Developing English language support for immigrant students in Irish post-primary schools:

A corpus linguistics approach

Ph.D. Thesis in Applied Linguistics

2012

Stergiani Kostopoulou

University of Dublin, Trinity College

Centre for Language and Communication Studies

Declaration

I declare that this doctoral thesis has not been submitted as an exercise for a degree at this or any other University and it is entirely my own work.

I agree to deposit this thesis in the University's open access institutional repository or allow the Library to do so on my behalf, subject to Irish Copyright Legislation and Trinity College Library conditions of use and acknowledgement.

S.K. 30/4/2012

Stergiani Kostopoulou Date

Summary

The present research was conducted as part of the English Language Support Programme (ELSP) of the Trinity Immigration Initiative (2007-2010), Trinity College Dublin (TCD). The aim of the ELSP was to investigate and enhance the provision of English language support for immigrant students in Irish post-primary schools. In line with the scope of the ELSP, the present research has two aims. First, it sets out to analyse the language of post-primary subject textbooks and examination papers in order to understand the demands that curriculum language makes of immigrant students. Secondly, it explores the pedagogical implications of the findings of this linguistic analysis and their application in post-primary English language support.

Chapter 1 establishes the general context and aims of the research. It begins by presenting the measures and supports that were adopted to facilitate the linguistic integration of immigrant students into Irish post-primary schools. A critical overview of official language support policy is provided, which outlines the major deficits, needs and challenges as these have been reported by research (Lyons and Little, 2009; Smyth et al., 2009). Specific reference is then made to the contribution of Integrate Ireland Language and Training (IILT), a not-for-profit campus company of TCD, and the role it played in migrant language education in Ireland from 2000 to 2008. The chapter concludes by setting out the aims of the present research as part of the ELSP, which sought to build on IILT's work.

Chapter 2 describes the methodology that was employed to address the primary aim of the research. It specifically argues for an applied corpus linguistics approach to the analysis of the language of six Junior Cycle (lower secondary) subjects (English, geography, history, CSPE (civic, social and political education), mathematics, and science) as this is manifested in textbooks and Junior Certificate examination papers. The methodological procedure involved: i) the compilation of twelve subject-specific corpora, two for each of the six subjects, one comprising four textbooks and the other comprising ten examination papers, and ii) frequency-based lexical analysis of corpora and the qualitative interpretation of findings facilitated by the WordList and the Concord tools of WordSmith 4 (Scott, 2004) respectively.

Chapter 3 reports and interprets the findings of this corpus analysis. The discussion is divided into two parts. In the first part, corpus-derived lexical profiles (meeting particular frequency and range restrictions) are presented for each of the twelve corpora. These profiles consist of i) the commonest function and lexical words, ii) the most significant collocates of the commonest lexical words, iii) the most significant collocations as single units, iv) the commonest 4-word clusters, analysed in respect of their densities, forms, structures,

pragmatic meanings and functions, and complementation patterns (i.e. adjacent words/phrases). Concordances are also provided for 4-word clusters to display their use in context. The second part of the discussion compares and contrasts the semantic and functional associations first of lexical words, based on their collocates and concordances, and second of 4-word clusters which appear to recur across six or five corpora of the same register (i.e. textbooks/examination papers), based on their complementation patterns and concordances. The interpretation of corpus findings in this chapter suggests a number of pedagogically useful insights into subject-specific language which constitute the first empirical information provided so far on the language of the Irish post-primary curriculum.

Chapter 4 addresses the second aim of the research, discussing the pedagogical implications and applications of the empirical findings. It is argued that corpus-derived language descriptions can inform and facilitate the provision of English language support in multiple ways by contributing to: i) principled decision making about the content and methodology of language teaching, ii) raising the awareness of teachers, iii) developing whole-school approaches to cross-curriculum language learning, iv) elaborating the current curriculum framework and its conversion into a lexically-based pedagogic agenda, v) developing context-specific materials, and vi) constructing appropriate assessment instruments.

Chapter 5 recapitulates the aims of the research and explains how they were achieved. Then it evaluates the contribution of the research to i) an understanding of curriculum language, ii) the development of pedagogical practice in language support, iii) teacher education programmes for post-primary (subject and language support) teachers, iv) applied corpus linguistics and English for Specific Purposes research, and v) migrant education beyond the Irish context. The limitations of the research are also pointed out and some possible lines of future work are identified.

Overall, the thesis begins to remedy the lack of empirical research into the language of the Irish post-primary curriculum. The findings could be exploited in numerous ways in English language support to establish direct links with the mainstream subject classroom. Both corpus data and corpora could be further utilised in teacher education programmes to promote teachers' language and pedagogic awareness. Demonstrating as it does the value of language corpora, the research also responds to calls for more applied corpus linguistics research motivated by an interest in second language pedagogy specifically in secondary education.

Acknowledgements

This doctoral thesis represents an important milestone of my academic, professional and personal development. I owe my gratitude to many people who supported me throughout this project. First of all, I am extremely thankful to my supervisor, Professor David Little, for involving me in the English Language Support Programme (ELSP) to conduct the present research, which introduced me to the fascinating world of language corpora and corpus linguistics tools. I am grateful to David for his unfailing guidance and support during the past four years. Our stimulating discussions always provided me with plenty of food for thought and helped me to develop the ability to think rigorously and to grow intellectually. By exercising the pedagogical principles of his learner autonomy theory (1991; 2007), i.e. encouraging me to take charge of the planning and monitoring of my research work and to scaffold my critical reflection on and self-evaluation of all aspects of my academic progress, David has moulded me into the autonomous researcher I am today.

I would also like to thank my colleagues from the ELSP project, Zachary Lyons and Rachael Fionda, who have been a source of good advice and collaboration. Special thanks go to Zach for his invaluable help with the design and development of the language corpora that were analysed in this research.

I am also grateful to several members of the Centre for Language and Communication Studies (CLCS) for helping me in different ways. John Saeed helped me in my application process for a Ph.D. at CLCS and a research award by Trinity College. John Harris offered me his insights on the pedagogical context of the present research which informed my original research proposal. Elaine Uí Dhonnchadha and Denise O'Leary provided answers to my questions on corpus linguistics and statistics respectively. Lorna Carson offered me the opportunity to teach the Applied Linguistics of English Language Teaching module for three years, as part of my research studentship requirements. David Singleton granted me the privilege to deliver some of his lectures on first and second language acquisition and vocabulary learning to undergraduate and post-graduate students at CLCS. I am finally indebted to the CLCS Department for employing me as the Instructor for the English for Academic Purposes Induction course for M.Phil. students at CLCS for the past four years and I am most grateful to Breffni O'Rourke for coordinating this course. All of the above posts were truly rewarding and have contributed immensely to my professional and academic development.

I gratefully acknowledge the financial assistance I received from Trinity College Dublin in the first three years of my doctoral studies and from Dublin City Council in my fourth year. I am

also thankful to the ELSP project for funding my participation in a number of conferences in Europe and my attendance at a Corpus Linguistics Summer School in Aston University (Birmingham, 2009).

For their friendship and emotional support, I would like to thank two very good friends: my study companion, Deirdre Murphy, who made the endless hours of Ph.D. work in Room 4095 less mundane, and Elpida Tzanis-Piperigou for her constant encouragement and words of wisdom, which sustained my motivation and my determination to achieve my goals, and encouraged me to aim even higher.

I owe my deepest gratitude to my in-laws and my fiancé, Antonis Terzakis, for their generous support and understanding particularly in the closing stages of my studies. I reserve my final and most heartfelt thanks for my family for their unwavering love and patience, and especially for my sister, Ioanna, who has been a greater support than I would have ever hoped to deserve. I very lovingly dedicate this thesis to her.

Table of Contents

Volume I

Volume II

List of Appendices

List of Tables

List of Figures

Introduction

Following the economic boom of the 1990s and the expansion of the European Union in 2004, Ireland became another European country that experienced an influx of students in the school sector for whom the language of education, i.e. English, was a second language (henceforth ESL students). Irish educational policy regarding the integration of these students into schools complies with the declarations of European policy (e.g. European Convention on the Legal Status of Migrant Workers (Council of Europe, 1977); Article 16 of the United Nations Convention (1989) on the Rights of the Child, the conclusions from the 2000 Lisbon European Council of the European Parliament, etc.), granting all students access to education on the same basis as their native peers, regardless of legal status. The implementation of this policy, however, poses a number of challenges to Irish schools.

ESL students' educational integration is a multi-dimensional and complex process, which can be influenced by various factors, for instance, linguistic, social, personal, academic, situational and psychological (see Ward, 2004; Devine, 2005; Devine and Kelly, 2006). It is the language issue, however, that has been consistently identified as the key to successful integration but also as the greatest single barrier (Keogh and Whyte, 2003; Devine et al., 2004; INTO, 2004; Ward, 2004; Devine, 2005; DES, 2006; Lazenby-Simpson, 2006; Fionda, 2011; Lyons and Little, 2009; Smyth et al, 2009). On these grounds, ESL students' integration specifically into post-primary education, and especially into Junior Cycle, is examined purely from a linguistic perspective in this thesis.

To support ESL students' linguistic integration into the mainstream classroom, English language support is provided on a withdrawal basis. That is, students follow the designated curriculum but are withdrawn from their mainstream class, usually for one or two periods each weekfora maximum of two years (DES, 2009), for additional English language classes. Language support aims to equip students with the language knowledge and skills they need to access the curriculum, engage in classroom interaction, tasks and activities, and meet the requirements of public examinations. ESL students are, in other words, faced with a double challenge. They must 'simultaneously develop a form of double plurilingualism' both external, i.e. learning English as a second language, and internal, i.e. subject-specific academic language learning (Vollmer, 2007). Helping students to gain control of the 'sub-languages' that different curriculum subjects operate (Rosen, 1972) is fundamental to this end.

In practice, however, it appears that language support provision fails to address this aim effectively. Research into current practices reveals that learning and teaching in the language support class are not sufficiently or appropriately linked to the specialized language demands

of the mainstream subject classroom (Lyons and Little, 2009: 35; ESRI, 2009). Morover, because many teachers lack a clear understanding of curriculum-specific language, language support is conceived and delivered as a general English course (Lyons and Little, 2009). Teachers report a lack of appropriate teaching and learning materials designed to meet ESL students' specific needs (ibid. 37), and the use of inappropriate materials from the internet and materials for teaching English as a foreign language (ibid.) which do not address the communicative domain of post-primary education. In addition, although subject teachers' responsibility to support ESL students' linguistic access to subject classrooms is clearly stated in the official language support policy (DES, 2009: 2), many subject teachers appear not to recognise and accept this responsibility. As a result, there is usually no collaboration between language support and subject teachers and a systematic and coherent whole-school approach to supporting ESL students is rare. Finally, the lack of teacher training for post-primary ESL support and the need for a whole-school approach to continuous professional development evidently appear to be among the pressing concerns of teachers and principals (Lyons and Little, 2009; OECD, 2009; Smyth et al., 2009).

Taken together, these deficits, gaps, and needs result in haphazard language support and a disintegration of practices at a school level. These most probably have a detrimental effect on ESL students' language development that inevitably impinges on their academic achievement. Failure to address these needs and enable students to access and benefit from mainstream education could, in the long term, have a negative impact on students' professional and economic development in Irish society and give rise to social problems whose seriousness cannot be overestimated (e.g. unemployment, criminal and antisocial behaviour, negative impact on social cohesion etc.).

It may be argued that a possible common denominator linking the above issues might be the lack of a clear understanding of the linguistic demands that are imposed on ESL students by the post-primary classroom. Detailed information on curriculum language that accurately describes the linguistic repertoire required of students could address these issues and help teachers to establish direct links between language support and the mainstream subject classroom. More specifically, it would i) help ESL teachers to establish direct links between language learning/teaching in the language support class and disciplinary learning in the subject class, ii) inform the development of context-specific pedagogical materials, iii) inform teacher education to raise subject teachers' awareness of the language of the subject areas they teach, and iv) provide a common platform for the collaboration of all teachers for a whole-school approach to ESL students' language development.

To date, there has been no investigation of the nature of the language that students encounter in Irish post-primary education, and the English Language Proficiency Benchmarks developed by Integrate Ireland Language and Training (IILT, 2003a; see Appendix A1) are the only systematic attempt to capture the linguistic demands of the post-primary curriculum.

The Benchmarks constitute the official curriculum for language support (DES, 2009). They are functional descriptions of English language proficiency scaled according to the first three levels (A1-B1) of the *Common European Framework of Reference for Languages* (CEFR; Council of Europe, 2001) in relation to listening, reading, spoken interaction, spoken production and writing. These descriptions are formulated in positive 'can do' statements of behavioural capacity linked to core subjects of the post-primary curriculum. The Benchmarks, however, represent a general map for the planning, design and delivery of language support. They were designed to reflect the functional development that is expected of ESL students in post-primary education in terms of the communicative tasks they need to be able to perform; they are not intended to serve as a pedagogic agenda that specifies in detail the linguistic repertoire that students must develop to perform these tasks. In other words, the Benchmarks provide only a broad understanding of the language that is required for subject-specific learning.

It should be clear from the preceding discussion that there is a need for concrete information on the language of the Irish post-primary curriculum that could help teachers to facilitate ESL students' subject-specific language learning and address the aforementioned challenges. The research that is reported in this thesis constitutes a first attempt to fill this knowledge gap by undertaking the first empirical analysis of post-primary curriculum language.

The present research was conceived within the English Language Support Programme (ELSP) of the Trinity Immigration Initiative (TII), Trinity College Dublin (TCD; www.tcd.ie/immigration); a research programme on diversity, integration and policy (2007-2010). Acknowledging immigration as one of the major challenges faced by Irish society, the TII brought together six interacting projects (Children, Youth and Community Relations; English Language Support Programme; Migrant Careers and Aspirations; Migrant Networks – Facilitating Migrant Integration; National Survey of Immigrants in Ireland; National Policy Impacts) with the double mission to a) address key challenges posed by the unprecedented numbers of migrants arriving in Ireland in recent years, and b) help Irish society develop appropriate policies and practices for the new reality. Recognising that language is the key to integration into the school community and Irish society, and taking the existing deficiencies and needs of English language support as its starting point, the ELSP aimed to address some of the major challenges in this context and enhance language support practices.

In line with the scope of the ELSP, this research has two aims. The primary aim is to analyse the language of six Junior Cycle subjects, namely English, geography, history, CSPE (civic, social and political education), mathematics and science, as this manifests itself in two important school registers: commonly used textbooks and Junior Certificate examination papers (i.e. papers from the State examination taken at the end of the three-year Junior Cycle). For the envisaged linguistic analysis, a corpus linguistics methodology was employed which involves the analysis of subject-specific textbooks and examinations corpora using WordSmith Tools (Scott, 2004). The second aim of the research is to discuss and exemplify the pedagogical implications of the corpus analysis and their specific applications to language support provision.

The thesis consists of five chapters. Chapter 1 establishes the general context and aims of the research. It begins by presenting the measures and supports that were adopted to facilitate the linguistic integration of immigrant students into Irish post-primary schools. A critical overview of official language support policy is provided, which outlines the major deficits, needs and challenges as these have been reported by research (Lyons and Little, 2009; Smyth et al., 2009). Specific reference is then made to the contribution of Integrate Ireland Language and Training (IILT), a not-for-profit campus company of TCD, and the role it played in migrant language education in Ireland from 2000 to 2008. The chapter concludes by setting out the aims of the present research as part of the ELSP, which sought to build on IILT's work.

Chapter 2 describes the methodology that was employed to address the primary aim of the research. It specifically argues for an applied corpus linguistics approach to the analysis of the language of six Junior Cycle (lower secondary) subjects (English, geography, history, CSPE (civic, social and political education), mathematics, and science) as this is manifested in textbooks and Junior Certificate examination papers. The methodological procedure involved: i) the compilation of twelve subject-specific corpora, two for each of the six subjects, one comprising four textbooks and the other comprising ten examination papers, and ii) frequency-based lexical analysis of corpora and the qualitative interpretation of findings facilitated by the WordList and the Concord tools of WordSmith 4 (Scott, 2004) respectively.

Chapter 3 reports and interprets the findings of this corpus analysis. The discussion is divided into two parts. In the first part, corpus-derived lexical profiles (meeting particular frequency and range restrictions) are presented for each of the twelve corpora. These profiles consist of i) the commonest function and lexical words, ii) the most significant collocates of the commonest lexical words, iii) the most significant collocations as single units, iv) the commonest 4-word clusters, analysed in respect of their densities, forms, structures, pragmatic meanings and functions, and complementation patterns (i.e. adjacent

words/phrases). Concordances are also provided for 4-word clusters to display their use in context. The second part of the discussion compares and contrasts the semantic and functional associations first of lexical words, based on their collocates and concordances, and second of 4-word clusters which appear to recur across six or five corpora of the same register (i.e. textbooks/examination papers), based on their complementation patterns and concordances. The interpretation of corpus findings in this chapter suggests a number of pedagogically useful insights into subject-specific language which constitute the first empirical information provided so far on the language of the Irish post-primary curriculum.

Chapter 4 addresses the second aim of the research, discussing the pedagogical implications and applications of the empirical findings. It is argued that corpus-derived language descriptions can inform and facilitate the provision of English language support by contributing to: i) principled decision making about the content and methodology of teaching, ii) raising the awareness of teachers, iii) developing whole-school approaches to cross-curriculum language learning, iv) elaborating the current curriculum framework and its conversion into a lexically-based pedagogic agenda, v) developing context-specific materials, and vi) constructing appropriate assessment instruments.

Chapter 5 recapitulates the aims of the research and explains how they were achieved. Then it evaluates the contribution of the research to i) an understanding of curriculum language, ii) the development of pedagogical practice in language support, iii) teacher education programmes for post-primary (subject and language support) teachers, iv) applied corpus linguistics and English for Specific Purposes (ESP) research, and v) migrant education beyond the Irish context. The limitations of the research are also pointed out and some possible lines of future work are identified.

It should be noted that the complete corpus data sets which are explicitly referred to in Chapter 3 are included in the appendix of the thesis (Volume II). References to the relevant sections are indicated in the text where appropriate.

Overall, the thesis provides the first empirical insights into the linguistic characteristics of the six post-primary curriculum subjects under examination and thus begins to fill the vacuum of research in the Irish post-primary context. The findings of the present research confirm some intuitive beliefs about the nature of post-primary curriculum language and assumptions about appropriate pedagogical practices in language support and they refute others. As explained above, the novel information that is revealed by the research can be profitably used in post-primary language support in order to address some of the existing challenges and facilitate English language learning, teaching and assessment that are firmly embedded in the mainstream curriculum subjects. Taking into account the instrumental role of

specialised school language in accessing and constructing educational knowledge, it may be argued that empirical descriptions of the preferred language features, patterns and conventions of different subjects could equally benefit native speaking students who have learning difficulties due to a 'restricted code' (Bernstein, 1971).

Moving beyond the Irish context, and considering that the linguistic integration of immigrant students is, and will remain, a common challenge for all European educational systems, the design and methodology of this research has the potential to be adapted for use in other migrant education settings with similar pedagogical needs. By demonstrating the fruitful interface between corpus linguistics and language pedagogy, the thesis responds to the calls for more applied corpus linguistics studies which are motivated by an interest in second language education (e.g. Coxhead, 2010) as well as tothe need for more localised ESP corpora to examine additional language varieties. The thesis also responds to recent calls for applied corpus linguistics research specifically into the language of secondary education to support students in their school-based studies and in preparation for further studies (e.g. Flowerdew, 2010: 345; Coxhead, 2010: 466).

1 Facilitating the linguistic access of ESL students to Irish post-primary education

1.1 Introduction

This introductory chapter provides the background to the provision of English language support to students with English as a second language (henceforth, ESL students) in Irish post-primary schools and explains how the present research aims to address some of the existing needs and challenges in this pedagogical context.

The discussion begins with a brief introduction to the profile of the post-primary ESL student population in Ireland and the aims and structure of Irish post-primary education. Viewing language as the key to educational integration and academic achievement, the challenge of helping ESL students master the language of school education, and specifically its manifestations in the different curriculum subjects, is examined in some detail. In response to this challenge, the English language support policy that is currently in place is then described together with other supporting measures for ESL students. Moving from rhetoric to actual practice, some critical insights are offered into current language support provision, and deficits, needs, and challenges are outlined, as these have been reported by research studies (e.g. Lyons and Little, 2009; Smyth et al., 2009). The contribution of Integrate Ireland Language and Training (IILT) to language support pedagogy in the school sector forms an integral part of the discussion because IILT was an institution which played a pioneering role in migrant education in Ireland from 2000 until its closure in 2008. The English Language Support Programme (ELSP) of the Trinity Immigration Initiative (TII), Trinity College Dublin (TCD) was an attempt to build upon and move beyond IILT's work. The present study was conceived as part of the ELSP, and the chapter ends by setting out its aims and scope.

1.2 ESL students in Irish post-primary schools and the challenge of learning the language of education for academic achievement

Like other European countries, Ireland is a destination country for migratory movements. It was, in fact, transformed from an emigration country to an immigration country by the economic boom at the end of 1990s and the enlargement of the European Union in 2004. Its membership of the European Community and the dramatic increase in employment attracted large numbers of workers and asylum seekers, many of whom brought their families with them.

Irish policy towards the educational integration of migrant pupils and students (Education Act 1998, Education Welfare Act 2000, Equal Status Act 2000, Education for Persons with Special Educational Needs Act 2004,) complies with the declarations of the Council of Europe and the European Union (e.g., European Convention on the Legal Status of Migrant Workers (Council of Europe, 1977); Article 16 of the United Nations Convention (1989) on the Rights of the Child, the conclusions from the 2000 Lisbon European Council of the European Parliament, etc.). It acknowledges that all students are entitled to access primary and post-primary educationon the same basis as their native peers, regardless of legal status. In contrast with other countries which require proof of residence status (e.g. Denmark, Lithuania, Poland, Sweden and Iceland; Eurydice, 2004: 34), Ireland admits the children of irregular immigrants to its schools, according to Education Act 1998. Ireland is also one of the few countries which provide for the educational support of unaccompanied minors seeking asylum in the country (Separated Children Education Service of the Irish Refugee Council, set up in 2003).

