contrast in this property between the Data Report and all other genres. The Data Report mainly describes, informs, compares but does not explain.
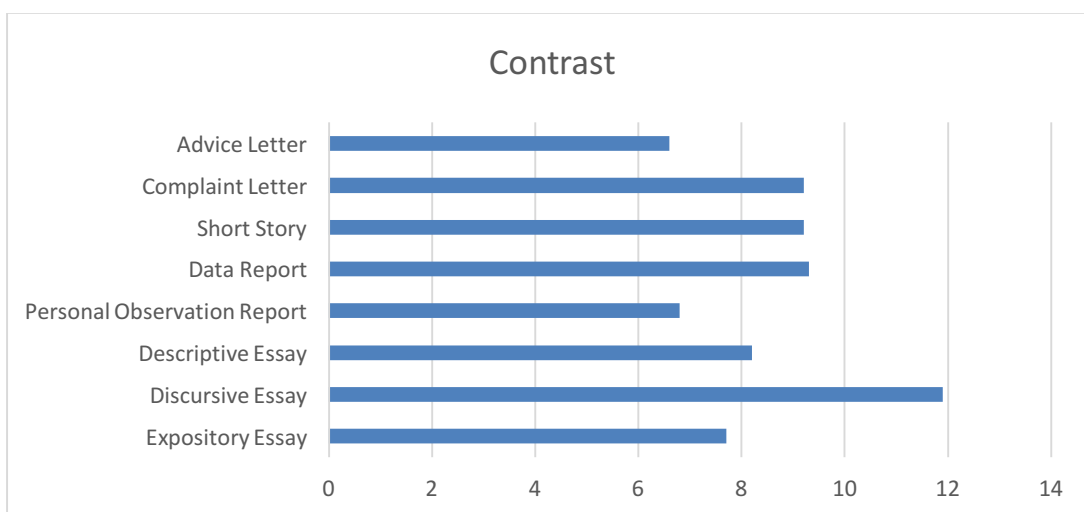
Figure 6.11 Genres and text property 12: 'Contrast'
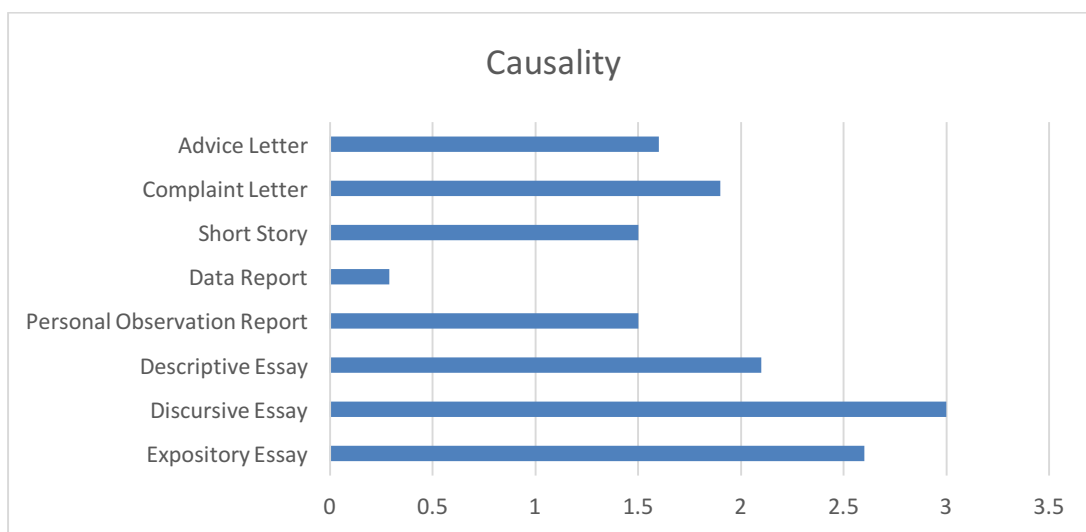
Figure 6.12 Genres and text property 13: 'Causality'

'Authorial stance' in the sense that the authors express obligation, possibility and necessity is very prevalent in the Advice Letter and then in the Discursive and Expository Essay (figure 6.13). The difference between the two Reports is also supported here. Results show that the Data Report informs and describes in a more assertive style than the Personal Observation Report. A similar distance is noticed between the two Letter genres. The Complaint Letter is more factual and assertive even though the writer is involved (figure 6.13, 6.14). Writer 'Involvement' is also evident in the Short Story and much less in the Advice Letter and the Descriptive Essay. There is clearly more 'Involvement' in the Descriptive than the other two Essay genres (figure 6.14). Writers address their readers much less in the Complaint Letter than in the Advice Letter where reference to the readers' problems and needs is central. 'Reader engagement' is very low in the rest of the genres. The difference between the two Reports is also evident here. Figure 6.14 shows that there is almost no 'Involvement' or 'Reader engagement' in the Data Report while there is some 'Reader engagement' and greater 'Involvement' in the Personal Observation Report.
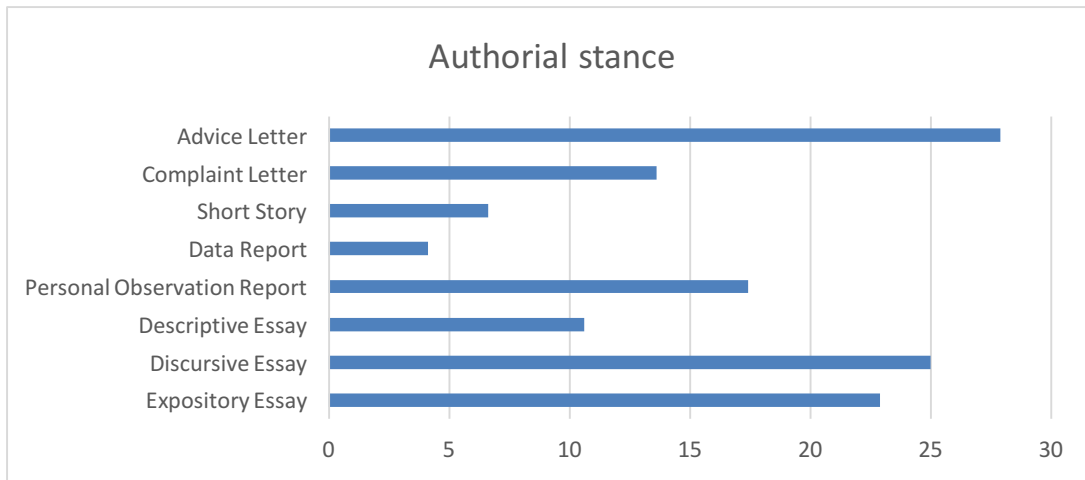
Figure 6.13 Genres and text property 14: 'Authorial stance'
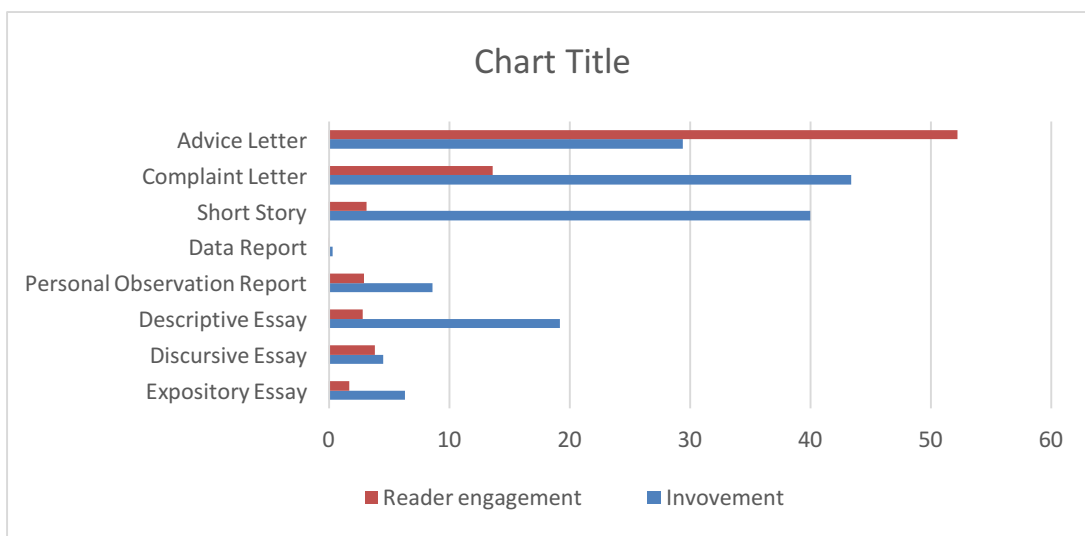
Figure 6.14 Genres and text properties 15, 16: 'Involvement' and 'Reader engagement'

Table 6.4 presents the descriptive statistics on which the calculation of z scores was based. Minimum refers to the smallest value of the variable and Maximum refers to the largest value. The Mean is the arithmetic mean across the observations. The Std. (Standard) deviation is the square root of the variance and measures the spread of this set of observations. The larger the standard deviation is, the more spread out the observations are. Looking at the standard deviation for example one can see that at this set there are four text properties that deviate a lot across genres. These refer to the use of present and past simple verbs as well as the use of the first singular and the second person pronouns. This finding verifies the intial observation during the discourse analysis that time reference, personal

173

involvement and reader engagement are indeed factors that cause significant variation and that they can be seen as genre markers.

Table 6.4 Descriptive statistics for the calculation of z-scores

Descriptive Statistics

	#of genres	Minimum	Maximum	Mean	Std. Deviation
Passive_v_verbs	8	1.9	12.9	7.938	3.5209
Lexical_Density	8	45.2	53.4	49.538	3.3166
Nominalisation	8	4.4	27.1	16.950	8.5530
STTR	8	57.7	71.8	68.763	4.6614
Word_length	8	3.87	4.75	4.3725	.31685
Words_per_sentence	8	12.95	20.43	17.2600	2.71094
Lexical_bundles	8	.6	4.9	2.588	1.3076
S_present_verbs	8	8.7	63.3	41.213	18.3955
S_past_verbs	8	3.6	102.2	31.388	31.8760
Temporal_connectors	8	3.9	19.8	8.737	5.1556
Adding_connectors	8	23.0	37.8	31.388	4.9976
Contrastive_connectors	8	6.6	11.9	8.613	1.7033
Caus_conseq_connectors	8	.3	3.0	1.813	.8149
modals	8	4.1	27.9	16.013	8.7592
I	8	.3	43.4	18.963	16.7926
you	8	.1	52.2	10.025	17.5214

Figures 6.15 - 6.22 show how each genre is positively or negatively marked for each text property providing an evidence-based description of what these genres consist of, on a common scale that allows comparisons among text properties and among genres.

+ Objectivity and Elaboration (1.18671)

+ Causality (.96473)

+ Reference to timeless present (.86910)

+ Authorial stance (.78631)

+ Lexical complexity (.78113)

+ Informational density (.68217)

+ Lexical variety (.65163)

+ Syntactic complexity (.61233)

+ Addition (.24262)

- Objectivity and Formality (-.23786)

- Reader engagement (-.47513)

- Contrast (-.53572)

- Involvement (-.75405)

- Standardised language   (-.83168)

- Reference to past (-.84664)

-Events set in time (-.93831)

Figure 6.15 Text properties: The 'Expository Essay'

+ Contrast (1.93007)

+ Standardised language (1.76851)

+ Causality (1.45398)

+ Reference to timeless present (1.20070)

+ Authorial stance (1.02606)

+ Objectivity and Elaboration (1.02303)

+ Lexical complexity (.87581)

+ Lexical variety (.50146)

+ Informational density (.47112)

+ Syntactic complexity (.23608)

+ Addition (.18259)

- Reader engagement (-.35528)

- Objectivity and Formality (-.46508)

- Involvement (-.86124)

- Reference to past (-.87174)

- Events set in time (-.88012)

Figure 6.16 Text properties: The 'Discursive Essay'

+ Syntactic complexity (1.16934)

+ Addition (.50275)

+ Lexical variety (.45856)

+ Causality (.35318)

+ Reference to past (.19490)

+Events set in time (.10911)

+ Reference to timeless present (.04281)

+ Informational density (.01884)

+ Involvement (.01414)

- Objectivity and Formality (-.09586)

- Contrast (-.24218)

- Lexical complexity (-.35506)

- Reader engagement (-.41235)

- Objectivity and Elaboration (-.54367)

- Authorial stance (-.61792)

- Standardised language (-1.51996)

Figure 6.17 Text properties: The 'Descriptive Essay'

+ Objectivity and Formality (1.40943)

+ Addition (1.28313)

+ Informational density (1.16460)

+ Syntactic complexity (.82628)

+ Contrast (.40363)

+ Lexical complexity (.40240)

+ Objectivity and Elaboration (.19291)

+ Standardised language (.08604)

+ Events set in time (.03152)

- Reference to past (-.09372)

- Reference to timeless present (-.49537)

- Reader engagement (-.56645)

- Involvement (-1.11135)

- Authorial stance (-1.35999)

- Causality(-1.86067)

- Lexical variety (-2.37323)

Figure 6.18 Text properties: The 'Data Report'

+ Lexical complexity (1.19142)

+ Informational density (1.01384)

+ Addition (.96297)

+ Objectivity and Formality (.69939)

+ Objectivity and Elaboration (.69566)

+ Reference to timeless present (.17328)

+ Authorial stance (.15840)

+ Syntactic complexity (.12542)

+ Standardised language (.00956)

- Lexical variety (-.09922)

- Causality (-.38070)

- Reader engagement (-.40665)

- Reference to past (-.40744)

- Involvement (-.61709)

- Events set in time (-.72495)

- Contrast (-1.06411)

Figure 6.19 Text properties: The 'Personal Observation Report'

+ Reference to past (2.22150)

+ Events set in time (2.14574)

+ Involvement (1.25278)

+ Contrast (.34492)

- Lexical variety (-.09922)

- Standardised language (-.37282)

- Causality (-.38070)

- Reader engagement (-.39523)

- Addition (-.47773)

- Objectivity and Formality (-.57868)

- Authorial stance (-1.07458)

- Lexical complexity (-1.17564)

- Informational density (-1.27766)

- Objectivity and Elaboration (-1.46732)

- Syntactic complexity (-1.58985)

- Reference to timeless present (-1.76742)

Figure 6.20 Text properties: The 'Short Story'

+ Involvement (1.45525)

+ Objectivity and Formality (.98340)

+ Standardised language (.85079)

+ Lexical variety (.45856)

+ Events set in time (.41945)

+ Contrast (.34492)

+ Reference to past (.32352)

+ Reader engagement (.20404)

+ Objectivity and Elaboration (.15784)

+ Causality (.10855)

+ Syntactic complexity (.02582)

- Lexical complexity (-.13413)

- Authorial stance (-.27542)

- Reference to timeless present (-.84872)

- Informational density (-1.30782)

- Addition (-1.67832)

Figure 6.21 Text properties: The 'Complaint Letter'

+ Reader engagement (2.40706)

+ Authorial stance (1.35714)

+ Reference to timeless present (.82561)

+ Involvement (.62155)

+ Lexical variety (.50146)

+ Standardised language (.00956)

- Events set in time (-.16245)

- Causality (-.25838)

- Reference to past (-.52038)

- Informational density (-.76509)

- Addition (-1.01800)

- Contrast (-1.18153)

- Objectivity and Elaboration (-1.24517)

- Syntactical complexity (-1.40542)

- Lexical complexity (-1.58593)

- Objectivity and Formality (-1.71474)

Figure 6.22 Text properties: The 'Advice Letter'

Each genre was contrasted to the rest of the genres based on the 16 text properties in order to measure its distance from each of the genres. Figures 6.23-6.28 present each genre and its relation to the rest of the genres ranked according to mean distance.
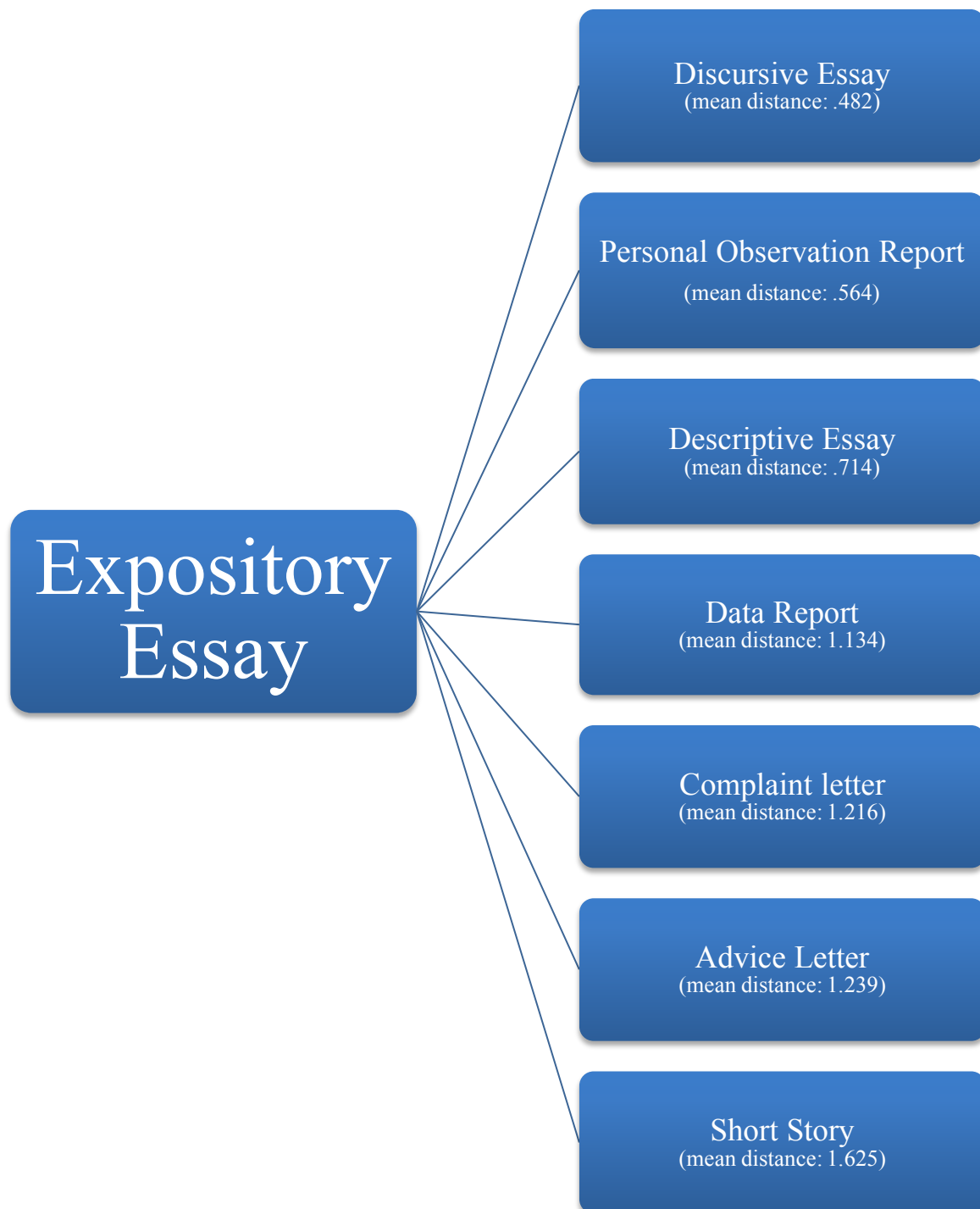
Figure 6.23 Genre relations: The Expository Essay
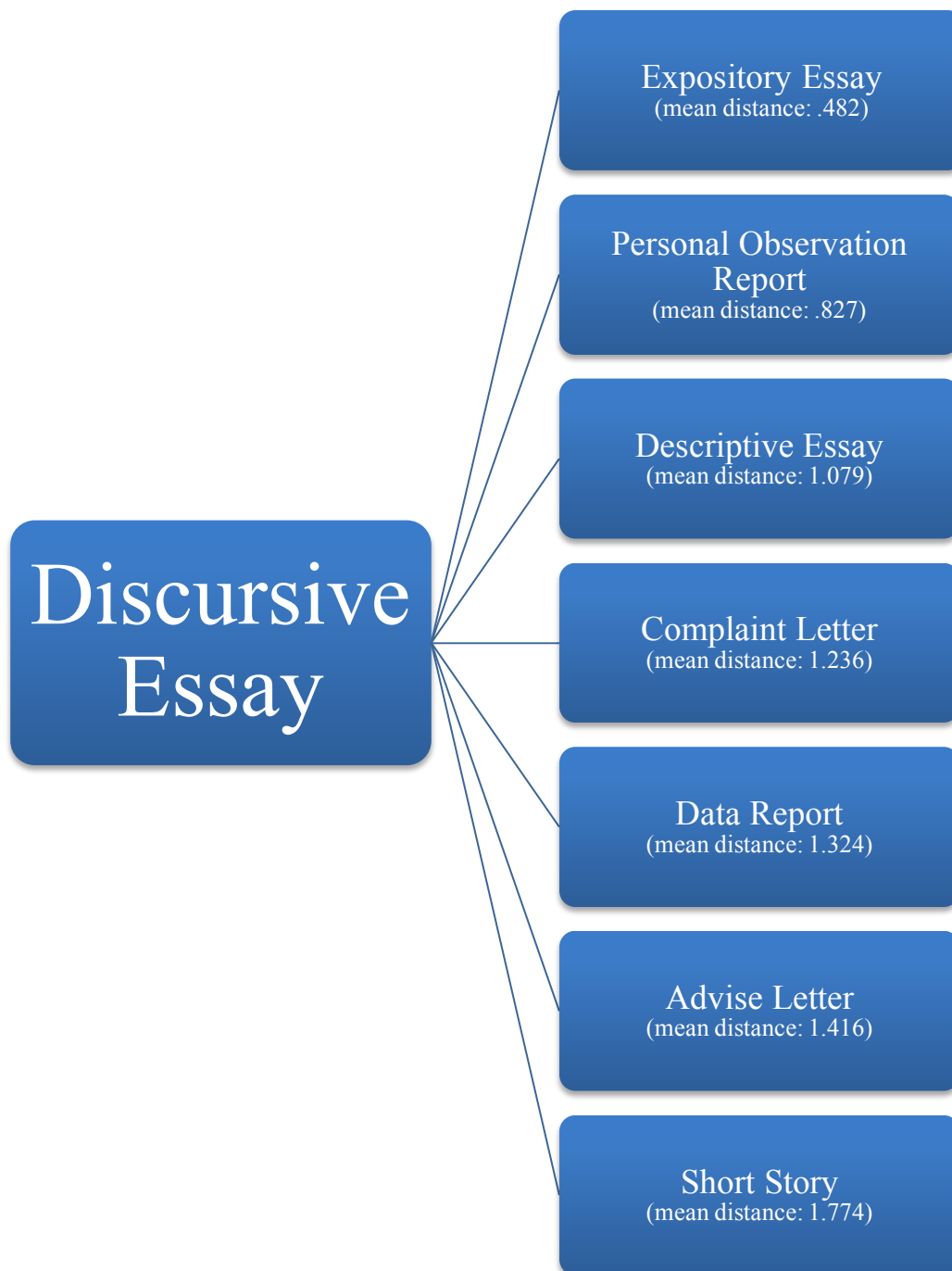
Figure 6.24 Genre relations: The Discursive Essay
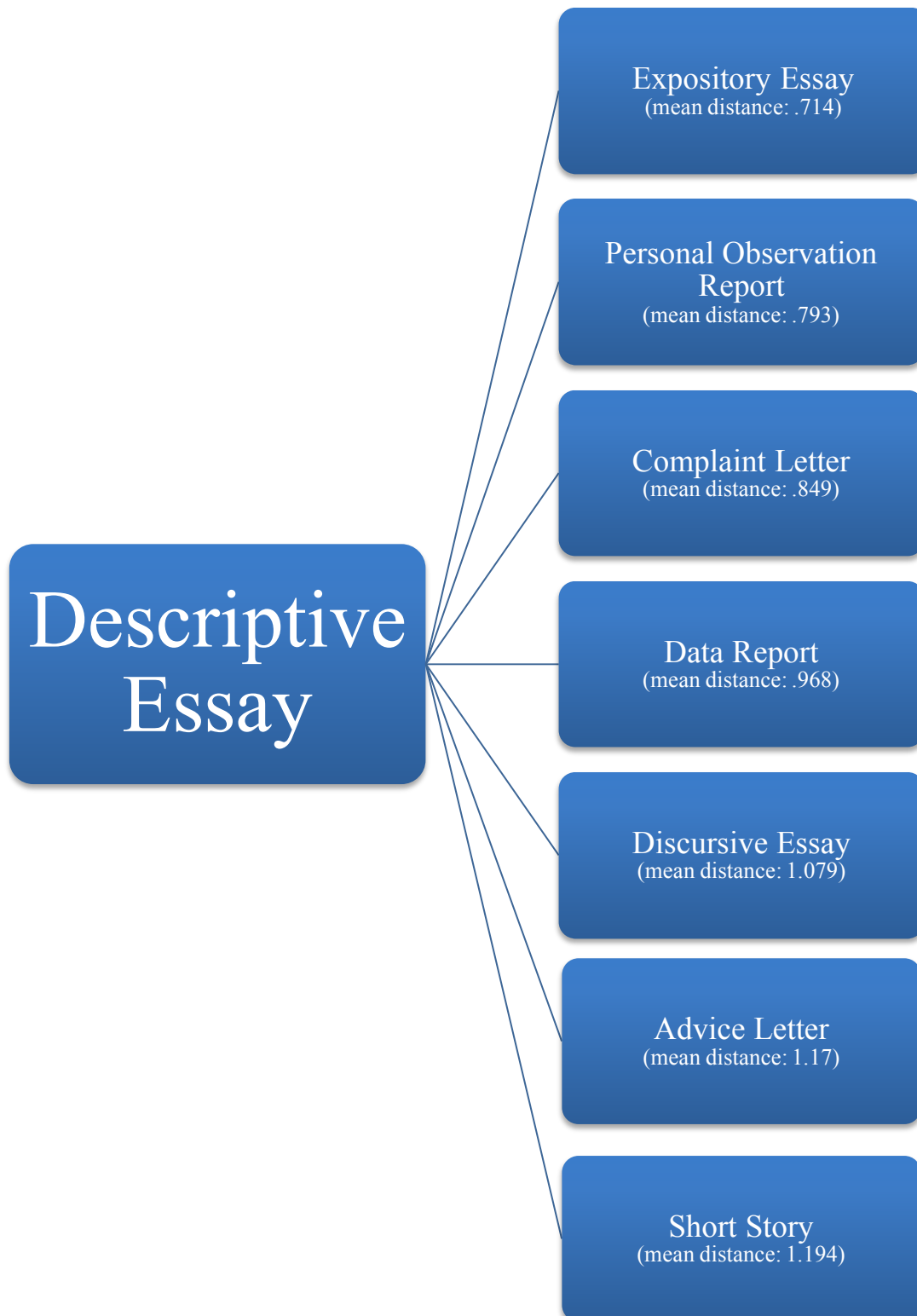
Figure 6.25 Genre relations: The Descriptive Essay
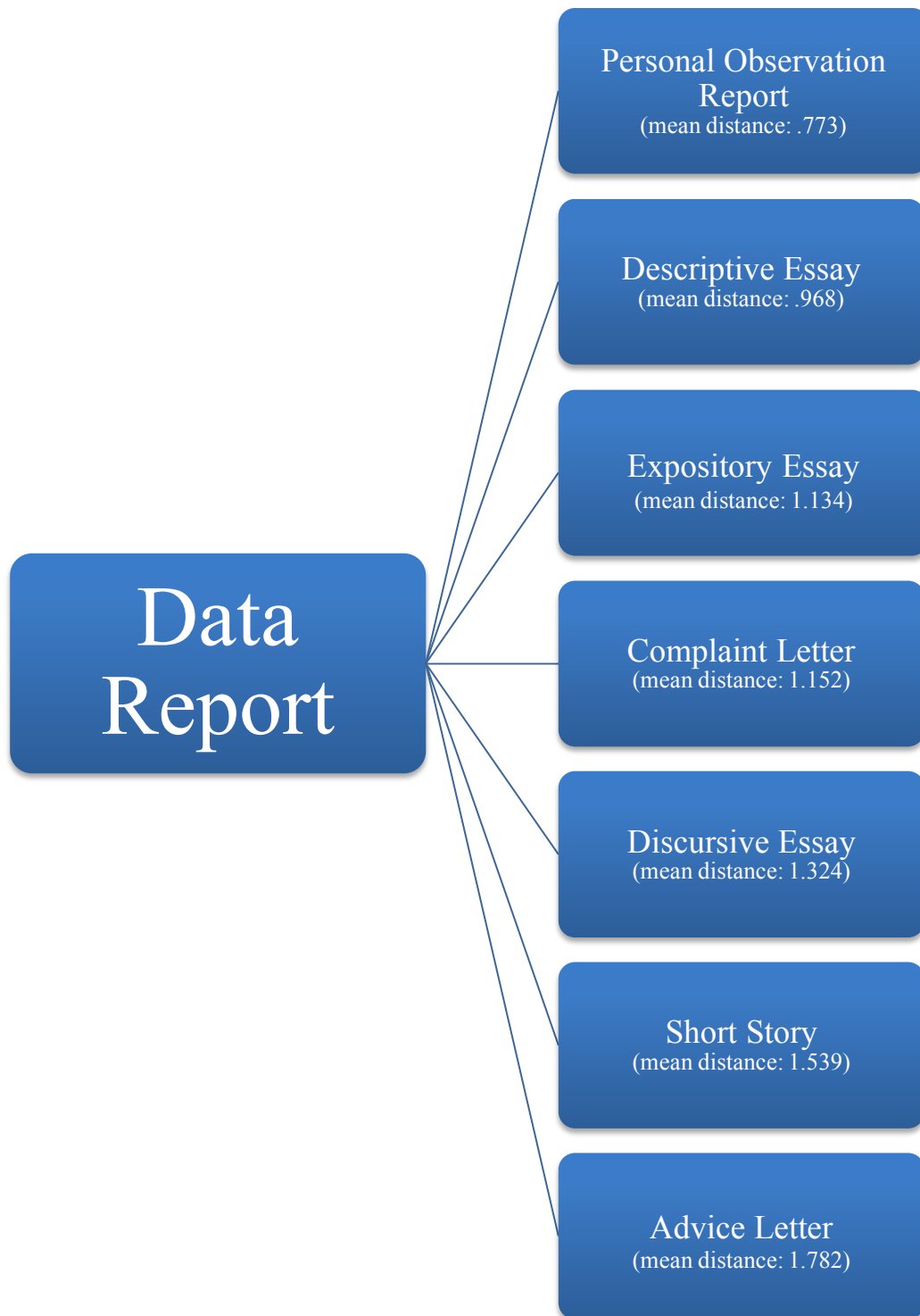
Figure 6.26 Genre relations: The Data Report