Successful student integration into mainstream education is a multi-dimensional process which can be influenced by various factors, for instance, linguistic, social, personal, academic, situational and psychological (see Ward, 2004; Devine, 2005; Devine and Kelly, 2006). It is the language issue, however, which constitutes the key to all facets of students' integration and the factor that has been consistently identified as the greatest barrier (e.g. Keogh and Whyte, 2003; Devine et al., 2004; INTO, 2004; Ward, 2004; Devine, 2005; DES, 2006; Lyons and Little, 2009; Smyth et al., 2009). On these grounds, this discussion focuses on the linguistic dimension of ESL students' educational integration with particular reference to post-primary level.

Accessing post-primary education is considerably more challenging than accessing primary education for various reasons. First, it is common sense that the later ESL students enter the education system of the host country, the more difficult the transition will be,

because they will have to compensate for a greater educational loss to bridge the gap and reach the level of their English-speaking peers. The nature and delivery of the post-primary curriculum makes this endeavour further challenging for ESL students due to its fragmented and specialised nature, as opposed to the more holistic conceptualization of knowledge enacted in the primary curriculum. The technical and subject-specific post-primary curriculum, coupled with the demands of examinations, require more demanding and complex linguistic skills and registers compared to the more concrete and conversational English used in primary education. Tied to this, is the fact that the post-primary curriculum is delivered by teachers who are subject specialists with no knowledge of language acquisition theories or formal linguistic training in relation to the language of the subjects they teach.

This section begins with a brief profile of ESL students in Ireland which is followed by background information on the aims and structure of Irish post-primary education. Then the challenge of learning the language of school education as a second language is discussed in detail, with an emphasis on the importance of curriculum-specific language learning.

1.2.1 Profile of post-primary ESL student population

Post-primary ESL students have a diverse profile which needs to be accommodated within the school community. In 2009, the post-primary population in Ireland amounted to over 342,000 students of whom 17,000 represent more than 160 nationalities, in addition to Irish (Irish Department of Education and Science, 2008). Contrary to immigrant students in many European countries, the non-Irish student population is evenly dispersed in post-primary schools as nearly all schools have students from immigrant backgrounds (OECD, 2009: 18) but there is no school with over 50% immigrant students (Smyth et al., 2009: 46). A high concentration of newcomer students, however, can be found in disadvantaged schools with limited resources (i.e. vocational, community and comprehensive schools with students from low socio-economic background; OECD, 2006; Lyons and Little, 2009: 20); a fact which can be detrimental to students' educational achievement.

The largest groups of ESL students come from East European countries (Smyth et al., 2009: 68). The languages that are currently spoken by significant minorities in Ireland, apart from the official languages (i.e. English and Irish), include: Chinese, Polish, Lithuanian, Romani, Vietnamese, Yoruba, Albanian, Moldovan, Arabic and Russian (DES, 2006: 6). As regards students' English language proficiency, there is great variation. No official information is

available about students' proficiency levels; nevertheless, language competence has been identified as a determining factor and major challenge to students' integration (Keogh and Whyte, 2003; Devine et al., 2004; INTO, 2004; Ward, 2004; Devine, 2005; DES, 2006; Lazenby-Simpson, 2006; Lyons and Little, 2009; Smyth et al., 2009). Consistent with the results from PISA studies at a European level (OECD, 2006), research in the Irish context shows that language difficulties have a negative impact on student achievement, hindering their grasp of subjects (Keogh and Whyte, 2003; Vekic, 2003) and influencing their self-esteem (Vekic, 2003).

As regards the prior educational experience of ESL students in their first language, this is not only different from that of their native peers but, in certain cases, it can also be limited, interrupted, intermittent or even non-existent (Lyons and Little, 2009). This may be another factor that could hinder students' successful integration into mainstream education, in addition to the English language barrier. The situation is further complicated by the irregular arrival of students, who may enter post-primary education at different points in the school year. On a more positive note, however, post-primary ESL students appear to be motivated learners (Smyth et al., 2009: 151) with high educational aspirations (ibid.: 155) and positive attitudes towards their teachers (Devine et al., 2002; Vekic, 2003; Devine, 2005; Darmody, 2007).

The linguistic plurality of ESL students and the multi-faceted diversity it represents (e.g. in terms of ethnicity, culture, values, beliefs, religions, etc.) could be a source of enrichment for the student population because it offers access to different cultures and it also encourages a principled means of learning to respect difference. In other words, school diversity prepares students for the 'increasingly diverse Ireland, Europe and global society' (NCCA, 2006: ii). On these grounds, it should be safeguarded and celebrated within the school community. At the same time, however, this diversity creates major challenges for the Irish educational system for pedagogical, social, cultural, political, ideological, and economic reasons. Addressing students' needs and promoting their academic success is nevertheless essential as it benefits not only students (personal fulfilment, lifelong learning, employment, critical stance to public affairs, etc.), but also Irish society, and, by extension, Europe (active and responsible citizens, social cohesion, economic growth and progress, etc.).

1.2.2 Irish post-primary education: aims, structure, curriculum, examinations

Education in Ireland is divided into primary and post-primary levels and it is compulsory for students from five to 16 years old. The post-primary level comprises the Junior Cycle (equivalent to lower-secondary education) and the Senior Cycle (upper-secondary). There are three types of post-primary schools: a) voluntary secondary (394), b) vocational (247), and c) community/comprehensive (91). Their difference lies in management arrangements and funding and also in their student population (more middle-class students attend voluntary secondary schools; Hannan et al., 1996; Smyth et al., 2004).

The present discussion is concerned specifically with the Junior Cycle level which concerns students between the ages of 12 and 15. According to the website of the National Council for Curriculum and Assessment (NCCA; www.ncca.ie), the broad aims of Junior Cycle education are to:

- reinforce and further develop in the young person the knowledge, understanding, attitudes, skills and competencies acquired at primary level;
- extend and deepen the range and quality of the young person's educational experience in terms of knowledge, understanding, skills and competencies;
- develop the young person's personal and social confidence, initiative and competence through a broad, well-balanced, general education;
- prepare the young person for the requirements of further programmes of study, of employment or of life outside full time education;
- contribute to the moral and spiritual development of the young person and develop tolerance and respect for the values and beliefs of others;
- prepare the young person for the responsibilities of citizenship in the national context and in the context of the wider European and global communities.

The Junior Cycle curriculum is structured in a way that contributes to the above aims. It builds on the programme of study followed in primary education and its general aim is 'to provide students with a broad and balanced programme of study across a wide range of curriculum areas in order to prepare them for transition to Senior Cycle education' (www.ncca.ie). The curriculum comprises a wide variety of subjects to ensure the provision of a broad and balanced education that introduces students to all the areas of experience relevant to the needs of learners at this age and stage of their development. The 28 subjects it offers, listed in Table 1.1 below, result in a varied programme of study.

Table 1.1 Curriculum subjects of Junior Cycle education.

Irish	French	Technical Graphics
English	Spanish	Business Studies
Mathematics	Italian	Typewriting
History	Science	Environmental and Social Studies
Geography	Technology	Civic, Social and Political Education
Latin	Home Economies	Religious Education
Greek	Music	Physical education
Classical Studies	Art, Craft, Design	Social, Personal and Health Education
Hebrew Studies	Materials Technology (Wood)	
German	Metalwork	

It should be noted that all three types of school follow the same curriculum. Only the subjects of English, Irish and mathematics are compulsory. Details about the aims, objectives, course description (teaching units, topics, concepts etc.) and assessment of each subject are provided in the NCCA subject syllabuses and guidelines for subject teachers (www.ncca.ie). Students can study all subjects at either Ordinary or Higher Level while for Mathematics, Irish and English, a Foundation Level is also offered. As explained in subject syllabuses, this differentiation exists to meet the needs of all students (e.g., exceptionally able students, students with learning disabilities) and it is manifested in a) the content and learning outcomes (the depth of study and degree of abstraction), b) the process of teaching and learning, and c) the assessment arrangements associated with examinations.

As regards the assessment of student performance, continuous assessment takes place in many forms in the different curriculum subjects throughout the school year (details can be found in the NCCA subject syllabuses and guidelines for subject teachers) and school-based summative assessment takes place at the end of each school year. At the end of the third year of Junior Cycle, students are assessed in the curriculum subjects they studied in a State examination for the Junior Certificate which is designed and administered by the State Examinations Commission. The subjects which are excluded from the Junior Certificate Examination are Physical Education, Social, Personal and Health Education and Computer Studies, while Religious Education is available both as an examination and a non-examination subject.

ESL students follow the standard Junior Cycle curriculum and are subject to the same assessment procedures as their native peers. To fully access the curriculum, meet the requirements of the subject classroom and pass the Junior Certificate examinations, students need to develop the level of English language proficiency that is required in the formal context of schooling. The next two sections shed some light on the challenge of mastering the

language of school education for academic achievement, with an emphasis on the importance of the language of curriculum subjects.

1.2.3 Learning the language of school education as a second language

1.2.3.1 English language development for academic achievement

Accessing the curriculum and developing advanced academic literacy in post-primary education can be primarily viewed as a linguistic activity since 'developing knowledge and understanding in school subject areas and developing control of the linguistic resources that construct that knowledge and understanding are essentially the same thing' (Hasan, 1996). The correlation between the academic language and literacy development of all students in school education has been extensively documented in the relevant literature (e.g., Halliday, 1993; Halliday and Martin, 1993; Ravid and Tolchinsky, 2002; Schleppegrell and Colombi, 2002) and in school policy documents (e.g. the Bullock Report, 1975; the Swann Report, 1985, etc.). In the same vein, it has also been argued that educational failure can be attributed to language failure (cf. Bernstein's (1971) theory of 'restricted' and elaborated' codes and his sociological theory of pedagogical practices and the relationship between educational achievement, language and social structure (1973; 1977; 1990)).

On this evidence, the need for a 'visible pedagogy' (Bernstein, 1990) which focuses explicitly on learning and teaching the language of schooling has been consistently emphasised for native and non-native speaking students. A recent example can be found in the Languages of Schooling project launched by the Council of Europe's Language Policy Division in 2005 to highlight the importance of valuing the language varieties present in school education and acknowledge their diversity at a theoretical and practical level. The failure of educational systems to highlight the linguistic demands that are critical to students' academic success is also underlined in these projects: 'school itself is often responsible for this failure because the linguistic and communicative demands implied in subject learning are not made explicit enough to learners. They are part of a hidden agenda, a hidden curriculum' (Vollmer, 2009).

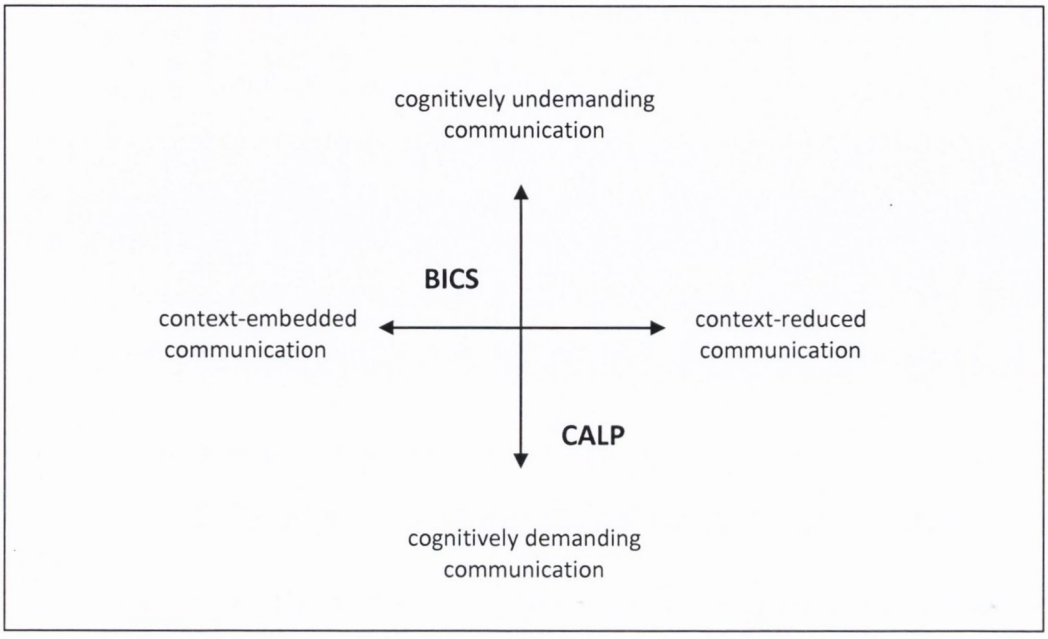
To effectively support ESL students' learning of the language of school education, various conceptualisations and theoretical models have been put forward aiming to provide an understanding of the nature and process of expected language development (e.g. Chamot and

O'Malley, 1987; 1994; Cummins and Swain, 1986; Snow, Met and Genesee, 1989; Cummins, 2000; Scarcella, 2003, etc.). Due to space limitations, only the theoretical framework proposed by Cummins is briefly considered here. This conceptualizes the development of the language of school as a second language (L2) in relation to academic achievement and can thus be used 'to estimate the language and cognitive demands of a variety of communicative situations in school' (Leung, 2003: 5).

According to Cummins, the type of language proficiency that is associated with school-related use of language can be conceptualised along two continua: the one referring to the range of contextual support that is offered for expressing or receiving meaning and the second relating to the amount of cognitive involvement required by the individual to carry out an activity (1984: 12-13; see Figure 1.1). Literacy activities taking place in formal education are, by definition, more context-reduced and cognitively demanding than face-to-face (classroom) communication because 'the message (meaning) is carried by a smaller *range* of cues' (ibid., emphasis in the original). In terms of English language proficiency in this context of language use, Cummins (1979, 1981, 2000) emphasizes the difference between certain salient rapidly developed aspects of communicative proficiency for face-to-face communication, i.e., basic interpersonal communication skills (BICS), and cognitive academic language proficiency (CALP), i.e. academic aspects of L2 development. He argues that it takes immigrant students approximately two to three years to develop BICS and five to seven years of L2 study to reach native-speaker levels of CALP (Cummins, 1981). These findings are supported by other studies (e.g., Collier, 1987; 1989) which underscore the complexity of language learning for immigrant students in mainstream education of the host country.

It should be pointed out, however, that valid as they may be, these findings do not necessarily apply to all educational contexts. For instance, it may take up to seven years for post-primary ESL learners to develop the same level of CALP as their native speaking peers but it may take considerably less time for ESL learners in primary education to achieve this because they have to compensate for a shorter period of 'loss' of educational experience. It should also be borne in mind that English language proficiency development for academic achievement is generally influenced by a number of factors and their potential interrelation; e.g. first language acquisition, second language acquisition, student age on arrival in the host country and age at the time of exposure to a second language, the age that schooling in the second language began, the length of residence in the host country, etc. (Collier, 1989). All these factors should

Figure 1.1 Range of contextual support and degree of cognitive involvement in communicative activities (Cummins, 1984: 12).



be taken into consideration in language support pedagogy; their analysis, however, falls beyond the scope of this thesis.

With specific regard to the Irish post-primary context, newly-arrived ESL students have to undergo three increasingly demanding language learning stages (IILT, 2003b) which take account of the distinction between BICS and CALP. As a starting point, students with minimum or total lack of English language proficiency need to acquire the basic linguistic resources which are necessary for everyday interaction and social inclusion in the school community (ibid.: 7). This language socialisation perspective promotes a view of students primarily as social agents and of the L2 as a tool for communicative integration in school life. This preliminary stage of core language learning is linked to BICS and must precede all other kinds of language learning but it may not be necessary for students who are already able to cope with basic everyday communicative needs in English.

Having addressed their immediate language needs for communication in the daily school routine, students should then start to develop school learning skills (e.g., using strategies for accessing text, developing grammatical accuracy, using learning supports and reference tools such as dictionaries, etc.; ibid.: 16). In other words, students in this phase, should begin to gradually develop literacy skills in the target language such as reading and writing through an engagement with activities and tasks that are typical of the subject classroom and which will most probably differ across subject areas (ibid.: 37).

The third, and perhaps most challenging, phase of language learning is crucial to students' academic achievement because it involves subject-specific language learning (ibid.: 20) and has implications for advanced literacy development. In this phase, students are expected to master an academic linguistic repertoire that will enable them to access and acquire curriculum knowledge, cope with content-based classroom tasks and activities and, ultimately, pass school examinations. All of these depend on the successful development of CALP. In other words, English language learning at this stage should be interwoven with the linguistic demands of the different curriculum subjects.

These stages of English language learning reflect a phased engagement of students with the mainstream subject classroom through a) core language learning, b) school learning skills and c) subject-specific language learning (IILT, 2003b). This learning does not take place in an additive or incremental fashion but rather in a holistic way and it involves an overlap between the three phases (particularly between school learning skills and subject-specific language learning) as well as recycling of language. To successfully support students' progression through the aforementioned phases, it is essential to help them master the particular linguistic resources required at each stage. What would be crucial to students' academic achievement, however, is mastery of the language that defines the different content areas of the curriculum. The distinctive characteristics of curriculum-specific language in general and its instrumental role in academic literacy development are discussed in the following section.

1.2.3.2 The language of curriculum subjects

The language of school education has been extensively discussed in the relevant literature under various labels: academic language, the language of education (Halliday, 1993), the language of school, the language of schooling, the language that reflects schooling (Schleppegrell, 2001), advanced literacy (Colombi and Schleppegrell, 2002), scientific language (Halliday and Martin, 1993) or more specifically, academic English (Scarcella, 2003; Bailey, 2007). Similarly, several definitions of academic language (e.g. Barnes, 1972: 47; Chamot and O'Malley, 1994: 40; Bailey, 2007: 10-11) have been formed. This section brings together some of the key characteristics of subject-specific language and their implications for ESL students' language learning.

What makes the language used within a school context distinctive is its role as a means of accessing, developing, demonstrating and evaluating educational knowledge. This formal

knowledge involves higher order thinking skills and the acquisition of scientific concepts (Vygotsky, 1986). In addition, educational knowledge:

- is a way of generalising about experience, of constructing meaningful sets of relationships between contexts
- gives us resources to go beyond the local and directly experienced – to learn concepts, to reason abstractly, to generalise, to predict, to hypothesise, to explain things which are counter-intuitive etc.
- is closely related to the world of written language and to public, institutional life
- is related to particular bodies of knowledge (scientific, legal, religious, humanities, bureaucratic etc.) and associated with 'professional' employment
- is built up consciously and rapidly
- is systematically presented, logically sequenced within a topic
- maintains disciplinary boundaries

(Based on Bernstein, 1990 and Painter's summary, 1999: 7)

To perform the above functions, educational knowledge at secondary level can only be constructed through advanced academic linguistic resources which differ significantly from the conversational language of primary education and the informal language used to mediate everyday knowledge (e.g., in terms of lexis, grammar, syntax, discourse, style, etc.; see also Bernstein's distinction between horizontal and vertical discourse (1999); Halliday's discussion of 'commonsense' and 'educational learning' (1988), and 'everyday' and 'school knowledge' (1994)). Furthermore, because school knowledge is linked to specific disciplines, the academic language used to construct and mediate this knowledge has, in essence, a subject-specific nature. Research into the nature of subject-specific language reveals the following characteristics:

- it is more specific, it is embedded into semantic fields and networks of concepts;
- it uses a far more formal register and style (e.g. 'reduce' instead of 'becoming less');
- it is more abstract or generalised in word choice: verbs, adverbs, collocations (e.g. a 'curve increases sharply' instead of 'goes up strongly');
- it is more precise and succinct (e.g. 'precipitation' instead of 'rain');
- it is more explicit and detailed ('from January till March the sales figures rise, whereas from April until September they stay even – at a high level');
- it is more cohesive (explicitly linking ideas, sentences and parts thereof);
- it is more rationally structured (concerning the logic of sequencing, arguing, evidencing);
- it is more coherent or goal-oriented in terms of the overall structuring of a discourse or text.

(Vollmer, 2007: 2)

These general features of subject-specific language make up what Halliday (1993) refers to as 'the syndrome of scientific discourse'.

Beyond these general characteristics, wide variation exists in language use across the different curriculum subjects in terms of i) content specific subject-matter, as mentioned above ii) grade level and iii) mode of language use (spoken, written, graphic). Of immediate interest to the present discussion are the English language varieties that are manifested across the Junior Cycle curriculum subjects in the written medium (the focus on written and not on oral language is explained in Chapter 2, section 2.2). The linguistic variation across these subjects can be demonstrated by the following three samples of written academic discourse taken from Junior Cycle textbooks for English, mathematics and science respectively:

> Short story writers are especially good at creating atmosphere. The setting, weather, landscape or imagery are important in getting the reader into the mood of the story.
>
> (Kelly, 2004: 74)

> In a right-angled triangle, the square on the hypotenuse is equal to the sum of the squares on the other two sides.
>
> (Humphrey, 2001: 319)

> A covalent bond is formed when atoms combine by sharing electrons so that each atom has a stable outer shell of electrons.
>
> (Cullen, 2003: 171)

The first extract is part of a discussion of writing short stories in English with specific reference to writer's 'mood'. The Pythagorean Theorem, stated in the second text, can be found in an algebra unit of a mathematics textbook, and chemical bonding is the topic of the third sentence, which was taken from the chemistry section of a science textbook. Although they comprise only one or two sentences, the above texts are revealing of characteristic differences across the three subjects, namely in thematic content and the specialised vocabulary items that mediate that content. Even these short sentences illustrate the fact that 'all subjects operate sub-languages which are encrusted with linguistic conventions' (Rosen, 1972: 120).

These different linguistic conventions emerge from the distinctive thematic concerns and the particular communicative functions that need to be fulfilled in each discipline (Conrad, 1996: 300), as well as the different kinds of argumentation each discipline favours. For instance, in science subjects students are asked to describe experiments and report results, while in history they are required to critically examine the causes and consequences of past events among other things. Accordingly, different genre types emerge; science favours reports while history favours historical recounts. Therefore, each subject area of schooling has its own

expectations and conventions which are realized through different linguistic means. It may also be added that differences exist within the sub-disciplines of a single discipline (e.g. differences among the science subjects of biology, chemistry and physics). As van Lier remarks: 'Every academic subject has its own registerial and discoursal variety that requires specialised forms of language use and understandings' (2002: 144), explaining that these 'do not flow automatically from being a proficient speaker' (ibid); an argument which has direct implications for native English-speaking students as well as for non-native speakers. What is important to note here is that academic language is necessary for achievement not only in school subjects that have been traditionally associated with language (e.g. literature, English, etc.) but in all other subjects (e.g., mathematics, science, etc.; Snow and Uccelli, 2009: 114).

All of the above characteristics and functions of curriculum language attest to its complex and challenging nature. The manifestation of diverse English language varieties across curriculum subjects, in particular, poses a double challenge to ESL students as they are required to 'simultaneously develop a form of double plurilingualism' both external, i.e. learning English as an L2, and internal, i.e. subject-specific academic language learning (Vollmer, 2007). Challenging though it may be, this double plurilingualism has a gate-keeping function for ESL students, as it is required for 'long-term success in public schools, completion of higher education and employment with opportunity for professional advancement and financial rewards' (Rumberger and Scarcella, 2000: 1).

The next section describes the Irish policy response to the need to help post-primary ESL students to develop this double plurilingualism if they are to be fully integrated into mainstream education.

1.3 English language support policy and some criticisms of current provision

To cater for the needs of ESL students in post-primary (and primary) schools, the integration pattern in Ireland involves their direct integration within mainstream classes with additional support in English language learning where appropriate (DES, 2009). English language support is provided on a withdrawal basis; i.e. students are withdrawn from their mainstream class (most often from Irish and Religion classes; Lyons and Little, 2009: 27) for usually one or two periods each week. Language support is provided for a maximum period of two years but it can be extended beyond the two-year period when necessary. A small number of schools with large proportions (over 20 %) of non-national students choose to run 'initiation/immersion'

classes in which students spend most of their time joining their mainstream peers for the least language-dependent subjects, in the weeks immediately following enrolment.

According to the Department of Education and Skills Circular 0015/2009 for post-primary language support (DES, 2009), all schools have to delineate their policy on the provision of language support to specify i) the arrangements for the identification of students requiring support, ii) the assessment of students' English language proficiency, iii) programme planning, iv) the recording and monitoring of students' progress and v) the roles of school personnel. Although language support teachers are appointed to deliver the language support programme, it is emphasized in the circular that 'The EAL [English as an Additional Language] pupil remains the responsibility of the [...] subject specialist teachers at post-primary level who will work closely with the EAL support teachers' (ibid.). This clearly calls for a whole-school approach to supporting ESL students' integration.

Research into the education of immigrant students in the Irish post-primary sector and the implementation of the language support policy is rather limited, providing little information at a national level. The existing studies, however, criticise the policy response and identify several deficits in its implementation in post-primary schools (e.g. Lyons and Little, 2009; Fionda, 2011). Three recurrent issues are consistently identified by teachers as key challenges: i) the lack of teacher education, including professional knowledge about linking ESL support pedagogy to the demands of the mainstream subject classroom, ii) the lack of appropriate pedagogical materials and resources, and iii) the lack of a collaborative whole-school approach to supporting ESL students' linguistic integration.

The lack of teacher training for post-primary ESL support and the need for a whole-school approach to continuous professional development systematically appear to be the top concerns of teachers and principals (Lyons and Little, 2009; OECD, 2009; Smyth et al., 2009). It is specifically reported that very few language support teachers have qualifications for or experience of language teaching in the given context (Smyth et al., 2009: 120) and even fewer subject teachers have received training in dealing with linguistically and culturally diverse students (OECD, 2009: 41). Some of the specific areas of language support in which teachers feel they need training include: i) ESL teaching methodologies (Lyons and Little, 2009: 65), ii) strategies to link ESL learning with curriculum content (ibid.); iii) the teaching of vocabulary (ibid.: 77), iv) speaking skills and written language for different curriculum subjects (ibid.), v) supporting day-to-day spoken English and academic English (ibid.). This absence of professional knowledge and skills has detrimental effects on teaching practice and students' learning alike and thus the need for training and professional development has been repeatedly emphasized (Lyons and Little, 2009: 64; OECD, 2009: 46; Smyth et al., 2009: 187).

The dissatisfaction with the lack of materials and resources tailored to the aims and needs of ESL support pedagogy is expressed by the vast majority of teachers (Lyons and Little, 2009; Smyth et al., 2009: 129). Lyons and Little report that commonly used materials in the language support class come from the website of Integrate Ireland Language and Training (see section 1.4), the internet, and books used for teaching English as a foreign language (2009: 37). They further add that teachers criticise many of the available materials as being inappropriate to ESL students' needs, age or interests (ibid.) and underscore the lack of language support materials aligned to curriculum subjects (ibid.: 75). The same study also reveals that only 5% of the teachers surveyed made explicit use of curriculum subject textbooks in language support (ibid.). This failure to exploit subject textbooks demonstrates teachers' disregard of the fact that it is the content-based language of subject textbooks that ESL students need to master. In short, the lack of context-specific materials and the use of inappropriate resources create a gap between language learning in the language support class and the linguistic requirements of the subject classroom.

Despite the fundamental importance of curriculum language for students' integration, socialisation, and academic performance, and the need to teach it explicitly in the language support class, in actual practice, English language instruction appears to be insufficiently or inappropriately linked to the specialized language demands of the post-primary curriculum (Lyons and Little, 2009; Smyth et al., 2009). More specifically, Lyons and Little (2009: 35) found that:

> 45% of the teachers surveyed did not consider that language support in their schools was linked to the mainstream curriculum that their newcomer students were required to follow. Further reflecting this discontinuity between language support and the curriculum, only 41% included the use of subject-specific keywords in their language support classes.

What is even more striking is the fact that:

> Remarkably, only 17% of the teachers surveyed explicitly linked their language support provision to the literacy aims of the Junior and Senior Cycle English syllabuses despite the fact that 43% of the language support teachers surveyed had English as their mainstream subject.
>
> (ibid.)

At the same time, as already mentioned, all teachers recognise language as being ESL students' greatest barrier to demonstrating and fulfilling academic potential and report a number of language difficulties linked to the subject classroom experienced by students; e.g.,

the reading level of written texts (Lyons and Little, 2009; Smyth et al., 2009), the complex subject-specific language (INTO, 1998; OECD, 2009: 25), the different acquisition of oral and written skills (Smyth et al., 2009: 78), the specialized vocabulary for curriculum subjects and grammar causing academic difficulties (ibid.: 79), the language, structure and readability of tests and examinations (Lyons and Little, 2009: 62). Despite the overwhelming importance of the specific language demands of the post-primary curriculum, these have not been researched to inform language support teachers so that they can help students overcome the aforementioned language difficulties. As a result, 'the absence of a more explicit and rigorously defined pedagogy often leaves the ESL teacher without a language explicit working agenda on the everyday level, and, indeed, the professional language to talk about their work' (Leung and Franson, 1997: 174).

In addition, contrary to official policy (DES, 2009), language support teachers are usually the ones who are charged with the full responsibility of supporting ESL students (Lyons and Little, 2009: 22). Most post-primary subject teachers fail to understand the role they have to play in supporting students' language development in the subject classroom and refuse to collaborate with ESL support teachers, even when the former have large numbers of ESL students in their class (ibid.: 40). The absence of a whole-school approach, coupled with the limited duration of language support, i.e. between 60 and 90 minutes of language support weekly (ibid.: 32), cannot adequately support students' language development and progress. What is even more disappointing is that many post-primary subject teachers seem to ignore ESL students in their classrooms, viewing them from a 'deficit perspective' (Devine, 2005: 53; Kearney, 2008; Lyons and Little, 2009: 70) and equating, in some cases, language support with special educational needs (Lyons and Little, 2009: 70).

The above are only few of the many challenges and deficits reported by research. Others include, for example, the lack of coordination of language support and dissatisfaction with the temporary contracts of language support teachers (Lyons and Little, 2009), racism and bullying (Smyth et al., 2004; Devine, Kenny and McNeela, 2008; Molcho et al., 2008), no support of mother tongue (INTO, 1998) etc. Overall, these findings show that current language support provision does not represent best practice. On the contrary, they provide evidence of an 'unsound, inconsistent and inadequately funded' policy (Lyons and Little, 2009: 75) which fails to provide the infrastructure required to successfully meet ESL students' needs. Based on the deficits and shortcomings revealed by their survey of 70 post-primary schools and the concerns expressed by 85 teachers, Lyons and Little (ibid.: 6) conclude that: 'While this [DES] policy grants newcomer students equality of presence, according to the research presented here it does not necessarily secure equality of participation or achievement'.

Careful consideration of the above deficiencies suggests that they may all be attributed to the lack of a clear understanding of the language learning demands imposed on ESL students by the Irish post-primary curriculum. It is, in other words, argued here that explicit information on curriculum language could i) inform teacher training and help ESL teachers establish multiple direct links between the language support and the subject class, ii) feed into the development of context-specific pedagogical materials, iii) raise subject teachers' awareness of the language of the subject areas they teach and facilitate collaboration between language support and subject teachers for a whole-school approach to supporting students' language development.

A noteworthy contribution to mapping the linguistic landscape of the Irish post-primary curriculum with a view to informing language support was made by IILT, a not-for-profit campus company of Trinity College Dublin, which provided exemplars of good pedagogical practice, teacher training and context-specific materials. For this reason, IILT forms an integral part of this discussion.

1.4 Integrate Ireland Language and Training

1.4.1 Introducing IILT

IILT played a leading role in the integration of immigrants (adults and students) through language learning in Ireland, providing also a role model for best practice at a European level. It was established in 1999 as the 'Refugee Language Support Unit' by DES in the context of a two-year pilot project under the auspices of the Centre for Language and Communication Studies of Trinity College Dublin. Its designated role was to co-ordinate the provision of English language support for adult refugees admitted to Ireland. Having successfully completed this pilot project, it officially became 'Integrate Ireland Language and Training Ltd', a not-for-profit campus company of Trinity College Dublin in 2001, and in due course 16 outreach centres were established throughout the country (IILT, 2007a: 1).

In order to implement its mission statement: 'through education and training, to empower people of other cultures and languages to achieve a place in Irish society' (ibid.), IILT promoted a learner-centred pedagogy with high levels of participant involvement and encouraged learner autonomy and lifelong learning. These principles were manifested in

language learning, teaching and assessment processes. A learner-centred curriculum was proposed for all classes, negotiated by learners based on the immediate needs of their life in Ireland. Classroom materials were created in-house in response to the needs of learners and were mapped to the Common European Framework of Reference for Languages (CEFR; Council of Europe, 2001).

IILT gained a reputation as a role model of excellent pedagogical and organisational practice in Ireland as well as at a European level. Receipt of various awards provides evidence for its important contribution to migrant education (e.g., the European Award for Languages - 'European Language Label', the AONTAS Star Award by the National Adult Learning Organisation).

1.4.2 IILT's contribution to language support in the post-primary school sector

IILT provided support to Irish post-primary, and primary, schools where non-English-speaking immigrant students were enrolled by developing a substantial bank of pedagogical materials and resources and providing nationwide induction seminars and on-going in-service training for school principals and language support teachers. More specifically, its action in this field aimed to achieve the following:

> a) To analyse the linguistic demands of the primary and post-primary curricula and identify the language needed by non-English-speaking non-national pupils in order to participate fully in the educational process.
> b) To develop materials to support the learning of English as a second language in schools.
> c) To present materials, methodology and supplementary aids via an ongoing in-service training programme for language support teachers.
>
> (IILT, 2007a: 30)

IILT's wide range of materials (available on NCCA's website; www.ncca.ie) became popular very quickly and were widely used by teachers. These include an immersion programme (2003b); a handbook for schools (2003c); English Language Proficiency Benchmarks (2003a); European Language Portfolio (2004); a resource book for language support teachers (2007b); the Post-primary Assessment Kit (2009), etc. All pedagogical resources were designed to address the specific learning needs of ESL students and are underpinned by the principles of modern language pedagogy mentioned above (section 1.4.1).

Among these resources, the curriculum framework for post-primary language support, comprising English Language Proficiency Benchmarks and a version of the European Language Portfolio, becomes immediately relevant to the concerns of the present discussion because it provides an ESL perspective on the post-primary curriculum and it represents the first systematic attempt to capture the communicative demands of the Irish post-primary curriculum. This curriculum framework is presented in the next section.

1.4.2.1 A curriculum framework for post-primary language support: English Language Proficiency Benchmarks and European Language Portfolio

1.4.2.1.1 English Language Proficiency Benchmarks: origins and description

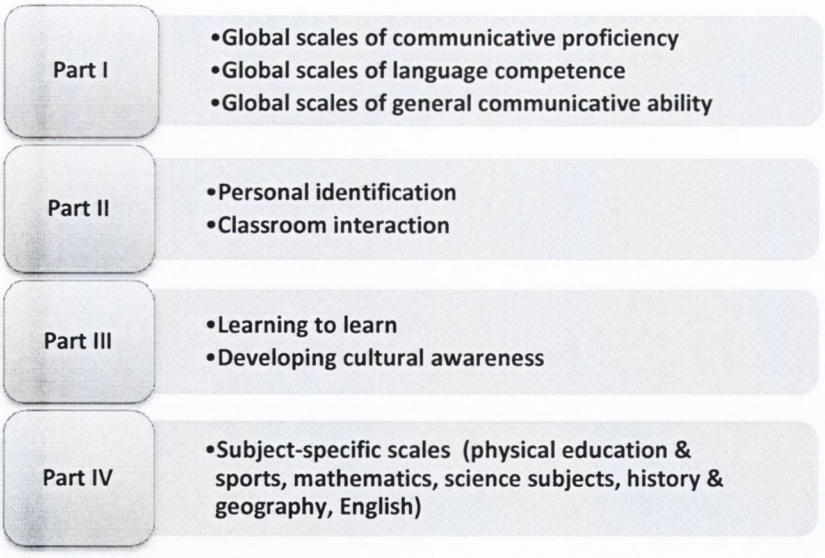
The English Language Proficiency Benchmarks (or Benchmarks; IILT, 2003a; see Appendix A1) represent the first and, so far, the only principled attempt to provide a context-specific conceptual framework of the target English language proficiency that ESL students need to develop to access post-primary education. Together with a version of the European Language Portfolio (or ELP; IILT, 2004; see Appendix A2) that mediates their ethos and content to students, the Benchmarks serve as the foundation of language support, facilitating appropriate English language learning, teaching and assessment.

The Benchmarks were derived from the CEFR (Council of Europe, 2001). The CEFR's description of language use (domains of language use, communicative language activities and strategies, and learner/user competences) was used to describe ESL students' target communicative repertoire, i.e. the English language knowledge, skills and competences which are directly rooted in the major topics, themes and communicative and linguistic demands of the post-primary curriculum. These are mediated through functional statements of the behavioural capacity ('can do' descriptors) that should be manifested in the performance of domain-specific tasks. The 'can do' statements are categorised into receptive (listening, reading) and productive (spoken production/interaction, writing) activities and encompass a description of language knowledge and the degree of skill in using it. Further, the descriptors follow the criteria of positiveness, definiteness, clarity, brevity and independence (North, 2000: 343-346; Schneider and Lenz, 2001: 47), while remaining simultaneously user-friendly (Council of Europe, 2001: 21), i.e., simple descriptions of language proficiency.

Learning outcomes in the Benchmarks are scaled according to the first three CEFR levels of foreign language (FL) proficiency: A1 (Breakthrough), A2 (Waystage) and B1 (Threshold). Differentiation across the three levels is manifested in the degree of ability to perform the particular tasks and the amount of external support offered by the interlocutor(s). It should be noted that these three levels 'specify the minimum proficiency required for full participation in mainstream schooling' (IILT, 2003a: 3) and do not cover the full spectrum of English language proficiency students may need to develop. It is thus acknowledged that language learning is a process with no identifiable endpoint and, to this end, the need to foster students' capacity for autonomous learning should be a principal aim of language support.

As can be seen in Figure 1.2, the Benchmarks comprise four components; progression of English language proficiency is described in respect of: i) general communicative ability and language competence, ii) personal identification and classroom interaction, iii) learning to learn and developing cultural awareness, and iv) the communicative demands of postprimary curriculum subjects (the full contents can be viewed in Appendix A1).

Figure 1.2 The components of the English Language Proficiency Benchmarks for post-primary ESL students (IILT, 2003a).



The first part of the Benchmarks offers an introduction to the first three levels of English language proficiency in a summary form. This part comprises three scales: i) global scales of communicative proficiency, ii) global scales of linguistic competence, and iii) global scales of general communicative ability. First, the global Benchmarks of communicative proficiency

provide a compact description of the communicative behaviour expected of learners at the three proficiency levels. The language development that is sketched in these behavioural descriptions resembles the developmental paths that are normally followed by native speakers of English (ibid.: 5). This scale is essentially 'a domain-specific interpretation of the first three levels of the self-assessment grid' (ibid.: 3) from the CEFR (Council of Europe, 2001). Second, the global scales of linguistic competence specify progression in the range and control of vocabulary, grammatical accuracy, phonological and orthographical control. Thirdly, the global scales of general communicative ability specify progression in spoken fluency, planning, compensating, monitoring and repair, flexibility, and turn-taking.

The second part of the Benchmarks describes the communicative language functions students need to be able to perform in order to demonstrate their ability to i) give an account of themselves and ii) participate appropriately in classroom interaction (IILT, 2003a: 10). Helping students develop the linguistic resources necessary for their personal identification and successful communication in the classroom is clearly 'the point of departure for language support but also its most important guiding threads' (ibid.: 3) since students' ability to talk about themselves and to participate in everyday classroom interaction underpins the development of communicative skills required by all subjects. Developing the ability for personal identification in particular is essential for students' socialization with peers and for accessing the life of the school community.

The third part of the Benchmarks is oriented both to linguistic and educational development. The focus is on two abilities which are highly important for students' educational progress: learning how to learn and developing cultural awareness. Descriptors for the ability of learning how to learn express skills divided into four categories: i) identifying learning activities, ii) setting learning objectives, iii) organizing learning, and iv) monitoring learning progress (ibid.: 14). The skills involved in the ability of learning-to-learn support students' goal-setting and self-assessment skills (ibid.), both of which are essential prerequisites for developing language learner autonomy (cf. Little, 1991; 2007).

The development of cultural awareness is linked to four areas in the Benchmarks: i) patterns of daily life, ii) cultural expectations, iii) values and attitudes, and iv) cultural events, and relevant language-based tasks are summarised across the three proficiency levels (IILT, 2003a: 13-14). Helping students to develop a clear awareness of their own cultural identity and to discuss and express opinions about cultural issues becomes increasingly important in the context of migrant education. This is because culture forms an integral part of students' linguistic identity and also because cultural awareness 'is highly important for some subject learning' (IILT, 2007b: 37) and thus its development should be promoted for all students by the

mainstream curriculum (cf. the NCCA (2006) guidelines for incorporating intercultural education in the post-primary curriculum).

The final component of the Benchmarks is crucial to supporting ESL students' linguistic access to post-primary curriculum subjects as it includes subject-specific scales of English language proficiency. These reflect the communicative tasks students encounter in the different subject areas of the post-primary curriculum, namely, i) physical education and sports, ii) mathematics, iii) science subjects, iv) history and geography, and v) English (IILT, 2003a: 15-20). Like all descriptors included in the Benchmarks, those in the subject-specific scales were developed based on a judicious combination of commonsense intuition and familiarity with the typical topics, themes and practices of the different curriculum subjects. Similarities in these areas between certain subjects resulted in assigning the same sets of descriptors for subject clusters, i.e., science subjects (chemistry, physics, biology) and history and geography.

1.4.2.1.2 Supporting ESL learning using the Benchmarks

The Benchmarks were designed with specific reference to the developmental phases of ESL learning that students are expected to follow in their engagement with the mainstream class (cf. section 1.2.3.1). The initial phase of core language learning is supported by the scales of personal identification and classroom interaction; areas which should be the priority for newly admitted students with minimum or very limited English language proficiency. The development of school learning skills is facilitated by the modules of learning to learn and developing cultural awareness. The phase of subject-specific language learning is supported by the subject-specific scales of the Benchmarks. It can also be observed that the ordering of the subject areas reflects a degree of increasing difficulty in the language demands imposed by the different subject areas, with physical education located at the least demanding end of the continuum and English at the most demanding end. Finally, the first part of the Benchmarks, i.e. the global scales, serves as a useful summary guide throughout the process of language support.

The distinction between the constructs of BICS and CALP and their interrelation (i.e. the dependence of CALP on BICS) is also reflected in the Benchmarks as follows:

> The first part presents an integrated view of communicative proficiency at levels A1, A2 and B1; the second part has to do with learners' ability to give an account of themselves and engage in general interaction (BICS and the beginnings of CALP); the third part is concerned with the development of learning skills and cultural awareness (CALP firmly embedded in BICS); and the fourth part is concerned with subject-specific communication (CALP).
>
> (Lyons and Little, 2009: 83)

The Benchmarks imply non-sequential language learning, acknowledging also the different learning paths and pace among students, their English language proficiency level and previous educational experience: 'the individual students' progress through the Benchmarks will rarely be entirely predictable and consistent; some skills may develop more quickly than others, and at any particular level some of the tasks articulated in the Benchmarks may prove to be more difficult than others' (IILT, 2003a: 4). In every case, however, achieving the subject-specific language learning outcomes which are described in the subject scales requires a significant period of time for all students.