Figure 6.27 Genre relations: The Personal Observation Report
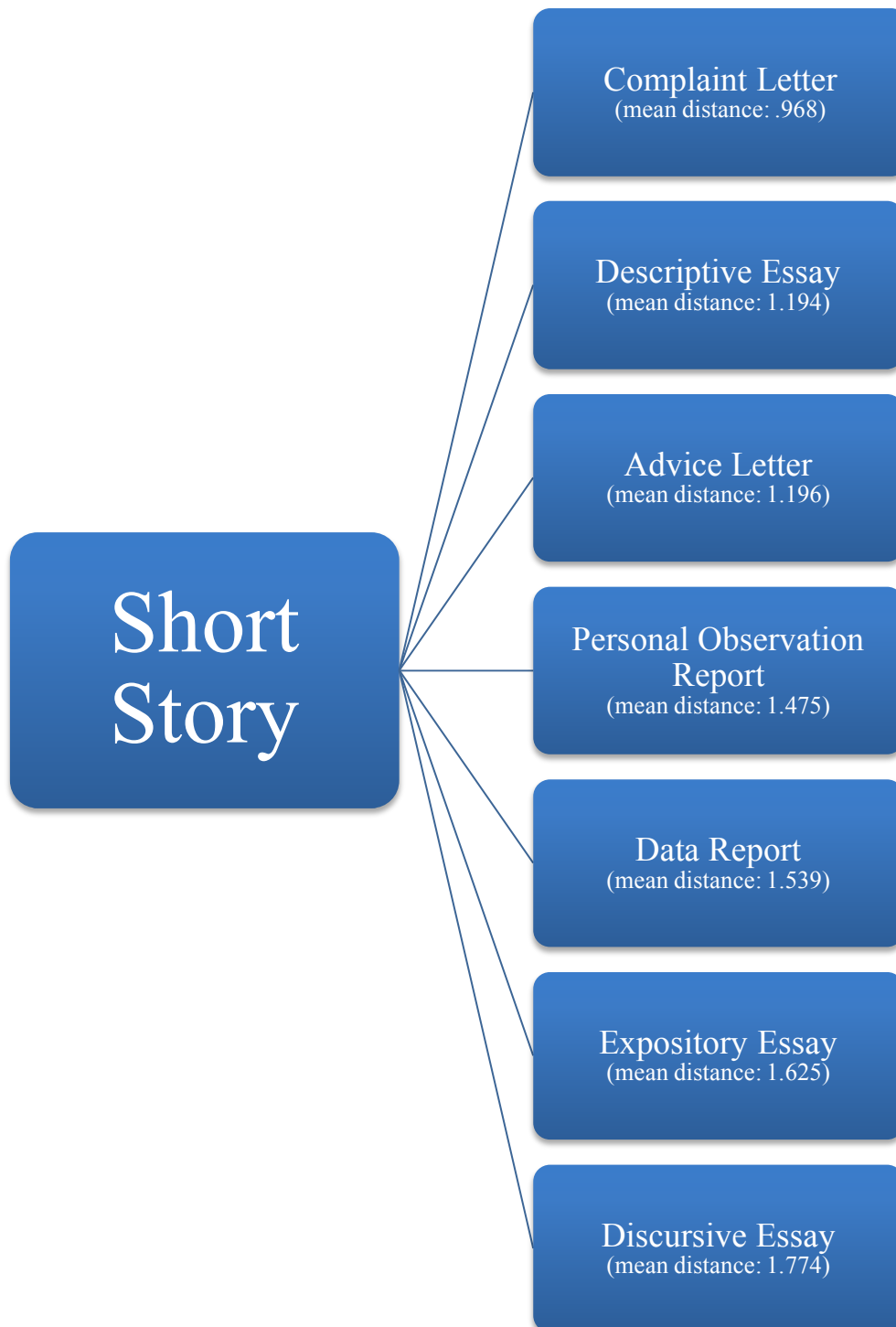
Figure 6.28 Genre relations: The Short Story
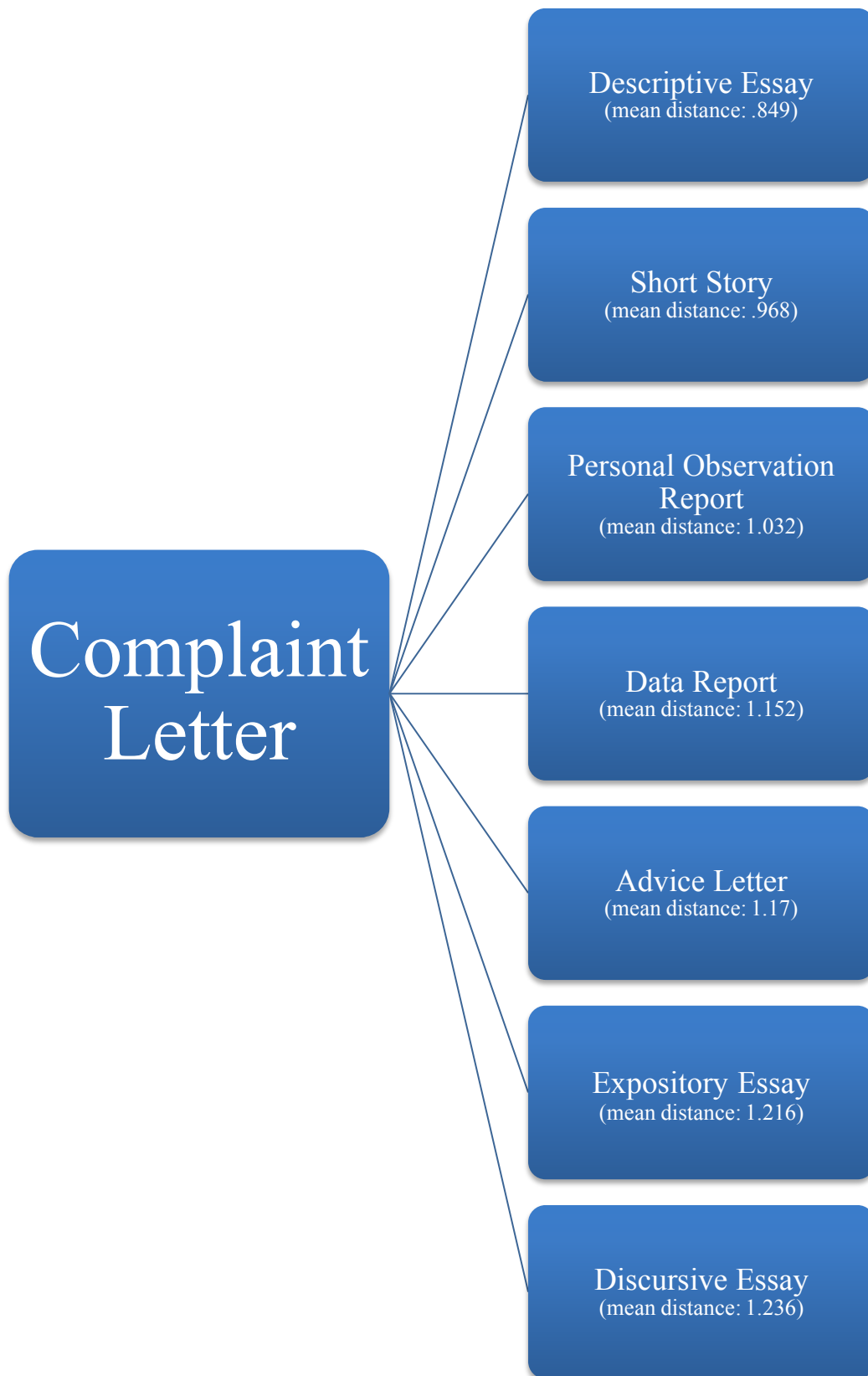
Figure 6.29 Genre relations: The Complaint Letter
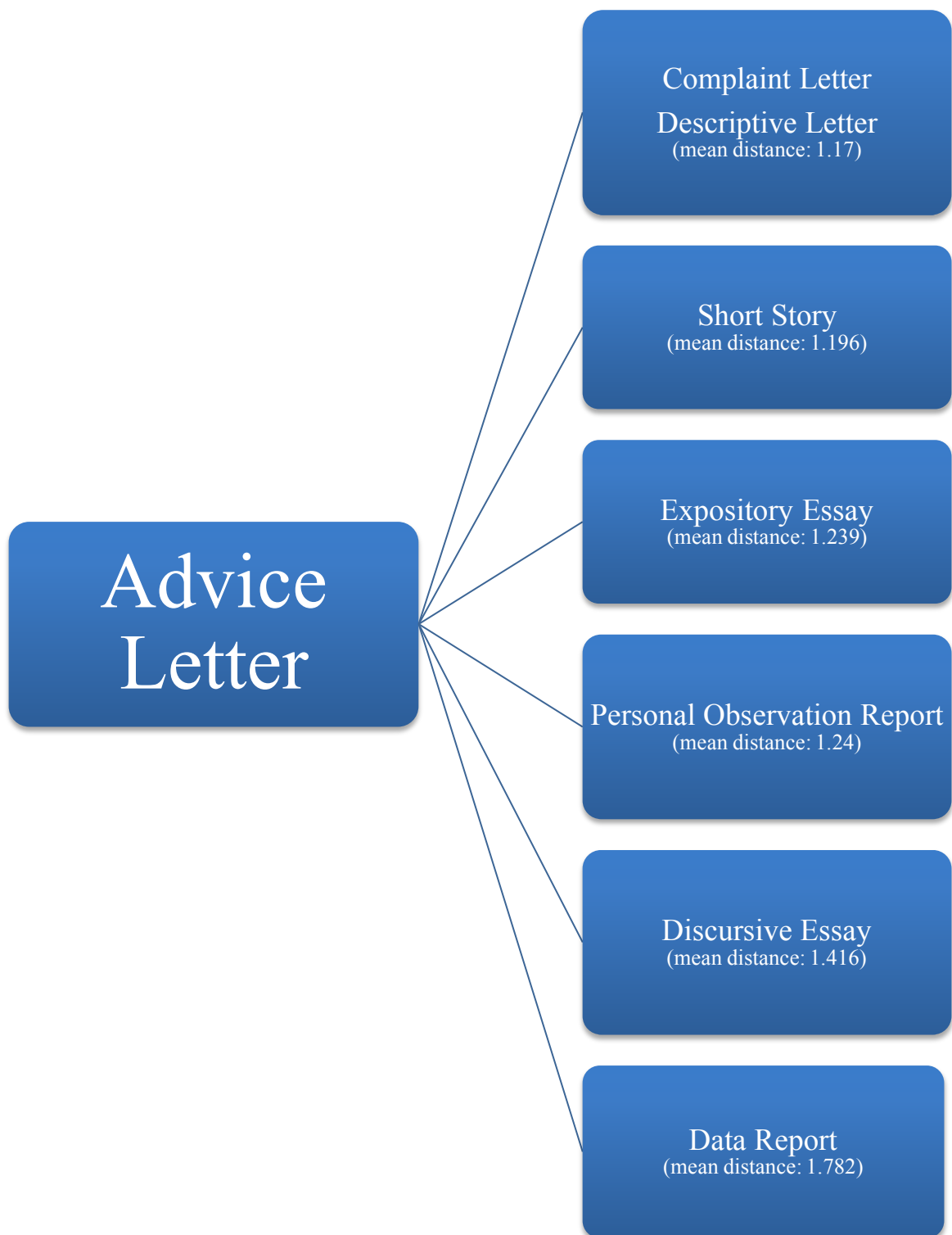
Figure 6.30 Genre relations: The Advice Letter

Table 6.5 Genre relations from two perspectives

Rank according to mean distance (small to large distance)	Pairs of genres	Mean distance (based on 16 text properties)	Rank according to strength of relation (large to small overlap)	Strength of relation (% of lexical overlap)
1.	Expository Essay - Discursive Essay	.482	1.	58.3
2.	Expository Essay - Personal Observation Report	.564	4.	46
3.	Expository Essay - Descriptive Essay	.714	8.	43.7
4.	Personal Observation Report - Data Report	.773	11.	40.3
5.	Descriptive Essay - Personal Observation Report	.793	2.	47.6
6.	Discursive Essay - Personal Observation Report	.827	3.	47.4
7.	Descriptive Essay - Complaint Letter	.849	14.	38.4
8.	Short Story - Complaint Letter	.968	12.	38.9
	Descriptive Essay - Data Report	.968	19.	36
9.	Personal Observation Report - Complaint Letter	1.032	6.	44.6
10.	Discursive Essay - Descriptive Essay	1.079	7.	44.3
11.	Expository Essay - Data Report	1.134	13.	38.7
12.	Data Report - Complaint Letter	1.152	20.	35
13.	Complaint Letter - Advice Letter	1.17	9.	42.4
	Descriptive Essay - Advice Letter	1.17	10.	40.8
14.	Descriptive Essay - Short Story	1.194	5.	45.9
15.	Short Story - Advice Letter	1.196	8.	43.7
16.	Expository Essay - Complaint Letter	1.216	17.	36.4
17.	Discursive Essay - Complaint Letter	1.236	21.	34.8
18.	Expository Essay - Advice Letter	1.239	15.	38
19.	Personal Observation Report - Advice Letter	1.24	9.	42.4
20.	Discursive Essay - Data Report	1.324	16.	37.9
21.	Discursive Essay - Advice Letter	1.416	18.	36.2
22.	Personal Observation Report - Short Story	1.475	10.	40.8
23.	Data Report - Short Story	1.539	25.	31.8
24.	Expository Essay - Short Story	1.625	23.	33.7
25.	Discursive Essay - Short Story	1.774	22.	34.3
26.	Data Report - Advice Letter	1.782	24.	33.3

Table 6.5 shows how the eight genres investigated here relate to each other seen from two different perspectives: one based on shared vocabulary and the other based on text properties.

Pairs of genres are ranked starting from strongest relations based on text properties while their ranking based on shared vocabulary appears on the right. As can be seen, the Expository and the Discursive Essay come first in similarity seen from both perspectives. Figure 6.31 illustrates the distance between these two Essays in all the text properties investigated here. Even though the distance between these two Essay genres in most properties is really small there is a remarkable distance in Standardized language (distance: 2.60019) and Contrast (distance: 2.46579) as the Discursive Essay was found to include a lot more 3-word lexical bundles and contrasting connectors.
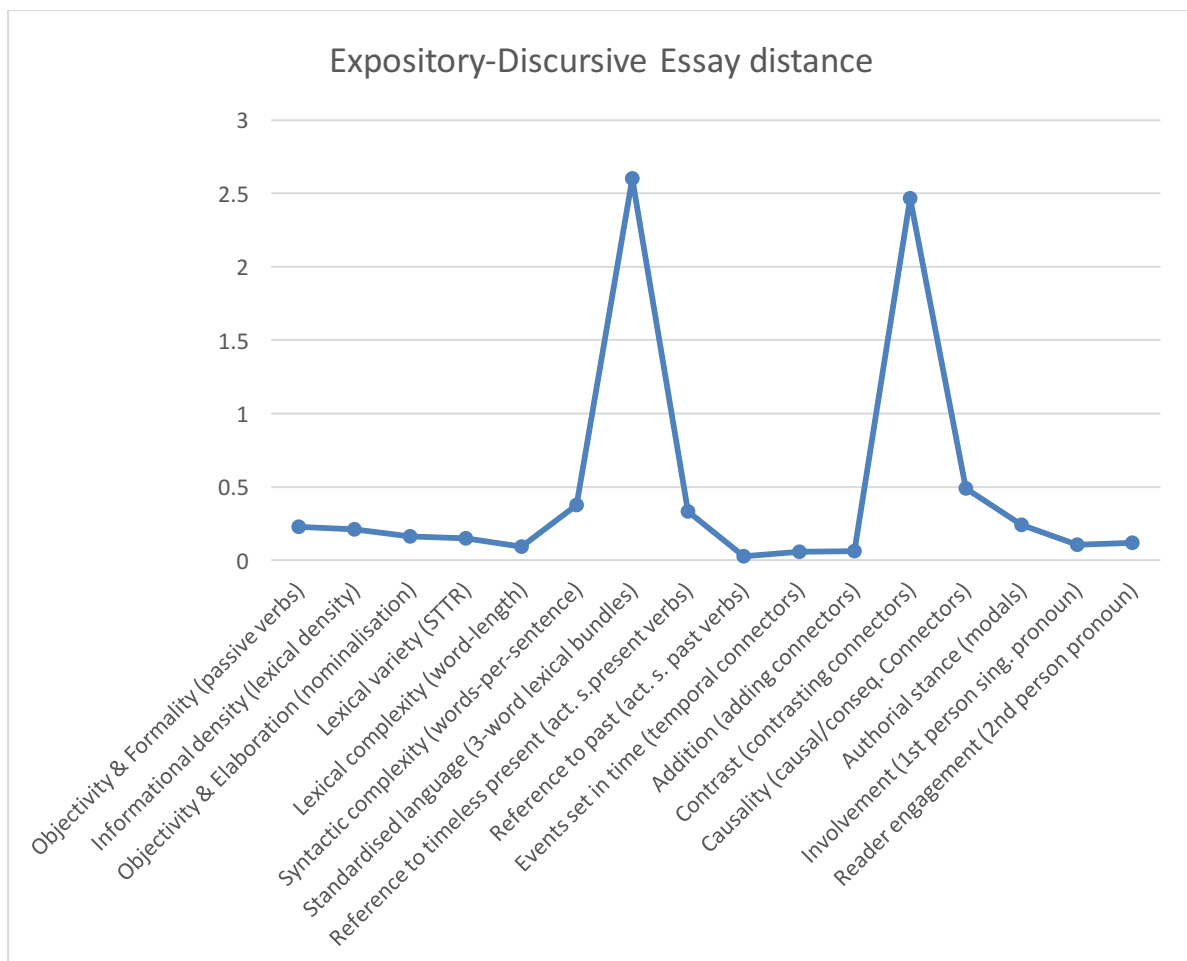
Figure 6.31 The Expository-Discursive Essay distance

Based on text properties the Descriptive Essay relates strongly to the Expository Essay (3rd position), the Personal Observation Report (5th position), the Complaint Letter (7th position) and the Data Report (8th position) while its similarity to the Discursive Essay is weaker (10th position). Figure 6.32 shows that the distance between the Descriptive and the Discursive Essay is increased especially in Standardised language (distance: 3.28847) and Contrast (distance: 2.17225) due to the high number of 3-word lexical bundles and contrasting connectors observed in the Discursive Essay (two factors that caused distance between the Expository-Discursive Essay, seen also in figure 6.31).
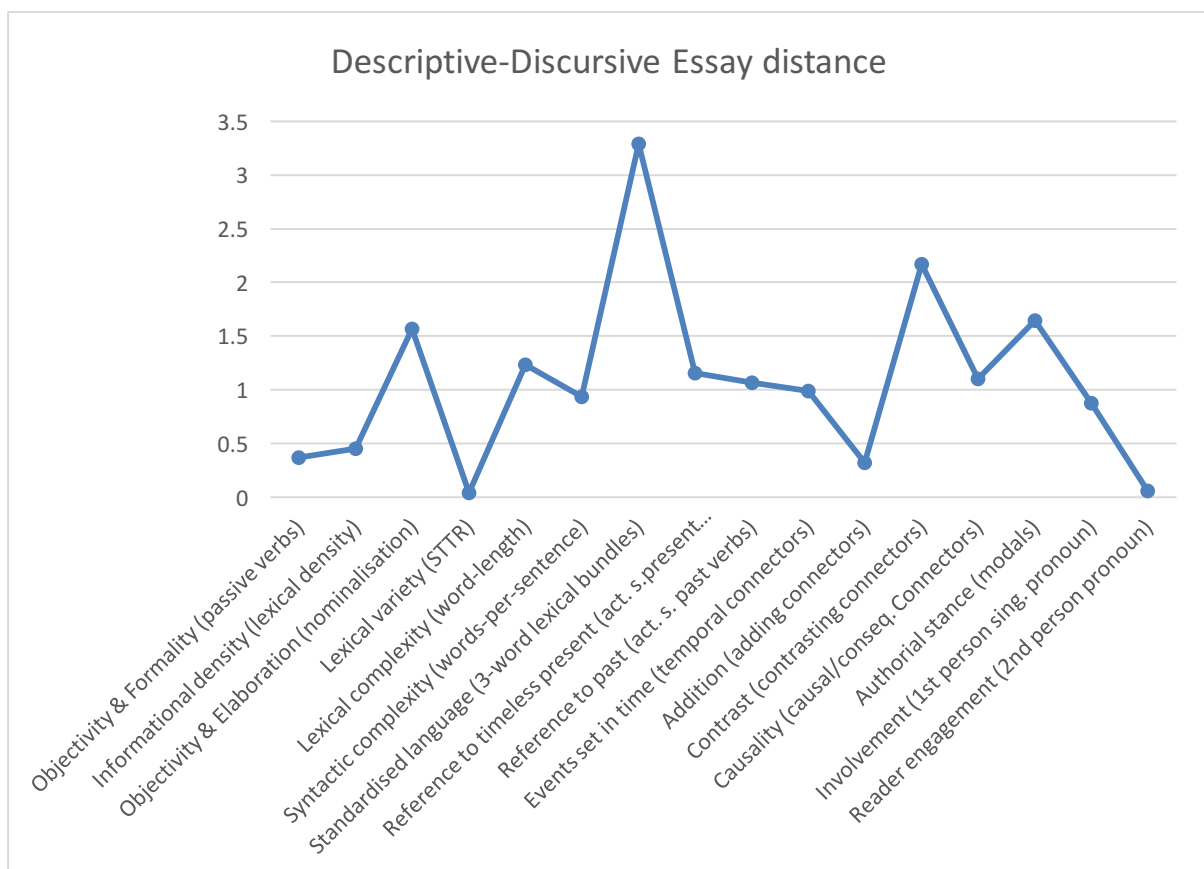
Figure 6.32 The Descriptive-Discursive Essay distance

It is interesting that the Descriptive Essay and the Personal Observation Report are the second strongest pair in terms of shared vocabulary with 47.6% of lexical overlap. The Descriptive Essay also shares more vocabulary with the Short Story (position 5, lexical overlap 45.9%) than it does with the Discursive Essay (position 7, lexical overlap 44.3%) and the Expository Essay (position 8, lexical overlap 43.7%).