To mediate the Benchmarks to ESL students and to encourage them to become actively involved in the pedagogical process and keep a record of their learning progress, an ELP model was developed; this is briefly presented in the following section.

1.4.2.1.3 Mediating the Benchmarks through the European Language Portfolio

The Irish ELP for post-primary ESL students (IILT, 2004) was primarily designed 'to support students whose mother tongue is not English as they meet the challenge of learning English in order to participate fully in the mainstream classroom' (IILT, 2007b: 523). To this end, the ELP reflects the English language demands of the Irish post-primary curriculum (IILT, 2004: 32) and mediates the Benchmarks to ESL students. Effective ELP use in the language support class can serve different functions with many pedagogical benefits for students. The present discussion, however, is confined to explaining its role as a mediation tool of the Benchmarks, following a brief presentation of its structure (for background information on the ELP see Little and Perclová, 2001; Little, 2002, for the pedagogical value of ELP use for immigrant students see Little and Lazenby-Simpson, 2004; Kostopoulou, 2009).

The structure of the Irish post-primary ELP for immigrant students follows the Council of Europe's designated three-part ELP structure comprising a language passport, a language biography, and a dossier, with some adaptations tailored to the aims of ESL support (the full

ELP contents can be viewed in Appendix A2). In the language passport, students are asked to draw their overall language profile reflecting on their skills in all languages they know, including their mother tongue. This linguistic overview is based on the CEFR's self-assessment grid and the scales of the global Benchmarks of communicative proficiency (IILT, 2004: 4-9).

The language biography occupies the largest part of the ELP (ibid.: 10-28), reflecting its important pedagogical role as the link between the passport and the dossier: 'From a pedagogical point of view the language biography plays a pivotal role, providing a focus for the reflective processes that mediate between the language passport and the dossier' (Little and Perclová, 2001: 2). The biography is divided intosections which focus on i) students' intercultural experiences and intercultural awareness gained from these (cf. Little and Lazenby-Simpson, 2003), ii) their English language learning linked to post-primary curriculum school subjects, and iii) the self-managing of their learning process (goal-setting, reflection on and recording of learning). This second ELP component is of particular relevance to the concerns of this discussion.

The increasing value of the ELP biography stems fromthe series of self-assessment checklists of 'I can' descriptors. These were derived from the Benchmarks and are categorised as follows (IILT, 2004: 22-28): i) personal identification, ii) class interaction, iii) physical education, iv) mathematics and science, v) history and geography, and vi) English. Like the Benchmarks scales, these checklists are related to the five language skills of listening, reading, spoken interaction, spoken production, and writing, and they are scaled according to the first three levels of A1, A2 and B1. Taken together, these thematically-based checklists make up a task-based description of the target linguistic and communicative repertoire required for full participation in the Irish post-primary curriculum.ESL students can use the 'I can' statements for goal-setting, reflection, planning, monitoring and assessing learning outcomes (for a detailed discussion of the cyclical process of goal-setting, learning and assessment stimulated by the ELP in school practice, see Little and Perclovà, 2001). The biography checklists become further useful to teachers as they can be used to inform teaching objectives and assessment criteria and to suggest communicative tasks and language activities.

Finally, the ELP dossier allows students to demonstrate their language learning progress and achievements by collecting samples of their work. Materials inserted in this section should be carefully selected in order to provide concrete evidence that supports attestations of learning outcomes made in the language biography and passport. This is also one way to validate subjective judgements which emerge from learner self-assessment in the passport and biography.

The ELP structure clearly shows its close interaction with the Benchmarks. More specifically, the overall scope of English language development conveyed through the global scales of communicative proficiency in the Benchmarks (IILT, 2003a: 6-7) is mediated to students through the self-assessment grid and the same global Benchmarks of communicative proficiency also included in the ELP passport (IILT, 2004: 5-8). The module of learning to learn in the Benchmarks (IILT, 2003a: 14) is manifested in the ELP biography in the section which includes tips on monitoring learning and recording successful language learning techniques (IILT, 2004: 16-18). Developing cultural awareness - which appears in the same part with learning to learn in the Benchmarks (IILT, 2003a: 14) - is also present in the ELP biography, in the sections where students record their intercultural experiences and cultural awareness (IILT, 2004: 11-13). The Benchmarks modules of personal identification, classroom interaction (IILT, 2003a: 11-12) and the subject-specific modules (ibid.: 15-20) are directly reflected in the 'I can' checklists in the ELP biography (IILT, 2004: 22-28). This overlap of language learning content of the Benchmarks and ELP results in a fruitful interface between ESL teaching and learning sinceinstructional objectives are translated into learners' goals and both are oriented to the same target language learning outcomes.

By mediating the content of the Benchmarks, the ELP supports essentially the same three phases of English language development of core language learning, school learning skills and subject-specific language learning (cf. section 1.2.3.1). The BICS and CALP dimensions are also incorporated across the three CEFR levels of A1, A2, B1 in the thematic 'I can' checklists of the language biography. Thus, for instance, 'I can talk with my school friends in a natural way about the sports we play together' (IILT, 2004: 24) is a descriptor of spoken interaction at B1 level linked to physical education which depends on the development of BICS, whereas a statement describing a cognitively-demanding and context-reduced task of spoken production at B1 level such as 'I can give a clear explanation of a mathematical problem' (ibid.: 25) is clearly indicative of CALP.

Taken together, the Benchmarks and ELP mediate to language support teachers and ESL students respectively the communicative repertoire that is required for full participation in the Irish post-primary curriculum and, as such, they can be used in the language support class in many different ways.

1.4.2.1.4 The multi-functionality of the Benchmarks

The Benchmarks have multiple applications in the ESL classroom. They are, first of all, an indispensable tool for planning language instruction. As a three-step scaled curriculum, mapping students' language learning progress from entry to language support to the point of full integration into the mainstream classroom, they offer teachers an overview of expected learning outcomes across A1, A2 and B1 levels. The modules of the Benchmarks help teachers set short-term and long-term teaching objectives using the 'can do' descriptors. Teachers can further define their task on a daily basis by analysing and breaking down the descriptors.

By mediating communicative tasks that students encounter in the subject classroom, specifying also the degree of skill, the 'can do' descriptors inform materials design. Teachers can use the descriptors to plan and design a sequence of graded language activities and tasks which encourage the use of the target language for the expansion of students' proficiency. In this way, the pedagogical activities and materials introduced in the ESL support class encourage authentic communication and language use, drawing students into the reality of the mainstream classroom.

In combination with other pedagogical instruments developed by IILT, the Benchmarks can be also used for assessment purposes. They can facilitate diagnostic assessment for the placement of students newly admitted to school in combination with observation of student classroom performance and the Post-primary Assessment Kit (IILT, 2009; i.e., a suite of CEFR-based language tests administered for placement purposesand for the measurement of proficiency, especially at the end of a student's two-year allocation of language support). They can also be used to analyse individual students' learning needs and for continuous assessment purposes (i.e. monitoring students' progress on a regular basis).

Underpinned as it is by the Benchmarks, the ELP can be used for all of the above purposes, taking account of individual student needs. In other words, it can be used to set learning goals and introduce tasks which reflect the target communicative repertoire that is described in the language biography. It could further facilitate subjective needs analysis and learner self-assessment, which can be combined with objective needs analysis and teacher assessment. This combination of teacher and learner perspectives should yield more reliable and complete specifications and evaluations of language learning.

1.5 The need for further development

The discussion has so far examined the provision of English language support to ESL students in Irish post-primary schools, outlining some of the existing deficits and needs based on the testimony of teachers. It was argued that an explicit understanding of curriculum language is urgently needed to promote a principled and rigorously defined approach to language support and address some of the challenges facing current provision. The present curriculum framework of language support was presented, i.e. IILT's Benchmarks (2003a) and ELP (2004), as this constitutes the first systematic attempt to articulate the demands that are imposed on ESL students by the post-primary curriculum.

It was explained, however, that the Benchmarks and ELP, or any other materials currently available, do not provide detailed linguistic descriptions with subject-specific substance. Although the subject-specific scales of the Benchmarks mediate the communicative demands of curriculum subjects, they do not capture the actual linguistic repertoire that ESL students need to develop in the different subject areas. For instance, the Benchmarks descriptors for students' reading skills at B1 level in science subjects state that a student: 'Can understand textbook *explanations* and *examples* provided that new *words* and *concepts* have been prepared in advance' (IILT, 2003a: 18, emphasis added), or that he/she 'Can follow written *directions* for simple *observational* or *experimental procedures* that he/she has not encountered before' (ibid., emphasis added). These 'can do' descriptors inform teachers about the communicative and language-based tasks that are typical to the science classroom but they do not specify the language that students are expected to use to express these explanations, examples, key words, concepts, directions and observational or experimental procedures in science subjects.This is because the Benchmarks were designed as 'a general map of the ground to be covered' (ibid.: 2) in language support, defining the external context of target language use in a holistic manner (i.e., in terms of domain-specific communicative themes, tasks and activities). Describing as they do the functional development that is expected of ESL students, they provide only a broad understanding of the language of curriculum subjects which is based on commonsense intuition.

Considering the overwhelming importance of an explicit focus on subject-specific language learning in the language support class, there appears to be a need for detailed and empirically-based language specifications tied to curriculum subjects. Analyzing the lexis of the target language domain is a further step that may be taken in objective needs analysis to obtain a detailed picture of students' precise language learning needs, in addition to analysing

the external context of language use in terms of target domains, topics, tasks etc. (Council of Europe, 2001: 44-100). A close scrutiny of the post-primary curriculum could provide a concrete description of the target linguistic spectrum which, coupled with the broad map of the terrain sketched by the Benchmarks, could more effectively support language learning embedded in curriculum subjects.

The need to formulate an empirically-based understanding of curriculum language and to continue and further develop IILT's contribution to the provision of post-primary language support, following its closure in 2008, acted as the springboard for the English Language Support Programme and the present research.

1.6 The ELSP and the aims of this research

Acknowledging immigration as one of the major challenges faced by Irish society, Trinity College Dublin launched the Trinity Immigration Initiative (TII) in 2007 (www.tcd.ie/immigration). The TII was a three-year Research Programme on Diversity, Integration and Policy (2007-2010) which brought together six interacting multidisciplinary projects (Children, Youth and Community Relations; English Language Support Programme; Migrant Careers and Aspirations; Migrant Networks – Facilitating Migrant Integration; National Survey of Immigrants in Ireland; National Policy Impacts) with the double mission a) to address key challenges posed by the unprecedented numbers of migrants arriving in Ireland in recent years, and b) to help Irish society develop appropriate policies and practices for the new reality. The research and actions of each of the six projects contributed in their own ways to the core themes of the TII, 'Diversity, Integration and Policy'. The present research study was conceived within the English Language Support Programme (ELSP) for post-primary schools.

Acknowledging that language is the key to integration in the school community and Irish society, and taking the existing deficiencies and needs of post-primary language support provision as its starting point, the ELSP aimed to address some of the urgent issues in this context. Building on the work of IILT, the general aims of the project were:

- To work with teachers in a number of post-primary schools to develop approaches to English language support that can be generalized throughout the sector.
- To support this activity by empirical research that identifies dominant patterns of language acquisition and recurrent problems for language pedagogy.

- To investigate at a more general level the linguistic challenges that surround the longer-term integration of non-English-speaking students in mainstream Irish education.

In line with the ELSP's objective to investigate the linguistic challenges faced by post-primary ESL students, the present research aims to provide concrete empirically-based information on the language of the Junior Cycle curriculum, as this is manifested in commonly used subject textbooks and Junior Certificate examination papers. The research focuses specifically on the linguistic analysis of six curriculum subjects, namely, English, geography, history, CSPE (civic, social and political education), mathematics and science, and it seeks to address the following questions:

1. How is the content of each curriculum subject 'languaged' in Junior Cycle textbooks and Junior Certificate examination papers?
2. Are there any linguistic similarities among the six subjects?
3. How can this information on subject-specific language be exploited in the post-primary language support classroom?

By addressing the above questions, this research aims to provide i) distinct linguistic profiles of the above six subjects which will reflect their typical characteristics, as these are enacted in textbooks and examinations, and ii) a set of language features which cut across content areas. On the basis of these findings, pedagogical implications and possible practical applications will be explored in relation to ESL learning, teaching, the language support curriculum framework and assessment.

1.7 Conclusion

This introductory chapter established the pedagogical context and the aims of this research. Following a brief description of ESL students' profile and Irish post-primary education, the provision of language support was critically examined. It was argued that DES policy towards ESL students' educational integration does not adequately respond to the gravity of the situation, causing several shortcomings and deficits in current language support provision. The lack of in-depth and empirically-based research into the linguistic demands of post-primary education was underlined as a serious limitation with detrimental effects on the quality of language support practices. Mediating as it does the communicative and linguistic demands of

the post-primary curriculum, IILT's curriculum framework for language support (IILT, 2003a; 2004) was presented as a useful contribution to this end.

It was nevertheless argued that curriculum language warrants a very great deal more investigation than it has received to date in order to acquire more concrete and detailed information about the linguistic requirements of the Irish post-primary curriculum. This information would provide a valuable foundation for a better-informed and more effective language support pedagogy. In response to this need, the aims and scope of this study were presented, embedded within the ELSP of the TII. The discussion now proceeds to the methodological procedure and tools that were adopted to achieve the aims of this research.

2 A corpus linguistic analysis of Junior Cycle textbooks and Junior Certificate examination papers

2.1 Introduction

The aim of this chapter is to justify and describe the research methodology that was used to obtain empirical and detailed information on the language of Junior Cycle textbooks and Junior Certificate examination papers with a view to informing language support pedagogy.

The discussion is divided into three main parts. First, the decisions involved in the design of the study are justified; It is explained why the language analysis i) focuses on six curriculum subjects (English, geography, history, CSPE, i.e. civic social and political education, mathematics, and science), ii) specifically at Junior Cycle level, and iii) why it is confined to the written academic discourse of subject textbooks and Junior Certificate examination papers. Then, the rationale for a corpus linguistics methodology is provided for the envisaged linguistic analysis based on a critical examination of possible methodological approaches. A brief description of the successive stages of the methodological procedure follows and, finally, background information is provided on: i) the design rationale and milestones in corpus development and ii) the lexical analysis software used in corpus analysis (WordSmith tools 4; Scott, 2004). The chapter concludes by summarising the key points of the discussion.

2.2 Rationale for analysing the language of textbooks and examination papers of six Junior Cycle curriculum subjects

In has been so far made explicit in this discussion that ESL students' access to Irish post-primary education and advanced academic literacy development largely depend on successful linguistic literacy development (e.g. Ravid and Tolchinsky, 2002). This entails gaining control of the linguistic variation that exists across curriculum subjects which can be pedagogically facilitated through an explicit focus on the linguistic features that characterise the different subject areas within a content-specific communicative context. To draw a complete picture of the full communicative and linguistic requirements of Irish post-primary education the ideal

would be to examine all types of students' exposure to academic language in the mainstream post-primary classroom; that is, both through a) spoken academic discourse (manifested in teacher instruction, oral classroom activities and tasks, etc.), and b) written academic discourse (enacted in subject textbooks, workbooks, tests and examination papers, etc.). Furthermore, to be comprehensive, this analysis would also have to encompass all six grade levels of post-primary education, i.e., covering both Junior and Senior Cycles (cf. section 1.2.2). The required time framework and resources for such an extensive and large-scale research, however, extend far beyond the possibilities of the present research.

Within its available resources, this research aims to make a contribution to the above ambitious goal by undertaking the first empirical examination of the language of six core subjects of the Junior Cycle curriculum (English, geography, history, CSPE, mathematics, science), as these are manifested in 24 commonly used subject textbooks and in the most recent Junior Certificate examination papers (of the last decade 2001-2010).

It seemed reasonable to begin with an examination of the Junior Cycle curriculum as a starting point as this could be subsequently extended to the Senior Cycle curriculum by future research. Preference for Junior Cycle subjects was also based on the fact that this level of education is compulsory for all students as opposed to the Senior Cycle which some students may not attend (if they decide to attend Vocational schools or to leave school). The above six subjects were selected for analysis, out of the 28 of Junior Cycle curriculum, because they are considered fundamental to post-primary education, they are formally assessed through terminal examinations, and they are also the ones typically most valued throughout adult life.

Although the importance of oral language input (received through teacher instruction and classroom discourse) is recognized, priority was given to analyzing the written language of subjects textbooks and Junior Certificate examination papers because of the variability of teacher discourse style (Bailey, 2007: 148). This would most probably yield inconsistent information and conclusions which could not be generalised. Working with focus groups of subject teachers as informants to arrive at conclusions about the linguistic dimensions of curriculum subjects was also avoided because, as previously mentioned (Chapter 1, section 1.3), teachers appear to be unaware of the language of the subject they teach and they tend to focus exclusively on content knowledge. On these grounds, teachers' expectations and views about the language demands of the subjects they teach tend to be impressionistic and thus not highly reliable.

Subject textbooks and examination papers, on the other hand, were selected as more consistent and robust sources of linguistic information. Both have a prominent role in the post-primary school life and the academic language they contain has a gate-keeping function for ESL

students' access to knowledge and educational performance. Subject textbooks in particular are 'omnipresent in all kinds of literacy activities' taking place in the content classroom while simultaneously 'interfacing with spoken language in the same mental space of the language user' (Ravid and Tolchinsky, 2002: 230). More specifically, textbooks:

- are the primary source of disciplinary knowledge
- underpin teacher talk in the subject classroom
- inform the spoken production and spoken interaction required of students
- contain written language essential forstudents' academic literacy
- model the academic language standards which define assessment criteria (in tests and examinations)

Subject textbooks are, in short, the basis of classroom learning, teaching and assessment and thus accessing their language is fundamental for students' academic achievement. The state examinations of Junior Certificate, on the other hand, have admittedly a gate-keeping function for students because exam performance defines students' academic profile and determines the path they will follow subsequently. Examinations-specific language is highly critical because ESL students need to successfully decode and comprehend the language of instructions in exam papers in order to demonstrate their academic potential.

In addition to the widely recognised value of textbooks and examinations in post-primary education, the difficulties that ESL students and language support teachers face with these materials further motivate their analysis. As previously explained (Chapter 1, section 1.3), teachers report ESL students' difficulties with the complex subject-specific language (INTO, 1998; OECD, 2009: 25) and the specialised vocabulary of curriculum subjects (Smyth et al., 2009: 79), as well as their own need for strategies to teach written language for different curriculum subjects (Lyons and Little, 2009: 77). Additionally, exam-specific language, and end-of-year examinations and public examinations (Junior Certificate) in general, are considered 'significantly stressful challenges for teacher and student alike' (ibid.: 62), while language support teachers express the need to learn how to help students cope with the language, structure and format of examinations and use appropriate strategies (ibid.). On these grounds, the linguistic analysis of both subject textbooks and examination papers would be of the highest pedagogic value as it could make explicit their academic language demands so that these can be subsequently 'focused upon through conscious didactic effort and support measures' (Vollmer, 2009: 9) in the language support classroom. For all of the above reasons, the present research focuses on the examination of the above two school registers.

Prior to the linguistic analysis of textbooks and examination papers, subject syllabuses and teacher guidelines for the six curriculum subjects mentioned above were studied to gain insights into curriculum subject aims, content and practice and the way patterns of language use are positioned in these official documents (official in the sense that they are expected to be used by all post-primary teachers and they therefore reflect official expectations of student communicative skills and language use). Information from these sources was considered relevant to this study because it can provide 'the communicative background to the types of discourse in which ... [linguistic] patterns occur' (Hunston, 2002: 99). Accordingly, this information proved to be useful for the interpretation of some of the linguistic findings.

The linguistic analysis of Junior Cycle subject textbooks and Junior Certificate examination papers could be conducted within different theoretical frameworks and by using several methodologies which would yield different types of insights into subject-specific language. The discussion that follows provides a critical overview of possible analytical frameworks and explains why a corpus linguistics methodology was considered the most appropriate for the purposes and needs of this research.

2.3 The need for a corpus-based methodology

2.3.1 Analytical frameworks and methodological approaches to written academic discourse analysis: an overview

In methodological terms, the linguistic analysis of subject textbooks and examination papers of the six Junior Cycle curriculum subjects envisaged in this research falls into the paradigm of applied discourse analysis. Both 'discourse' and 'discourse analysis' can be interpreted differently, depending on the theoretical prism through which they are viewed. Discourse refers here to 'any instance of the use of written language to communicate meaning in a particular context' (Bhatia, 2004: 19) and discourse analysis can be broadly understood as 'an umbrella-term for all issues that have been dealt with in the linguistic study of text and discourse' (Östman and Virtanen, 1995). Applied discourse studies are those which are motivated by concern with language teaching, e.g. register analysis, etc. (Celce-Murcia, 2000: 5).

If the present research is classified as an applied discourse study, ESL students' linguistic access to the Junior Cycle subject classroom can be viewed as acquiring membership of the different subject-specific discourse communities which have shared goals and values and particular ways of communication (Paltridge, 2006: 24). To effectively participate in these communities, students need to have the appropriate linguistic repertoires and a number of language varieties the choice of which is, in every instance, determined by the domain in which the language variety is used (ibid.: 29). To gain a clear understanding of these language varieties required by each subject discourse community, and specifically by its written academic medium, the discourse analysis of subject textbooks and examination papers would be valuable.