The two Reports are on the 4th position based on text properties but they are on the 11th based on shared vocabulary. In fact, the Personal Observation Report has a stronger relation with the Expository Essay seen from both perspectives and a stronger relation with the Descriptive and the Discursive Essay when seen from the shared vocabulary perspective (positions 2 and 3 respectively). Figure 6.33 shows the distance between the Personal Observation Report and the Data Report as well as between the Personal Observation Report and the Expository Essay. As can be seen, the distance between the two Report genres is bigger in nine text properties than the distance between the Personal Observation Report and the Expository Essay. The mean distance between the Personal Observation Report and the Data Report is a bit larger (.773) than that with the Expository Essay (.564). This is because the two report genres are remarkably distant in Lexical variety (distance: 2.27401), Authorial stance (distance: 1.51839], Causality (distance: 1.47997), and Contrast (distance: 1.46774).
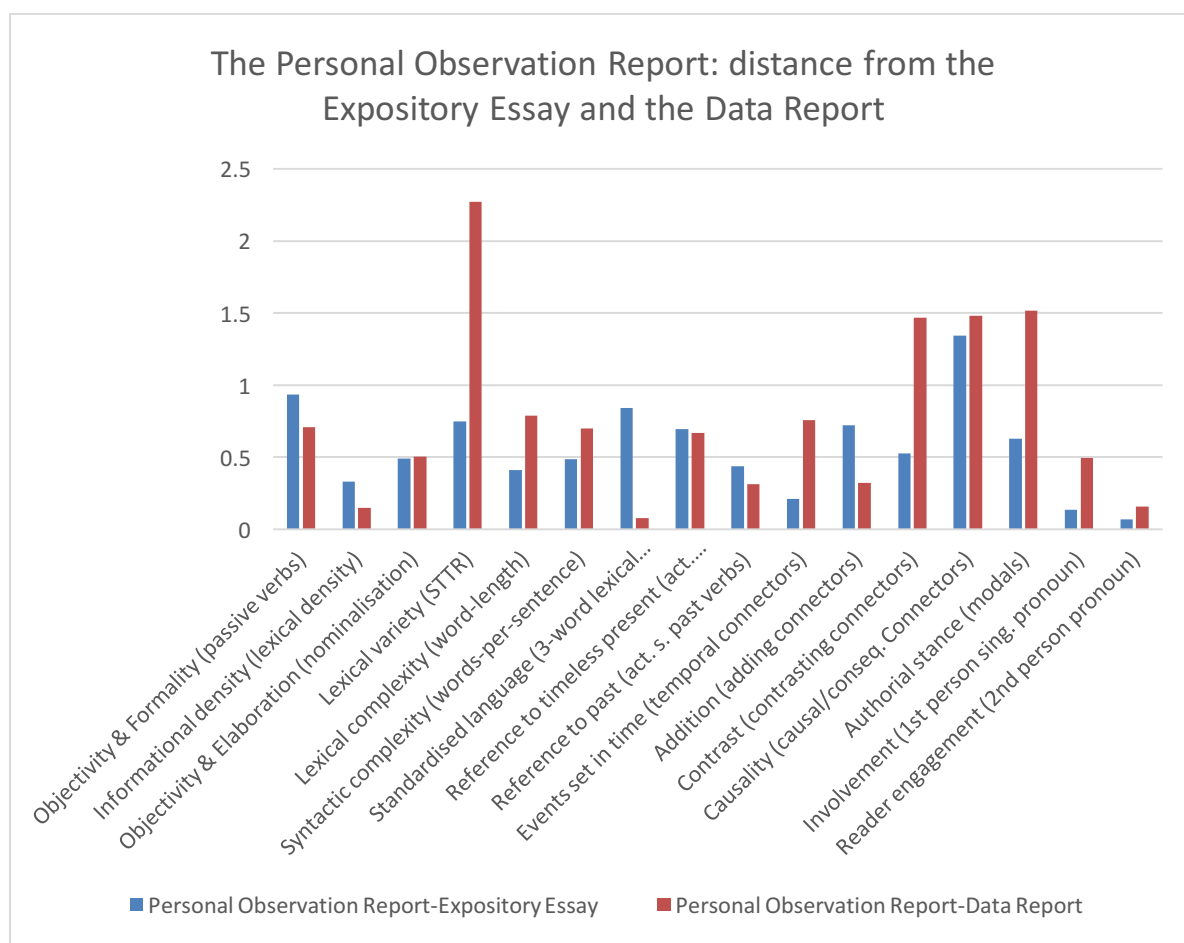
Figure 6.33 The Personal Observation Report: distance from the Expository Essay and the Data Report

The two Letter genres also match with other genres before they match with each other (13th position based on text properties and 9th position based on shared vocabulary). In terms of text properties, the Complaint Letter relates more strongly to the Description Essay (position 7), the Short Story (position 8), the Personal Observation Report (position 9) and the Data Report (position 12) than it does to the Advice Letter. The Advice Letter has the same degree of relation with the Complaint Letter as it has with the Descriptive Essay and then starts to relate to the rest of the genres. Although in exam task types they are distinguished by the Formal/Informal Letter label, it is clear that there is much more difference between them. Besides the distinction between Formal and Informal is not clearly defined in the literature. In figure 6.34 the distance between the two letter genres is big in Objectivity and Formality (distance: 2.69814), Reader engagement (distance: 2.20302), Reference to timeless present (distance: 1.67433), Authorial stance (distance: 1.63256), Contrast (distance: 1.52645), Lexical complexity (distance: 1.4518), Syntactic complexity (distance: 1.43124) as well as in Objectivity and Elaboration (distance: 1.40301).

Figure 6.34 The Complaint-Advice Letter distance

In terms of shared vocabulary, the Complaint Letter has a stronger relation with the Personal Observation Report (position 6) than it does with the Advice Letter (position 9). Even though the lexical overlap is big in both cases (44.6% and 42.4% respectively), this type of comparison also corroborates the claim that the two letter genres are quite distinct from each other to the extent that they may be more similar to different genres and that grouping them together in corpus categories can misguide corpus analysts.

To sum up, there is great similarity between the Expository and the Discursive Essay seen from both perspectives while the Descriptive Essay is closer to the Short Story in terms of vocabulary and although it is similar to the Expository Essay in terms of text properties it forms stronger relations to other genres before it matches the Discursive Essay. Neglecting to clarify the similarity and difference between the Descriptive and the other two Essays during teaching may lead to low quality writing. Evidence shows that genres involved in the Essay text type category are not necessarily more similar to each other than they are with other genres. The same has been shown for the Report category.

There is some similarity between the two Letter genres studied here but of medium strength. Therefore, the plain distinction in writing tasks between formal and informal letters is not very accurate and certainly not helpful to novice writers. Furthermore, the extent of differentiation between these two genres suggests that the tendency in big corpora to group Letters in the same category can affect the analyses based on these corpora negatively.

CHAPTER SEVEN: DISCUSSION

In this chapter I restate the four research questions explaining how each research question has been answered. I discuss the main findings showing how they connect to previous studies and how they add knew knowledge in areas where there has been a gap in the literature.

7.1 Teaching and testing second language writing: The learners' viewpoint.

The aim of this part of the research has been to identify the learners' difficulties, needs, preferences as well as their views on teaching practices related to EFL exams. This study based on a questionnaire handed to a large number of students in Rhodes, Greece attempted to fill the gap of the learners' own perspective being underrepresented in the literature resulting in lack of confidence in teachers' guidance (Hamp-Lyons, 2003: 168). Apart from being a student-focused research it is also the first to relate second language writing to more than one international EFL exams providing a realistic picture of preparation courses for various exams at the same time. Rather than focusing on university students as the majority of the studies have done, this study places young learners at the centre of interest addressing the particularity of the Greek context where young students seek foreign language certification participating in exams designed to address adult learners (Papaefthymiou-Lytra, 2012: 24, 25).

As testing is usually a 'daunting' experience (Gebril & Plakans, 2015: 1) raising anxiety levels due to time limitations and associated high-stakes decisions for the candidates (Ferris, 2008; Hamp- Lyons, 1991), gaining more knowledge on this area from the learners themselves adds crucial knowledge for the improvement of both teaching and assessment practices.

Findings show that 65% of the participants have already been certified. B2 level has 'the lion's share' (48%) followed by level C2 (24%) whereas both B1 and C1 certification refers to 12% of the students for each level. Only 4% of the certified students relate to levels lower than B1. The findings coincide with the B2 level priming in the Greek context (Dendrinos et al., 2013: 11, 16). C2 certificates are also valued for years due to the Greek legislation not only recognising foreign language proficiency to its holders but also seeing it as an official teaching qualification (Papaefthymiou-Lytra, 2012; Sifakis, 2009). The much smaller

numbers for <B1, B1 and C1 levels can be explained easily by the fact that these levels have been introduced and promoted more recently in the Greek context. Results show an upward trend for certification at these levels in young ages (age-group 12-17).

Writing has been considered a difficult skill by various researchers (Bereiter & Scardamalia, 1983; Lines, 2014; Raimes, 1994). This study shows that this view is also shared by students and refers to testing contexts too. Compared to the other skills (Reading, Listening and Speaking), 42% of the participants rated Writing as the most difficult skill. This has been much bigger than any other percentage. Writing preoccupies mostly males and ages 12-17, information that shows where the teaching of writing should place emphasis.

Concerning time spent on preparation especially for Writing in classrooms, 37% of the learners estimates that 1/3 of total time is spent on Writing and 30% think this time represents 1/4 of the total classroom time. In the learners' eyes therefore, time seems to be allocated fairly even though teachers think they place little if any emphasis on writing in previous research (Dendrinos et al., 2013:44).

The Review is rated as the most difficult text type, followed by the Report and the Formal Letter. Learners show a completely different attitude towards the two Letter text types. The Formal Letter gathers the biggest number of choices in the first position whereas for the Informal Letter this is seen in the sixth position. Most students consider the time spent on preparation for the Formal Letter, the Informal Letter, the Essay and the Short Story to be 'enough'. For the Report the majority of the answers refers to 'some' time whereas the Review has most answers at 'little time'. Clearly, the Review and the Report are both seen as difficult and disprivileged in the classroom whereas the Formal Letter is considered difficult even though enough time is spent on its preparation. This expressed need for more help regarding Reviews, Reports and Formal Letters is something that both teachers and educational material writers should take into consideration.

The learners' main difficulty in writing concerns grammar/syntax (31%) followed by vocabulary/appropriate phrases (28%). The analysis of the answers given by learners who are certified at C1 and C2 levels shows that 'none of the above' is the first choice at these levels suggesting less problems with writing as proficiency levels increase. This finding adds up to the previous observation of a higher need for help in Writing at young age groups.

'Vocabulary/appropriate phrases' attracts more answers than 'grammar/syntax' by C level students. The fact that basic grammar has usually been taught by the time students reach level B2 and that levels C1 and C2 involve advanced vocabulary justifies this difference.

Almost 35% of the students, state that they are affected moderately by time limits when writing. The second largest percentage refers to a slight effect (33%). Options related to more serious effects have not be chosen by many students which may be interpreted in two ways. Either writing tasks in textbooks and exams are reasonably timed or students have had adequate timed-writing practice in the classrooms.

In the writing classroom, learners express the need to be given model answers (22% of total answers), the desire for more writing practice represents 19% of total answers and the same is true for the need to discuss specific topics before writing. Their next choice concerns the need for more time in general (17%). 14% of them ask for more detailed feedback and only 9% want to know the assessment criteria beforehand. The use of model texts in the classroom is supported by many researchers (Charney & Carlson, 1995; Derewianka, 1990; Flowerdew, 1993; Hyland, 2004; Knapp, 1989; Tardy, 2006; 2009) and Tardy (2006: 94) has especially mentioned this as a student need. This study supports this claim.

7.2 The WriMA corpus

Building a specialized corpus for this study has been immensely beneficial. Based on strict criteria set by the researcher such as the inclusion of model answers targeting EFL exams, coming from many different sources with texts classified according to text type with detailed prompts and CEFR levels in metadata, it has been a valuable tool for the biggest part of this research. It has made it possible to develop and present a genre-identification method, find out how both text types and genres are represented in educational material, conduct genre analysis on individual genres based on large numbers of texts and contrast these genres with each other to assess their relations.

Nothing of the above would have been possible without the specific corpus as the type of research chosen could not be based on available general corpora. The experience has shown that a specialized corpus built by the researchers themselves can add new and original

knowledge and although its creation can be time-consuming it is worthwhile. This research therefore strongly agrees with the stressed significance and contribution of small specialized corpora in the literature (Flowerdew, 1998; 2004) and presents a step-by-step guide on how to build such a corpus, a process not adequately described in the literature (Rizzo, 2010).

7.3 Genres and text types

Due to the fuzziness around the term 'genre' and the time and effort this procedure may involve, the identification of genres is often avoided. This is seen both in the educational area studied here but also in corpus building. In the current study I have provided a review of the terms 'genre', 'register' and 'text type' in the literature, showing the lack of consensus among researchers and have positioned myself as to the distinction of the terms on a theoretical level. Through the following genre identification process and the contrastive analysis of eight genres I have offered evidence which supports the adopted position on the use of these terms.

On the theoretical level, I have sided with researchers who see register as encompassed and defined by genre rather than text type. I have also supported the view that text-type classification is too broad and may involve genre and register variation which goes unnoticed. The evidence provided discussed below shows that analyses based on corpora which are classified according to text type may affect the results as they disregard internal genre variation. Taking it to the educational context it shows that learners preparing for writing tasks grouped according to text types are not well-equipped to distinguish among genres even though this skill is a presupposition in language testing (Hamp-Lyons, 2003; Cambridge English First, 2015; Cambridge English Advanced, 2015; Cambridge English Proficiency, 2015). This is firstly because several genres are underrepresented in teaching material and secondly because linguistic variation can sometimes be greater between texts of the same text type group rather than texts of different text type labelling.

7.3.1 Genre identification

As the aim of this study is to explore written genres in EFL testing contexts and provide a basis for the teaching of these genres, the process of identification is a basic step. The genre identification process adopted in this study has enabled the classification of genres in

different sub-corpora within the WriMA corpus facilitating genre analysis. Although many researchers have stressed the importance of this stage and see the need for explaining the criteria used for text categorisation in corpora (Lee, 2001; Sharoff, 2015; Stamatatos et al. 2001) the description of the process of genre identification and the linguistic framework on which it has been based is a rare case in the literature. Despite the active interest in automatic genre recognition/identification (e.g., Stamatatos et al., 2000, 2001; Santini, 2006) there is no widely accepted method of categorisation yet.

What has been proposed here is an approach for manually identifying genre without the need for advanced computational knowledge. It is therefore time-consuming and more appropriate for small corpora. It is however, less time-consuming than approaches which require reading whole texts as this approach exploits the information given in task prompts and makes use of whole texts only in cases this is information is not adequate. It is, thus, appropriate for corpora with rich metadata. The solid linguistic framework of Systemic Functional Linguistics on which it has been based combining functional purpose and register variables (Halliday, 1978; Martin, 1985, 1983; Thompson 2014) as well as self-referential and inferential genre markers (Countinho & Miranda, 2009) provides a reliable method for grouping texts that actually represent a genre.

The additional stage of 'labelling' these genre categories mainly based on purpose and mode is helpful and practical for anyone using the corpus or anyone with a need to understand genre requirements such as second language novice writers in the case of this study.

7.3.2 'Text type' versus 'genre' representation in educational material

An initial observation during the compilation of the WriMA corpus has been the treatment of emails in several textbooks as similar to letters. Although the Email has been studied as an individual text type/genre in the literature (Wollman-Bonilla, 2003; Wright, 2013) several textbooks include advice and model answers under the comprehensive title Formal Letter/email or Informal Letter/email. The same has been observed in official exam guides (Cambridge English First, 2015: 30). The classification process adopted here, keeps separate categories for formal and informal letters/emails the same way they are presented in educational material and differentiates letters from emails only in cases where specific genres occur only as letters or emails (e.g. Reference Letter, Resignation Letter).

The seven text type categories in which texts were initially classified in the WriMA corpus are represented in the educational material as follows: 36% for Essays, 17% for Formal Letters/emails, 15.3% for Reports, 13.9% for Informal Letters/emails, 7.6% for Articles, 6.8% for Stories and 3.4% for Reviews.

Overall, Essays are overrepresented, a finding which coincides with Tribble's review (2009: 114) of 27 academic writing textbooks. Articles, Stories and Reviews on the other hand, are clearly underrepresented. Especially the Review with only 39 texts out of 1151 in the corpus is an interesting case as the previous questionnaire analysis revealed that learners rate this as the most difficult text type. In addition, most students state they have spent little time on this text type during classroom preparation.

The results after the genre identification show thirty-three different genres deriving from these text type categories. Apart from a great variety this process reveals a largely uneven distribution of genres with several of them poorly represented in educational material. This finding is an alert for material designers as they clearly fail to offer guidance in several genres asked for in testing contexts. This is especially observed in Formal Letters with twelve different genres when seven of them have a less than 1% representation. Among the seven Essay genres found, the 'Discursive Essay' predominates in the educational material which seems justified looking at the range of writing tasks posed in such exams as this is summarised in table 1.1.

This information on generic requirements is found in the official guides of some testing bodies and in most cases it is to be inferred by the students when reading the prompts. Teaching comes to play an important role here, filling the gap between implied genre requirements in exam tasks and preparation material which as this study has shown fails to highlight these issues and represent a large number of genres adequately. At the same time, contacts with second language teachers both in professional settings as well as at the conferences I have presented parts of this research show a lack of genre awareness and a serious need for more knowledge that is specific, evidence-based and that can be effectively transferred to the large number of teaching practioners and material writers around the world.

7.4 Genre analysis

7.4.1 Mixed methodology for genre analysis

The combination of both quantitative and qualitative analysis in this study addresses the problem observed in the literature of studies commonly being of one or the other type (Biber, 2004; Stubbs, 1996). Largely descriptive discourse analyses that are based on one or a small number of texts without specific measurements on the phenomena discussed run the risk of being subjective on what is chosen to be discussed and lack the contrastive character which can reveal the extent to which an element is actually typical of the genre at hand compared to other genres. Strictly quantitative studies on the other hand, no matter how statistically verified they may be, fail to provide a full picture of a genre or multiple genres if they do not provide an interpretation of the contextual factors, linking them to linguistic features and illustrating their use with several examples.

The way this analysis has been organized leads to an all rounded investigation of genres from various perspectives: first by seeing lexis and grammar as interdependent in the first part but also features from each category, measured separately in the second part. Second, by studying individual genres in depth and contrasting them in the second part to notice the extent of similarity and difference among them. The metrics used for comparison also come to enrich the information concerning individual genres presented in the first part. Third, by combining a variety of quantitative indices with qualitative analysis of their functional properties. Fourth, by referring to rhetorical moves together with lexicogrammatical features when most studies do either a detailed moves analysis or a register analysis. Rhetorical stages have been discussed where the use of certain words or phrases seemed to form patterns closely related to specific aims and common positions in the text. Figure 7.1 shows the mixed methods research for genre analysis used in this study:
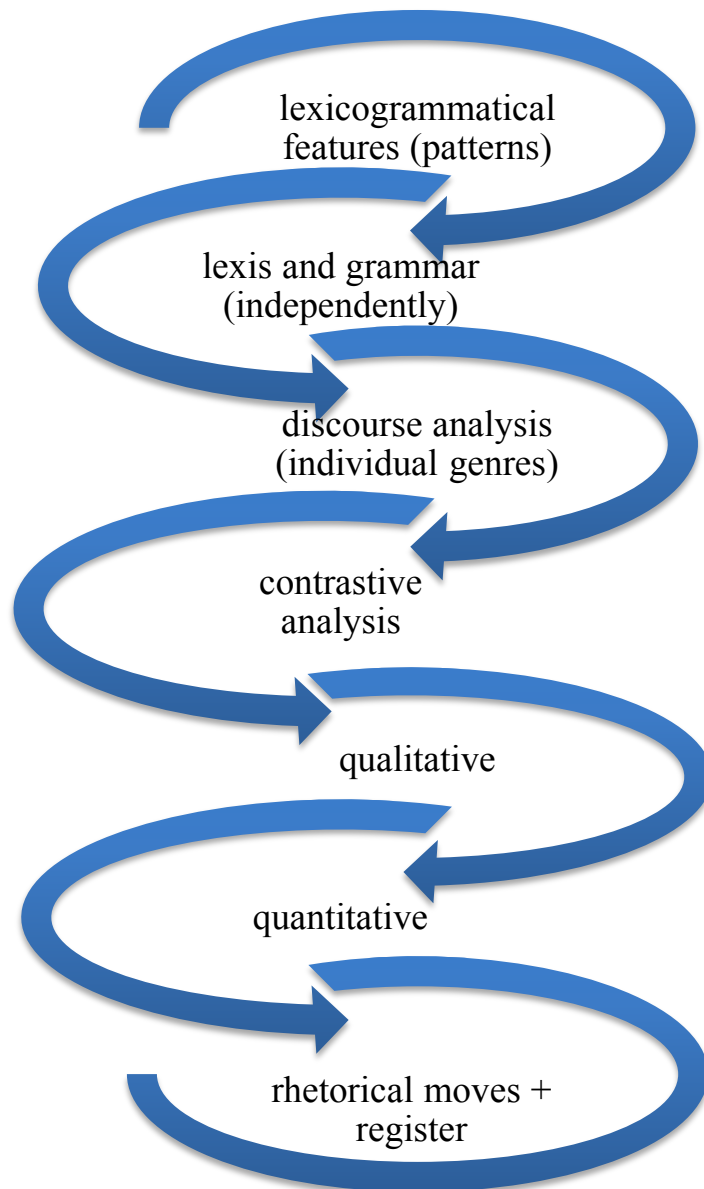
Figure 7.1 Mixed methods research used for genre analysis in this study

7.4.2 Understanding individual genres and relations among genres

This chapter begins with highlighting the main and typical features of each genre. I refer to main findings both from the individual genre analysis (chapter 5) and the contrastive analysis (chapter 6) and discuss genres in groups based on their text type as they are currently presented both in educational material as well as in exam tasks. This allows for points of convergence and divergence between genres to be discussed and seen in the larger frame of

Essays, Reports and Letters. The Short Story is discussed separately as both its name and the absence of other narrative genres under the same text type cause no confusion. Discussing its typical features and its relation to all other genres however, raises awareness and supports a genre-based pedagogy.

7.4.2.1 The Short Story

The Short Story, being part of the larger narrative genre, aims to entertain and/or instruct through a real or imaginative pattern of unexpected events leading to a problem which is usually resolved. The moves in the structure follow the pattern: ^ Orientation ^ Complication ^ Resolution.

Specific rather than general reference is made with named characters from familiar contexts (friends, relatives). The settings are also familiar. There is fast-paced action which is achieved by increased use of verbs, a great deal of which are action and mental-behavioural verbs. It is the only genre in this study with a noun/verb ratio of 1:1 leaning more towards spoken rather than written genres (Biber, 2006: 47, Ferencik, 2004, ch. 4.4). The use of past simple verbs is a lot more frequent than in the rest of the genres in this study. This fast pace is accentuated with the use of contractions, question and exclamation marks and dialogue. Temporal connectors set events in time and contrasting connectors introduce the unexpected in the story. Emotions are constantly described but there is an emphasis on the characters' psychological state at the 'Resolution' stage. Writer involvement is high as most texts in this corpus belong to the first-person narration type. This type of narration is frequently required in task prompts and is preferred by the majority of writers in this context when the task leaves room for choice.

Contrasted to the rest of the genres in this study, the Short Story is found to be closer to the Descriptive Essay in terms of shared vocabulary and the Complaint letter in terms of text properties. The next most similar genre is the Advice letter. It has distant relations with the rest of the genres.

7.4.2.2 The Expository, the Discursive and the Descriptive Essay

The purpose of the Expository Essay is to put forward a viewpoint, provide arguments in defense of or as objections to the proposition made, justify this position and reach a conclusion. A common structure is: ^ Introduction of the issue ^ Thesis statement ^ Arguments ^ Conclusion. The first paragraph includes the first two moves followed by separate paragraphs for each argument before the conclusion paragraph.

The Expository Essay refers to social issues and human concerns that can be debatable. These concerns often transcend local boundaries and refer to large groups of people or the world in general.

Main features are the general reference to groups of people ('they') instead of specific reference ('he', 'she', 'I', 'you'). Writers generalize about what people do what they need and/or what they should do using the simple present tense to imply that what is said applies generally. Adding and causal/consequential connectors are used extensively to add arguments on the same side and justify the position chosen. Increased use of nominalization and content over functions words (lexical density) create 'packed', meaningful sentences which are syntactically complex (increased ratios of words per sentence). There is also increased lexical variety (STTR) and lexical complexity (word length) in these texts. Writers take stance using modals and try to avoid direct involvement or reader engagement except for the moves where they need to position themselves (Thesis statement and Conclusion). They sometimes include themselves when referring to people in general, using the pronoun 'we' to make readers part of the group in an attempt to add strength to arguments about shared beliefs, needs and attitudes.

In the Discursive Essay the purpose is to discuss two opposing views presenting arguments for and against and form an opinion based on these arguments. Its structure commonly involves the following moves: ^ Introduction of the issue ^ Argument in favour of one side ^ Argument in favour of the other side ^ Summary of pros and cons and conclusion in favour of one side. A four paragraph essay is common with one paragraph for each move.

Socially significant issues are also raised in this genre but writers need to discuss different views before they position themselves. Therefore, apart from adding, causal and

consequential connectors writers also use contrasting connectors here. In fact, this genre uses more contrasting connectors than any other genre in this study. Although reference to general participants is also made in this genre the need for contrast leads writers to separate people in two groups in order to describe two opposing views. The word 'others' and the connectors 'on the one hand', 'on the other hand' are very typical of the genre. As this separation is based on perception rather than facts, people are separated using fractions rather arbitrarily (some/many/the majority/large numbers of people). During this contrast there is a tendency to refer to positive points more often than negative ones (advantages/benefits over disadvantages/drawbacks).

Another typical feature of this genre is the increased use of lexical bundles. In fact, the Discursive Essay has been found to be the most conventionalised genre in this study in the sense of formulaic, standardized expressions. The fact that the other two Essay genres come last in this text property is an interesting point of divergence within this set of genres. Writers use modals as a stance-taking technique here too, avoiding involvement and reader engagement.

The third genre studied as part of this set is the Descriptive Essay. Its purpose is to describe a person, place or event and in most instances this is combined with the need to justify why this person, place or event has impressed the writer. A common structure includes the following stages or moves: ^ Introduction of the subject of description ^ Extended description ^ Explanation of the writer's feelings.

In contrast to the other Essay genres, the Descriptive Essay makes specific rather than general reference using names and third person singular pronouns. Description is enhanced by the use of adjectives most frequently stressing positive characteristics of people or pleasant experiences. Another typical element of this genre is the connection of past to present time as reference to past experiences leads to feelings at present. Reader engagement is also low here but writers are involved and express their thoughts directly rather than taking stance in a discreet style in the form of modals. Even though involvement is higher than in the Expository and the Discursive Essay it is not as high as in other genres (the Complaint Letter and the Short Story).

The Descriptive Essay is strongly related to the Personal Observation Report in terms of shared vocabulary and to the Expository Essay in terms of text properties. Overall, this analysis has showed that there is great similarity between the Expository and the Discursive Essay less similarity between the Expository and the Descriptive Essay and much less between the Discursive and the Descriptive Essay. This makes sense as the Expository and the Discursive Essay belong to the larger argumentation genre family whereas description is a different genre category. It is the fact that both exam writing tasks and educational material treat these texts as an entity under the umbrella term 'Essay' that can cause problems as this grouping hides considerable variation among these task types as this study has shown. The possible misguidance of the candidate due to this classification in international testing contexts where failure is associated with critical decisions affecting the candidates' lives is a matter which has triggered this research. The initial suspicion that text type classification involves writing tasks which have different generic requirements has been confirmed through multiple types of analyses.

The initial discourse analysis has shown how the Expository and the Discursive Essay share similar purposes, moves and main linguistic features. It has also pointed to the features that cause finer variation between the two Essays. The analysis of the Descriptive Essay has verified the suspicion of greater variation in the typical linguistic features of this genre in relation to the other two Essays. It has also shown its clearly different purpose and move structure.

The contrastive analysis has examined these genres from two perspectives. It has revealed how similar they are in terms of the vocabulary used and in terms of basic text properties. The fact that this comparison has been based on large sets of texts and normalised values to control for size effect gives confidence as to the evidence provided and arms interested parties with quantified proof that can support or oppose long-standing beliefs and attitudes in teaching and assessment based on perception. The fact that this contrastive analysis has included other genres of the same context has given the researcher the ability to assess whether genres belonging to the same text type labelling set in this context share more similarity with the genres included in the same set or with genres included in other sets. Quantitative evidence through this analysis has verified the diversity of the Descriptive Essay in this category and its weaker relations with the other two Essay genres shown also in the first type of analysis. It has also shown that the Descriptive Essay is similar to genres that

belong to other text type categories. At the same time, both types of analysis have shown that the relation between the Expository and the Discursive Essay is very strong.

7.4.2.3 The Data and the Personal Observation Report

The purpose of the Data Report is to present information by summarising and contrasting the data provided in the prompt. The stages commonly followed are: ^ Introducing the subject and stating the purpose of writing ^ Description and comparison of key points ^ Conclusion.

Topics involve issues of public concern. Typical features of the genre include specific reference and precise information. People are grouped according to age, nationality or other criteria set by the data provided. Events are set in time and time is specified in years, hours or periods. Numbers, fractions, percentages but also approximation phrases are used which are also derived from the data. Comparison is a key element used to stress similarity and difference between trends in different time periods. Comparative adjectives, contrasting connectors and verbs indicating movement are frequently used at this stage. Present as well as past reference is made to describe transition and change. The language is fairly standardized with fixed expressions and low lexical variety. Texts are dense in information with high ratios of content over function words and increased use of nouns. Words are long and sentences are syntactically complex. The style is formal and objective with a lot of passive verbs and a distant tone lacking involvement, reader engagement and expression of authorial stance.

The Personal Observation Report on the other hand, is a report based on personal views or observations rather than data. The issues discussed have a local interest rather than a general one and the person reporting does so because of his/her proximity to the place or his/her experience of the service provided. Topics are usually relevant to young people's lives and interests. The ability to 'read' visualised data and the basic world knowledge needed in the Data Report are not necessary here.

The purpose of the Personal Observation Report is to describe the key features of a facility, service or product, assessing strengths and weaknesses leading (in most cases) to its recommendation to a person of higher status. The structure usually involves the following moves/stages: ^ Introducing the subject and stating the purpose of writing ^ Description of

key features (in different sections) ^ Conclusion (± Recommendation + Justification of recommendation). Sub-headings based on thematic units (e.g. Cost, Staff, Food) and/or moves (e.g. Introduction, Recommendation, Conclusion) guide the readers and signpost the information they are looking for.

There is specific reference here too but people are less often seen as groups. The recipient is usually named. Quantification is typical in this genre too with numbers and percentages but it can also be more subjective and general (e.g. many, several, few). Adding connectors are also used but contrasting connectors are less frequent than in the Data Report and other genres and comparison is achieved mainly by comparative adjectives. Causality however, is much more evident here than in the Data Report as there is a need to justify personal views which are not based on data and in cases where there is a recommendation stage it is necessary to explain the reasons for this recommendation.

Texts are also dense in information with long words and full sentences. There is standardisation in the sense of increased lexical bundles but there is more lexical variety than in the Data Report. There is no special interest in the past as is common in the Data Report. Although the Personal Observation Report is also objective and formal passive verbs are less often used than in the Data Report and there is some expression of authorial stance.

Contrastive analysis based on text properties has suggested a close relationship between these two Reports. There is however, greater similarity between the Personal Observation Report and the Expository Essay from both perspectives, text properties and shared vocabulary. The lexical overlap between the Personal Observation Report and all the other genres in this study has been greater than the one observed with the Data Report.

Based on these findings it is necessary to stress that although the two Reports are similar from a text property point of view, the fact that the Personal Observational Report is closer to the Expository Essay and the fact that its vocabulary is more similar to all other genres proves that text categories defined only by text type conceal significant variation between texts. The inclusion of these Reports under the same label implies more similarity within the group and less similarity with texts belonging in other groups while a variation in labelling even if this contains the word 'Report' indicates that there is some difference between the

two genres which should be explained in the classroom. Awareness of the similarities mentioned above can eliminate fears concerned with the difficulty level of Reports seen in the questionnaire analysis and due to the great similarity observed in this study, the Personal Observation Report can be taught close to the Expository Essay genre.

7.4.2.4 The Complaint and the Advice Letter

Letter genres are those that will most certainly be part of the students' lives and therefore their purpose is more easily understood by learners. Writing tasks in international language exams make a distinction between formal and informal letters referring this way to one aspect of style which has not been specified clearly in the literature yet. This study has focused on one letter from each category, the 'Complaint Letter' from formal letters and the Advice Letter from the informal letter category.

The purpose of the Complaint Letter is to complain to a person of high status about a faulty product, poor facilities or unsatisfactory service and ask for a response. The structure usually follows this pattern: ^ Purpose of writing ^ Reasons for complaining (description of the problem) ^ Expectations.

Typical features of this genre include conventionalised formal greetings in opening and closing positions and standardized expressions at specific stages. There is specific reference rather than general. Writers are involved and address the reader directly. They complain in a formal and polite way addressing the reader by surname, using a great deal of passive verbs and choosing to express trust that the situation will be resolved even though the relationship between writer and reader is not close. Contrasting connectors show the difference between what was expected and what has been offered while consequential connectors introduce the reasons of the disappointment and justify the necessity for a response. Even though past experience is connected to present feelings the focus is in the past as the description of the events leading to the problem is the biggest part of the letter.

A basic difference between the two letters in this study besides the formal/informal tone is the fact that the Complaint Letter initiates communication and asks for a response while the 'Advice Letter' is a response to previous communication.

The purpose of the Advice Letter is to respond to a friend asking for advice on how to deal with a problem. The structure usually involves the following moves: ^ Reference to previous communication stating the problem ^ Offer of advice and justification ^ Expression of hope for resolution ^ Request for further communication and updating.

Topics refer to everyday problems affecting young people and there is a close relationship between the writer and the reader. Because of this relationship, writers are involved, empathise with the readers and show sincere interest for their problems. The Advice Letter scores higher in authorial stance and reader engagement than all the other genres in this study. Although greetings are characteristic of this genre too, there is great variation in closing phrases. Standardised, formulaic expressions involve a great deal of phrasal verbs common in spoken genres. Informality is also evident in the use of contractions, punctuation (question and exclamation marks) and imperatives. In contrast to the Complaint Letter the focus here is in the present, the current problems and feelings.

The relationship between the two letter genres studied here is of medium strength based on text properties and a bit stronger if based on shared vocabulary. The Complaint Letter however, forms closer relationships with other genres and the Advice Letter is as similar to the Complaint Letter as it is to the Descriptive Essay.

Considering the dispersion of genres in Letters found in the genre identification process (12 under the formal Letter/email label and 6 under the Informal Letter/email label) this category seems to be the most problematic. The contrastive analysis of two letter genres has shown that the distinction between formal/informal text type refers to only one of many variables affecting register. Looking again at the extent of underrepresentation of Letter genres in educational material (chapter 4) and the findings in chapters five and six one could say that learners take part in these exams largely unprepared concerning letter-writing.

CHAPTER EIGHT: CONCLUSION

8.1 Conclusions and implications for teaching, material writing and assessment

The majority of the learners in the first part of this study rated 'writing' as the most difficult skill tested in international language exams. They also stated that they have spent one third or one fourth of the total classroom time on writing preparation which seems enough if one considers there are three more skills tested (Reading, Listening and Speaking) which require preparation too. Even though the need for more time in the writing classroom is expressed by the students it is not the first answer. These findings suggest that although time spent on preparation for writing in such contexts is allocated fairly, learners still find this skill difficult. Since the organisation of specialised writing courses is not a choice in this context due to time limitations, the teaching of writing needs to be more efficient, targeting student weaknesses and offering specific guidance.

Concerning CEFR levels, nearly half of the students have gained B2 level certification and one fourth have gained C2 level certification. Although B1 and C1 levels have attracted only 12% of the certified students in this study there has been indication of upward trends for these levels especially regarding the 12-17 age group. Interest for certification centers between levels B2 and C2 with exams below B1 level attracting only a minority of students.

The results have shown that writing difficulties preoccupy mostly males and young age-groups (12-17) while learners seem to overcome major difficulties as they grow up. Connecting these findings to everyday teaching practices it seems that the increased availability in courses and teaching material which address levels B2 and C2 is justified by the increased interest (also described by Dendrinos et al., 2013; Papaefthymiou-Lytra, 2012) but more courses and resources targeting B1 and C1 levels for non-adults should become available in time. Emphasis on the teaching of writing should especially be placed on groups of young students with teachers responding promptly particularly to male students' needs.

Regarding the traditional dichotomy between grammar and vocabulary still present in every day teaching practices and material, learners choose both as main difficulty with a slightly more increased percentage for grammar/syntax. As students at more advanced levels do not seem preoccupied with these difficulties it seems that more attention should be placed on

language demands related to writing in B level courses and teaching should focus on developing actual writing skills from then on.

The Review followed by the Report are the most difficult text types according to learners and the less privileged in terms of time spent in preparation. Combining this finding with the underrepresentation of Reviews in educational material, this research suggests the inclusion of more guidance on Reviews in material and the increase of teaching time and emphasis on both 'Reviews' and 'Reports' in the classroom.

The need for model answers has been the option with most positive answers regarding students' needs in the writing classroom. As the use of model answers in the classroom is supported by many researchers (Charney & Carlson, 1995; Derewianka, 1990; Flowerdew, 1993; Hyland, 2004; Knapp, 1989; Tardy, 2006; 2009) material writers need to offer as many writing model answers as possible and teachers need to allow time for the analysis of these models in the classroom. This study shows that feedback is not that important to students and supports previous suggestions (Hyland, 2003b; Scardamalia & Bereiter, 1986) that teachers should offer more explicit guidance and limit their reliance on feedback as a teaching method, expecting learners to 'demistify' the writing skill through implied and often vague advice.

The text type classification and labelling of texts has prevailed in exam writing tasks causing a subsequent adoption of the same labelling in second language educational material. I have attempted to clarify the terms 'genre', 'register' and 'text type' reviewing their use in the literature, an area characterized by lack of consensus, and have taken a position on the distinction of these terms.

Apart from the theoretical view this research has also applied genre identification processes in texts previously categorized according to text type. Through this process it has been shown that the term 'text type' is a broad grouping of texts involving many different genres. Through the following contrastive analysis of genres, it has been shown that the notion of 'register' is better connected to 'genres' rather than 'text types' in the sense that it is sometimes possible to find greater similarity between two genres which belong to different text type categories than between two genres of the same text type category. Through the

contrastive analyses cases of such internal variation have been observed in the Essay, the Report as well as the Letter category.

A lot of work needs to be done in order for the whole attitude on text classification and labelling to change and move towards a genre perspective the same way perhaps this has been achieved in primary and secondary schools in Australia (Board of studies, 1998; Knapp, 1989, 2002; Knapp & Watkins, 1994; Martin, 1985) and New Zealand (Glasswell et al. 2001), but this time taking advantage of the recent contribution of corpus linguistics in research. A consensus needs to be reached in this context on what exactly is meant by the term 'genre', what method should be followed in identifying main genres and what type of information should be included in curricula and teaching guides. The whole process in Australia and New Zealand was a national effort however, as it involved public schools. This is a different area with independent testing bodies, teaching material editors and publishers with no obligation to cooperate. As associations of language testers (such as ALTE – Association of Language Testers, EALTA- European Association for Language Testing and Assessment, ILTA-International Language Testing Association) constantly emerge around the world however, getting involved in research or 'spreading the word' to practitioners, there is reason to believe that all of the stakeholders will eventually be affected by new knowledge and new Schools of thought in applied linguistics. Even if this connection takes time, researchers should keep trying to make their findings publicly known and to promote interaction both among fellow researchers as well as between applied linguists and all those involved in second language teaching and assessment.

Teachers involved in this type of exam preparation can benefit from research such as this one and subsequently help their students become aware of the typical features of the genres involved in these tasks. Fighting against the time lag observed in the incorporation of applied linguistics findings into educational material (Burton, 2012; Littlejohn, 1992; McCarthy & O'Keefe, 2010; Sampson, 2013) and the 'hidden curriculum' concerning genre requirements (Devit et al. 2004; Nesi & Gardner, 2012; Reppen, 2002; Rothery, 1985) they can adjust their teaching to encompass genre-based teaching.

Respecting the learners' need for more model answers shown in this study as well as in previous research (Charney & Carlson, 1995; Derewianka, 1990; Flowerdew, 1993; Hyland, 2004; Knapp, 1989; Tardy, 2006, 2009) teachers should bring texts in the classroom which

are prototypical of the genre explored each time. The method followed for genre identification proposed here is clear and can be applied by people with basic knowledge of linguistics. Although it takes time it can arm teachers with examples of 'occluded' genres (Nesi & Gardner, 2012; Swales, 2004) such as the Review and some genres under the 'Formal Letter' label which were both found in this study to be underrepresented in educational material and considered difficult by learners.

In the genre analysis that follows the 'modelling' stage, the main features of each genre (presented in chapter 5) can be discussed and highlighted so that students become conscious of the conventions of each genre. Typical features presented here, are not to be used as a constraining formula for second language writers. They should not be treated as rigid templates but carefully brought to the attention of the students in order to make generic conventions more specific. Even though this specificity is beneficial to all language learners it is crucial to EFL learners who lack sociopragmatic knowledge due to their distance from naturalistic settings causing delay in the long process of situated acquisition (Hyland, 2007; McNamara & Roever, 2006). When these obstacles are combined with premature candidates, as is the case of Greek candidates, of exams which are designed to address adults (Papaefthymiou-Lytra, 2012) then the provision of such explicit guidance becomes essential.

Learners need to have mastered a basic or 'core vocabulary' according to CEFR level descriptions but there is more than one core vocabulary in a language depending on the communicative needs of the speaker and the selection of this vocabulary based on intuition is not the safest option (Ιακώβου et al., 2003). Using the vocabulary lists provided in this study (genre key words from chapter 5, frequent common words as they occur in each genre from appendix 6 and frequent lexical bundles from appendix 7) in classrooms can shortcut the learning process and provide explicit guidance in genre-based writing courses.

Finally, knowledge about the strength of relation between pairs of genres (chapter 6) can inform the ordering of genres both in teaching and in material design, a process traditionally based on intuition and subjective criteria. Teaching can thus move from known genres to those which are less known but most similar to the ones already taught in order to reduce teaching time and enhance effectiveness. This should be combined with the learners' world knowledge and cognitive abilities varying with age as discussed for certain genres. The required skill of 'reading' visualized data and quantifying information in the Data Report for

example, as opposed to the Personal Observation Report, or the socially significant issues requiring argumentation in the Discursive and the Expository Essay as opposed to the Descriptive Essay and the Short Story should be considered. As in this context of international language exams the progression of CEFR levels does not necessarily coincide with age - a student group of advanced proficiency level for example, could be younger than another group of lower level - this ordering of genres also requires careful thinking and adjustment according to the population of students.

Additionally, these findings can aid both everyday classroom assessment and formal assessment in testing contexts. Awareness of the typical features of genres in this context can lead to faster and less subjective evaluations of learners' work. Even carefully constructed and well- researched rating scales seem too general. Terms like 'adequate', 'reasonable', 'flow', 'formality' and 'cohessiveness' remain subjective and ambiguous. There is lack of definable and measurable criteria to assist the raters' work. Instead of remembering which aspects should be in the text, based on intuition, raters could base their judgment on elements that have proven to be basic and appropriate for the type of question and mark according to their presence or absence in the text. Consiousness about what should be sought for during assessment can also improve feedback offering students specific advice for each genre.

8.2 Limitations of this study and suggestions for further research

For the part of the study where learners' views are investigated, one should be aware that the 'aged 25 and above' sample is rather small as the focus has been on test-takers as close to the time of testing as possible. Therefore, conclusions for those ages may not be very reliable. The closed questions may have also affected results. In some cases, perhaps students had a limited choice of ready-made answers, which may to an extent be subjective not covering all possible answers. This is a common problem when one tries to create a short questionnaire. However, since most of the participants were young and restless I opted for large participation numbers rather than detailed questionnaires from less participants.

Genres analysed in detail in this study do not include the Review as not enough model answers were found for reliable findings based on corpus analysis. The underrepresentation of this genre in the content of the WriMA corpus is justified by the learners' statement that little time is spent on preparation for this genre in classrooms which leaves room for

assuming that this underrepresentation relates to the material in general. Keeping in mind the learners' expressed need for help in this genre, future corpus-based genre analyses could focus on this genre collecting model answers from a larger sample of educational material in order to provide a description and raise awareness of this genre too.

In the genre analysis, CEFR levels have been taken into consideration only as a criterion of representativeness in building the WriMA corpus not as variables in the analysis. Model answers written by expert writers are seen as instances of quality performance with the focus being on genre-specific properties, an under-researched area in the specific testing context. Within the testing context however, research related to CEFR levels is very active. Recently, there has been "an increasing awareness among researchers of the need to add language-specific lexical and grammatical details to the functional characterizations of the proficiency levels in the CEFR" (Callies & Gotz, 2015: 2). The notion of 'criterial features', central to research conducted by the English Profile programme, aspires to create an inventory, a 'set of specifications of lexis and grammar', in order to enrich and enhance the CEFR (Hawkins & Filipović 2012; Tono, 2013). Even though I support the view that learners do not move linearly in terms of writing competence from one genre to the other and that their performance depends on their awareness of specific genres (Fu, 2009; Torrance, 1996a), I understand the need for the connection of various competences to proficiency levels for practical reasons of assessment in high-stakes exams. Researchers with such a focus could use findings from this study as a basis for comparison with learner performance on specific genres, using learner corpora. This way genre performance could be related to CEFR levels. I believe that the present findings can be a contribution to this effort too.

The shared vocabulary lists provided in appendix six and the three-word lexical bundles per genre provided in appendix seven could be the basis of in-depth future analyses with a focus on vocabulary and/or lexical bundles. The pedagogic value of formulaic sequences has recently gained attention (Jung et al., 2015; Martinez & Schmitt, 2012; Simpson-Vlach & Ellis, 2010) challenging many English learning resources, such as dictionaries and vocabulary books which usually present language points word by word (Huang, 2013: 186).

Finally, I would like to stress the value of specialized corpora for educational purposes, especially those which are created by educators themselves aiming to address specific questions related to teaching and assessment. The use of ready-made general corpora in

research up to now has often resulted in studies which are either too narrow with unclear pedagogic value or too general with no specified context or intended audience in mind. Corpora are powerful tools, able to provide answers to a great number of research questions. If, however, researchers decide on the type of question to ask based on the limited availability of corpora instead of trying to address practical problems and under-researched areas using tailor-made corpora, this tool will not be fully exploited and findings will be less interesting than they could have been.

References (in English)

Abbs, P. & Richardson, J., (1990). *The forms of narrative. A practical study guide for English.* Cambridge, UK: Cambridge University Press.

ALTE, (2009). *Writing: Individual components checklist.* Association of Language Testers in Europe. Available online at http://www.alte.org/attachments/files/writing_check.pdf

Amogne, D. (2013) Enhancing students' writing skills through the genre approach. *International Journal of English and Literature,* 4(5), 242-248.