In discussing the concerns of discourse analysis, Bhatia explains that:

> As a general term, discourse analysis therefore can focus on lexicogrammatical and other textual properties, on regularities of organization of language use, on situated language use in institutional, professional or organizational contexts, or on language use in a variety of broadly configured social contexts, often highlighting social relations and identities, power asymmetry and social struggle.
>
> (2004: 3)

The above definition names three distinct foci of discourse analysis: i) lexico-grammatical and other textual properties, ii) organizational patterns of language use, and iii) social dimensions of language use, e.g. social relations, identities, power asymmetry and social struggle. These foci reflect in fact the systematic development of discourse analysis which has extended its scope from an initial examination of discourse at a textual level, i.e. discourse as text, to a more enlarged view of discourse as genre, encompassing structure and organization, and finally to discourse perceived from the wider socio-cultural context of use. Accordingly, the analysis of curriculum subject discourse could follow three different directions, depending on the research questions posed and the focus of analysis that is selected.

Focusing on the textual level of subject specific discourse, a possible analysis might seek to investigate the lexico-grammar of subject texts and the functional values its features represent. This type of study would be in line with the first tradition of discourse analysis which was largely influenced by formal, or structural, linguistics, text linguistics and language variation and register studies (Barber, 1962; Halliday, McIntosh and Strevens, 1964). Originally, discourse studies focused on 'statistically significant features of lexico-grammar used in a particular subset of texts associated with a particular discipline' (Bhatia, 2004: 4). Due to the absence of computational analytical procedures, however, they were limited to only a small

set of significant features, rather than a complete analysis of the corpus, and mainly within clause boundaries without consideration of discourse organization (ibid.: 5). In the present context, this type of text analysis would reveal the formal and functional aspects of the discourse of individual subjects, that is, phonological, lexico-grammatical and semantic elements without taking account, however, context in a broad sense.

A different possibility would be to examine the structure and patterns of organization of information or general organizational patterns that are favoured by subject specific discourses. This type of study would, in other words, examine discourse as genre, i.e. 'a recognizable communicative event characterized by a set of communicative purpose(s) identified and mutually understood by the members of the professional or academic community in which it regularly occurs' (Bhatia, 1993: 13). Genre theory has been the focus of interest of many different scholarly traditions in relation to second language instruction, notably the British School of English for Specific Purposes (ESP), American Genre Studies and Australian Systemic Functional Linguistics (SFL).

In the British ESP School represented by Swales (1981, 1990) and Bhatia (1993), genre has been used as a tool for analyzing the spoken and written language that is required of non-native speakers. This entails identifying text types in terms of their formal properties and communicative purposes. Models of genre analysis for ESP have focused on predictable formulaic aspects of certain genres through structural move analyses to describe global organizational patterns in genres (e.g. Swales, 1981; 1990) and genre staging (initially inspired by Halliday and Hasan, 1980). The American New Rhetoric and Genre studies, as represented by the works of Miller (1984, 1994), Bazerman (1994) and Berkenkotter and Huckin (1995), were focused more on 'the social purposes or *actions*, that [...] genres fulfill' within specific situations (Hyon, 1996: 696, emphasis in the original; e.g. Miller, 1984). The Australian genre theories and SFL theory elaborated by Halliday (1994) and further developed by others (Martin, Cristie and Rothery, 1987; Martin, 1992; Matthiessen, 1995 and Hasan, 1996) were centred specifically in primary and secondary education contexts and later in adult education for immigrants (the Literacy and Education Research Network, cf. Hyon, 1996). It is worth elaborating on the SFL approach to linguistic description at this point, as it has been extensively used for the analysis of school subject language from a functional perspective, giving rise to some interesting insights (e.g., Lemke, 1990; Halliday and Martin, 1993; Hasan and Williams, 1996; Christie and Misson, 1998; Martin and Veel, 1998; Christie, 1999, 2002; Schleppegrell, 2001; 2004 etc.).

SFL theory is concerned with describing how and why language varies in relation to groups of users and social context (Halliday and Hasan, 1989). Focusing on the functional role

of language, that is enacted primarily through grammar (cf. functional grammar, Halliday, 1994), SFL theory facilitates the analysis of 'whole texts in ways that clarify the relationship between language and context and highlight the role of social experience in the linguistic choices made by speakers and writers' (Schleppegrell, 2004: 19). These concerns of SFL reveal its two fundamental principles: that human experience and learning are constructed in the form of language (Halliday, 1993: 9) and that language cannot be understood divorced from its social context. Thus, in relation to school education, SFL theory advocates a language development approach to education which evidently acknowledges the language-based nature of subject learning. To examine the language of education using SFL, the concept of register is adopted which represents the constellation of lexical and grammatical features that realizes a particular situational context (Halliday and Hasan, 1989). Three parameters are used to describe a register: field - the social action, tenor - the role structure, and mode - the symbolic organization (Halliday and Martin, 1993: 33). In this way, 'what is happening' (field), 'who are taking part' (tenor), and 'what part the language is playing' (mode) in school subject registers can be described, and the linguistic functions and particularly the grammatical choices and structures which create meanings in the different school texts can be identified (Schleppegrell, 2004).

A third approach to the analysis of subject discourse would be to examine 'the external aspects of genre construction' (Bhatia, 2004: 11), i.e., the social context of discourse. Increasingly, academic discourse-based studies have been concerned with the social, cultural and institutional context of language use dealing with issues such as social practices and ideological assumptions. Critical discourse analysis is one paradigm which 'explores the connections between the use of language and the social and political contexts in which it occurs' (Paltridge, 2006: 179). Insights from such a study in this case would demonstrate how language construes social and power relations in the subject classroom. Representative examples of this kind of critical discourse analysis for students' empowerment can be found in the Australian genre-based literacy programs (e.g. Martin, 1993; Cope and Kalantzis, 1993; Hasan and Wiliams, 1996). This approach clearly moves beyond the linguistic dimension of discourse to more sociological perspectives (cf. Foucault's (1972) view of discourse and sociological theories of pedagogy, e.g. Bernstein's (1999) distinction between horizontal and vertical discourse).

Undoubtedly, all three approaches to discourse analysis could offer relevant insights into the discourse of the six curriculum subjects examined in this research. A multi-perspective model of discourse analysis could also be profitably used, encompassing a holistic examination of discourse as text, genre and social practice (Bhatia, 2004: 18). This appears to be fairly

broad, however, considering the scope and time constraints of the present research. What is of immediate interest to post-primary ESL support, and hence to the concerns of this research, is the first approach which would examine the textual level of subject discourse and its instantiation in lexico-grammatical features and patterns. It is the specification of the linguistic (lexical and grammatical) characteristics of the different curriculum subjects that could address the research questions established in Chapter 1 (section 1.6) and provide an explicit agenda for language support.

The SFL approach may have contributed significantly to analysing the language of schooling in general but this analytical approach cannot fulfill the objectives of the present research because it is primarily concerned with the language functions and grammatical choices of curriculum subjects in connection with the social context and culture of education. As such it places less emphasis on concrete descriptions of the linguistic substance of subject-specific language varieties which are precisely what this research seeks to provide. As for an examination of subject discourse through the prism of its socio-cultural context, this may have implications for empowering students in the host community, but it is irrelevant to this study as this approach to discourse analysis is 'more contextually rather than linguistically grounded' and would be thus more appropriate for 'a pedagogy developed along the lines of general consciousness raising rather than overt didactics' (Flowerdew, 2002: 3).

Accordingly, this research is concerned with analysing the textual dimension of subject-specific discourse, focusing on the lexico-grammatical features of the six curriculum subjects. In other words, subject-specific language use is examined here from the perspective of register which is not understood as a semantic unit, as in SFL theory, but as a 'situationally-defined variet[y] described for [its] characteristic lexico-grammatical features' (Biber, 2006: 11). In addition, the research aims to demonstrate that 'subject specific registers employ different constellations of features to achieve their particular communicative goals' (ibid.: 30). Therefore, register becomes very useful in this discussion as a tool for text analysis and description of textual variation across curriculum subjects. The specification of the register features of subject-specific discourse and aspects of linguistic variation could have considerable benefits for language support. It could inform ESL instruction and the elaboration of the subject-specific scales of the English Language Proficiency Benchmarks (IILT, 2003a) and it could further provide a language-rich basis for materials design and language tests development. In this way, ESL students' acquisition of register characteristics and the ability to switch across registers could be promoted both of which are fundamental for full membership of the discourse community of the post-primary subject classroom (Biber, Conrad and Reppen, 1998: 135).

The most appropriate research methodology for identifying the features of the subject textbooks and examination papers registers and for bringing out elements of linguistic variation is one which uses corpus linguistics techniques and tools. Applied corpus linguistics in general and the value of a corpus-based methodology for this research in particular are discussed in the next section.

2.3.2 Applied corpus linguistics and the need for a corpus-based methodology in this research

Like other approaches to linguistics, corpus linguistics is concerned with the description, examination and explanation of the nature, structure and use of language. Its 'applied' aspect lies in the practical applications of descriptive corpus-based research to language pedagogy. Corpus-based studies do not follow one single methodology; rather, there are various methodologies which are consistent with the overall characteristics of a corpus-based approach (Conrad, 1996: 304). All methods are consonant with the theory of empiricism which asserts that knowledge arises from experience and, as such, corpus-based studies depend on observations of natural language in use through the medium of corpus.

A corpus is defined as 'a large and principled collection of natural texts' (Biber, Conrad and Reppen, 1998: 12) which can be based on either written texts or transcriptions of recorded speech, or a combination of both. It is nowadays taken for granted that such a body of linguistic data comes in an electronic form. What distinguishes a corpus from other collections of texts is that the corpus includes samples of authentic language which are representative of a language variety and have been assembled and organized according to specific criteria in order to answer pre-defined research questions. Accordingly, there are different types of corpora: general, specialized, comparable, parallel, multilingual, historical (or diachronic) and monitor corpora.

The distinguishing features of corpus-based studies are that they i) are empirical investigations of language structure and use, ii) make use of large and principled collections of natural texts (i.e. corpora) as the basis of analysis, iii) utilise computer programs for automatic analysis and iv) depend on both quantitative and qualitative analytical techniques (ibid.: 4). These features are clearly suggestive of the advantages of corpus linguistics methodologies compared to other methods of discourse analysis. The strength of corpus-based studies lies in the extensive use of specialised software programmes. These facilitate fast, accurate,

consistent and complex analyses of language which are not possible through any other means, allowing, as a result, 'access to a quality of evidence that has not been available before' (Sinclair, 1991: 4). More specifically, corpus-based studies have identified a number of linguistic features and patterns which might have been impossible to describe purely on an intuitive basis; for instance, collocations, frequency of ready-assembled chunks, etc. (Adolphs, 2006: 97, O'Keeffe, McCarthy and Carter, 2007: 21). Perhaps the pioneering achievement of corpus-based linguistic description is its identification of the pre-patterned nature of language and of the English language in particular which was revealed to its full extent first by Sinclair and his arguments for the idiom and open-choice principles (1991). In addition to new and rich information about language patterns in use, data from corpus-based studies also offer 'the communicative background to the types of discourse in which such patterns occur' (Hunston, 2002: 99). Offering examples of contextualised natural language use is essential for the conceptual understanding of language and hence valuable for ESL teaching/learning purposes.

Perhaps the strongest argument in favour of using a corpus linguistics methodology for linguistic analysis is provided by the fact that the corpus provides 'an objective frame of reference' (Bowker and Pearson, 2002: 20) and 'electronic text analysis is a more replicable process and any analysis can be verified by other researchers' (Adolphs, 2006: 7). Thus, from the point of view of scientific method, corpus-based methodology conforms to standards commonly ascribed to the 'scientific model': falsifiability, completeness, simplicity, strength, and objectivity (Leech, 1992). On these grounds, corpus-based studies are valid and reliable because they are 'open to objective verification of results' (ibid.). It should be underlined that a synergy between quantitative analysis, for statistically reliable and generalisable results, and qualitative interpretations, which provide greater richness and precision, is essential to enhance the quality of results.

Certain concerns have, nevertheless, been expressed about corpus-based studies. One drawback is that electronic text analysis leaves out certain components which contribute to the meaning of a text. These can be, for instance, images, particular graphic styles, or elements that are linked through hypertext (Sinclair, 2005). In other words, 'corpus data is usually only language data' whereas 'discourses are not confined to verbal communication' (Baker, 2006: 17). Although there is presently no standardized way of encoding images in corpora, there have been attempts to address this issue (for example Smith et al. (1998) have created and encoded corpora of visual materials in a corpus of children's posters). Similarly, the construction of multimodal corpora, which combine video, audio and textual records of naturally occurring discourse (e.g. Knight et al., 2009), can counter the same limitations of corpora of spoken discourse (see Allwood, 2008; Wittenburg, 2008).

Further to the previous criticism, another argument against corpus-based analyses is that, because they apply bottom-up rather than top-down methodologies, they do not consider the socio-cultural context of language use as they are concerned withdecontextualized examples of language (Widdowson, 1998; 2002). Baker supports, however, that not knowing the ideologies of the text producers in a corpus can be a methodological advantage (2006: 18). Flowerdew (2005), on the other hand, suggests the integration of corpus-based and genre-based approaches to text analysis in English for Academic Purposes (EAP)/English for Specific Purposes (ESP) to counter these criticisms.

The generalizability of data that emerge from corpus analysis is another issue that should be seriously considered. It should be taken for granted that 'corpus research can only ever produce results that reflect the particular corpus that is being used for study' (Adolphs, 2006: 12). On these grounds, it may not be possible in all cases to argue that results are representative of a language variety. The issue of representativeness largely depends on the size and design of the corpus. It is equally important to note that a corpus does not itself offer information about language; rather it is the 'software [that] offers us new perspectives on the familiar' (Hunston, 2002: 3). Consequently, any conclusions that are drawn from corpus analysis in order to make generalisations are extrapolations (ibid.: 23). For this reason, the interpretation of data is of crucial importance as it may influence the objectivity of conclusions.

The role of intuition cannot be totally dismissed as it has an instrumental role to play in 'extrapolating important generalizations from a mass of specific information in a corpus' (ibid.). In his discussion of bias in the analysis and interpretation of corpus-based results, Baker notes that 'Corpus researchers can theoretically be just as selective as anyone in choosing which aspects of their research to report or bury. And their interpretations of the data they find can also reveal bias' (2006: 12). He goes on to add, however, that 'at least with a corpus, we are starting (hopefully) from a position whereby the data itself has not been selected in order to confirm existing conscious (or subconscious) biases'. When it comes to the presentation of corpus findings, Baker further underlines the importance of reporting 'exceptions alongside the overall patterns or trends, but not to over-report them either' (ibid.: 12).

These limitations of corpus-based methodologies should not, however, devalue corpus linguistics since this has ushered in a new era in linguistic research by providing original insights into language and new research approaches towards language exploration. The significant contribution of corpus-based studies to language education, among other fields (e.g.lexicography, stylistics, translation, forensic linguistics, sociolinguistics, language and

ideology), can be demonstrated by considering the applications these studies have to i) language description (e.g., Sinclair 1991 on phraseology and Biber et al. 1999 on variation), ii) data-driven learning (e.g. Johns, 1991), iii) parallel concordancing and reciprocal learning (e.g. Johns, 1991), iv) language teaching methodology (see Hunston, 2002), v) language testing and assessment (e.g. Alderson, 1996; Barker, 2004; Taylor and Barker, 2008), vi) syllabus design (e.g. Sinclair and Renouf, 1988; Flowerdew, 1993), vii) ESP/EAP (e.g. Coxhead, 2000; Biber, 2006; Hyland, 2008, etc.), and viii) learner corpora (e.g. Granger, 1998; Nesselhauf, 2004) and teacher corpora (e.g. Farr, 2008, 2010; see also O'Keeffe et al., 2007: 220-245). The aforementioned contributions of corpus linguistics to L2 education clearly demonstrate the potential of a corpus-based methodology for the present research.

In short, a corpus-based methodology was selected for this empirical research on account of its advantages compared to other methodologies (i.e. high-quality analytical techniques and tools, rich and colourful descriptions of linguistic data, reliability and objectivity), but also because it is the one that could most effectively address the needs of the envisaged analysis (i.e.the collection of a large number of texts, the examination of a variety of linguistic features and comparisons across subject registers).

The successive stages involved in the methodological procedure, the text corpora that were developed, and the software tools that were used for corpus analysis are all described in the discussion that follows.

2.4 Overview of methodological procedure: computational analysis of subject-specific text corpora

The corpus-based methodology that was employed in this research encompassed the quantitative and qualitative analysis of twelve subject-specific text corpora; six of them were built on commonly used Junior Cycle textbooks and six were built on Junior Certificate examination papers. The research involved the following three stages: i) the setting of specific objectives to guide corpus analysis, ii) the design and construction of twelve subject-specific corpora, and iii) corpus analysis and retrieval of quantitative and qualitative data. Each stage is briefly described in this section.

In the first stage, specific objectives were established in direct correspondence with the general aims of the research to guide corpus analysis. To examine how subject content is 'languaged' in the textbooks and examination papers of individual subjects, the primary

objective was to conduct frequency analysis of the twelve corpora (meeting minimum frequency and range restrictions). This analysis identified the most commonly used language features which convey important meanings of individual subject areas. The choice of a frequency approach to language analysis was motivated by the significance of lexical frequency data in language learning and pedagogy, which was valued even before the emergence of corpus linguistics (e.g. Zipf, 1935, 1949; West, 1953). Frequency information on language indicates language items that are important for students since they repeatedly appear in their subject textbooks and examination papers. Greater frequency is further significantly associated with a greater dispersion and coverage (Leech, 2001: 332). Also, according to Zipf's Law of least effort (1935, 1949), which states that the more complex a word is, the less frequent it will be, high lexical frequency appears to correlate with less complexity and thus greater learnability (see Milton, 2009 for empirical evidence on vocabulary frequency and learnability). From the above it follows that information on the most frequent language can be used as an important criterion in determining teaching priorities in language support specifically in terms of content, sequence and gradation. In addition, the commonest lexical features in subject textbooks and examination papers can considerably facilitate ESL students' comprehension of these texts as well as the construction of appropriate subject-specific texts. For these reasons, and because frequency is easily measurable, a frequency-based approach to linguistic analysis was selected in this research. The second objective was to compare the linguistic characteristics of the six corpora of each register in order to identify common features, and to subsequently compare and contrast their meanings and uses across subject-specific contexts. In this way, the second aim of this research, i.e. to examine aspects of linguistic similarity among subjects, was addressed.

To achieve the aforementioned objectives, twelve subject-specific corpora were constructed that represent each of the six curriculum subjects under investigation in the two school registers of textbooks and examination papers. The design and construction of these corpora, including decisions about sampling, representativeness, balance, size, and pre-processing, are discussed in detail in section 2.4.2.1, and section 2.1.2.2 describes their composition.

Following the development of corpora, a quantitative analysis was conducted which was integrated with a qualitative analysis of the underlying functions of selected quantitative data. The combination of these two types of linguistic analysis is consistent with the common consensus that exists in the corpus linguistics literature which contends that 'quantitative techniques are not sufficient' and that 'qualitative interpretations are necessary to examine the functional bases underlying patterns of linguistic features' (Biber, Conrad and Reppen,

1998: 139). As Sinclair explains, 'different methods of observation provide multiple views of the text, which highlight some patterns and obscure others, such methods are exploratory and corpus data are essentially indirect (2004: 7). A multi-method approach combining both qualitative and quantitative perspectives on the same phenomena (McEnery and Wilson, 1996: 63) therefore allows one to compile comprehensive linguistic profiles, in this case, of curriculum subjects.

This multi-method corpus analysis was focused on the lexical dimension of subject-specific language and it was facilitated by the lexical analysis software WordSmith Tools 4 (Scott, 2004) and namely by the WordList and Concord tools (see section 2.6 for background information to WordSmith and a description of the two tools). Three types of lexical items were extracted from individual corpora: words, collocations and 4-word clusters. These features were selected for analysis on account of their pedagogical value for language support which is discussed in Chapter 3. More specifically, frequency analysis using the WordList tool revealed the most frequent and consistently used i) function and content words, ii) collocates of the commonest content words, iii) significant collocations, and iv) 4-word clusters in the twelve corpora. All the derived corpus data were discussed with reference to their semantic and functional associations in relation to their context of use. The 4-word clusters were further analysed in respect of their lexical density, structural and complementation patterns (words that precede/follow them). The Concord tool was used in order to i) illustrate and closely examine the use of linguistic features in authentic contexts, ii) identify the functional role of specific linguistic features and patterns (e.g. the discourse functions of word clusters), and iii) to compare the occurrences of a common feature across the different subject corpora. Having drawn detailed linguistic profiles of individual corpora, comparisons of the lexical words and 4-word clusters across the six corpora of the same register took place to identify common features. The meanings and uses of these features across subjects were subsequently compared and contrasted. The empirical findings from the quantitative and qualitative analysis of the twelve corpora are described and discussed in detail in Chapter 3.

The research methodology employed here is i) exploratory, in the sense that no apriori assumptions were made about expected results from the analysis, and ii) data-driven, because corpus analysis 'provides a way into the data that is informed by the data itself' (Adolphs, 2006: 19). Contrary to a corpus-based analysis, a corpus-driven analysis proceeds in a more inductive way – the corpus itself is the data and the patterns in it are noted as a way of expressing regularities (and exceptions) in language use (Tognini-Bonelli, 2001: 85, rephrased by Baker, 2006: 16). As a result, no predefined linguistic categories are imposed nor are any judgements made in advance about what may or may not be important. The exploratory and

data-driven nature of the approach stems also from the 'lack of a theoretical foundation for the interpretation of results prior to the analysis' which explains why 'most corpus research has been of a post-hoc nature, looking at the frequency counts and deciding what can be said about these results' (Grabe and Kaplan, 1996: 46). To prevent bias, no specific framework of linguistic analysis was employed. The information obtained from the manual examination of subject syllabuses and guidelines for subject teachers was used to explain and interpret the corpus data, assuming that the different linguistic choices of the six subject areas are not random but determined by disciplinary aims and practices. Some references to the relevant literature on ESP and the language of school subjects are also made where appropriate.

To conclude, it is worth explaining why Biber's multi-dimensional (or MD) approach (1988) was not adopted, despite being one of the most powerful methodological instruments that has been extensively used to address similar questions (e.g. for the investigation of spoken versus written English language, Biber 1988; Biber and Finegan, 1994; for the characterization of the written registers of university language, Biber and Conrad and Cortes, 2004; Biber 2006 etc.). Biber's approach is based on the 'comparison of frequency distributions across genres, but use[s] a multi-feature, multi-dimensional methodology, grouping sets of linguistic features associated with a number of factors called text dimensions' (Rayson, 2008: 525). This approach requires, however, considerable expertise in data extraction and statistical analysis and it is also very time-consuming. It has been further criticized as 'being linguistically and statistically unsound' (Lee, 2000) on the grounds that the pre-defined linguistic/functional categories restrict the data that can be extracted. It is worth adding here that Tribble (1999) and Xiao and McEnery (2005) demonstrate that an approximate effect of Biber's MD approach can be achieved through the Wordlist (and Keyword) functions of WordSmith Tools (Scott, 2004).

Following the overview of the research methodology, detailed information is provided in the sections that follow on the design and development of the twelve subject-specific text corpora and on the use of corpus query tools for the computational analysis.

2.5 Milestones in corpus development

2.5.1 Design rationale: sampling, representativeness, balance and size

The creation of new corpora involves several considerations which are important for the validity and reliability of the research that is conducted (Kennedy, 1998: 60); for instance, corpus size, balance, representativeness, data capture, corpus mark-up, and character encoding (McEnery et al., 2006: 76) as well as medium, authorship and publication date (Bowker and Pearson, 2002: 45). Decisions related to all of these issues are discussed in this section as none of the twelve corpora used in this research study was readily available but had to be constructed in accordance with the aims and objectives of the research.

It is important to begin by defining the sampling frame (Biber, 1993), that is, to define the entire population of texts from which samples are collected for inclusion in the corpus as well as the boundaries of the selected samples. The sampling frame for the corpora built on textbookscomprises all textbooks used at Junior Cycle level for the six curriculum subjects (English, geography, history, CSPE, mathematics, science). To compile the six textbooks-based corpora, the 24 most commonly used textbooks were sampled, covering all three years of Junior Cycle education. These subject textbooks were selected for corpus compilation in consultation with post-primary subject teachers based on the criterion of frequency of use. The sampling frame for the examinations-based corpora comprises all examination papers administered by the State Examinations Commission for the six subjects. In this research, however, linguistic analysis is confined to the examination papers of the last decade, i.e. from 2001 to 2010, due to the major syllabus changes which occurred in 2000 and which affected the structure and content of examination papers. Therefore, each of the six examinations-based corpora comprises the ten most recent examination papers of the corresponding curriculum subjects (see the details on the composition of the twelve subject corpora in section 2.5.3).

As regards the boundaries of the selected texts, the text collections of the 12 corpora comprise the entire texts of subject textbooks and examination papers. Specifically in the textbooks-based corpora, whole documents were included in each corpus: i.e. table of contents, preface, whole chapters and exercises, captions of illustrations and diagrams, appendices and indexes. Images, acknowledgments and information on publishers were excluded from each file as these do not contain subject-specific language. The examinations corpora comprise entire texts with the exception of some captions which accompany images

that are included in source papers (usually provided in history and CSPE examinations). The inclusion of whole texts in the corpus is important because in corpora created for specific purposes (in this case, post-primary subject- and examinations-specific English) 'the concepts, terms, patterns and contexts that interest you might appear in any section of the text' (Bowker and Pearson, 2002: 49). It is thus not possible to arbitrarily select text fragments for inclusion in the corpora because there is 'a *conceptual* interest in the text, as well as a linguistic one' (Bowker, 1996: 316, emphasis in the original) and, therefore, terms can appear anywhere in the text and, in case a text is cut, an important conceptual description may be omitted either partially or completely (ibid.). As Sinclair remarks, it is 'an unsafe assumption that any part of a document or conversation is representative of the whole – the result of research for decades of discourse and text analysis make it plain that position in a communicative event affects the local choices' (2005: 6).

It is also important to note that, for the compilation of the textbooks-derived corpora, the use of more than one textbook for each subject (written by different authors) is of great significance for the purposes of this research as this allows the examination of the terms and concepts commonly used in each discipline, whereas in the case of a single-author corpus, the terms that are preferred by that particular author would be identified (ibid.: 49). In other words, the construction of multi-author corpora 'helps neutralise bias that may result from the idiosyncratic style of one writer' (Sinclair, 1991; Atkins et al. 1992) and it further increases the number of lexical items in the corpus (Sinclair et al., 1994). The publication date of textbooks was also taken into consideration for corpus inclusion, and an effort was made to sample the most recent publications available (with the exception of one science textbook published in 1989).

Another consideration related to the content of the corpora is that, no distinction was made in terms of a) the Junior Cycle year (1st, 2nd, 3rd) of the textbooks and b) the ability levels to which textbooks and examination papers correspond (i.e., foundation, ordinary, higher levels; cf. Chapter 1, section 1.2.2). Although both grade levels and syllabus differentiation are dimensions of language variation, this fact makes little difference to English language learning in the language support classroom. This is because the majority of ESL students enter post-primary education at the grade level which corresponds to their age and, further, their arrival can take place at any point in the school year. This automatically means that the English language support that ESL students receive does not entirely correspond to the language requirements of the syllabus and subject textbooks of the school year in which they were registered. Language support provision is normally targeted at a broad coverage of the key English language demands of Junior Cycle education and of the curriculum subjects which are

compulsory for students in all three years. On these grounds, specifying language learning needs according to grade and syllabus levels would have no significant impact on language support pedagogy.

Representativeness, the extent to which a sample includes the full range of variability in a population (Biber, 1993: 243), is another important issue in corpus design. The aim of the present research is to capture the full range of linguistic variability that exists in the written academic discourse of the six subjects under examination. From a statistical point of view, to achieve corpus representativeness one should 'secure a sample which, subject to limitations of size, will reproduce the characteristics of the population, especially those of immediate interest, as closely as possible' (Yates, 1960: 9). The degree to which this can be achieved is determined by balance, i.e., the range of genres included in a corpus, and sampling, i.e., the way texts for each genre are selected (McEnery et al., 2006: 13). 'To achieve corpus representativeness [...] requires knowledge of which genre is used and how often by the language community in the sampling period' (ibid.: 73). Accordingly, for the construction of corpora for this research, full texts of similar size were collected for each subject area, representing the most up-to-date and commonly used text types used in the Irish post-primary classroom for the six curriculum subjects under investigation. Further, the criteria for text selection were external, that is, they were 'derived from an examination of the communicative function of [texts]' rather than internal, based on an examination of the language of texts (Sinclair, 2005: 1).

As regards the size of a corpus, this depends in principle on the predefined purposes of the research but it can also be influenced by the availability of data and the amount of time that is available to the researcher (Bowker and Pearson, 2002: 45). It is generally accepted (e.g. Engwall, 1994: 51; Bergenholtz and Tarp, 1995: 95) that corpora which are to be used as a basis for a language for general purposes (LGP) study must be larger (i.e., running into millions of words) than corpora needed for language for specific purposes (LSP) work, i.e. 'the language that is used to discuss specialized fields of knowledge' (Bowker and Pearson, 2002: 25). Sinclair (1992), however, supports that selecting from a large amount of data is preferable to restricting the amount of data that can be available. A potential drawback of small corpora in LSP can be that they may not contain all the concepts, terms and linguistic patterns that are relevant to the language variety that is being investigated, thus preventing any valid generalizations (Bowker and Pearson, 2002: 48). Nevertheless, Aston (1997) and Tribble (1997) recommend the use of a small corpus when a very specialised language register is being examined, as is the case in the present research. O'Keeffe et al. (2007: 198) also argue that

'specialised lexis and structures [in LSP corpora] are likely to occur with more regular patterning and distribution, even with relatively small amounts of data'.

Taking the above views into account, it could be argued that the size of the textbooks-based corpora (two million words) and the examinations-based corpora (approximately 340,000 words; see section 2.5.3 for details) in question is sufficient to obtain insights into the specialized language use of the restricted domains of commonly used textbooks and recent examination papers. Not surprisingly, the twelve corpora do not have the same size. This difference in sample size is not important however because, as Sinclair argues, 'the integrity and representativeness of complete artefacts is far more important than the difficulty of reconciling texts of different dimensions' (2005: 6). Due to size variation, normalization was necessary to ensure that it was possible to draw valid conclusions from linguistic comparisons of corpora. Normalization is 'a process which makes frequencies from samples of markedly different sizes comparable by bringing them to a common base' (e.g. per mille or per one million words; Baker, Hardie and McEnery, 2006). As a final point, it is important to repeat the limitation that is inherent in corpus-based studies and which also applies to the corpora of this research that 'no corpus, no matter how large, how carefully designed, can have exactly the same characteristics as the language itself' (Sinclair, 2005: 2).Thus, the findings that emerge from the analysis of the textbooks-based and examinations-based corpora cannot be generalizable to subject-specific language use in these registers as a whole; they are only valid for the language of the twelve corpora in question.

Finally, an equally important issue that also needs to be addressed here concerns copyright issues. There are no hard and fast rules in the corpus literature as copyright law varies from country to country. It is generally accepted, however, that 'both the legal and moral aspects of copyright must be taken into account before including texts in a corpus' (Bowker, 1996: 316). Irish copyright law permits the copying of material for not-for-profit educational use, as is the case in the present research. It should be made clear that the scope of this research is limited to using subject textbooks widely available in the market and examination papers freely available online in order to computationally analyse the language they contain; a process that in principle could also be done manually. No other usage of these educational materials is made, e.g. reproducing, modifying, photocopying or distributing parts of or the entire textbooks/examination papers. Finally, the digitized versions of the textbooks are not made available to any individual or agency external to the ELSP of TII.

2.5.2 Pre-processing corpora

Following the decisions involved in corpus design and data collection, the pre-processing of the selected texts was essential to produce the final versions of the text corpora. All six textbooks corpora were self-constructed in this research. The examinations-based corpora were more easily built and required less time and effort because all the examination papers that were used are available in pdf form on the State Examinations Commission website (www.examinations.ie) and therefore they only had to be downloaded and converted from pdf to plain text files (using Adobe Reader 9).

For the compilation of the textbooks corpora the 24 subject textbooks that comprise them had to be first selected in consultation with subject teachers and they were then purchased and finally converted into digital files. The entire contents of all 24 subject textbooks were scanned using the Optical Character Recognition (OCR) Software Able2Extract Professional 4 and were initially stored as pdf files. These were subsequently converted into word documents containing text only using the OCR ReadIris 11 (all images, diagrams, acknowledgments were removed). Finally, plain text files for each textbook were created which were then merged to form the six corpora. Manual proofreading, spellchecking and post-editing of scanned texts are often required because scanners and OCR software are prone to error in data capture (e.g. misreadings of letters such as *d* for *cl*, and *m* for *in*, *ni* or *ir*, also changes in font size, page layout the occurrence of tables, figures and running heads can all affect straightforward scanning). In this study, a small sample of scanned material was proofread to establish whether all files needed to be proofread and edited. Comparisons of corpus analysis data between proofread and unedited text collections revealed negligible differences in the quality and quantity of the derived data and, on these grounds, no human intervention took place to any of the text collections to achieve 100% accuracy due to time constraints. Therefore, a small degree of noise (corrupted data) still exists in the text corpora namely in the mathematics textbooks corpus which contains a great deal of numerical data.

Another possible type of text pre-processing could have been corpus mark-up which refers to 'a system of standard codes inserted into a document stored in electronic form to provide information *about* the text itself and govern formatting, printing or other processing' (McEnery, Xiao and Tono, 2006: 22). The aim of mark-up is to recover as much contextual information about the corpus data as possible to facilitate their interpretation and corpus access and use by other researchers for future study. Given that the corpora used in this research cannot be made available to other researchers due to copyright issues there was no real need to add any meta-data to the corpora.

Overall, every effort was made to ensure the optimum quality of the twelve corpora (e.g. different types of OCR software were tested for data capture, some textbooks which were initially included in certain subject-specific corpora were in the end excluded because they did not meet the set criteria, etc.). The composition of the final corpora used in the research is presented in the following section together with relevant statistical information.

2.5.3 The composition of the subject-specific text corpora

Following the discussion of the design rationale, data collection and pre-processing of the subject-specific text corpora that were developed in this research, this section offers detailed information on their composition. All the information presented here was obtained from the statistics computed by WordSmith Tools 4 (Scott, 2004).

Tables 2.1 - 2.8 below provide details on the text collections that comprise each corpus including the following information: i) the subject textbook or examination paper that was included in the corpus, ii) the number of tokens, i.e. running words in a text, iii) the number of types, i.e. different words in the text, alternatively word forms, and iv) the standardized type/token ratio (STTR; see explanation below). It should be explained here that the hash symbol (#) signifies the word 'number'.

Table 2.1 The English textbooks corpus.

Textbook code	#oftokens	# of types	STTR
EN1	88,684	8,828	44.21
EN2	88,310	7,758	41.65
EN3	129,427	9,483	34.23
EN4	145,363	10,677	36.49

Table 2.2 The geography textbooks corpus.

Textbook code	# of tokens	# of types	STTR
GEO1	100,652	7,660	38.66
GEO2	107,370	8,423	38.76
GEO3	19,300	2,051	32.21
GEO4	102,935	7,632	38.13

Table 2.3 The history textbooks corpus.

Textbook code	# of tokens	# of types	STTR
HIST1	145,501	11,315	42.58
HIST2	54,207	6,811	43.03
HIST3	40,626	5,354	41.58
HIST4	127,374	9,310	39.00

Table 2.4 The CSPE textbooks corpus.

Textbook code	# of tokens	# of types	STTR
CSPE1	62,675	6,270	38.71
CSPE2	63,353	6,865	40.26
CSPE3	37,160	4,387	41.39
CSPE4	50,152	5,743	40.51

Table 2.5 The mathematics textbooks corpus.

Textbook code	# of tokens	# of types	STTR
MATH1	128,542	3,535	17.93
MATH2	113,624	3,432	18.58
MATH3	41,537	1,565	18.58
MATH4	53,035	1,707	18.24

Table 2.6 The science textbooks corpus.

Textbook code	# of tokens	# of types	STTR
SC1	112,291	5,036	29.99
SC2	148,949	8,453	36.42
SC3	35,469	2,409	28.34
SC4	31,717	3,446	33.52

Table 2.7 The six subject-specific corpora that comprise the Junior Cycle textbooks corpus.

Curriculum subject	# of textbooks	# of tokens	# of types	STTR
English	4	451,784	19,941	38.36
Geography	4	330,257	13,865	38.18
History	4	367,708	16,041	41.29
CSPE	4	213,340	13,213	40.06
Mathematics	4	336,738	5,818	18.37
Science	4	328,426	10,188	33.06
Textbooks-based corpus	24	2,028,253	42,838	35.39

Table 2.8 The six subject-specific corpora that comprise the Junior Certificate examinations corpus.

Curriculum subject	# of exam papers	# of tokens	# of types	STTR
English	10	95,143	7,834	39.68
Geography	10	48,178	2,756	33.50
History	10	46,351	4,053	38.46
CSPE	10	25,811	2,360	33.16
Mathematics	10	55,521	1,873	26.71
Science	10	68,617	2,967	28.63
Examinations-based corpus	60	339,621	12,790	33.94

The tables above exhibit the expected variation in size that exists among the twelve corpora but also in the size of the individual textbooks and examination papers which comprise each corpus. It should be explained here that the considerably smaller size of some textbooks included in certain textbooks-based corpora results from them being, in fact, workbooks and not course books. For example, the shortest text in the text collection of the geography corpus, i.e. 19,300 tokens, is a handbook (*Discovering maps and photographs*) which explains how maps and photographs can be studied and used by students in geography classes.

As can be seen in Table 2.7, the size of the entire Junior Cycle textbooks corpus slightly exceeds two million words; a size which, based on the discussion of LSP corpus size in section 2.5.1 above, meets the requirements for a corpus-based study of specialised language use. The same is also true for the size of the Junior Certificate examinations corpus which, as shown in Table 2.8, has 339,621 tokens. The smaller size of the examinations-based corpora can be

justified considering that they comprise short instructions and exercises and only in few cases they contain extended text (namely in source papers which accompany history and English examination papers). A remark that becomes relevant to the small size of the examinations-based corpora in question is made by Bowker and Pearson in their practical guide of working with specialised language corpora who argue that 'well-designed corpora that are anywhere from about ten thousand to several hundreds of thousands words in size have proved to be exceptionally useful in LSP studies' (2002: 48).

What is particularly interesting from the information that is presented in the above tables is the STTR for each corpus. WordSmith calculates the type/token ratio (TTR) of a corpus which is defined as 'the proportion of the text made up of lexical word tokens (nouns, lexical verbs, adjectives, and adverbs)' (Biber et al., 1999: 62). This is a good indicator of the level of complexity, or lexical density, of the subject-specific text collections that comprise the corpora. Specifically, TTR as a percentage can be calculated as follows: (types/tokens) x 100 (ibid.: 53). However, as Scott explains in the WordSmith (2004) online help manual, 'The conventional TTR is informative, of course, if you're dealing with a corpus comprising lots of equal-sized text segments', which is not the case with the corpora used of this study. For this reason, the standardised type/token ratio (STTR) was calculated and examined here; this is calculated in the Wordlist program of WordSmith as follows:

> The standardised type/token ratio (STTR) is computed every **n** words as Wordlist goes through each text file. By *default,* n = 1,000. In other words the ratio is calculated for the first 1,000 running words, then calculated afresh for the next 1,000, and so on to the end of your text or corpus. A running average is computed, which means that you get an average type/token ratio based on consecutive 1,000-word chunks of text. (Texts with less than 1,000 words (or whatever n is set to) will get a standardised type/token ratio of 0.).

Thus, contrary to the TTRs, the STTRs of the twelve corpora can be directly compared to draw conclusions about lexical density. The different STTRs of text collections in each corpus can be explained by considering that the STTR varies with the length of the text and specifically: 'longer texts have many more repeated words and therefore a much lower TTR [which is also true for STTR]. In longer texts, there is a greater chance that words which have already been used will be repeated. This is true both of the most frequent words which recur in all kinds of texts (*the, of, and*, etc,) and of the words which are connected with the topic of a particular text' (Biber et al. 1999: 53).

A comparison of the STTRs of the six textbooks-based corpora shows that history and CSPE have the highest STTR (41.29 and 40.06 respectively), followed by English (38.36) and

geography (38.18) and then science (33.06) with mathematics having by far the lowest (18.37). This confirms what might be viewed as commonly shared impressionistic assumptions according to which the language of mathematics textbooks is made up of the lowest proportion of lexical word tokens (i.e. nouns, lexical verbs, adjectives, and adverbs) compared to the other subject textbooks, because a large proportion of mathematical discourse comprises numbers and symbols. With regard to the STTR of the examinations-based corpora, English appears to have the highest STTR (39.68) followed by history (38.46). Geography and CSPE come next with very similar STTRs (33.50 and 33.16 respectively). Science with 28.63 and Mathematics with 26.71 have again the lowest STTRs compared to the other four subject corpora. This may be because the aforementioned subjects except mathematics and science are accompanied by source papers which contain extended text. As a final remark, it should be stressed, however, that STTRs alone may provide only some information about textual complexity and that, in order to make any robust statements, 'a closer linguistic analysis of grammatical structure and semantic fields of [...] individual terms that occur is necessary' (Adolphs, 2006: 40). In other words, STTRs can be used as a starting point for more elaborate analysis.

Having presented the subject-specific corpora developed and analysed in this research, the discussion proceeds to a brief introduction to the corpus query tools which were used for corpus analysis and comparison.

2.6 Software tools for corpus analysis: WordSmith 4

WordSmith 4 (Scott, 2004) is a lexical analysis software which examines 'how words behave in texts'. It comprises a suite of three tools: a) the WordList tool that is a programme that generates word lists shown either in alphabetical or frequency order, b) the Concord tool which displays concordances based on a search word, and c) the KeyWords tool which locates and identifies key words in one or more texts. Only the first two tools were utilised in this research. All computations in WordSmith are based on the use of Dunning's Log Likelihood statistic (Dunning, 1993; Oakes, 1998: 172). Although the above tools can perform a number of functions, only the ones that were exploited in this research are discussed here.

The WordList tool was used as a starting point in the analysis of the twelve corpora. A wordlist in WordSmith shows i) how often each word occurs in the text files, ii) what that is as a percentage of the running words in the text (frequency), and iii) how many text files each

word was found in (range). In addition to deriving frequency-based lists of single words, the Wordlist was used to compute collocations and multi-word units, referred to as clusters. Of all terms available in the literature for multi-word expressions (see Wray, 2002 for an overview), Scott (1999) uses the term 'clusters' because 'groups and phrases already have uses in grammar and because simply being found together in software doesn't guarantee they are true multi-word *units'*. The frequency analysis of words, collocations and 4-word clusters using WordList resulted in detailed lexical profiles of the twelve corpora. More elaborate functional types of analyses were facilitated by the Concord tool.