Anthony, L. (2003). *AntMover* (Version 1.0.0) [Computer Software]. Tokyo: Waseda University. Available from http://www.laurenceanthony.net/

Anthony, L. (2014). *TagAnt* (Version 1.1.2) [Computer Software]. Tokyo: Waseda University. Available from http://www.laurenceanthony.net/

Baker, P. (2004). Quering keywords. Questions of difference, frequency and sense in keyword analysis. *Journal of English Linguistics*, 32(4), 346-359.

Baker, P., Hardie, A. & McEnery, T. (2006). *A Glossary of Corpus Linguistics.* Edinburgh: Edinburgh University Press.

Barker, F. (2008). Exploring lexical differences in general English reading papers. *Research Notes,* 31, 22-26.

Bazerman, C. (2012). Genre as social action. In J. P. Gee & M. Handford (Eds.), *The Routledge handbook of discourse analysis* (pp. 226-238). USA and Canada: Routledge.

Beck, S.W. & Jeffery, J.V. (2007). Assessing Writing. *Science Direct,* 12, 60-79.

Bereiter, C. & Scardamalia, M. (1983). Does learning to write have to be so difficult? In A. Freedman, I. Pringle & J. Yalden (Eds.), *Learning to Write: First Language/ Second Language.* London: Longman.

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

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

Biber, D. (1989). A typology of English texts. *Linguistics,* 27(1), 3-43.

Biber, D. (1990). Methodological issues regarding corpus-based analyses of linguistic variation. *Literary and Linguistic Computing* 5(4), 257-69.

Biber (1993). Representativeness in Corpus Design. *Literary and Linguistic Computing* 8(4), 243-257.

Biber, D. (1995). *Dimensions of Register Variation: A Cross-Linguistic Comparison.* Cambridge: Cambridge University Press.

Biber, D. (2004). A corpus linguistic investigation of vocabulary-based discourse units in university registers. In U. Connor and T. A. Upton (Eds.) *Applied corpus linguistics: A multidimensional perspective* (pp. 53- 72). Amsterdam-New York, NY: Rodopi.

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

Biber, D. (2010). What can a corpus tell us about registers and genres? In A. O'Keeffe and M. McCarthy (Eds.), *The Routledge Handbook of Corpus Linguistics* (pp. 241- 254). London and New York: Routledge.

Biber, D. (2012). Register and discourse analysis. In J. P. Gee & M. Handford (Eds.), *The Routledge Handbook of Discourse Analysis* (pp. 191-208) USA and Canada: Routledge.

Biber D., Conrad S. & Reppen R. (1994). Corpus-based approaches to issues in applied linguistics. *Applied Linguistics,* 15(2), 169-189.

Biber, D., Conrad, S. & Reppen, R. (1998). *Corpus Linguistics: Investigating Language Structure and Use.* Cambridge, UK: Cambridge University Press.

Biber, D., Johansson, S., Leech, G., Conrad, S. & Finegan, E. (1999). *Longman Grammar of Spoken and Written English.* London: Longman.

Board of studies. (1998). *K-6 English syllabus: Modules.* Sydney: Board of Studies.

Bondi, M. & Danni, Y. (2015). Forward-looking statements in CSR reports: a comparative analysis of reports in English, Italian and Chinese. In Formato F. & Hardie A. (Eds.), *Corpus Linguistics 2015 Conference Abstract Book (pp.* 58-59). Lancaster: UCREL.

Boulton, A. & Tyne, H. (2015). Corpus-based study of language and teacher education. In M. Bigelow & J. Ennser-Kananen (Eds.), The Routledge *Handbook of Educational Linguistics* (pp. 301-312). New York: Routledge.

Boulton, A. (2010). Learning Outcomes From Corpus Consultation. In M. Moreno Jaen, F. Serrano Valverde & M. Calzada Perez (Eds.), *Exploring New Paths in Language Pedagogy: Lexis and Corpus-based Language Teaching* (pp.129-144). London: Equinox.

Brown, H.D. (2007). Teaching by Principles. An Interactive Approach to language pedagogy (3rd edition). White Plains, NY: Pearson Longman.

Burton, G. (2012). Corpora and coursebooks: destined to be strangers forever? *Corpora,* 7(1), 91–108.

Callies, M. & Gotz, S. (2015). Learner corpora in language testing and assessment: prospects and challenges. In Marcus Callies & Sandra Gotz (Eds.), *Learner Corpora in language testing and assessment* (pp.1-9). Amsterdam: John Benjamins.

Cambridge English First: Handbook for teachers, (2015). Cambridge English Language Assessment. Available at: http://www.cambridgeenglish.org/images/cambridge-english-first-handbook-2015.pdf

Cambridge English Advanced: Handbook for teachers, (2015). Cambridge English Language Assessment. Available at: http://www.cambridgeenglish.org/images/cambridge-english-advanced-handbook-2015.pdf

Cambridge English Proficiency: Handbook for teachers, (2015). Cambridge English Language Assessment. Available at: http://www.cambridgeenglish.org/images/168194-cambridge-english-proficiency-teachers-handbook.pdf

Carter, R., Hughes, R., McCarthy, M. (2000). *Exploring Grammar in Context.* Cambridge, UK: Cambridge University Press.

Carter, R. & McCarthy, M. (2006). *Cambridge Grammar of English: A Comprehensive Guide.* Cambridge, UK: Cambridge University Press.

Carter, R., McCarthy, M. Mark, G. & O'Keefe, A. (2011). *English Grammar Today: An A-Z of Spoken and Written Grammar.* Cambridge, UK: Cambridge University Press.

Casanave, C. P. (2004). *Controversies in Second Language Writing: Dilemmas and Decisions in Research and Instruction.* Ann Arbor, MI: University of Michigan Press.

Chang, J. Y. (2010). Postsecondary EFL students' evaluations of corpora with regard to English writing. *The SNU Journal of Education Research,* 19, 57-85.

Charney, H. D., & Carlson, A. R. (1995) Learning to Write in a Genre: What Student Writers Take from Model Texts. Research in the Teaching of English, 29(1), 88-125.

Chomsky, N. (1957). *Syntactic Structures.* The Hague/Paris: Mouton.

Chomsky, N. (1965). *Aspects of the Theory of Syntax.* Cambridge, Massachusetts: MIT Press.

Christie, F. (1984). *Varieties of written discourse (Section 1, Children Writing: B.Ed. Study Guide)* (pp.11-51). Geelong, Victoria: Deakin University Press.

Christie, F. (1993). Curriculum Genres: Planning for Effective Teaching. In B.

Cope and M. Kalantzis (Eds.), *The Powers of Literacy: A Genre Approach to Teaching Writing* (pp. 154-178). Pittsburgh: University of Pittsburgh Press.

Cohen, A. D. & Cavalcanti, M. C. (1990). Feedback on compositions: Teacher and student verbal reports. In B. Kroll (Ed.), *Second language writing. Research insights for the classroom* (pp. 155-177). Cambridge: Cambridge University Press.

Conrad, S. (2010). What can a corpus tell us about grammar? In A. O'Keeffe & M. McCarthy (Eds.), *The Routledge Handbook of Corpus Linguistics* (pp. 227-240). London: Routledge.

Conrad, S. & Biber, D. (Eds.). (2001). *Variation in English: Multi-dimensional studies.* Harlow: Pearson.

Cope, B. & Kalantzis, M. (1993). *The Powers of Literacy: A Genre Approach to teaching Writing.* Pittsburgh: University of Pittsburgh Press.

Council of Europe (2001). *Common European Framework of Reference for Languages: Learning, Teaching, Assessment.* Cambridge: Cambridge University Press.

Coutinho, M. A. & Miranda, F. (2009). To describe genres: Problems and strategies. In C. Bazerman, A. Bonini & D. Figueiredo (Eds.), *Genre in a changing world* (pp. 35-55). West Lafayette, IN: Parlor Press.

Crossley, S.A, Allen, L.K & McNamara D.S (2014). A Multi-Dimensional analysis of essay writing. In T. B. Sardinha & M. V. Pinto (Eds.), *Multi-Dimensional Analysis, 25 years on: A tribute to Douglas Biber* (pp. 197-237). Amsterdam- Philadelphia: John Benjamins Publishing Company.

Cumming, A. (2003). Experienced ESL/EFL writing instructors' conceptualizations of their teaching: Curriculum options and implications. In B. Kroll (Ed.), *Exploring the dynamics of second language writing* (pp. 71-92). Cambridge, UK: Cambridge University Press.

Cummings M. (2003). Iterative and Lexical Densities within Genres. In D. W. Coleman, W. J. Sullivan & A. Lommel (Eds.), *LACUS Forum XXIX: Linguistics and the Real World (pp. 193-204).* Houston: LACUS.

Czerniewska P. (1992). *Learning about Writing.* Oxford: Blackwell.

Dendrinos, B., Karavas, E. & Zouganelli, K., (2013). *European Survey on Language Competences: Greek National Report.* Athens: RcEL publications & The Institute of Educational Policy, Ministry of Education.

Derewianka, B. (1990). *Exploring How Texts Work.* Newtown, Australia: Primary English Teaching Association.

Derewianka B. (1996). *Exploring the Writing of Genres.* Royston, Hertfordshire: UK Reading Association.

Devitt, A. J., Reiff, M. J. & Bawarshi, A. (2004). *Scenes of writing: Strategies for composing with genres.* New York, NY: Longman.

Devitt, A. (2009). Teaching Critical Genre Awareness. In C. Bazerman, A. Bonini & D. Figueiredo (Eds.), *Genre in a changing world* (pp. 337-351). West Lafayette, IN: Parlor Press.

Dunning, T. (1993). Accurate Methods for the Statistics of Surprise and Coincidence. *Computational Linguistics* 19(1), 61-74.

ECCE (2015). *ECCE Report 2014*. Cambridge Michigan Language Assessments. Available online at: https://www.cambridgemichigan.org/wp-content/uploads/2015/04/ECCE-2014-Report.pdf

ECPE (2015). *ECPE Report 2014*. Cambridge Michigan Language Assessments. Available online at https://www.cambridgemichigan.org/wp-content/uploads/2015/04/ECPE-2014-Report.pdf

ECCE (2012). *ECCE Sample Test Guide 2013*. Cambridge Michigan language Assessments. Available online at http://www.cambridgemichigan.org/wp-content/uploads/2014/11/ECCE-2013-Sample-Test-Guide.pdf

Ellis, R., Johnson, K. E., Henry, A., & Roseberry, R. L. (1998). An evaluation of a genre-based approach to the teaching of EAP/ESP writing. *TESOL Quarterly*, 32, 147-156.

ETS, TOEFL (2011). *A guide to the TOEFL test for institutions*. Educational Testing Service. Available online at http://www.ets.org/s/toefl/pdf/guide_toefl_test_institutions.pdf

ETS TOEFL iBT, (2015). *TOEFL iBT Test Questions.* Educational Testing Service. Available online at http://www.ets.org/Media/Tests/TOEFL/pdf/SampleQuestions.pdf

Fang, Z., Schleppegrell, M.J. & Cox, B.E. (2006). Understanding the Language Demands of Schooling: Nouns in Academic Registers. *Journal of Literacy Research,* 38(3), 247-273.

Fathman, A. & Whalley, E. (1990). Teacher response to student writing: focus on form versus content. In B. Kroll (Ed.), *Second language writing: Research insights for the classroom* (pp. 178-190). Cambridge: Cambridge University Press.

Feez, S. (1998). *Text-based syllabus design*. Sydney: McQuarie University/ AMES.

Feez, S. (2002). Heritage and innovation in second language education. In A. Johns (Ed.) *Genre in the classroom: Multiple perspectives* (pp. 47-68). Mahwah, NJ: Lawrence Erlbaum.

Ferenčík, M. (2004). A Survey of English Stylistics. Prešovská univerzita v Prešove. Available online at http://www.pulib.sk/elpub2/FF/Ferencik/INDEX.HTM

Ferris, D. R. (2008). Myth 5: Students must learn to correct all their writing errors. In J. Reid (Ed.), *Writing myths: Applying second language research to classroom teaching* (pp.90-114). Ann Arbor, MI: University of Michigan Press.

Ferris, D. (2011). *Treatment of error in second language student writing* (2nd edition). Ann Arbor, MI: University of Michigan Press.

Fialho, O., Miall, D.S. & Zyngier, S. (2012). Experiencing or interpreting literature: Wording instructions. In M. Burke et al., (Eds.) *Pedagogical Stylistics: Current Trends in Language, Literature and ELT* (58-74). London, New York: Continuum.

Finegan, E., & Biber, D. (1994). Register and social dialect variation: An integrated approach. In D. Biber & E. Finegan (Eds.), *Sociolinguistic perspectives on register* (pp. 315-347). New York: Oxford University Press.

Firth, J. R. (1957). A synopsis of linguistic theory, 1930-1955. Studies in linguistic Analysis, [Special Volume], 1-32.

Flowerdew, J. (1993). An educational, or process, approach to the teaching of professional genres. *ELT Journal,* 47(4), 305-16.

Flowerdew, J. (2002). Genre in the classroom: A linguistic approach. In A. Johns (Ed.), *Genres in the Classroom: Multiple Perspectives* (pp. 91-102). Mahwah, NJ: Lawrence Erlbaum.

Flowerdew, J. (2011a). Some dichotomies in applied genre analysis. In V. K. Bhatia, P. Perez-Paredes and P. Sanchez Hernandez (Eds.), *Researching Specialized Languages* (pp. 139-153). Amsterdam: John Benjamins.

Flowerdew, J. (2011b). Reconciling approaches to genre analysis in ESP: The whole can equal more than the sum of the parts. In D. Belcher, A. Johns and B. Paltridge (Eds.), *New directions in ESP research* (pp. 119-144). Ann Arbor, MI: University of Michigan Press.

Flowerdew, J. (2015). John Swales's ideas on pedagogy in Genre Analysis: A view from 25 years on. *Journal of English for Academic Purposes*, 19, 102-112.

Flowerdew, J. & Wan, A. (2010). The linguistic and the contextual in applied genre analysis: The case of the company audit report. *English for Specific Purposes,* 29(2), 78-93.

Flowerdew, L. (1998). Concordancing on an Expert and Learner Corpus for ESP. *CALL Journal,* 8(3), 3-7.

Flowerdew, L. (2000). Investigating referential and pragmatic errors in a learner corpus. In L. Burnard & T. McEnery (Eds.), *Rethinking Language Pedagogy from a Corpus Perspective* (pp. 145-154). Frankfurt: Peter Lang.

Flowerdew, L. (2004) The argument for using English specialized corpora to understand academic and professional settings. In U. Connor & T. Upton (Eds.) *Discourse in the professions: Perspectives from corpus linguistics* (pp. 11-36). Amsterdam: John Benjamins.

Flowerdew, L. (2005). An integration of corpus-based and genre-based approaches to text analysis in EAP/ESP: Countering criticisms against corpus-based methodologies. *English for Specific Purposes* 24, 321-332.

Foley, J. (1990). Genre: Verbal tailoring from ready-made cloth? *Language Sciences,* 12(2), 221-242.

Freedman, A. (1993). Show and Tell? The Role of Explicit Teaching in the Learning of New Genres. *Research in the Teaching of English,* 27(3), 222-251.

Frodesen, J. & Holten, C. (2003). Grammar and the ESL writing class. In B. Kroll (Ed.), *Exploring the dynamics of second language writing* (pp. 141-161). Cambridge, UK: Cambridge University Press.

Fu, D. (2009). Writing Between Languages: How English Language Learners Make the Transition to Fluency, grades 4-12. Portsmouth, NH: Heinemann.

Gabrielatos C. (2005). Corpora and Language Teaching: Just a fling or wedding bells? *TESL-EJ,* 8(4), A1, 1-37.

Gardner, S. & Holmes, J. (2009). Can I use headings in my essay? Section headings, macrostructures and genre families in the BAWE Corpus of student writing. In M. Charles, S. Hunston & D. Pecorari (Eds.), *Academic Writing: At the Interface of Corpus and Discourse* (pp. 251-271). London: Continuum.

Gardner, S. and Holmes, J. (2010). From section headings to assignment macrostructure in undergraduate student writing. In E. Swain (Ed.), *Thresholds and potentialities of systemic functional linguistics: Applications to other disciplines, specialized discourses and languages other than English* (254-276). Trieste: Edizioni Universitarie Trieste.

Gebril, A. and Plakans, L. (2015). *Assessment myths.* Ann Arbor, MI: University of Michigan Press.

Glasswell, K., Parr, J., Aikman, M. (2001). *Development of the asTTle writing assessment rubrics for scoring extended writing tasks.* Technical Report 6, Project asTTle. Auckland: University of Auckland.

Grabe, W. (2003). Reading and writing relations: Second language perspectives on research and practice. In B. Kroll (Ed.), *Exploring the dynamics of second language writing.* Cambridge, UK: Cambridge University Press.

Granger, S. (2002). A Bird's-eye View of Learner Corpus Research. In S. Granger, J. Hung & S. Petch-Tyson (Eds.), *Computer Learner Corpora, Second Language Acquisition and Foreign Language Teaching* (pp. 3-33). Amsterdam & Philadelphia: John Benjamins.

Granger, S., Meunier, F., & Dagneaux, E. (2002). *The International Corpus of Learner English. Handbook and CD-ROM.* Louvain-la-Neuve: Presses Universitaires de Louvain.

Gries, S., Newman, J. & Shaoul, C. (2011). N-grams and the clustering of registers. *Empirical Language Research Journal, 5*(1), 1-13.

Halliday, M.A.K. (1978). *Language as social semiotic: The social interpretation of language and meaning.* London: Edward Arnold.

Halliday, M.A.K. (1985). *An Introduction to Functional Grammar*. London: Edward Arnold.

Halliday, M.A.K. (1992). Language as system and language as instance: The corpus as a theoretical construct. In J. Svartvik (Ed.), *Directions in corpus linguistics: Proceedings of Nobel Symposium 82 Stockholm*, volume 65 (pp. 61-77). Berlin: Mouton de Gruyter.

Halliday, M.A.K. & Ruqaiya Hasan (1985). *Language, context, and text: Aspects of language in a social-semiotic perspective.* Oxford/Geelong: Oxford University Press/Deakin University Press.

Halliday, M.A.K. & Matthiessen, C. (2004). *An Introduction to Functional Grammar (3ʳᵈ edition).* London: Edward Arnold.

Hammond, J., and Derewianka, B. (2001). An introduction to genre. In D. Nunan & R. Carter (Eds.), *The ELT Handbook* (pp.32-57). Cambridge: Cambridge University Press.

Hamp-Lyons L. (1991). Pre-text: Task- related influences on the writer. In L. Hamp-Lyons (Ed.), *Assessing second language writing in academic contexts* (pp.87-107). Norwood, NJ: Ablex.

Hamp-Lyons L. (2003). Writing teachers as assessors of writing. In B. Kroll (Ed.), *Exploring the dynamics of second language writing.* Cambridge, UK: Cambridge University Press.

Handford, M. (2010). What can a corpus tell us about specialist genres? In A. O'Keeffe & M. McCarthy (Eds.), *The Routledge Handbook of Corpus Linguistics* (pp. 255- 269). London: Routledge.

Harklau, L. & Pinnow, R. (2009). Adolescent Second Language Writing. In L. Christenbury, R. Bomer & P. Smagorinsky (Eds.), *Handbook of Adolescent Literary Research* (pp. 126-39). New York: The Guilford Press.

Hawkins, J. A. & Filipović, L. (2012). *Criterial Features in L2 English*: *Specifying the Reference Levels of the Common European Framework*, English Profile v.1. Cambridge: Cambridge University Press.

Henry, A. & Roseberry, R. L. (1996). A corpus- based investigation of the language and linguistic patterns of one genre and the implications for language teaching. *Research in the teaching of English* 30, 472-492.

Henry, A., & Roseberry, R. L. (2001). A narrow-angled corpus analysis of the move and strategies of the genre: Letter of Application. *English for Specific Purposes Journal, 20*, 153-167.

Henry, A., & Roseberry, R. L. (2007). Language errors in the genre-based writing of advanced academic ESL students. *RELC Journal*, *38*(2), 171-198.

Hunston, S. (2002). *Corpora in Applied Linguistics.* Cambridge: Cambrdige University Press.

Huang, Z. (2013). *The Core and Periphery of Lexical Bundles in Modern English Dialogues: A Comparative Study of English Varieties.* PhD Thesis. The Chinese University of Hong Kong. Available online at http://repository.lib.cuhk.edu.hk/en/item/cuhk-327650

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

Hyland, K. (2001). Humble servants of the discipline? Self-mention in research articles. *English for Specific Purposes, 20*, 207–226.

Hyland, K. (2002). Genre: Language, Context and Literacy. *Annual review of Applied Linguistics*, 22, 113-135.

Hyland, K. (2003a). *Second Language Writing.* Cambridge University Press, USA.

Hyland, K. (2003b). Genre pedagogy: Language, literacy and L2 writing instruction. *Journal of Second Language Writing,* 16(3), 148-164.

Hyland, K. (2004). *Genre and second language writing.* Ann Arbor, MI: University of Michigan Press.

Hyland, K. (2005). Stance and engagement: a model of interaction in academic discourse. *Discourse Studies* 7(2), 173-192.

Hyland, K. (2007). Genre pedagogy: Language, literacy and L2 writing instruction. *Journal of Second Language Writing,* 16, 148–164.

Hyland, K. (2009). Corpus Informed Discourse analysis: The case of academic engagement. In M. Charles, S. Hunston & D. Pecorari (Eds.) *Academic Writing: At the Interface of Corpus and Discourse* (pp. 110-128). London: Continuum.

Hyland, K. (2012). English for academic purposes and discourse analysis. In J. P. Gee & M. Handford (Eds.), *The Routledge Handbook of Discourse Analysis* (pp. 412- 423). London: Routledge.

Hymes, D. H. (1972). On communicative competence. In J. B. Pride & J. Holmes, *Sociolinguistics: selected readings* (pp. 269–293). Harmondsworth: Penguin.

Hyon, S. (1996). Genre in Three Traditions: Implications for ESL. *TESOL Quar- terly, 30*(4), 693-722.

IELTS (2016a). *IELTS Test format.* Available online at https://www.ielts.org/about-the-test/test-format

IELTS (2016b). *IELTS Sample test questions.* Available online at https://www.ielts.org/about-the-test/sample-test-questions

Johns, A. M. (1995). Genre and Pedagogical Purposes. *Journal of Second Language Writing,* 4(2), (181-190).

Johns, A. M. (2003). Genre and ESL/EFL composition instruction. In B. Kroll (Ed.), *Exploring the dynamics of second language writing.* Cambridge, UK: Cambridge University Press.

Johns, A. M. (2011). The future of genre in L2 writing: Fundamental, but contested, instructional decisions. *Journal of Second Language Writing,* 20, 56–68.

Johns, A. M. (2015). Moving on from Genre Analysis An update and tasks for the transitional student. *Journal of English for Academic Purposes 19*, 113-124.

Johns, A. M., Bawarshi, A., Coe, R. M., Hyland, K., Paltridge, B., Reiff, M.J. &

Johns, T. (1991a). Should you be persuaded: Two examples of data-driven learning. In T. Johns and P. King (eds.), Classroom Concordancing. *English Language Research Journal*, 4: 1–16.

Johns, T. (1991b). From printout to handout: Grammar and vocabulary teaching in the context of data-driven learning. *CALL, Austria,* 10, 14-34.

Jung, Y., Crossley S. A. & McNamara. D. S. (2015). Linguistic Features in MELAB writing performances. *CaMLA Working Papers,* No. 2015-05, 1-17. Retrieved from Cambridge Michigan Language Assessments website: http://www.cambridgemichigan.org/about-us/research/working-papers/

Kanaris, I., & Stamatatos, E. (2007). Webpage Genre Identification Using Variable-Length Character n-Grams. In proceedings of the 19th IEEE International Conference on Tools with Artificial Intelligence (ICTAI 2007).

Knapp, P. (1989). Teaching Factual Writing: The Report Genre. Sydney: Metropolitan East Disadvantaged Schools' Program, NSW Department of School Education.

Knapp, P. (2002). Disembodied Voices: The Problem of Context and Form in Theories of Genre. In C. Richard, L. Lingard & T. Teslenko (Eds.), *The Rhetoric and Ideology of Genre: Strategies for Stability and Change* (pp. 275-296). Cresskill, NJ: Hampton Press.

Knapp, P., & Watkins, M. (1994). *Context-Text-Grammar: Teaching the Genres and Grammar of School Writing in Infants and Primary Classrooms*. Sydney: Text productions.

Kopaczyk, J. (2012). Long lexical undles and standardization in historical legal texts. *Studia Anglica Posnaniensia*, 47, 2-3.

Koprowski, M. (2005). Investigating the usefulness of lexical phrases in contemporary coursebooks. *ELT Journal*, 59/4, 322-332.

Kress, G. and Knapp, P. (1992). Genre in a social theory of language. *English in Education,* 26, 4–15.

Kroll, B. (1990). The Rhetoric/Syntax split: Designing a curriculum for ESL students. *Journal of Basic Writing,* 9(1), 40-55.