Concord is a program which can demonstrate the immediate context of use of lexical items (e.g. content words) or lexical combinations (word clusters) through the display of concordances, i.e., 'alphabetical indexes of a search pattern in a corpus, showing every contextual occurrence of the search pattern' (Baker, Hardie and McEnery, 2006). Concordance lines displaying occurrences of words and clusters within authentic contexts allow a closer analysis of the lexical, semantic and grammatical behaviour of the node word/cluster under examination. This analysis provides insights into i) collocates, ii) semantic prosody, i.e. the 'consistent aura of meaning with which a form is imbued by its collocates' (Louw, 1993: 157), iii) patterns of repeated phraseology and iv) colligation, i.e. 'the collocation between a lexical word and a grammatical one' (Hunston, 2002: 12). In this research, concordance lines are provided for subject-specific 4-word clusters to display their complementation patterns and their semantic and functional characteristics in individual corpora. Concordances for content words and 4-word clusters which occur across corpora are also examined to compare and contrast the uses and meanings of these features across subject-specific contexts.

Overall, the quantitative and qualitative analysis of subject-specific corpora facilitated by WordSmith 4 results in extensive and detailed descriptions of language features and patterns. These language specifications yield two types of linguistic description: subject-specific and cross-curricular. At the level of individual subjects, insights can be obtained into i) vocabulary use (the most frequent content and function words and collocations), ii) phraseological patterns (the most frequent 4-word clusters, their complementation patterns, structures and meanings). At a cross-curricular level, lexical comparisons among subject-specific corpora of the same register illuminate i) common lexical words with the same or different uses and sense(s), and ii) common 4-word clusters. In addition, the examination of the lexical features of individual subjects in the two different registers of textbooks and examination papers further reveals any differences that might exist between these two fundamental school registers. Taken together, these corpus-based findings offer a rich database of empirically-derived linguistic information for the language support classroom.

2.7 Conclusion

The purpose of this chapter was to justify the need for the present empirical research, to delineate its purposes and scope and describe the methodology and tools that were used for its implementation. The need to analyse the language of Junior Cycle subject textbooks and Junior Certificate examination papers was justified by the fundamental role these registers play in students' academic performance (cf. section 2.2; Chapter 1, section 1.2.3) and the difficulties they cause to ESL students and language support teachers (cf. Chapter 1, cf. section 1.3). Based on a consideration of different possible theoretical frameworks, it was argued that the analysis of subject-specific discourse from a textual perspective could address the concerns of this research, focusing on the register features of subjects and their functional values in discourse. Then, based on a brief overview of the contributions of corpus linguistics to L2 education, it was explained why the linguistic analysis that is envisaged in this study could not be more effectively conducted by using any other methodology. The successive stages of the adopted methodological procedure were outlined next, placing particular emphasis on corpus design and development milestones and the software program that was used for corpus analysis.

An important point that has not been yet stressed, but becomes evident in Chapter 3, is the indispensable role of the human analyst to the successful implementation of the procedures described in this chapter. An unquestionable fact for all corpus-linguistic studies, including the present one, is that 'the human analysis of text extracts [is] central at every stage: despite the use of sophisticated computational and quantitative techniques, all stages of the analysis [are] shaped and guided by human observation of [language] features in natural discourse contexts' (Biber et al., 1999: 38). Human intuition is 'an essential tool for extrapolating important generalizations from a mass of specific information in a corpus' (Hunston, 2002: 22), and technical knowledge and skills and critical analytical abilities are essential to evaluate corpus design, methodology and findings, avoid bias and exemplify the practical applications of findings.

3 The linguistic features of Junior Cycle subject textbooks and Junior Certificate examination papers

3.1 Introduction

Following the description of the corpus methodology employed to analyse the language of Junior Cycle subject textbooks and Junior Certificate examination papers in Chapter 2, this chapter reports the findings of the quantitative and qualitative analysis of subject-specific textbooks and examinations corpora.

The discussion is divided into two parts. The first part draws frequency-based lexical profiles of the twelve subject-specific corpora. These comprise vocabulary items and lexis, namely: i) the commonest function and content words of individual corpora, ii) significant collocates of the most frequent content words, iii) significant collocations extracted as single units from each corpus, and iv) 4-word clusters, which are analysed in respect of their densities, forms, structures, complementation patterns and pragmatic meanings and functions in subject-specific discourse. The second part of the discussion establishes lexical linkages among textbooks corpora and among examinations corpora by identifying content words and 4-word clusters which are (in)variably used across the six corpora of each register. The meanings and contextual uses of cross-curricular content words in different subject-specific contexts are described with reference to their central collocates and concordance lines. The semantic and functional associations of cross-curricular clusters are analysed based on their complementation patterns revealed by concordancing. A critical commentary on the results forms an integral part of the discussion which often draws upon subject syllabuses and guidelines for post-primary subject teachers as well as upon the relevant literature on ESP and the language of school subjects in general for the interpretation and explanation of results.

Overall, the findings from the empirical analysis of the twelve corpora demonstrate the lexical variation that exists across the six subjects and between the two registers and reveal that the preferred lexical realisation of each corpus is largely influenced by the communicative and thematic needs of particular texts. The chapter concludes by drawing together the key findings from corpus analysis, and by adding a caveat for the interpretation and generalisability of these findings for the characterization of subject-specific language registers as a whole.

3.2 The commonest vocabulary items: words, collocations and 4-word clusters

This section draws frequency profiles of the twelve subject-specific corpora which comprise single words, collocations and 4-word clusters together with information on their contextual use. The rationale behind frequency information is that it indicates the most important language to be taught in language support, the assumption being that, the more frequent certain language items are, the greater the need is for ESL students to use these in the subject classroom. Frequency of language also correlates with learnability, on the evidence that, the more times students encounter a word, the more easily and faster they will understand it in their textbooks and learn how to use it productively. Research reveals, in fact, that second language vocabulary learning is frequency-driven (e.g., Milton, 2009; see also Hoey's 2005 theory of lexical priming), i.e. students learn the most frequent words first, regardless of the learning burden of individual words, i.e. 'the amount of effort required to learn it' (Nation, 2001: 24).

Highly-frequent words also normally have wide text coverage; a fact which increases their pedagogical value on the grounds that ESL students' learning of the most frequent lexical items can facilitate their linguistic access to and comprehension of subject textbooks and examination papers (see Nation, 2001 and Milton, 2009 for a discussion of the strong relationship between vocabulary text coverage and comprehension). When applied to the study of word clusters in particular, an additional strength of a frequency-driven approach is that it reveals 'the extent to which a sequence of words is stored and used as a prefabricated chunk, with higher frequency sequences more likely to be stored [in the mental lexicon] as unanalyzed chunks than lower frequency sequences' (Biber et al., 2004: 376). This has direct implications for ESL teaching and learning since learning how to use these chunks contributes to fluent linguistic production (e.g. Pawley and Syder, 1983).

In the following sections, the most frequent subject-specific vocabulary and lexis are decribed and discussed separately for textbooks- and examinations-derived corpora from the following perspectives: i) function and content words (section 3.2.1.1), ii) content words (section 3.2.1.2), iii)the central collocates of the top content words (section 3.2.2.2), iv) significant collocations (section 3.2.2.3), and v) 4-word clusters (section 3.2.3). The most frequent 4-word clusters of the twelve corpora are specifically analysed in respect of i) their density, forms and frequency (section 3.2.3.1), ii) their structural correlates (section 3.2.3.2), and iii) their complementation patterns (i.e. items that follow them), meanings and functions

in subject-specific texts (section 3.2.3.3). Prior to the presentation of findings from frequency analysis, it is essential to note some methodological considerations at this point.

It is, first of all, important to define what counts as a word as this is 'a key research issue in all vocabulary analyses' (Biber, 2006: 31). In this research, the basic units of analysis are types (distinct words) rather than tokens (running words in the text), or the more abstract units of lemmas (i.e. 'the base word and its inflections'; Schmitt, 2000: 2) and heads of word families (i.e. 'the base word, all of its inflections, and its common derivatives'; ibid.). The amount of data that were extracted and analysed was controlled by restrictions of frequency which were set for reasons of practical manageability, making sure at the same time that language data would be adequate to be exploited for pedagogical purposes in the language support classroom. More specifically, a relatively high cut-off point of frequency of occurrence was set (a minimum frequency of 40 times per million words adjusted to the size of individual corpora) for the analysis of both words and word clusters. The frequency counts of all words and clusters presented in the following sections are normalized rates of occurrence per 1000(‰) words of text based on the following formula: (number of occurrences of the type/number of tokens in entire corpus) x 1000. Normalized rather than raw frequencies are provided to compensate for the uneven size of the twelve corpora and allow direct comparisons of data. It is important to further add that although reporting findings as percentages demonstrates 'proportional use' and does not reveal exact frequencies, it is still a better approach because 'most people find it easier to understand and compare figures such as percentages than fractions of unusual numbers' (McEnery and Wilson, 1996: 68).

In addition to minimum frequency requirements, range (i.e. consistency of use across texts in a corpus) was also used as a criterion for selecting the words and clusters to be closely examined. An occurrence in at least 75% of the texts that comprise a single textbooks-based corpus (i.e. in at least three out of four textbooks) was thus set as a minimum and a minimum range of 60% in the examinations-based corpora (i.e. in at least six out of ten examination papers). This was essential to ensure that the idiosyncratic language use of the authors of textbooks and examination papers does not influence the results. Taken together, frequency and range restrictions served to isolate the most frequently and consistently used language features.

To refine the final version of the wordlists to be examined and used for language support purposes, certain items (most of which have narrow range) were manually removed, namely, proper nouns - names of places (e.g. *Dublin*) and people (e.g. *Shakespeare*) - , Irish (e.g. *Taoiseach*) and other non-English words (e.g. *conquistadors*), symbols (#), single letters (e.g. *O, T,* etc.), contractions (e.g. *isn't*), genitive cases with an apostrophe (e.g. *people's*),

numbers (e.g. IV). Following the deletion of non-word items (symbols, numbers, single letters), the frequency counts of all remaining words which had been automatically computed by WordSmith, were manually re-calculated using the Microsoft Office Excel functionality. These were subsequently normalized to their occurrence per mille words of text, as explained above. The shortcoming of the corpus-derived wordlists in this research is that homographs, i.e., 'different words spelt alike' (Leech et al., 2001: 6) are not distinguished and therefore some entries are ambiguous at the level of lexical word categories (nouns, verbs, adjectives, adverbs; e.g. *writing, thought, lie*, etc. function both as verbs and nouns; similarly, auxiliaries are not separated from lexical verbs, i.e. *have, be* and *do*). Word-sense disambiguation did not take place either (e.g. the word *may* can either function as an auxiliary modal verb denoting ability or permission, or refer to the fifth month of the calendar). Finally, phrasal verbs were not detected as the analysis captured single word items only. These ambiguities remain because, as explained in Chapter 2, the corpora that were analysed with the WordList tool of WordSmith 4 (Scott, 2004) were not annotated and to resolve these ambiguities manually would be a prohibitively labour-intensive task.

Due to space restrictions, the presentation of data in this chapter is necessarily selective. The corpus data which are explicitly referred to in this chapter are included in the second volume of this thesis which constitutes the appendix with the complete data sets that derived from the analysis of the twelve corpora.

3.2.1 Words

3.2.1.1 Function and content words

The examination of the most frequently-used vocabulary items was the starting point for the macroscopic analysis of the twelve corpora. As Sinclair argues, 'anyone studying a text is likely to need to know how often each word-form occurs in it' (1991: 30). Word frequency analysis provides a preliminary orientation to the text collections that form the subject-specific corpora, facilitating an understanding of vocabulary which can be further developed by more elaborate qualitative analysis at the level of concordancing (cf. section 3.3.1). Departing from words rather than grammar and syntax is consistent with the lexical approach (Lewis, 1993) to language description and analysis that is adopted in this research. This emphasizes the 'utility'

and 'power' of frequent words because, as explained by Willis: 'The commonest patterns in English occur again and again with the commonest words in English' (1990: 38) and '... the commonest and most important, most basic meanings in English are those meanings expressed by the most frequent words in English' (ibid.: 47-48), therefore, 'If we are to provide learners with language experience which offers exposure to the most useful patterns of the language, we might well begin by researching the most useful words in the language' (ibid.: 38).

The initial frequency-based wordlists that were computed using the WordList tool of WordSmith 4 (Scott, 2004), comprising all types in the twelve corpora, confirmed what is typical of the wordlist of almost any corpus (based on at least a few hundreds of words); that it contains 'a very small number of very highly used items, and a long declining tail of items which occur infrequently, with roughly half occurring once only as hapax legomena' (Scott and Tribble, 2006: 27-29)[1]. This latter point can be illustrated by considering, for example, that there is a total of 8,688 hapax legomena out of the 19,941 types in the English textbooks corpus which is the largest corpus in this research; but in the smallest textbooks-corpus of CSPE too, its 6,440 hapax legomena also account for approximately half of its size (13,213 types).

Although the study of low-frequency words and hapax legomena can often reveal interesting facts about the language and content of a text, the present research is concerned only with the most frequent vocabulary items because of their value (cf. section 3.2). This is the reason why the original wordlists are not examined here and are thus not included in the appendix of this thesis. To exclude low frequency items, new wordlists were computed using the minimal frequency and range restrictions mentioned above. The wordlists that emerged contained approximately 2,000 words each (the top 100 words in the twelve corpora are provided in Appendices B1 and B2 together with information on frequency and range). Tables 3.1 and 3.2 below show the 20 most frequent words in the six textbooks- and the six examinations-based corpora in a descending rank order.

Perhaps the most striking finding from examining the words in the two tables below, as well as the wordlists in Appendices B1 and B2, is that the most frequent items in the twelve corpora are,in their vast majority, closed-class or function words such as articles (*the, a*), prepositions (*in, on, to*), conjunctions (*and*) and auxiliary verbs (*is, are*), with few exceptions of content, or lexical, words, i.e. nouns (*water*) , lexical verbs (*find*), adjectives (*new*) and adverbs

[1]This phenomenon has been explained by the American philologist George Kingsley Zipf who examined the statistical regulations of language use. According to his best known Law (1965), there is 'a constant relationship between the rank of a word in a frequency list, and the frequency with which it is used in a text' (Crystal, 2003: 87) which applies regardless of subject matter, author, and language.

(*well*). It can be seen from the complete frequency wordlists (Appendices B1 and B2) that it is specifically after the top 50 items that function words begin to become scarce and lexical words denser. The floating of function words to the top of the frequency wordlists is a common phenomenon in almost all vocabulary analyses. This is evidently because 'such grammatical words are necessary to the structure of English regardless of the topic' (Schmitt, 2000: 73).

Table 3.1 The 20 most frequent words in the six textbooks-based corpora (cut-off frequency point - 40 per million words, range - 100%).

English	the, a, and, to, of, in, you, is, I, it, he, that, was, for, on, this, his, with, what, are
Geography	the, of, and, in, a, to, is, are, on, as, for, this, from, that, it, or, by, at, have, they
History	the, of, and, in, to, a, was, were, that, for, by, they, on, as, he, from, their, people, had, with
CSPE	the, of, to, a, and, in, is, for, that, are, you, on, people, I, it, what, have, be, or, this
Mathematics	the, of, a, and, is, in, to, find, for, cm, are, number, calculate, on, if, at, as, each, by, line
Science	the, of, a, and, to, is, in, water, are, it, for, on, from, by, with, that, this, as, be, or

Table 3.2 The 20 most frequent words in the six examinations-based corpora (cut-off frequency point - 40 per million words, minimum range - range 100%).

English	the, of, a, and, you, to, in, I, this, for, is, your, or, it, on, that, answer, was, what, he
Geography	the, of, a, in, and, to, is, page, correct, box, this, map, tick, at, answer, one, which, or, on, are
History	the, of, in, a, and, to, one, was, that, two, from, answer, an, picture, during, or, name, by, what, on
CSPE	the, of, to, in, a, marks, you, and, for, your, page, this, that, is, section, people, on, name answer, one
Mathematics	the, of, a, in, and, to, is, page, find, for, cm, box, on, text, that, at, number, your, be, each
Science	the, of, a, in, to, is, name, diagram, and, for, what, one, an, use, on, give, that, this, from, water

The high frequency of function words is responsible for their wide text coverage which, as mentioned earlier, increases their pedagogical value. The approximately 270 function word types (176 word families) that exist in the English language, in fact, account for 43-44% of the running words in most texts (Johansson and Hofland, 1989; Francis and Kučera, 1982) and they are thus considered a prerequisite for effective language use. In addition, although function words do not carry any real semantic value, their unusually high frequency in a corpus 'may indicate important features of the discourse' (Nation, 2001: 206). One example is the use of the personal pronoun *you* in the English textbooks- and exams-based corpora which has a normed frequency of 13.11‰ and 21.68‰ in the two corpora respectively, and which appears

to be considerably less frequent in the other corpora (below 5‰; except in the CSPE textbooks and exams corpora with a frequency of 8.38‰ and 9.55‰ respectively). A possible explanation for the dense use of this pronoun could be that 'the writer typically addresses his message directly to the reader to involve the reader in the topic' (ibid.), contributing to a more personal and conversational tone of language use in this particular subject register.

Another observation that can be made based on the most frequent words presented above concerns the content words that are featured in the listings. The few content words that appear among the top 20 items of the six textbooks-based corpora (Table 3.1, e.g. *people, find, cm, number, calculate, line, water*) provide some clues for the information content of these texts; while the slightly bigger number of content words that are included in the commonest 20 words of the six exams-based corpora (Table 3.2) appear to be exam-specific (e.g. *answer, page, name, marks*) rather than linked to subject-specific content.

On the whole, frequency-based lists of function and content words shed some light on vocabulary use in subject-specific corpora. A more focused and pedagogically useful vocabulary analysis, however, could be based on content words because of their increased value compared to function words, as argued in the following section.

3.2.1.2 Content words

Content or lexical words (nouns, lexical verbs, adjectives and adverbs) are the 'main carriers of meaning' in language and they accordingly function as 'the main building blocks of texts' (Biber et al., 1999: 55); as opposed to function words (such as determiners, prepositions, pronouns, conjunctions and auxiliary verbs) which are used to make language grammatical and add no real semantic substance to spoken or written texts. More importantly, lexical words comprise the specialised vocabulary of curriculum subjects which has a gate-keeping function for students because it provides access to entire knowledge categories (Corson, 1985). These facts highlight the 'utility' and 'power' (Willis, 1990: v) of the lexical words of post-primary curriculum subjects for subject-specific language learning in language support and render them indispensable elements of ESL students' target lexical repertoire.

A stop list of function words was therefore used (based on Nation, 2001: 430-431, see Appendix B3) as a filter to eliminate 'noise' (i.e. to exclude the closed-class items) and the most frequent content words were in this way extracted from each corpus using WordSmith's WordList tool (Scott, 2004). The 20 most frequent content words in the subject-specific

textbooks and examinations corpora are presented in Tables 3.3 and 3.4 (the top 100 content words in the twelve corpora can be viewed in Appendices B4 and B5 with information on frequency and range).

Table 3.3 The 20 most frequent content words in the six textbooks-based corpora (cut-off frequency point - 40 per million words, range - 100%).

English	story, like, write, said, think, answer, poem, words, people, time, make, read, good, give, see, writing, just, use, new, well
Geography	people, map, population, areas, water, area, countries, city, river, land, following, figure, South, world, sea, large, photograph, new, high, explain
History	people, war, Irish, government, century, British, new, world, became, used, made, called, year, land, army, great, church, time, following, work
CSPE	people, action, community, rights, local, world, work, school, government, project, European, make, Irish, class, children, council, day, group, groups, think
Mathematics	find, centimetres, number, calculate, line, area, value, equation, point, example, given, following, solution, angle, sides, length, graph, diagram, solve, sine
Science	water, energy, used, light, air, food, test, heat, experiment, called, plants, oxygen, tube, gas, place, method, carbon, plant, acid, current

Table 3.4 The 20 most frequent content words in the six examinations-based corpora (cut-off frequency point - 40 per million words, minimum range - 80%).

English	answer, page, section, think, poem, write, give, like, questions, following, question, paper, explain, story, extract, read, turn, describe, play, name
Geography	page, correct, box, map, tick, answer, population, name, photograph, diagram, area, explain, shows, survey, ordnance, describe, shown, questions, following, provided
History	answer, picture, name, page, following, give, people, write, source, mention, marks, questions, account, war, paper, explain, document, reasons, book, Irish
CSPE	marks, page, section, people, name, answer, question, questions, action, write, class, poster, CSPE, describe, school, rights, campaign, give, community, European
Mathematics	page, find, cm, box, number, marks, calculate, value, write, answer, graph, shown, work, point, line, area, total, triangle, diagram, length
Science	name, diagram, use, give, water, page, labelled, section, question, shows, used, answer, experiment, examiner, part, shown, right, describe, energy, food

What can be immediately seen in Tables 3.3 and 3.4, as well as in the content word lists in Appendices B4 and B5, is that the commonest content words in all corpora are in close relationship with the information content of the texts in which they are found. Specifically the

content words in textbooks-based corpora are clearly topic-related, reflecting the distinct thematic concerns and common concepts of the different subject areas in which they occur, and they are thus markedly different across subjects. The limited sample of the top 20 subject-specific words provided in Table 3.3 can illustrate this point which is confirmed by a brief glance at the full content wordlists of subject textbooks (Appendix B4). Thus, in the English textbooks corpus, the commonest content words highlight the literary dimension of the subject which is primarily concerned with *stor[ies], and poem[s], words, writing*, etc., inviting students to *write, think, answer, read*, etc. The majority of the commonest words in the geography textbooks corpus reveal the focus of geographical study on people (*people, population*) and the environment (*area(s), country (-ies), city, sea, river, land*, etc.), reflecting the aim of the subject, as stated in the Junior Cycle syllabus: 'the study of people and their relationships with the environment'. Similarly, content words in the history textbooks corpus are associated with the study of people (*army, government*) and events (*war*) from the past (*century, year, time*). The highly frequent occurrence of past tense verbs in the wordlist of this same corpus reflects the pivotal role of past tense to historical meaning-making (e.g. studying past events, people, ways of life, etc.). CSPE content words denote civic, social and political agents (*community, government, council*, etc.) while the most frequent items in mathematics describe major mathematical concepts (*number, value, equation, sine*) and actions (*find, calculate, solve*). The commonest science-specific words reveal concepts important to scientific study (i.e. substances and materials, e.g. *water, energy, light, air, oxygen, tube, gas*, etc.) and terms that attest to its experimental nature (e.g. activity verbs such as *test, experiment, heat, place, used*, etc.). Finally, the occurrence of 'people' in the top 20 words of the textbooks corpora of English, geography, history and CSPE but not in the mathematics and science textbooks corpora is not accidental either, as it is the humanities and social science textbooks that normally discuss people. In short, the most frequent content words in the six textbooks-based corpora are characteristic of individual subject areas.