Larsen-Freeman, D. (2001). Grammar. In D. Nunan & R. Carter (Eds.) *The Cambridge Guide to teaching English to Speakers of Other Languages.* Cambridge: Cambridge University Press.

Lee, D. (2001). Genres, Registers, Texts types, Domains, and Styles: Clarifying the Concepts and Navigating a Path through the BNC Jungle. *Language Learning & Technology,* 5(3)*, 37-72

Lee, D. (2010). What corpora are available? (pp.107- 121). In M. McCarthy & A. O'Keeffe (Eds.), *The Routledge Handbook of Corpus Linguistics* (pp.107-121). London: Routledge.

Leech, G. (1992). Corpora and theories of linguistic performance. In J. Svartvik (Ed.), *Directions in Corpus Linguistics* (pp. 105-122). Berlin: Mouton de Gruyter.

Leech, G. (1998). Preface. In S. Granger (Ed.), *Learner English on computer* (pp. xiv–xxii). London: Longman.

Lewis, M. (1993). *The lexical approach*. Hove, England: LTP Publications.

Lim, S. G. (2012). Developing and validating a mark scheme for Writing Cambridge ESOL Research Notes, 49, 6-10.

Lines, H. E. (2014). *It's a matter of individual taste, I guess: Secondary school English teachers' and students' conceptualisations of quality in writing.* PhD thesis, University of Exeter. Available online at https://ore.exeter.ac.uk/repository/handle/10871/14806?show=full

Littlejohn, A. (1992). *Why are ELT materials the way they are?* Phd thesis, Lancaster University. Available online at: http://www.andrewlittlejohn.net/website/books/phd.html

Malinowski, B. (1923). The problem of meaning in primitive languages. Supplement 1 to C. K. Ogden and I. A. Richards (Eds.), *The meaning of meaning: a study of the influence of language upon thought and of the science of symbolism* (8th edition, 1946) (296-336). NY: Harcourt Brace and World.

Martin, J. R. (1985). *Factual writing: exploring and challenging social reality.* Oxford University Press (republished 1989).

Martin, J. R. (1991). Nominalization in science and humanities: Distilling knowledge and scaffolding text. In E. Ventola (Ed.), *Functional and systemic linguistics* (pp. 307–337). Berlin: Mouton de Gruyter.

Martin, J. R. (1992). *English text: system and structure.* Amsterdam: John Benjamins.

Martin, J. R. (1993). Genre and Literacy: Modeling Context in Educational Linguistics. *Annual Review of Applied Linguistics,* 13, 141-172.

Martin, J.R. (2009). Genre and language learning: A social semiotic perspective, *Linguistics and Education*, 20, 10–21.

Martinez, R. & Schmitt, N. (2012). A phrasal expressions list. *Applied Linguistics, 33*(3), 299-320.

McCarthy, M. (2004). *Touchstone: From corpus to course book.* Cambridge: Cambridge University Press. Available at: http://www.cambridge.org/gb/cambridgeenglish/catalog/adultcourses/touchstone/resources/

McCarthy, M., McCarten, J. & Sandiford, H. (2005). *Touchstone 1-4: From corpus to course book.* Cambridge: Cambridge University Press.

McCarthy, P. M., Myers, J. C., Briner, S. W., Graesser, A. C., & McNamara, D. S. (2009). A psychological and computational study of genre recognition. *Journal for Language Technology & Computa tional Linguistics*, 24, 23-55.

McCarthy, M. and O'Keeffe, A. (2010). What are corpora and how have they evolved? In A. O'Keeffe & M. McCarthy (Eds.), *The Routledge Handbook of Corpus Linguistics* (pp. 3- 13). London: Routledge.

McEnery, T. & Wilson, A. (2001) *Corpus linguistics: An Introduction* (2nd edition)*.* Edinburgh: Edinburgh University Press.

McEnery, A. M. & Xiao, R. Z. (2007). Parallel and comparable corpora: What are they up to? In G. Anderman & M. Rogers (Eds.) *Incorporating Corpora: Translation and the Linguist* (pp. 18-31). Clevedon, UK: Multilingual Matters.

McNamara, T. & Roever, C. (2006). *Language Testing: The social dimension.* London: Blackwell.

Meara, P. & H. Bell. (2001). P_Lex: A simple and effective way of describing the lexical characteristics of short L2 texts. *Prospect,* 16(3), 5–19.

Meunier, F. & Gouverneur, C. (2009). New types of corpora for new educational challenges: collecting, annotating and exploiting a corpus of textbook material. In K. Aijmer (Ed.), *Corpora and Language Teaching* (pp.179-201). Amsterdam and Philadelphia: John Benjamins.

Meyer C. (2002). English Corpus Linguistics: An Introduction. Cambridge: Cambridge University Press.

Michos, S. E., Stamatatos, E., Fakotakis, N. & Kokkinakis G. (1996). Categorising Texts by Using a Three-Level Functional Style Description. In A. Ramsay (Ed.), *Artificial Intelligence: Methodology, Systems, Applications, Frontiers in Artificial Intelligence and Applications, 35 (pp.191-198).* IOS Press.

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

Moon, R. (1997). Vocabulary connections: Multi-word items in English. In N. Schmitt & M. McCarthy (Eds.), *Vocabulary description, acquisition and pedagogy* (pp.40-63). Cambridge: Cambridge University Press.

Mukherjee, J. (2004). Bridging the gap between applied corpus linguistics and the reality of English language teaching in Germany. In U. Connor & T. A. Upton (Eds.), *Applied corpus linguistics: A multidimensional perspective* (pp.239-250). Amsterdam: Rodopi.

Nagy, W. Townsend, D. (2012). Words as Tools: Learning Academic Vocabulary as Language Acquisition. *Reading Research Quarterly,* 47(1), 91–108.

Nation, I. S. P. (2006). How large a vocabulary is needed for reading and listening? *Canadian Modern Language Review*, *63*, 59–82.

Nesi, H. & Gardner, S. (2006). Variation in disciplinary culture: University tutor's views on assessed writing tasks. In R. Kiely, G. Clibbon, P. Rea-Dickins, & H. Woodfield (Eds.), *Language, culture and identity in applied linguistics* (British Studies in Applied Linguistics, vol.21) (pp. 99-117). London: Equinox Publishing.

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

O'Keefe, A. McCarthy, M. J. & Carter, R. A. (2007). *From Corpus to Classroom: Language Use and Language Teaching.* Cambridge: Cambridge University Press

Olson, R.D. (2012). Narrative, cognition, and rationality. In J. P. Gee & M. Handford (Eds.), *The Routledge Handbook of Discourse Analysis* (pp. 604-615) London: Routledge.

Ooi, V. (2001). Investigating and Teaching genres using the World Wide Web. In M. Ghadessy, A. Henry & R. Roseberry (Eds.), Small Corpus Studies and ELT (175-203). Amsterdam: John Benjamins.

Osborne, J. (2004). Top-down and Bottom-up Approaches to Corpora in Language Teaching. In U. Connor and T. A. Upton (Eds.), Applied Corpus Linguistics: A Multidimensional Perspective (pp. 251-265). London: Rodopi.

Paltridge, B. (1996) Genre, text type, and the language learning classroom. In *ELT Journal,* 50(3), 237-243.

Paltridge, B. (1997). *Genre, Frames and Writing in Research Settings.* Amsterdam/ Philadelphia: John Benjamins.

Paltridge, B. (2001). *Genre and the language learning classroom.* Ann Arbor, MI: University of Michigan Press.

Papaefthymiou-Lytra, S. (2012). Foreign Language Testing and Assessment in Greece: An Overview and Appraisal. *Research Papers in Language Teaching and Learning,* 3(1), (pp. 22-32). Available online at http://rpltl.eap.gr

Passoneau, R. J., Ide, N., Su, S. & Stuart, J. (2014). Biber Redux: Reconsidering Dimensions of Variation in American English. *Proccedings of the 25th International Conference on Computational Linguistics (COLING 2014)* (pp. 565-576). Dublin.

Poole, R. (2015). A corpus-aided approach for the teaching and learning of rhetoric in an undergraduate composition course for L2 Writers. In F. Formato & A. Hardie (Eds), *Corpus Linguistics 2015 Conference abstract book* (pp. 275-277). Lancaster: UCREL.

Pravec, N. A. (2002). Survey of learner corpora, *ICAME Journal, 26,* 81-114.

Raimes, A. (1994). Language proficiency, Writing ability, and Composing strategies: A study of ESL College Student Writers. In A. H. Cumming, (Ed.), *Bilingual Performance in Reading and Writing* (139-172). Ann Arbor MI: Research club in language learning, distributed by John Benjamins.

Reid, J. (2008). Myth(s) 9: Students' Myths about Academic Writing and Teaching. In J. Reid (Ed.), *Writing Myths: Applying Second Language Research to Classroom Teaching* (177-201). Ann Arbor, MI: University of Michigan Press.

Reppen, R. (2002). A Genre-Based Approach to Content Writing Instruction. In J. Richards & W. Renandya (Eds.), *Methodology in Language Teaching: An Anthology of Current Practice* (pp. 321-327). Cambridge: Cambridge University Press.

Rizzo, C. R. (2010). Getting on with corpus compilation: from theory to practice. *ESP World,* 9 (1), 1-23.

Römer, U. (2004). Textbooks: A corpus-driven approach to modal auxiliaries and their didactics. In J. Sinclair (Ed.), *How to Use Corpora in Language Teaching* (pp. 185-199). Amsterdam/Philadelphia: John Benjamins.

Römer, U. (2006). Looking at *Looking*: Functions and Contexts of Progressives in Spoken English and 'School' English. In A. Renouf & A. Kehoe (Eds.), The Changing Face of Corpus Linguistics (pp. 231-242). Papers from the 24th International Conference on English Language Research on Computerized Corpora (ICAME 24). Amsterdam: Rodopi.

Rothery, J. (1985). Two Varieties of Writing: Report and Exposition. In J.R. Martin 1985, Factual writing: Exploring and Challenging Social Reality (pp. 71-82). Deakin: Deakin University Press.

Sampson, G. (2013). The empirical trend: Ten years on. In A. Hardie & R. Love *Corpus linguistics conference 2013 abstract book* (pp. 256-258). Lancaster: UCREL.

Santini M. (2006). Some issues in Automatic Genre Classification of Web Pages, *Proceedings of JADT 2006 Conference*, Besançon.

Sardinha, B. & Pinto, V. (2014). Introduction. In B. Sardinha & V. Pinto (Eds.), *Multi-Dimensional Analysis, 25 years on: A tribute to Douglas Biber* (pp. xv-xxv), Amsterdam/Philadelphia: John Benjamins.

Scardamalia, M., & Bereiter, C. (1986). Research on written composition. In M. C. Wittrock (Ed.), *Handbook of Research on Teaching* (3rd ed.). New York/London: Macmillan Publishing Company.

Schmid, H. (1994). Probabilistic Part-of-Speech Tagging Using Decision Trees. In *Proceedings of International Conference on New Methods in Language Processing* (pp. 44-49), Manchester: UMIST

Scholfield, P. J. & Gitsaki, C. (1996). What is the advantage of private instruction? The example of English vocabulary learning in Greece. *System*, 24(1), 117-127.

Scott, M. (1997). PC analysis of key words and key key words. *System, 25*(1), 1-13.

Scott, M., (2012). *WordSmith Tools Help*. Stroud: Lexical Analysis Software.

Scott, M., (2015). *WordSmith Tools version 6*, Stroud: Lexical Analysis Software. Available at http://lexically.net/wordsmith/

Scott, M. & Tribble, C. (2006). *Textual Patterns: Keywords and Corpus Analysis in Language Education*. Amsterdam: John Benjamins.

Sharoff, S. Wu, Z. & Markert, K. (2010). The Web Library of Babel: Evaluating Genre Collections. In *Proceedings of the Seventh Conference on International Language Resources and Evaluation (LREC 2010)* (pp. 3063-3070). Malta.

Sharoff, S. (2015). Approaching genre classification via syndromes. In F. Formato & A. Hardie (Eds), *Corpus Linguistics 2015 conference abstract book,* Lancaster: UCREL

Sifakis, N. C. (2009). Challenges in teaching ELF in the periphery: The Greek context. *ELT Journal*, 63(3), 230-237.

Sifakis, N. C. & Fay, R. (2011). Integrating an EFL Pedagogy in a changing world: The case of Greek state schooling. In A. Archibald, A. Cogo & J. Jenkins (Eds.), *Latest Trends*

in ELF Research (pp. 285-297). Newcastle upon Tyne: Cambridge Scholars Publishing.

Sifianou M. (2006). *Discourse Analysis: An introduction*. Athens: Hillside Press.

Silva, T. (1993). Toward an understanding of the distinct nature of L2 writing: The ESL research and its implications. *TESOL Quarterly,* 27, 665-77.

Simpson-Vlach, R. & Ellis, N. (2010). An academic formulas list: New methods in

phraseology research. *Applied Linguistics, 31*(4), 487-512.

Sinclair, J. M. (R. Carter, Ed.) (2004a). *Trust the text: Language, Corpus and Discourse.* London: Routledge.

Sinclair, J. McH. (2004b). New evidence, new priorities, new attitudes. In J. McH. Sinclair (Ed.), *How to Use Corpora in Language Teaching* (pp. 271-299). Amsterdam: John Benjamins.

Sinclair, J. (2005). Corpus and Text: Basic Principles. In M. Wynne (Ed.), *Developing Linguistic Corpora: A Guide to Good Practice* (pp. 1-16). Oxford: Oxbow Books.

Smedegaard, A. (2015). Genre and writing pedagogy. In S. Auken, P. S. Lauridsen, & A. J. Rasmussen (Eds), Copenhagen Studies in Genre 2, *Genre and ....* Copenhagen: Forlaget Ekbátana.

So, B. P. C. (2005). From Analysis to Pedagogic Applications: Using Newspaper Genres to Write School Genres. *Journal of English for Academic Purposes* 4(1), 67-82.

Spillett, H. (2012). The revision of the Cambridge English: Proficiency Writing paper. *Cambridge ESOL Research Notes,* 49, 2-5. Available online at http://www.cambridgeenglish.org/images/23166-research-notes-49.pdf

Spiris, S. (2014). Investigating normalisation: Do teachers of English in Greece

integrate technology in their everyday teaching practice? *Research Papers in Language Teaching and Learning,* 5(1), 351-373. Available online at http://rpltl.eap.gr

Štajner, S., Mitkov, R. & Pastor, G. C. (2015). Simple or not simple? A readability question. In N. Gala, R. Rapp, and G. Bel-Enguix (Eds.), *Language Production, Cognition and the Lexicon* (pp. 379-398). Switzerland: Springer.

Stamatatos, E., Fakotakis, N. & Kokkinakis, G. (2000). Text genre detection using common word frequencies. In the *Proceedings of the 18th Conference on Computational Linguistics - Volume 2, COLING '00,* (pp. 808–814), Stroudsburg, PA: Association for Computational Linguistics.

Stamatatos, E., Kokkinakis, G. & Fakotakis, N. (2001). Automatic text categorization in terms of genre and author. *Computational Linguistics*, 26(4), 471–495.

Stubbs, M. (1996). *Text and corpus analysis*. Oxford: Blackwell.

Swales, J. M. (1990). *Genre Analysis: English in Academic and Research Settings.* Cambridge: Cambridge University Press.

Swales, J. M. (2004). *Research genres.* New York: Cambridge University Press.

Tardy, C. (2006) Researching first and second language genre learning: A comparative review and a look ahead. *Journal of Second Language Writing*, 15, 79-101.

Tardy, C. (2009). *Building genre knowledge*. West Lafayette, Indiana: Parlor Press.

Tardy, C. (2015). *Innovation and Creativity in the L2 Academic Writing Classroom: Toward Critical Genre Awareness.* Paper presented at the ROCTEFL 32nd International Conference on English Teaching & Learning, Taipei.

Thompson, G. (2014). *Introducing Functional Grammar* (3rd edition). London and New York: Routledge.

Tognini-Bonelli, E. (2010). Theoretical overview of the evolution of corpus linguistics. In A. O'Keeffe & M. McCarthy (Eds.), *The Routledge Handbook of Corpus Linguistics* (pp. 14-27). London: Routledge.

Tono, Y. (2013). 'Criterial feature' extraction from CEFR-based corpora: Methods and techniques. In A. Hardie and R. Love (Eds) *Corpus Linguistics 2013 conference abstract book* (pp. 280-282). Lancaster: UCREL.

Torrance, M. (1996a). Is writing expertise like other kinds of expertise? In G. Rijlaarsdam, H. van den Bergh & M. Couzijn (Eds.), *Theories, Models and Methodology in Writing Research* (pp. 3-9). Amsterdam: Amsterdam University Press.

Torrance, M. (1996b). Strategies for familiar writing tasks: Case studies of undergraduates writing essays. In G. Rijlaarsdam, H. van den Bergh & M. Couzijn, (Eds.), *Theories, Models and Methodology in Writing Research* (pp. 283-298). Amsterdam: Amsterdam University Press.

Tribble, C. (2001). Small corpora and teaching writing: Towards a corpus-informed pedagogy of writing. In M. Ghadessy, A. Henry & R. Roseberry (Eds.), *Small Corpus Studies and ELT: Theory and practice* (pp.381-408). Amsterdam/Philadelphia: John Benjamins.

Tribble, C. (2009). Writing academic English: A survey review of current published resources. *ELT Journal*, 63 (4), 400-417.

Tsagari, D., & Papageorgiou, S. (2012). Special Issue on Language Testing and Assessment in the Greek Educational Context. *Research Papers in Language Teaching and Learning*, 3(1), 4-7. Available online at http://rpltl.eap.gr

Tsui A. (2004). What teachers have always wanted to know – and how corpora can help. In J. McH. Sinclair (Ed.), *How to use Corpora in Language Teaching* (pp. 39-61). Amsterdam/Philadelphia: John Benjamins.

Tweedie, F. & Baayen, H. (1998). How variable may a constant be? Measures of lexical richness in perspective. *Computers and the Humanities,* 32 (5), 323-352.

Ure, J. (1971). Lexical density and register differentiation. In G. E. Perren & J. L. M. Trimm (Eds.). *Applications of Linguistics: Selected papers of the 2nd International Congress of Applied Linguists*, (pp. 443-452). Cambridge: Cambridge University Press.

Ventola, E. (1996). Packing and unpacking of information in academic texts. In E. Ventola & A. Mauranen (Eds.), Academic writing: Intercultural and textual issues (pp. 153-194). Amsterdam/Philadelphia: John Benjamins.

Viana, V., Giordani, A. & Zyngier, S. (2008). Empirical evaluation: Towards an automated index of lexical variety. In S. Zyngier, M. Bortolussi, A. Chesnokova & J. Auracher (Eds.), *Directions in empirical literary studies: In honor of Willie van Peer* (pp. 271-282). Amsterdam/Philadelphia: John Benjamins.

Vygotsky, L. S. (1978). *Mind in Society: The Development of Higher Psychological Processes.* (M. Cole, V. John-Steiner, S. Scribner & E. Souberman, Eds.) (A. R. Luria, M. Lopez-Morillas & M. Cole [with J. V. Wertsch], Trans.). Cambridge, MA: Harvard University Press. (Original manuscripts [ca. 1930-1934])

Vygotsky, L. S. (1987). Thinking and speech (N. Minick, Trans.). In R. W. Rieber & A. S. Carton (Eds.), *The collected works of L. S. Vygotsky. Vol. 1. Problems of general psychology* (pp.39-285). New York: Plenum Press.

Wollman-Bonilla, J. (2003). E-mail as Genre: A Beginning Writer Learns the Conventions. *Language Arts,* 81 (2), 126-134.

Wood, D., Bruner, J. S., & Ross, G. (1976). The role of tutoring in problem solving. *Journal of Child Psychology and Psychiatry*, 17(2), 89-100.

Wray, A. & Perkins, M. (2000). The functions of formulaic language: An integrated model. *Language and Communication, 20*(1), 1-28.

Wright, D. (2013). Stylistic variation within genre conventions in the Enron email corpus: developing a text sensitive methodology for authorship research. International Journal of Speech Language and the Law (IJSLL), 20(1), 45-75.

Yasuda, S. (2011). Genre-based tasks in foreign language writing: Developing writers' genre awareness, linguistic knowledge, and writing competence. *Journal of Second Language Writing,* 20, 111–133.

Yates, S. (1996). Oral and Written Linguistic Aspects of computer conferencing. In S. C. Herring (Ed.), *Computer-Mediated Communication: Linguistic, social, and cross-cultural perspectives* (pp. 29-46). Amsterdam/Philadelphia: John Benjamins.

References (in Greek)

Γεωργακοπουλου Α. & Γουτσος Δ. (2011). *Κείμενο και επικοινωνία.* Εκδόσεις Πατάκη.

Δενδρινού, Β., Καραβά, Ε. (2013) *Ξενόγλωσση Εκπαίδευση για την προώθηση της πολυγλωσσίας στην Ελλάδα σήμερα: Προσεγγίσεις και πρακτικές διδασκαλίας.* Αθήνα: ΕΚΠΑ, ΥΠΑΙΘ.

Εθνικό και Καποδιστριακό Πανεπιστήμιο Αθηνών (2007-2013). *Οικονομικά Στοιχεία για τις ξένες γλώσσες στην Ελλάδα. Η εκμάθηση της Αγγλικής σε πρώιμη παιδική ηλικία. Νέες πολιτικές ξενόγλωσσης εκπαίδευσης στο σχολείο.* Διαθέσιμο:

http://elp.enl.uoa.gr/fileadmin/elp.enl.uoa.gr/uploads/Oikonomika_stoicheia_gia_tis_X.G..pdf

Ιακώβου, Μ., Μαρκόπουλος, Γ., & Μικρός, Γ. Κ. (2003). Καθορισμός βασικού λεξιλογίου μέσω Ηλεκτρονικών Σωμάτων Κειμένων για τη διδασκαλία της Νέας Ελληνικής ως Ξένης Γλώσσας *Πρακτικά του 2ου Διεθνούς Συνεδρίου για τη διδασκαλία της Ελληνικής ως Ξένης Γλώσσας, 19-21 Σεπτεμβρίου 2002, Αθήνα.* Αθήνα.

Ζαμπετάκης, Λ. (2013). Η κανονική κατανομή. Έλεγχος κανονικότητας. Μη παραμετρικά τεστ. Διάλεξη 5. Ρέθυμνο: Πανεπιστήμιο Κρήτης. Διαθέσιμο: ftp://ftp.soc.uoc.gr/Psycho/Zampetakis/%D3%F4%E1%F4%E9%F3%F4%E9%EA%DE%20%C9/%C4%E9%E1%EB%DD%EE%E5%E9%F2/(5)_lecture5_27-03-2013.pdf

Κονδύλη, Μ. & Λύκου, Χ. (2009). Γλωσσολογική περιγραφή των θεμάτων του ΚΠΓ: Η οπτική του κειμενικού είδους και της λεξικογραμματικής. *RCeL. Διαθέσιμο:* http://rcel.enl.uoa.gr/periodical/research1.htm

Μικρός, Γ. Κ. (2003). Στατιστικές προσεγγίσεις στην αυτόματη κατηγοριοποίηση κειμένων της Νέας Ελληνικής: Μια πιλοτική αξιολόγηση υφομετρικών δεικτών και στατιστικών μεθόδωνΠρακτικά του 6ου Διεθνούς Συνεδρίου Ελληνικής Γλωσσολογίας, 18-21 Σεπτεμβρίου 2003, Ρέθυμνο. Retrieved from http://www.philology.uoc.gr/conferences/6thICGL/ebook/a/mikros.pdf.

Παιδαγωγικό Ινστιτούτο (2004-2011). Διδακτικά πακέτα γυμνασίου. Διαθέσιμο: http://www.pi-schools.gr/books/gymnasio/

Φραντζή Κ. (2012). Εισαγωγή στην επεξεργασία σωμάτων κειμένων. Αθήνα: Εκδοτικός όμιλος ΙΩΝ.

Appendix One: Questionnaire (in Greek)

Ερωτηματολόγιο

---

1. Στις εξετάσεις γλωσσομάθειας ή κατά την προτεοιμασία σας γι αυτές και ειδικότερα για την αγγλική γλώσσα, **ποιο μέρος σας δυσκολεύει περισσότερο**;
Βαθμολογήστε **με 1-4** ξεκινώντας με το πιο δύσκολο (1).
α. Κατανόση γραπτού λόγου (Reading Comprehension) ____
β. Κατανόηση προφορικού λόγου (Listening Comprehension) ____
γ. Παραγωγή γραπτού λόγου (Writing) _____
δ. Παραγωγή προφορικού λόγου (Speaking) _____

---

2. Στα μαθήματα που έχετε κάνει για την αγγλική γλώσσα στο παρελθόν (οπουδήποτε και συνολικά) **πόσος χρόνος** αφιερώθηκε **στην προετοιμασία της έκθεσης**; Κυκλώστε μία απάντηση:
α. το 1/3 του συνολικού χρόνου περίπου
β. το 1/4 »   »   »   »
γ. το 1/5 »   »   »   »
δ. το 1/6 με 1/10 »   »   »
ε. σχεδόν καθόλου

---

3. Από τα παρακάτω είδη κειμένων που ζητούνται συνήθως στις εξετάσεις των αγγλικών **ποιο σας δυσκολεύει περισσότερο**; Βαθμολογήστε **με 1-6** ξεκινώντας από το πιο δύσκολο (1). (Πρέπει να χρησιμοποιήσετε όλους τους αριθμούς)
α. Γράμμα επίσημο   ____
β. Γράμμα ανεπίσημο   ____
γ. Έκθεση-δοκίμιο (essay)   ____
δ. Μικρή ιστορία   ____
ε. Αναφορά (Report)   ____
ζ. Κριτική (βιβλίου, ταινίας)   _____

---

4. Τώρα σημειώστε **πόσο έχετε προετοιμαστεί για αυτά τα είδη, με τα γράμματα α-δ**
όπου : α. ικανοποιητικά β. μέτρια γ. ελάχιστα δ. καθόλου
α. Γράμμα   επίσημο   ____
β. Γράμμα ανεπίσημο   ____
γ. Έκθεση-δοκίμιο (essay)   ____
δ. Μικρή ιστορία   ____
ε. Αναφορά (Report)   ____
ζ. Κριτική (βιβλίου, ταινίας)   ____

---

5. Κατά την εξέταση της **έκθεσης** στα αγγλικά **με δυσκολεύει περισσότερο**: (Κυκλώστε **μόνο ένα**)
α. Το λεξιλόγιο και οι κατάλληλες εκφράσεις
β. Η γραμματική/ σύνταξη
γ. το περιεχόμενο (ιδέες)
δ. το να τηρήσω το όριο των λέξεων
ε. η κατανόηση του ερωτήματος
ζ. τίποτα από τα παραπάνω

6. Πόσο **σας αγχώνει το χρονικό όριο** που μπαίνει όταν γράφετε έκθεση στα αγγλικά; (εξετάσεις και/ ή στην ταξη) Κυκλώστε **μόνο μία απάντηση**:
α. καθόλου   δ. πολύ (*δρα αρνητικά στην ποιότητα του γραπτού μου*)
β. ελάχιστα   ε. πάρα πολύ (*με επηρρεάζει τόσο που δεν προλαβαίνω να τελειώσω*)
γ. αρκετά

7. Κατά τη **διδασκαλία** της έκθεσης **θα ήθελα**: (**κυκλώστε όλα όσα ισχύουν για σας**):
α. περισσότερο χρόνο γενικά
β. να γνωρίζω από πριν τα κριτήρια αξιολόγησης
γ. να γράφω πιο συχνά για εξάσκηση
δ. να μου δίνονται μοντέλα απαντήσεων
ε. πιο αναλυτικά σχόλια κατά την διόρθωση του γραπτού μου
ζ. να προηγείται συζήτηση του ερωτήματος που θα μου ζητηθεί να γράψω
η. Άλλο: _____
   _____
   _____

Παρακαλώ συμπληρώστε τα παρακάτω στοιχεία:

Ηλικιακή ομάδα :   12-17___,   18-24 ___,   25- 30 ____,   31 και πάνω ____
Φύλο: Α___ Θ ___
Επίπεδο γνώσεων στην αγγλική (δική σας εκτίμηση):
< Β1 ____
 Β1 (pre-lower) ___
 Β2 (lower) _____
 C1 (advanced) ____
 C2 (proficiency) _____
(*Οι όροι στις παρενθέσεις, αν και όχι όλοι κατάλληλοι, έχουν καθιερωθεί στο ελληνικό περιβάλλον και έχουν μπει για να σας βοηθήσουν να κατανοήσετε τα επίπεδα.*)
Σε ποιο **επίπεδο** από τα παραπάνω έχετε κάποια **πιστοποίηση**; ____ Σε κανένα ___.

Το παρόν ερωτηματολόγιο αποτελεί μέρος έρευνας που διεξάγεται στο Παν/μιο Αιγαίου και θα χρησιμοποιηθεί ανώνυμα.
                    Σας ευχαριστούμε πολύ για το χρόνο σας!

Appendix Two: Sample metadata sheet

A	B	C	D	E
ModEs	241	TbS10	B1	You have been talking in class about school sports lessons. The students had very difficult opinions. Now your teacher has asked you to write an...
	242	TbS11	B1	Mobile phones in cinemas.
	243	TbS11	B1	Write a composition expressing your opinion on students bringing their mobile phones to school. Give reasons and examples to support your ans...
	244	TbS6	B1	You have just had a discussion in class about homelessness. Now your teacher has asked you to write an essay suggesting ways to solve the prob...
	245	TbS6	B1	You had a class discussion about doing shopping on the Internet. Your teacher has asked you to write an essay discussing the pros and cons of sh...
	246	TbS6	B1	Your teacher has asked you to write an essay suggesting ways to reduce the problem of air pollution in big cities. Write your essay presenting you...
	247	TbS6	B1	Your teacher has asked you to write an essay, answering this question: "What are the advantages and disadvantages of mobile phones?"
	248	TbS12	B2	You have had a class discussion about shopping. Now your teacher has asked you to write an essay on the following question: What are the adva...
	249	TbS12	B2	More and more people nowadays prefer watching DVDs at home to going to the cinema. What are the pros and cons of watching DVDs at home?
	250	TbS12	B2	You have had a class discussion on how travel affects people's outlook. Now your teacher has asked you to write an essay on whether you agree
	251	TbS12	B2	You have had a class discussion on how television promotes violence. Now your teacher has asked you to write an essay on whether you agree
	252	TbS12	B2	Your teacher has asked you to write an essay based on the following question: Our town suffers from the negative effects of tourism. What are
	253	TbS12	B2	Your parents are considering leaving the town where you live and moving to the countryside. You are not sure whether you are happy with this i
	254	TbS12	B2	Your teacher has asked you to write an essay about the importance of tourism. Write an essay on this subject.
	255	TbS12	B2	Cars should be banned in city centres. Do you agree? Write an essay for your teacher.
	256	TbS12	B2	Your teacher has asked you to write an essay based on the question: What can we do to protect animals from becoming extinct? Write an essay
	257	TbS12	B2	You have had a class discussion about living abroad. Your teacher has asked you to write an essay, giving your views on the following statement:
	258	TbS12	B2	As the popularity of the internet grows, more young people are using internet chatrooms. What are the advantages and disadvantages of using i
	259	TbS5	B2	The Ministry of Education is considering removing the study of music from the national curriculum for high schools. The Ministry claims that the
	260	TbS7	ECCE	(prompt) Having a pet is a serious responsibility. What suggestions would you give to people who are considering getting a pet? Be specific.
	261	TbS5	B2	Your class has had a discussion about learning foreign languages. Your English teacher has now asked you to write a composition, saying whether
	262	TbS7	ECCE	Some people believe space is valuable in cities and that green areas should be used to develop business and commerce. Other people feel that th
	263	TbS11	B1	Write an essay in response to an article called: Teachers ban dictionaries in English exams.
	264	TbS5	B2	Should children learn about managing money at school?
	265	TbS5	B2	Is hiring a car the best way to travel in a foreign country?
	266	TbS13	B1	Some people believe that the weather can change our mood, while others think that the way we feel has nothing to do with the weather. Write a
	267	TbS13	B1	Write an essay sying what you think can be done to improve the health service in your country.

Appendix Three: Sample POS tagged text

^_SYM NONE_NONE *_SYM In_IN recent_JJ times_NNS our_PP$ lifestyle_NN has_VHZ changed_VVN dramatically_RB in_IN many_JJ ways_NNS ._SENT In_IN the_DT past_JJ people_NNS could_MD drive_VV freely_RB around_IN the_DT city_NN without_IN any_DT problem_NN ,_, but_CC nowadays_RB traffic_NN in_IN city_NN centres_NNS has_VHZ become_VVN a_DT serious_JJ headache_NN ._SENT The_DT authorities_NNS now_RB feel_VVP obliged_VVN to_TO take_VV urgent_JJ measures_NNS to_TO solve_VV this_DT issue_NN ._SENT *_SYM Most_RBS modern_JJ cities_NNS and_CC large_JJ towns_NNS have_VHP turned_VVN their_PP$ centre_NN into_IN pedestrian_JJ zones_NNS due_JJ to_TO the_DT increase_NN in_IN population_NN and_CC a_DT greater_JJR number_NN of_IN car_NN users_NNS ._SENT Parking_NN is_VBZ almost_RB impossible_JJ and_CC space_NN is_VBZ running_VVG out_RP ,_, not_RB to_TO mention_VV the_DT health_NN risk_NN from_IN pollution_NN and_CC the_DT question_NN of_IN road_NN safety_NN ._SENT *_SYM Although_IN banning_VVG cars_NNS from_IN the_DT city_NN centre_NN seems_VVZ to_TO be_VB an_DT inevitable_JJ consequence_NN of_IN all_PDT this_DT ,_, we_PP should_MD not_RB forget_VV about_IN those_DT people_NNS who_WP must_MD be_VB able_JJ to_TO access_VV these_DT areas_NNS without_IN restrictions_NNS ._SENT Pensioners_NNS ,_, disabled_JJ people_NNS and_CC local_JJ residents_NNS are_VBP an_DT exceptional_JJ case_NN ,_, and_CC lorry_NN and_CC van_NN drivers_NNS should_MD be_VB allowed_VVN to_TO deliver_VV to_TO shops_NNS or_CC offices_NNS ._SENT *_SYM In_IN conclusion_NN ,_, the_DT prohibition_NN of_IN cars_NNS from_IN city_NN centres_NNS is_VBZ probably_RB the_DT only_JJ practical_JJ solution_NN owing_VVG to_TO lack_NN of_IN space_NN ._SENT These_DT measures_NNS should_MD not_RB ,_, however_RB ,_, affect_VV those_DT who_WP live_VVP or_CC work_VVP there_RB ._SENT ^^_NN

Appendix Four: Tree Tagger Set

Tree Tagger Tag Set (58 tags)		
POS Tag	**Description**	**Example**
CC **CD**	coordinating conjunction cardinal number	*and, but, or, &* *1, three*
DT **EX**	determiner existential there	*the* *there is*
FW **IN**	foreign word preposition/ subordinating conjunction	*d'œuvre* *in, of, like, after, whether*
IN/that **JJ**	complementizer adjective	*that* *green*
JJR **JJS**	adjective, comparative adjective, superlative	*greener* *greenest*
LS **MD**	list marker modal	*(1),* *could, will*
NN **NNS**	noun, singular or mass noun plural	*table* *tables*
NP **NPS**	proper noun, singular proper noun, plural	*John* *Vikings*
PDT **POS**	predeterminer possessive ending	*both the boys* *friend's*
PP **PP$**	personal pronoun possessive pronoun	*I, he, it* *my, his*
RB **RBR**	adverb adverb, comparative	*however, usually, here, not* *better*
RBS **RP**	adverb, superlative particle	*best* *give up*
SENT **SYM**	end punctuation symbol	*?, !, .* *@, +, *, ^, |, =*
TO **UH**	*to* interjection	*to go, to him* *uhhuhhuhh*
VB **VBD**	verb *be*, base form verb *be*, past	*be* *was| were*
VBG **VBN**	verb *be*, gerund/participle verb *be*, past participle	*being* *been*
VBZ **VBP**	verb *be*, pres, 3rd p. sing verb *be*, pres non-3rd p.	*is* *am| are*
VD	verb *do*, base form	*do*

VDD	verb *do*, past	*did*
VDG	verb *do* gerund/participle	*doing*
VDN	verb *do*, past participle	*done*
VDZ	verb *do*, pres, 3rd per. sing	*does*
VDP	verb *do*, pres, non-3rd per.	*do*
VH	verb *have*, base form	*have*
VHD	verb *have*, past	*had*
VHG	verb *have*, gerund/participle	*having*
VHN	verb *have*, past participle	*had*
VHZ	verb *have*, pres, 3rd per. sing	*has*
VHP	verb *have*, pres, non-3rd per.	*have*
VV	verb, base form	*take*
VVD	verb, past tense	*took*
VVG	verb, gerund/participle	*taking*
VVN	verb, past participle	*taken*
VVP	verb, present, non-3rd p.	*take*
VVZ	verb, present 3d p. sing.	*takes*
WDT	wh-determiner	*which*
WP	wh-pronoun	*who, what*
WP\$	possessive wh-pronoun	*whose*
WRB	wh-abverb	*where, when*
:	general joiner	*;, -, --*
\$	currency symbol	*\$, £*

A screenshot of the "Concord" concordance application window showing a KWIC (key-word-in-context) concordance for the node word "people". The columns are: N, Concordance, Set, Tag, Word#, Sent/Sent/Para/Para/Head/Head/Sect/Sect, File, Date, %.

N	Concordance	Word #	File	Date	%
1	^ * As we move into the twenty- first century an increasing number of people are relying on new forms of technology. A possible consequence of this	16 0 68 0 5%	0 5% ModEs0405:	2015/Aug/16	4%
2	investment than space research because it could ease the suffering of billions of people. Furthermore, the Earth's resources are diminishing, thus time is of the	230 9 10 0 75	0 75 ModEs0323:	2015/Aug/16	73%
3	this century, the role of women within society has changed, and the majority of people feel that this change is for the better. More women work than ever	19 0 69 0 4%	0 4% ModEs0277:	2015/Aug/16	4%
4	about by the Internet has been the way in which people communicate. Millions of people are members of social networking sites such as <NM>, where they can	25 1 15 0 8%	0 8% ModEs0328:	2015/Aug/16	8%
5	as a waste of money. Firstly, if a language is only spoken by a small number of people, expensive education programmes will be needed to make sure that	71 3 35 0 26	0 26 ModEs0388:	2015/Aug/16	25%
6	of people and culture, at least in this way. * Hence, I conclude that mixing of people and their cultures, despite its critics and occasional blemishes, is taking	309 15 33 0 94	0 94 ModEs0442:	2015/Aug/16	94%
7	that exams are the most efficient way of comparing the abilities of a group of people and that an exam will often encourage people to compete to get better	150 6 60 0 64	0 64 ModEs0334:	2015/Aug/16	63%
8	more careful consideration. * It is true as the balance of power among group of people throughout history has shifted, languages have arisen, changed, and	81 4 56 0 28	0 28 ModEs0444:	2015/Aug/16	27%
9	new things are promoted because they have good impacts for the majority of people. A lot of innovations are made with the aim of making money for a few.	181 7 10 0 67	0 67 ModEs0454:	2015/Aug/16	68%
10	of the international community. * In conclusion, I understand the point of view of people who worry about cultural globalisation because it is a threat to national	270 12 48 0 88	0 88 ModEs0401:	2015/Aug/16	86%
11	entertainment. For example, in the nineteenth century only a small proportion of people could go to the ballet or the theatre. However, it is now possible for	217 10 62 0 68	0 68 ModEs0411:	2015/Aug/16	69%
12	^ * In this era of globalization, few countries are left from extensive mixing of people and their cultures. Countries are becoming mere political regions and	17 0 84 0 5%	0 5% ModEs0442:	2015/Aug/16	4%
13	are tempted to drive faster. This causes more traffic accidents and thousands of people die each year from reckless driving and speeding vehicles. * On the	96 5 53 0 54	0 54 ModEs0335:	2015/Aug/16	55%
14	the already cherished ones. It is foolish, therefore, to be critical of mixing of people and culture, at least in this way. * Hence, I conclude that mixing of	294 14 63 0 89	0 89 ModEs0442:	2015/Aug/16	90%
15	^ <MTLE> Winning is more important than taking part * There are two types of people; those who say that winning is everything and those who say it's the	17 0 48 0 6%	0 6% ModEs0271:	2015/Aug/16	6%
16	as cancer and AIDS will have been found. Therefore, the lives of thousands of people will be saved. * Also, new technological breakthroughs will make our	79 4 73 0 32	0 32 ModEs0337:	2015/Aug/16	31%
17	ago. * Enhancement of society, which is done easiest through blending of people, also has to be stated. A country may be conservative and grow	136 7 72 0 41	0 41 ModEs0442:	2015/Aug/16	42%
18	to be more expensive to purchase and to maintain. Indeed, a large majority of people choose to live in apartments because they cannot afford the mortgage to	171 11 32 0 58	0 58 ModEs0413:	2015/Aug/16	58%
19	you ever considered becoming your own boss? In recent years the number of people choosing to start their own business has risen significantly. Many claim	18 1 47 0 7%	0 7% ModEs0329:	2015/Aug/16	6%
20	that one of the main aims of university is to secure a better job. The majority of people want to improve their future career prospects and attending university is	72 3 14 0 24	0 24 ModEs0063:	2015/Aug/16	22%
21	have made life far easier and more convenient for large numbers of people and will continue to do so for decades to come. * Nevertheless, the	129 5 67 0 43	0 43 ModEs0060:	2015/Aug/16	43%
22	, there is a fashion for television programmes to show videos of such jokes or of people having "funny" accidents. If these reach the point where we are laughing	148 8 86 0 66	0 66 ModEs0175:	2015/Aug/16	66%
23	a familiar environment. They might also feel homesick and miss the company of people from their home country. * In conclusion, it seems to me that living	143 9 75 0 74	0 74 ModEs0257:	2015/Aug/16	75%
24	for this development. * On the other hand, city parks benefit a wide variety of people. For example, they offer workers a place to take a break from the stress	102 6 10 0 58	0 58 ModEs0262:	2015/Aug/16	57%
25	illegal to smoke in public places. This would improve the health of thousands of people, and that is most definitely a positive development. ^^	288 16 56 0 97	0 97 ModEs0079:	2015/Aug/16	96%
26	the middle classes. Last but not least, in many countries, there is a shortage of people to do manual jobs such as plumbing and carpentry, so making university	140 5 39 0 42	0 42 ModEs0077:	2015/Aug/16	42%
27	a heavy burden on the car drivers. At present, taxes are already high for a lot of people, and so further taxes would only mean less money at the end of the	191 11 30 0 67	0 67 ModEs0066:	2015/Aug/16	67%
28	produce more handicrafts to meet this demand. * On the other hand, the influx of people to tourist centres also brings problems, such as an increase in pollution.	120 7 45 0 71	0 71 ModEs0254:	2015/Aug/16	70%
29	will definitely be a big help, a factor that might swing the decision for a lot of people. * On the other hand, we must not ignore some difficulties that a subsidy	173 9 10 0 53	0 53 ModEs0095:	2015/Aug/16	53%
30	imposed on them. We have to accept, however, that there are a large number of people who smoke and will continue to smoke. Many of them enjoy the	167 7 67 0 53	0 53 ModEs0097:	2015/Aug/16	54%
31	slightly. While rules and laws have changed, it is the deep-rooted opinions of people within society which are taking a longer time to evolve. Needless to say,	414 15 58 0 92	0 92 ModEs0277:	2015/Aug/16	92%
32	the day wandering from place to place, penniless and hungry? For thousands of people in the western world this is an everyday experience - although, in my	40 1 14 0 9%	0 9% ModEs0278:	2015/Aug/16	8%
33	The advantages and disadvantages of online shopping * These days lots of people shop online. Everything from designer clothes to groceries can be	16 0 88 0 7%	0 7% ModEs0248:	2015/Aug/16	7%
34	fashion industry is undeniably a source of profit and income. It hires millions of people all over the world and generates millions of dollars every year.	78 3 35 0 31	0 31 ModEs0186:	2015/Aug/16	31%
35	industry, as it is today, has a harmful effect, because it values a minority of people in detriment to the majority. However, it has such a wide reach that, it	228 8 83 0 90	0 90 ModEs0186:	2015/Aug/16	91%
36	^ * These days, a large number of people have mobile phones. There are advantages and disadvantages to using	10 0 75 0 6%	0 6% ModEs0247:	2015/Aug/16	5%
37	^ * Nowadays a growing number of people are faced with the question whether to buy a car or a motorbike so as	9 0 23 0 4%	0 4% ModEs0160:	2015/Aug/16	3%
38	for us all. * On the other hand, a foreign country is not for everyone. Some people need the security of a familiar environment. They might also feel	125 8 30 0 65	0 65 ModEs0257:	2015/Aug/16	65%
39	nuclear fusion process, making it a source of energy for their power grids. Some people believe that this is really advantageous, whereas others would go to any	34 1 15 0 12	0 12 ModEs0178:	2015/Aug/16	11%
40	their budget. In this sense, advertising is a very useful and instructive tool. Some people, however, take advertising too seriously and are far too easily convinced	126 7 13 0 48	0 48 ModEs0460:	2015/Aug/16	49%
41	10 years, the Internet has slowly become a part of most people's lives. Some people think that the world will never be the same because of it. * There are	21 1 20 0 12	0 12 ModEs0187:	2015/Aug/16	10%
42	has more advantages or disadvantages depends upon the individual. For some people it may be an unpleasant experience, while for others it may be ideal. As	263 14 24 0 90	0 90 ModEs0332:	2015/Aug/16	90%

447 entries   Row 37   wadays a growing number of people are faced with the questio

Appendix Six: Shared vocabulary across genres (first 400 words)

Core shared vocabulary (8 genres)

	Joint words	Freq. #		Joint words	Freq. #		Joint words	Freq. #		Joint words	Freq. #
1.	the	8,244	20.	have	856	39.	one	431	58.	while	227
2.	to	5,111	21.	they	803	40.	were	428	59.	out	224
3.	and	4,476	22.	we	792	41.	other	425	60.	make	216
4.	of	4,145	23.	not	778	42.	when	406	61.	day	213
5.	a	3,701	24.	at	769	43.	so	393	62.	new	206
6.	in	3,347	25.	there	745	44.	also	389	63.	into	206
7.	is	2,424	26.	can	729	45.	about	369	64.	no	200
8.	that	2,168	27.	their	726	46.	time	369	65.	work	199
9.	I	2,111	28.	from	686	47.	if	369	66.	get	180
10.	it	1,769	29.	more	683	48.	many	349	67.	two	172
11.	for	1,666	30.	or	596	49.	most	318	68.	after	167
12.	are	1,401	31.	by	571	50.	only	317	69.	see	147
13.	be	1,281	32.	all	563	51.	do	315	70.	before	107
14.	on	1,172	33.	but	551	52.	than	314	71.	last	104
15.	this	1,095	34.	an	521	53.	very	297	72.	finally	98
16.	was	1,069	35.	had	482	54.	been	289	73.	did	67
17.	as	1,058	36.	which	445	55.	them	289			
18.	with	923	37.	some	436	56.	up	256			
19.	people	873	38.	will	435	57.	first	228			

Core shared vocabulary (7 genres)

	Joint words	Freq. #	Not in		Joint words	Freq. #	Not in
1.	you	1,001	DR	25.	well	167	CL
2.	<name>	923	DR	26.	place	166	CL
3.	would	548	DR	27.	home	163	AL
4.	my	745	DR	28.	take	155	SSt
5.	your	441	DR	29.	even	149	DR
6.	has	427	SSt	30.	around	149	CL
7.	should	407	DR	31.	food	147	SSt
8.	<location name>	376	DR	32.	different	141	CL
9.	our	359	DR	33.	being	141	AL
10.	who	347	AL	34.	feel	139	DR
11.	me	307	DR	35.	where	134	CL
12.	these	305	SSt	36.	any	133	CL
13.	however	303	AL	37.	made	132	AL
14.	school	287	CL	38.	find	129	DR
15.	what	250	DR	39.	too	125	DR
16.	because	243	DR	40.	city	123	CL
17.	could	242	DR	41.	each	122	AL
18.	over	230	AL	42.	another	119	DR
19.	way	225	DR	43.	now	116	CL
20.	<nationality>	214	DR	44.	little	75	CL
21.	both	212	AL	45.	small	69	AL
22.	money	206	DR	46.	until	69	AL
23.	like	198	DR	47.	quite	58	ExE
24.	how	188	CL				

[Abbreviations used in table: DR (Data Report), SSt (Short Story), AL (Advice Letter), CL (Complain Letter), ExE (Expository Essay)]

Core shared vocabulary (6 genres)

Joint words		Freq #	Joint in 6 genres							
			ExE	DisE	DesE	POR	DR	SSt	CL	AL
1.	he	390	+	+	+	+		+	+	
2.	such	265	+	+	+	+	+	+		
3.	her	227	+	+	+	+		+		+
4.	his	207	+	+	+	+		+	+	
5.	good	204	+	+	+	+		+		+
6.	much	191	+	+	+	+	+			+
7.	just	178	+	+	+		+	+		+
8.	then	167	+	+	+		+	+		+
9.	us	167	+	+	+	+		+	+	
10.	better	157	+	+	+	+		+		+
11.	think	155	+	+	+	+		+		+
12.	year	152		+	+	+	+	+		+
13.	years	152	+	+	+		+	+	+	
14.	help	151	+	+	+	+		+		+
15.	am	145	+	+	+	+			+	+
16.	go	143	+	+	+	+		+		+
17.	local	135	+	+		+	+	+		+
18.	country	135	+	+	+	+	+			+
19.	although	134	+	+	+	+	+		+	
20.	friends	132	+	+	+	+		+		+
21.	know	127	+	+	+	+		+		+
22.	music	126	+	+	+	+		+		+
23.	always	125	+	+	+	+		+		+
24.	best	119	+	+	+	+		+		+
25.	car	117	+	+	+	+	+	+		
26.	same	117	+	+	+	+	+	+		
27.	spend	112	+	+	+	+	+			+
28.	great	107	+	+	+	+		+		+
29.	back	104		+	+	+	+	+		+
30.	used	103	+	+	+	+	+		+	
31.	long	102	+	+	+	+		+		+
32.	must	101	+	+	+	+			+	+
33.	able	99	+	+	+	+		+		+
34.	centre	97		+	+	+		+	+	+
35.	course	96	+	+	+	+			+	+
36.	hours	95	+	+	+	+	+	+		
37.	few	95	+	+	+	+		+		+
38.	every	91	+	+	+	+		+		+
39.	give	89	+	+		+		+	+	+
40.	old	89	+	+	+	+		+		+
41.	visit	89		+	+	+		+	+	+
42.	still	87	+	+	+		+	+	+	
43.	having	87	+	+	+		+	+	+	

			ExE	DisE	DesE	POR	DR	SSt	CL	AL
44.	high	87	+	+	+	+	+	+		
45.	far	86	+	+		+	+	+		+
46.	look	81	+	+	+			+	+	+
47.	whole	76	+	+	+	+	+	+		
48.	almost	74	+	+	+	+	+	+		
49.	enjoy	71	+	+	+	+	+			+
50.	during	68		+	+	+	+	+	+	
51.	making	64	+	+	+	+	+		+	
52.	times	64	+	+	+	+	+	+		
53.	week	61		+		+	+	+	+	+
54.	looking	50		+	+		+	+	+	+

[Abbreviations used in table: ExE (Expository Essay), DisE (Discursive Essay), DesE (Description Essay), POR (Personal Observation Report), DR (Data Report), SSt (Short Story), CL (Complain Letter), AL (Advice Letter)]

Core shared vocabulary (5 genres)

	Joint words	Freq. #	Joint in 5 genres							
			ExE	DisE	DesE	POR	DR	SSt	CL	AL
1.	children	308	+	+	+	+	+			
2.	students	283	+	+	+	+	+			
3.	she	252	+	+	+	+		+		
4.	world	223	+	+	+	+	+			
5.	life	215	+	+	+			+		+
6.	may	213	+	+		+			+	+
7.	<name>	199		+	+	+		+	+	
8.	believe	190	+	+		+		+	+	
9.	important	174	+	+	+	+		+		
10.	number	174	+	+		+	+		+	
11.	those	165	+	+	+	+	+			
12.	hand	147	+	+	+	+	+			
13.	use	136	+	+		+	+			+
14.	its	134	+	+	+	+	+			
15.	part	117	+	+	+	+	+			
16.	public	114	+	+		+	+		+	
17.	between	109	+	+	+	+	+			
18.	fact	108	+	+	+	+			+	
19.	through	105	+	+	+		+	+		
20.	next	105		+	+	+	+	+		
21.	went	103			+	+	+	+	+	
22.	going	101		+	+	+		+		+
23.	family	99	+	+	+	+		+		
24.	lot	95	+	+	+	+				+
25.	really	95	+	+	+			+		+
26.	want	91	+	+		+		+		+
27.	three	89			+	+	+	+	+	
28.	experience	88	+	+	+	+		+		
29.	might	88	+	+	+	+				+
30.	town	85			+	+		+	+	+
31.	main	85	+	+	+	+	+			
32.	information	83	+	+		+	+			+
33.	never	81	+	+	+			+		+
34.	free	81	+	+	+	+				+
35.	soon	79		+	+			+	+	+
36.	popular	76	+	+	+	+	+			
37.	places	75	+	+	+	+				+
38.	order	75	+	+		+	+		+	
39.	why	74	+	+		+		+		+
40.	say	70	+	+	+			+		+
41.	large	70	+	+	+	+	+			
42.	end	69	+	+	+		+			
43.	working	69	+	+	+		+			+

#	word	count	1	2	3	4	5	6	7	8
44.	off	68		+	+		+	+	+	
45.	idea	67	+	+		+		+		+
46.	seen	66	+	+	+		+	+		
47.	everyone	66	+	+	+	+		+		
48.	several	65	+	+	+	+			+	
49.	offer	63	+	+	+	+			+	
50.	come	61	+	+	+			+		+
51.	days	58	+	+	+		+	+		
52.	buy	58	+	+		+		+		+
53.	taking	57	+	+	+		+		+	
54.	since	52	+	+	+	+			+	
55.	found	51	+	+	+			+	+	
56.	situation	51	+	+	+	+			+	
57.	though	51	+	+	+		+			+
58.	night	50			+	+		+	+	+
59.	stay	49		+		+	+	+		+
60	everything	49		+	+	+		+		+
61.	hard	48	+	+	+	+		+		

Core vocabulary (in the 400 most frequent words) used in 4 genres

	Joint words	Freq. #	Joint in 4 genres							
			ExE	**DisE**	**DesE**	**POR**	**DR**	**SSt**	**CL**	**AL**
1.	need	174	+	+		+				+
2.	young	169	+	+	+	+				
3.	conclusion	156	+	+		+	+			
4.	parents	152	+	+				+		+
5.	example	138	+	+	+	+				
6.	often	131	+	+	+	+				
7.	own	130	+	+	+	+				
8.	internet	110	+	+			+			+
9.	job	107	+	+				+		+
10.	less	103	+	+		+	+			
11.	problems	103	+	+	+	+				
12.	said	101		+	+	+		+		
13.	addition	96	+	+		+			+	
14.	things	92	+	+	+					+
15.	live	91	+	+	+					+
16.	modern	87	+	+	+	+				
17.	view	85	+	+		+			+	
18.	animals	83	+	+		+	+			
19.	play	81	+	+		+				+
20.	future	80	+	+		+			+	
21.	something	79	+	+	+			+		
22.	therefore	78	+	+		+			+	
23.	house	78		+	+			+		+
24.	problem	73	+	+					+	+
25.	furthermore	73	+	+		+			+	
26.	schools	72	+	+	+	+				
27.	down	71		+	+		+	+		
28.	felt	70		+	+	+		+		
29.	area	68	+	+		+	+			
30.	result	67	+	+			+		+	
31.	firstly	64	+	+		+			+	
32.	huge	64	+	+	+			+		
33.	under	62		+	+	+	+			
34.	difficult	60	+	+		+		+		
35.	art	59	+	+	+	+				
36.	right	59	+	+				+		+
37.	second	59	+	+		+	+			
38.	areas	59	+	+		+	+			
39.	visitors	56		+	+	+	+			
40.	makes	56	+	+	+	+				
41.	real	56	+	+	+			+		
42.	given	55	+	+		+	+			
43.	room	55			+	+		+	+	

44.	cost	55		+		+	+		+	
45.	again	55		+			+	+		+
46.	expensive	55	+	+		+				+
47.	start	54	+	+				+		+
48.	try	54		+	+	+				+
49.	least	53		+		+	+		+	
50.	holiday	52			+	+		+	+	
51.	away	51		+	+			+		+
52.	ever	50	+	+	+			+		
53.	put	50	+	+		+		+		
54.	quality	49	+	+		+			+	
55.	show	47	+	+		+	+			
56.	history	47	+	+	+	+				

Core shared vocabulary (3 genres)

Joint words	Freq. #	Joint in 3 genres							
		ExE	DisE	DesE	POR	DR	SSt	CL	AL
1. countries	138	+	+			+			
2. him	110	+		+			+		
3. become	106	+	+	+					
4. learn	99	+	+						+
5. without	95	+	+	+					
6. lives	92	+	+	+					
7. language	91	+	+		+				
8. age	90	+	+			+			
9. person	84	+	+	+					
10. means	82	+	+		+				
11. sports	81	+	+		+				
12. living	80	+	+	+					
13. provide	77	+	+		+				
14. teachers	77	+	+		+				
15. got	76			+			+		+
16. traditional	76	+	+		+				
17. women	76	+	+			+			
18. possible	75	+	+					+	
19. health	75	+	+		+				
20. saw	74			+		+	+		
21. process	73	+	+		+				
22. reasons	71	+	+		+				
23. higher	70	+	+			+			
24. water	67	+				+	+		
25. true	66	+	+				+		
26. products	65	+	+		+				
27. study	65	+	+		+				
28. transport	65		+		+	+			
29. foreign	65	+	+		+				
30. development	65	+	+			+			
31. today	64	+	+	+					
32. ways	64	+	+		+				
33. men	63	+	+			+			
34. positive	63	+	+		+				
35. group	63		+		+	+			
36. point	61	+	+		+				
37. computer	60	+	+			+			
38. took	59			+			+	+	
39. rather	59		+	+	+				
40. clear	58	+	+			+			
41. hope	58				+			+	+

42.	phone	57		+			+	+		
43.	enough	56	+	+				+		
44.	sure	56		+				+		+
45.	especially	56		+	+	+				
46.	level	55	+	+			+			
47.	amount	55	+	+			+			
48.	increased	54	+	+			+			
49.	using	54	+	+			+			
50.	later	53	+	+				+		
51.	teenagers	53	+	+		+				
52.	activities	52	+	+		+				
53.	variety	52	+	+		+				
54.	needs	51	+	+		+				
55.	begin	51	+	+		+				
56.	here	51	+					+		+
57.	shop	51		+		+			+	
58.	friend	51			+			+		+
59.	morning	51			+	+		+		
60.	reason	41	+	+	+					
61.	minutes	50					+	+	+	
62.	secondly	49	+	+		+				
63.	love	48	+		+					+
64.	range	48	+	+		+				
65.	looked	48			+			+	+	
66.	thought	47		+				+		+
67.	increase	47		+		+	+			

Core shared vocabulary (2 genres)

Joint words		Freq. #	Joint in 2 genres							
			ExE	DisE	DesE	POR	DR	SSt	CL	AL
1.	education	141	+	+						
2.	social	100	+	+						
3.	others	98	+	+						
4.	child	83	+	+						
5.	university	77	+	+						
6.	learning	73	+	+						
7.	does	72	+	+						
8.	cannot	71	+	+						
9.	shows	69		+			+			
10.	service	68				+			+	
11.	whether	66	+	+						
12.	mobile	65		+			+			
13.	games	63		+			+			
14.	case	61	+	+						
15.	technology	60	+	+						
16.	per	59				+	+			
17.	prices	57		+		+				
18.	restaurant	57				+			+	
19.	support	56	+	+						
20.	museum	56		+		+				
21.	facilities	56		+		+				
22.	international	56	+	+						
23.	environment	55	+	+						
24.	career	54	+	+						
25.	personal	53	+	+						
26.	cars	53		+			+			
27.	seems	52	+	+						
28.	issue	52	+	+						
29.	travel	52	+	+						
30.	improve	52	+	+						
31.	bar	51				+	+			
32.	pay	51	+	+						
33.	half	50			+		+			
34.	mother	50		+				+		
35.	shopping	49		+		+				
36.	business	49	+	+						
37.	research	49	+	+						
38.	started	48					+	+		
39.	knowledge	48	+	+						
40.	phones	48		+		+				
41.	watch	48		+				+		
42.	change	47	+	+						

Appendix Seven: Frequent 3-word lexical bundles per genre

3-word lexical bundles in the 'Expository Essay'

	LB	freq.		LB	freq.		LB	freq.
1.	I believe that	23	18.	it is a	7	35.	that they will	6
2.	in my opinion	22	19	on the other	7	36.	the quality of	6
3.	in order to	15	20	that there are	7	37.	the same time	6
4.	there is no	13	21.	that we should	7	38.	to sum up	6
5.	be able to	11	22.	the idea that	7	39.	a variety of	5
6.	the fact that	11	23.	the other hand	7	40.	a waste of	5
7.	first of all	10	24.	we need to	7	41.	as well as	5
8.	it would be	10	25.	a lot of	6	42.	I think that	5
9.	one of the	10	26.	a number of	6	43.	on their own	5
10.	in my view	9	27.	and it is	6	44.	people who are	5
11.	is no doubt	9	28	at the same	6	45.	that it would	5
12.	as a result	8	29.	in the world	6	46.	the lives of	5
13.	I do not	8	30.	more likely to	6	47.	there are many	5
14.	no doubt that	8	31.	need to be	6	48.	there is a	5
15.	of the world	8	32.	not believe that	6	49.	to have a	5
16.	in terms of	7	33.	part of the	6	50.	to me that	5
17.	in this way	7	34.	that it is	6			

	3-word lexical bundles in the 'Discursive Essay'						
	LB	**f**		**LB**	**f**		**LB**	**f**
1.	on the other	78	**89.**	are those who	8	**177.**	such as the	6
2.	the other hand	74	**90.**	believe that the	8	**178.**	the advantages of	6
3.	I believe that	38	**91.**	can be a	8	**179.**	the amount of	6
4.	in my opinion	31	**92.**	is no doubt	8	**180.**	the benefits of	6
5.	that it is	29	**93.**	is not a	8	**181.**	the development of	6
6.	the one hand	26	**94.**	is not the	8	**182.**	the issue of	6
7.	a lot of	26	**95.**	is that there	8	**183.**	the most important	6
8.	on the one	25	**96.**	it is clear	8	**184.**	the right to	6
9.	in order to	23	**97.**	needs to be	8	**185.**	there are several	6
10.	it would be	20	**98.**	on the Internet	8	**186.**	there has been	6
11.	more and more	20	**99.**	over the world	8	**187.**	there is some	6
12.	there is a	20	**100.**	the best way	8	**188.**	this can be	6
13.	be able to	20	**101.**	the number of	8	**189.**	this essay will	6
14.	as a result	18	**102.**	the opportunity to	8	**190.**	this is particularly	6
15.	to sum up	18	**103.**	this is because	8	**191.**	this is the	6
16.	it is the	17	**104.**	those who are	8	**192.**	this means that	6
17.	that there is	17	**105.**	to become a	5	**193.**	to start with	6
18.	one of the	16	**106.**	to me that	8	**194.**	to the Internet	6
19.	of the world	16	**107.**	all things considered	7	**195.**	while some people	6
20.	I think that	16	**108.**	an early age	7	**196.**	would have to	6
21.	there is no	15	**109.**	and disadvantages to	7	**197.**	would say that	6
22.	the fact that	15	**110.**	believe that it	7	**198.**	access to the	5
23.	advantages and disadvantages	15	**111.**	best way to	7	**199.**	advantage of the	5
24.	in favour of	14	**112.**	can be seen	7	**200.**	and this can	5
25.	it is not	14	**113.**	can be very	7	**201.**	argue that it	5
26.	as well as	14	**114.**	for several reasons	7	**202.**	argument in favour	5
27.	are able to	13	**115.**	for those who	7	**203.**	be encouraged to	5
28.	around the world	13	**116.**	I do not	7	**204.**	because it is	5
29.	it is true	13	**117.**	I feel that	7	**205.**	being able to	5
30.	it is important	13	**118.**	I would argue	7	**206.**	believe that this	5
31.	is that the	13	**119**	I would say	7	**207.**	both advantages and	5
32.	is that it	13	**120.**	in front of	7	**208.**	can be expensive	5
33.	to begin with	13	**121.**	in this way	7	**209.**	children should be	5
34.	a result of	12	**122.**	it is clear that	7	**210.**	could also be	5
35.	to be a	12	**123.**	it is very	7	**211.**	do not believe	5
36.	that there are	12	**124.**	it may be	7	**212.**	from an early	5
37.	the Internet has	12	**125.**	live and work	7	**213.**	have to pay	5
38.	and more people	12	**126.**	live in a	7	**214.**	in other words	5
39.	in terms of	12	**127.**	may not be	7	**215.**	in recent years	5
40.	people believe that	12	**128.**	no doubt that	7	**216.**	in the world	5
41.	this is not	12	**129.**	of the argument	7	**217.**	is more important	5
42.	in the past	12	**130.**	of the Internet	7	**218.**	is no longer	5
43.	it can be	12	**131.**	people feel that	7	**219.**	is not an	5
44.	that this is	12	**132.**	that it would	7	**220.**	is one of	5
45.	it is also	12	**133.**	the case that	7	**221.**	is that they	5

#	Phrase		#	Phrase		#	Phrase	
46.	all in all	12	134.	the chance to	7	222.	it is difficult	5
47.	what is more	12	135.	the health of	7	223.	make sure that	5
48.	need to be	11	136.	the responsibility of	7	224.	may be a	5
49.	it is a	11	137.	the use of	7	225.	more likely to	5
50.	it is often	11	138.	there are those	7	226.	more people are	5
51.	there are some	11	139.	think that the	7	227.	not be able	5
52.	there are also	11	140.	to be more	7	228.	not have to	5
53.	be argued that	11	141.	to go to	7	229.	of the city	5
54.	do not have	11	142.	to live in	7	230.	one argument in	5
55.	is not always	10	143.	we need to	7	231.	out of the	5
56.	both sides of	10	144.	would argue that	7	232.	outweigh the disadvantages	5
57.	in the first	10	145.	a matter of	6	233.	own view is	5
58.	is true that	10	146.	addition to this	6	234.	part of our	5
59.	the first place	10	147.	an important part	6	235.	people argue that	5
60.	some people believe	10	148.	argue that the	6	236.	people think that	5
61.	seems to me	10	149.	at home and	6	237.	so it is	5
62.	it seems to	10	150.	at the same	6	238.	some of the	5
63.	it seems that	10	151.	be said that	6	239.	some people think	5
64.	in my view	10	152.	because they are	6	240.	standard of living	5
65.	this is a	10	153.	can be done	6	241.	that the world	5
66.	there are many	10	154.	due to the	6	242.	that young people	5
67.	view is that	10	155.	first of all	6	243.	the importance of	5
68.	they do not	10	156.	have to be	6	244.	the only way	5
69.	they want to	10	157.	if they are	6	245.	the same time	5
70.	in the future	9	158.	in addition to	6	246.	there are advantages	5
71.	all over the	9	159.	is important that	6	247.	there are good	5
72.	a number of	9	160.	is important to	6	248.	there are no	5
73.	argument is that	9	161.	is often the	6	249.	there are two	5
74.	is also a	9	162.	is the most	6	250.	there is also	5
75.	a variety of	9	163.	it is easy	6	251.	this kind of	5
76.	it could be	9	164.	it is more	6	252.	to ensure that	5
77.	part of the	9	165.	it is possible	6	253.	to understand the	5
78.	point of view	9	166.	lead to a	6	254.	understanding of the	5
79.	quality of life	9	167.	many young people	6	255.	way of life	5
80.	should not be	9	168.	my own view	6	256.	we live in	5
81.	sides of the	9	169.	number of people	6	257.	will be able	5
82.	the majority of	9	170.	of the most	6	258.	will have the	5
83.	the quality of	9	171.	of these is	6	259.	with each other	5
84.	there are a	9	172.	on both sides	6	260.	would be a	5
85.	this is that	9	173.	on the whole	6	261.	would be better	5
86.	to have a	9	174.	people have different	6	262.	would be to	5
87.	while there are	9	175.	people who have	6	263.	young people to	5
88.	and it is	8	176.	some people argue	6	264.		

3-word lexical bundles in the 'Descriptive Essay'								
LB		freq.		**LB**	freq.		**LB**	freq.
1.	one of the	17	**6.**	the end of	7	**11.**	was one of	6
2.	a lot of	15	**7.**	the fact that	7	**12.**	I went there	5
3.	all over the	8	**8.**	to go to	7	**13.**	to have a	5
4.	out of the	8	**9.**	all in all	6			
5.	of the most	7	**10.**	in the world	6			

3-word lexical bundles in the 'Data Report'								
LB		freq.		**LB**	freq.		**LB**	freq.
1.	the number of	54	**17.**	over the next	10	**33.**	according to the	6
2.	the bar chart	21	**18.**	there was a	10	**34.**	at the end	6
3.	the proportion of	17	**19.**	bar chart shows	9	**35.**	in the first	6
4.	is clear that	17	**20.**	in terms of	9	**36.**	it is evident	6
5.	it is clear	17	**21.**	the pie chart	9	**37.**	stages in the	6
6.	the most popular	15	**22.**	part of the	9	**38.**	than the other	6
7.	over the period	15	**23.**	number of people	9	**39.**	the production of	6
8.	end of the	14	**24.**	be seen that	8	**40.**	at just under	5
9.	of the period	13	**25.**	in the number	8	**41.**	beginning with the	5
10.	of the population	13	**26.**	most of the	8	**42.**	in contrast to	5
11.	the percentage of	12	**27.**	the line graph	8	**43.**	in the process	5
12.	the end of	11	**28.**	we can see	8	**44.**	looking at the	5
13.	the figures for	10	**29.**	chart shows the	7	**45.**	on the other	5
14.	the amount of	10	**30.**	clear that the	7	**46.**	that there are	5
15.	can be seen	10	**31.**	of people who	7	**47.**	the pie charts	5
16.	it can be	10	**32.**	there is a	7	**48.**	to begin with	5

3-word lexical bundles in the 'Personal Observation Report'								
	LB	freq.		LB	freq.		LB	freq.
1.	this report is	43	17.	a number of	9	33.	there should be	6
2.	of this report	36	18.	in order to	8	34.	and it is	5
3.	report is to	36	19.	in this report	8	35.	I believe that	5
4.	purpose of this	22	20.	to assess the	8	36.	I would like	5
5.	the purpose of	21	21.	would like to	8	37.	if these recommendations	5
6.	aim of this	14	22.	in the area	7	38.	in the city	5
7.	the aim of	14	23.	is to assess	7	39.	is to outline	5
8.	as well as	13	24.	on the whole	7	40.	on the other	5
9.	to sum up	12	25.	the most popular	7	41.	recommendations are implemented	5
10.	wide range of	11	26.	there is a	7	42.	the majority of	5
11.	per cent of	11	27.	a variety of	6	43.	the other hand	5
12.	some of the	11	28.	offers a wide	6	44.	the suitability of	5
13.	the fact that	11	29.	report on the	6	45.	these recommendations are	5
14.	a wide range	11	30.	the number of	6	46.	this is a	5
15.	would be a	10	31.	the service is	6			
16.	a lot of	9	32.	there are many	6			

3-word lexical bundles in the 'Short Story'								
	LB	freq.		LB	freq.		LB	freq.
1.	it was a	19	17.	I was in	7	33.	and saw that	5
2.	out of the	14	18.	I was very	7	34.	as fast as	5
3.	as soon as	12	19.	I woke up	7	35.	going to be	5
4.	decided to go	11	20.	it was my	7	36.	I had a	5
5.	it was the	11	21.	that I was	7	37.	I looked at	5
6.	there was a	10	22.	to go to	7	38.	I realised I	5
7.	and we were	9	23.	as I was	6	39.	I realised that	5
8.	I had been	9	24.	back to the	6	40.	it was raining	5
9.	there was no	9	25.	go to the	6	41.	looked at the	5
10.	I decided to	8	26.	he had been	6	42.	that he had	5
11.	I was so	8	27.	I couldn't	6	43.	the first time	5
12.	in the morning	8	28.	I felt so	6	44.	was a beautiful	5
13.	that I had	8	29.	I had no	6	45.	was about to	5
14.	that it was	8	30.	I left the	6	46.	was getting dark	5
15.	and I had	7	31.	I went to	6	47.	was going to	5
16.	I didn't	7	32.	in front of	6	48.	what to do	5

	3-word lexical bundles in the 'Complaint Letter'							
	LB	**freq.**		**LB**	**freq.**		**LB**	**freq.**
1.	I am writing	38	**13.**	to express my	8	**25.**	included in the	6
2.	am writing to	26	**14.**	hearing from you	8	**26.**	make matters worse	6
3.	look forward to	19	**15.**	as soon as	7	**27.**	one of your	6
4.	I look forward	16	**16.**	first of all	7	**28.**	refund of the	6
5.	to complain about	14	**17.**	soon as possible	7	**29.**	that you will	6
6.	I would like	13	**18.**	the fact that	7	**30.**	the cost of	6
7.	dear sir or	12	**19.**	when I tried	7	**31.**	to make matters	6
8.	sir or madam	12	**20.**	writing to you	7	**32.**	forward to your	5
9.	writing to complain	12	**21.**	as a result	6	**33.**	I have been	5
10.	would like to	9	**22.**	I tried to	6	**34.**	I hope you	5
11.	forward to hearing	8	**23.**	complain about the	6	**35.**	that I had	5
12.	to hearing from	8	**24.**	I would be	6			

	3-word lexical bundles in the 'Advice Letter'							
	LB	**freq.**		**LB**	**freq.**		**LB**	**freq.**
1.	let me know	23	**13.**	a good idea	6	**25.**	great to hear	5
2.	first of all	14	**14.**	a lot of	6	**26.**	hear about your	5
3.	m sure you	13	**15.**	be able to	6	**27.**	how are you	5
4.	I were you	12	**16.**	I know you	6	**28.**	I can give	5
5.	if I were	12	**17.**	I think you	6	**29.**	me know what	5
6.	me know how	11	**18.**	it was great	6	**30.**	that you have	5
7.	hear from you	9	**19.**	you have a	6	**31.**	things you can	5
8.	to hear from	9	**20.**	was great to	6	**32.**	to go to	5
9.	I hope you	8	**21.**	you have to	6	**33.**	to see you	5
10.	thanks for your	8	**22.**	you will be	6	**34.**	would be a	5
11.	you can do	8	**23.**	and tell me	5	**35.**	write back soon	5
12.	you want to	7	**24.**	can give you	5			