A slightly different pattern can be observed in the case of exams-based content wordlists. The highest ranks in the content wordlists of examinations-based corpora (Table 3.4) are occupied by items which are more neutral than those appearing in the textbooks-based content wordlists, in the sense that they are not charged with subject-specific meanings. These are words inextricably linked to assessment tasks and instructions in examination papers (e.g. action verbs such as *answer, write, tick, describe)*, or words that refer to the components of an exam paper (e.g.*page, question, marks, section*). This trend results from the commonsense fact that the content of exam papers which comprise the corpora in question contain instructions, tasks, exercises of examinations which are typically formulated in more generic terms using

language expressions that are common in the examination papers of different curriculum subjects (the same does not apply, however, to the source texts provided in English, history, geography and CSPE papers). Although it cannot be displayed by the data provided in the appendix of this thesis due to space constraints, it is worth mentioning that after the first 100 words of the six lists, these exam-specific words disappear and all remaining words are once again content-based and, to a great extent, they overlap with those in the textbooks-based wordlists.

3.2.1.3 Conclusions from the word frequency analysis

The general conclusion that emerges from the word frequency analysis of the twelve corpora is that, regardless of the type of text in which they occur, words are not randomly selected and they are not infinitely various. Instead, word choice is conditioned and restricted by the information content of a particular text (e.g. Hyland and Tse, 2009). The close interrelation of vocabulary use and subject-matter demonstrated by the analysis of content words in the textbooks and examination papers of curriculum subjects disproves the widespread assumption that there is a type of 'general English' to be taught in language support (cf. Chapter 1, section 1.3). The present vocabulary analysis clearly demonstrates that curriculum language has a subject-specific nature as, even at the most basic level of words, there is noticeable variation across subjects and distinct sets of lexical items are used both in the textbooks and examination papers of individual subjects. More specifically, it was shown that the topic vocabulary that is used in the textbooks-based corpora offers hints to the thematic units of subject textbooks, and the assessment-related vocabulary that is prominently featured in the exams-based corpora serves to mediate examination instructions and tasks. It was further argued that the consistent appearance of function words to the top of the frequency lists of all twelve corpora shows that these are fundamental in any text and that their acquisition is essential for effective language use in general. The distinct sets of frequent content words that are valued by the six subjects, on the other hand, are fundamental to accessing disciplinary knowledge and to constructing subject-specific meanings.

On the whole, the vocabulary items that have been described and discussed so far begin to reveal characteristics of subject-specific language which provide the first type of pedagogically useful information offered by this research for post-primary language support.

Subject-specific wordlists reflect, however, only a tiny sample of the language they represent and they are therefore considered as a starting point for further analysis. Having identified the most frequent single word items in the twelve corpora, this research proceeds to providing further insights into the lexical dimension of corpora beyond single words to word pairs (collocations; section 3.2.2) and to multi-word patterns (4-word clusters, section 3.2.3).

3.2.2 Collocations

Although it is natural to begin the vocabulary analysis of subject-specific corpora with words, it would be misleading and ineffective to confine this empirical examination to single word items because subject-specific words do not exist in a vacuum but 'in the environment of textual context' (Schmitt, 2000: 96) within which 'every single word occurs in the company of other words' (McEnery and Wilson, 1996: 71) and 'the use of a word [...], conditions the choice of the next, and of the next again' (Sinclair, 2004: 19). This results in strong patterns of word co-occurrence, that is, 'combination[s] of two words [or, more precisely, of word types] that exhibit a tendency to occur near each other in natural language' (Evert, 2008: 1214) more frequently than would be the case due to chance (combinations of more than two words are also typical in natural language, however, resulting in recurrent multi-word patterns, as will be discussed in section 3.2.3).

The concept of collocation was introduced by Palmer (1933: 4) to refer to words whose meaning cannot be deduced from their parts: 'Each [collocation] ... must or should be learnt, or is best or most conveniently learnt as an integral whole or independent entity, rather than by the process of piecing together their component parts'. The term was later used by Firth to emphasize that the meaning of a word depends on its collocates: 'You shall know a word by the company it keeps' (1957: 179). The collocational behaviour of words was fully documented and impressively demonstrated by Sinclair (Sinclair et al., 1969; reprinted in Krishnamurthy, 2004; 1987; 1991) in his lexicographic analysis of English for the COBUILD Project (Sinclair, 1995). Today abundant research evidence from corpus studies highlights the pervasiveness of collocations in language (e.g., Williams, 1998; Biber et al., 1999 Chapter 13; Hunston and Francis, 2000; Gledhill, 2000; Marco, 2000; Stubbs, 2001; Oakey, 2002; Hoey, 1991; 2005), illustrating the system of relations words form (lexical, grammatical, semantic, discoursal; Sinclair, 1996; Stubbs, 2001). Studies also indicate the psycholinguistic reality of collocations (e.g. Kjellmer, 1990; Schmitt, 2000: 79) and argue that their use contributes to the fluent

linguistic production of both native and non-native speakers on the grounds that the use of pre-constructed word units minimises cognitive processing (e.g. Pawley and Syder, 1983; cf. Nation, 2001: 336ff.). On this evidence, collocations are considered fundamental for language learning, knowledge and appropriate use and have therefore become 'an accepted aspect of vocabulary description and pedagogy' (e.g. Lewis, 2000; McCarthy and O'Dell, 2005).

For all of the above reasons, the analysis of subject-specific collocations in subject textbooks and examinations corpora forms an integral part of the present research. Their identification becomes increasingly important here, compared to other studies of language for general purposes (LGP). This is because in LSP (Language for Specific Purposes) the basic unit of meaning is longer than a single word (Bowker and Pearson, 2002: 26) and further because 'LSP contains a number of specialized terms, combined in special ways [...] that differ from LGP'(ibid.).

Despite the consensus on the pervasiveness and importance of collocations, scholars have put forward different definitions of the term (for an overview see Bartsch, 2004), based on different understandings of what counts as a collocation. There is thus no consistent system of classification and many different methods are used to extract collocations from corpora. It is not the aim of this section, however, to delve into the relevant literature but instead to operationalise the concept as it is understood in the present research and to discuss important methodological choices involved in the computation and analysis of collocations.

This research adopts a data-driven approach to analysing collocations that is in line with the Firthian tradition (e.g. Firth, 1957; Sinclair, 1991) which views collocations as 'a directly observable property of natural language' (Evert, 2008: 1214) rather than a theoretical concept 'defined by linguistic tests and speaker intuitions' (ibid.). It is therefore important to explain that the empirical collocations analysed here are those which result from word co-occurrence by surface proximity, i.e. 'two words are said to co-occur if they appear within a certain distance or *collocational span*, measured by the number of intervening word tokens' (ibid.: 1220). They are therefore distinguished from textual collocations (words which co-occur in the same textual unit; ibid.: 1222) and syntactic collocations (i.e. when only a single type of syntactic relation is considered; ibid.: 1224). The collocations examined here also differ from the word associations that exist in the mental lexicon of language users (cf. Scott and Tribble, 2006: 34-40 for the difference between mental linkage of words and textual word co-occurrence). Empirical collocations typically have a dual manifestation: as 'a tendency of one word to attract another' or as a 'tendency of two words to co-occur' (Hunston, 2002: 68). The subject-specific collocations of the twelve corpora in the present research are accordingly analysed from two perspectives: i) as node-collocate pairs, i.e. the collocates of the most

frequent content words of textbooks and examinations corpora and ii) as single units, i.e. 'the most strongly collocated word pairs' in the twelve corpora (Evert, 2008: 1217).

In both cases, the collocations that are extracted are limited to occurrences within a five-word span. Although in early corpus research, a four-word proximity (both to the left and the right of a word type) was viewed as the optimal span size (e.g. Jones and Sinclair, 1974) today technology makes it possible to identify collocations that extend beyond 25 words or even more (e.g. cooccurrence within Web documents and Google searches) and therefore the collocational span can be varied, depending on the aim of the study. In this research, a span of five words on either side of the node is examined because, as, Sinclair et al. convincingly demonstrate, less useful information emerges when searching for collocates beyond the four or five word span and 'the wider the span, the lower is the significance in general' (2004: xxvii).

Although grammatical words are often eliminated from analysis in collocation studies (e.g. Lewis, 2000), because they are perceived as less interesting than lexical ones, this is not the case in the present research. Both lexical and grammatical collocates are taken into account because corpus studies have repeatedly shown that lexical items form strong collocational patterns with particular grammatical forms (e.g. Francis and Hunston, 2000: 96; Sinclair et al., 2004: 69; Hoey, 2005: 40). It would therefore be misguided to exclude function words from collocation analysis as these are equally important for an accurate description of the lexical behavior of words: 'Significance collocation [...] can be a result of grammar as much as of lexis, due to the frequency of certain grammatical structures in the language. Since every word enters into some kind of grammatical relationship, this will be reflected in its collocational pattern' (Sinclair et al., 2004: 62). Another important reason, from a pedagogical perspective, for not excluding grammatical collocations is that they are arguably challenging for second language learners and they consequently deserve explicit focus in the classroom to expose students to the most frequent grammatical patterns of the language.

Collocations were removed, however, if they included an acronym or abbreviation, contracted forms (e.g. *it's*), genitives formed with an apostrophe (e.g. *world's*), a number or ordinal (other than *one, two, three* and *first, second, third*), days, months, the hash symbol (#) and, finally, when they appeared twice in the same list of significant collocations (e.g. *questions-answer* and *answer-questions* are frequent in both directions so one is removed). In addition, the present analysis does not account for the possibility of a node co-occurring with itself within the specified span which is, in any rate, a rarity (e.g. *reason-reason*).

The number of possible collocations for even a limited number of words is colossal and many word pairings are formed due to chance or because they are simply high frequency words. To reduce the total to a more manageable number and, more importantly, to filter out

idiosyncratic word pairs, the raw frequency of co-occurrences cannot be trusted because it tends to overvalue probabilistic pairings of highly frequent words (McEnery and Wilson, 2001: 217). A more reliable measure is instead needed to calculate the relative strength of the association of two words and identify true collocations. The statistical association measure used for this purpose is discussed in the following section.

3.2.2.1 Collocation extraction with the Log Likelihood ratio

There are currently over 50 different association measures for the extraction of collocations in corpus linguistics (see Clear, 1993; Barnbrook, 1996: 87-101; McEnery and Wilson, 1996; Biber et al., 1998: 265-268; Oakes, 1998; Evert, 2008; Pecina, 2005, etc.) but they result in entirely different rankings of word pairs. Mutual Information (MI; Church and Hanks, 1990: 23) in particular but also the z-score (Dennis, 1965: 69; Berry-Rogghe, 1973: 104) and [Student's] t-score (Church et al., 1991 section 2.2) are three of the most commonly used formulae in collocation studies but the log-likelihood ratio (abbreviated as G^2; Dunning, 1993) seems to have gained ground more recently as the predominant measure of word association (e.g. Leech et al., 2001: 16; Rayson, 2008). It cannot be argued, however, that one metric is better than the other or that there is a single best one in all cases: 'Different measures highlight different aspects of collocativity and will hence be more or less appropriate for different tasks' (Evert, 2008: 1236). A strategy for identifying the most appropriate measure for the extraction of collocations in a particular research study is to examine the empirical linguistic properties of collocations (ibid.: 1238); in other words, to compare the kind of word pairs that are identified as collocations by the different association measures. In the present research, the G^2 was found to be most suitable, following its comparison with MI.

MI is a simple association measure which falls into the group of effect-size measures and it measures the strength of the attraction of two words while G^2 is a statistical association measure and belongs to significance measures and determines how high or low the probability is that the a word pair is formed due to chance (McEnery and Wilson, 1996: 69). To illustrate the different linguistic aspects of collocativity that the two association measures capture, Tables 3.5 and 3.6 provide listings of collocations according to MI and G^2 scores that were computed based on two corpora of this research. The higher the MI and G^2 scores are, the stronger the attraction between the two words. The tables present i) individual words with their raw frequency of occurrence, ii) the range of the word pair ('Texts'), iii) the number of

Table 3.5 The top 15 significant collocations in the English textbooks corpus based on MI and G^2 scores.

MI collocations									G^2 collocations							
N	Word 1	Freq.	Word 2	Freq.	Texts	Gap	Joint	MI	N	Word 1	Freq.	Word 2	Freq.	Texts	Gap	Joint	G^2
1	turnips	12	clover	10	3	2	11	14.9849	1	twentieth	280	century	746	4	1	275	3455.755
2	Mona	12	Lisa	10	4	1	10	14.8474	2	such	411	as	1982	4	1	337	3168.657
3	Ardagh	12	Chalice	11	4	1	10	14.70989	3	set	355	up	668	4	1	263	3013.938
4	Fulacht	10	Fiadh	17	2	1	12	14.60793	4	world	578	war	1155	4	1	293	2636.57
5	sans	14	culottes	16	3	1	15	14.53189	5	De	258	Valera	165	4	1	165	2524.005
6	mein	15	kampf	14	4	1	14	14.52547	6	northern	364	Ireland	1656	4	1	268	2502.074
7	Aos	15	dana	13	3	1	13	14.52547	7	during	610	the	27402	4	1	544	2387.07
8	solemn	10	covenant	15	4	3	10	14.52547	8	home	316	rule	293	4	1	176	2201.26
9	Basil	14	Brooke	16	4	1	14	14.43236	9	international	211	relations	228	3	1	156	2196.244
10	Marco	17	Polo	16	3	1	16	14.3449	10	Fianna	138	Fail	135	4	1	129	2184.879
11	Marie	17	Antoinette	13	4	1	13	14.3449	11	more	594	than	363	4	1	183	1899.124
12	Albert	16	Reynolds	15	3	1	14	14.33282	12	junior	136	history	296	4	2	126	1777.723
13	Ignatius	17	Loyola	12	4	2	11	14.21937	13	middle	216	ages	140	4	1	117	1686.349
14	Del	16	Cano	13	3	1	10	14.05385	14	prime	109	minister	177	4	1	105	1642.255
15	Trinity	13	College	16	3	1	10	14.05385	15	did	830	not	740	4	1	202	1634.366

Table 3.6 The top 15 significant collocations in the science examinations corpus based on MI and G^2 scores.

MI collocations									G^2 collocations							
N	Word 1	Freq.	Word 2	Freq.	Texts	Gap	Joint	MI	N	Word 1	Freq.	Word 2	Freq.	Texts	Gap	Joint	G^2
1	local	17	studies	17	5	1	17	11.8603	1	examiner	220	only	236	10	2	218	2758.7
2	cut	16	flowers	16	3	1	13	11.6482	2	examiner	220	use	454	10	1	219	2278.48
3	north	15	pole	20	4	1	15	11.6258	3	list	156	right	204	10	3	134	1528.91
4	scruduithe	21	examinations	21	8	3	21	11.5555	4	list	156	on	437	10	1	139	1325.69
5	stait	21	commission	21	8	3	21	11.5555	5	can	168	be	240	10	1	124	1268.75
6	coimisiun	21	stait	21	8	3	21	11.5555	6	describe	187	aid	96	8	3	96	1179.09
7	coimisiun	21	scruduithe	21	8	2	21	11.5555	7	with	278	aid	96	8	2	96	1079.84
8	scruduithe	21	commission	21	8	4	21	11.5555	8	carbon	85	dioxide	77	10	1	74	1056.25
9	scruduithe	21	stait	21	8	1	21	11.5555	9	aid	96	labelled	358	8	3	90	911.346
10	examinations	21	commission	21	8	1	21	11.5555	10	describe	187	with	278	10	1	103	905.204
11	stait	21	examinations	21	8	2	21	11.5555	11	experiment	239	to	1068	10	1	144	877.49
12	local	17	candidates	22	5	2	17	11.4883	12	aid	96	diagram	765	7	4	95	840.0773
13	studies	17	candidates	22	5	1	17	11.4883	13	shown	207	diagram	765	10	3	120	799.6541
14	candidates	22	afternoon	13	4	4	13	11.4883	14	used	286	to	1068	10	1	143	795.1831
15	grain	10	direction	22	4	1	10	11.4883	15	shown	207	in	1372	10	1	136	792.6407

intervening words ('Gap'), iv) the frequency of the word pair ('Joint'), and v) the corresponding association scores based on the two formulae.

Comparing the outcome of the two tests, it can be argued that the G^2 represents collocations better than the MI in that it reveals more frequent and more recognizable patterns. Looking at the MI list for the English textbooks corpus, it can be seen that 12 out of the 15 collocations are proper names (e.g. *Mona Lisa, Fulacht Fiadh* etc.) and the rest are fixed and highly specialised lexical collocations (e.g. *turnips clover*) - no grammatical collocations appear in the list. The MI collocations in the science examinations corpus are also fixed word combinations (e.g. the repetition of *state examinations* and *stait scruduithe*) and they further include some accidental co-occurrences (e.g. *local candidates*). It is also noteworthy that most, if not all, MI collocations in both the English textbooks and science examinations corpora are 'infrequent words with low co-occurrence frequency' (Evert, 2008: 1218) which is close to the frequency threshold that was set (frequency ≥ 10).

When looking at collocations according to G^2, on the other hand, a more balanced picture emerges. The listings for both corpora bring to the surface both grammatical and lexical collocations, with only three proper names occurring in the English textbooks corpus (*De Valera, Northern Ireland, Fianna Fail*) and none in the science examinations corpus. Although grammatical collocations (word pairs that include at least one non-lexical word such as prepositions, determiners, primary and modal verbs, conjunctions, subordinating adverbs, pronouns etc., e.g. *set up, write beside, describe with* etc.) may not seem very interesting from a linguistic point of view, they are nevertheless pedagogically useful features, as explained in the section 3.2.2. Unlike MI collocations, the frequencies of individual words that form the G^2 collocations and their joint frequencies are very high. It is also important to note that allG^2 collocations are widely distributed (with a range of occurrence in four subject textbooks and in at least eight examination papers), as opposed to the MI listings which include many collocations with narrower range. Comparisons of collocations according to MI and G^2 scores based on the other subject corpora that were developed for this research yield similar results.

The above conclusions echo the criticisms of the strengths and limitations of these two measures expressed by many other researchers, namely that 'MI brings up highly specialised terms but also many obviously accidental co-occurrences' (Evert, 2008: 1238); it 'gives too much prominence to rare combinations' (Lindquist, 2009: 76) and that, like z-score and t-score, MI tends to 'substantially overestimate the significance of infrequent words' (cf. e.g. Dunning, 1993; Hunston, 2002: 70-75; McEnery et al., 2006: 216-217; Lindquist, 2009: 76-78) while 'G^2 appears to provide more realistic collocation information' (McEnery and Wilson, 2001: 217). Considering that the present research aims to give a balanced view of significant collocations,

assigning therefore equal importance to grammatical and lexical collocations rather than focusing on specialised lexical collocations which have low frequency and narrow range, the G^2 ratio was used for the computation of node-collocate pairs and the extraction of the most significant collocations in the twelve corpora. Finally, to reduce data to a manageable load, further parameter values were set. These include a frequency threshold of ten occurrences of the two co-occurring words and of their joint frequency in the corpus and a minimum range of occurrence of significant collocations of 75% and 60% in individual textbooks and examinations corpora respectively. The top 20 collocates of the ten most frequent content words of the twelve corpora are displayed in section 3.2.2.2 and the top 20 most significant collocations in the twelve corpora are presented in section 3.2.2.3 (the full collocations listings are presented in Appendix C).

3.2.2.2 The central collocates of the most frequent content words in the textbooks and examinations corpora

Identifying the commonest content words of the twelve corpora is important but not sufficient in its own right for effective vocabulary teaching and learning in English language support. Information on the characteristic 'lexical and grammatical friends' of these content words in subject-specific contexts is an integral part of word description and knowledge. This section describes the collocational behaviour of the ten commonest content words of the twelve corpora, that is, 'their patterns of combination with other words' (Partington, 1998: 2). More specifically, a ranking approach is adopted according to which each of the ten most frequent words (i.e. node word) is characterized by ranked lists of its top 20 (grammatical and lexical) collocates (Evert, 2008: 1217) based on G^2 scores. The higher the G^2 score is, the greater the significance value of the node-collocate pair. The full profiles of the collocational behaviour of the 15 most frequent contentwords in the twelve corpora are presented in Appendices C1 and C2 with statistical information displayed in a table format similar to the following:

N	Word	With	Relation	Total	Total Left	Total Right
1	ELECTRICAL	energy	1779.05	181	140	41

The first column (N) indicates the rank of the node-collocate pairs which are sorted by descending significance value based on G^2 scores as explained above. The second and third

columns ('Words', 'With') present the collocate and the node word respectively and beside these, the strength of their relationship is specified ('Relation'). The total number of times the collocate co-occurs with the node in the corpus is presented in the fifth column ('Total'), and a total for Left and Right of the node is given next. Although it is not shown above, a detailed breakdown, showing the total number of times the collocate appears up to five words on either side of the node, is also provided in the final columns of each table in Appendices C1 and C2. These detailed collocational statistics are useful because they summarise information that would be difficult to obtain based on the examination of node-collocate pairs in concordance lines. It therefore becomes very easy, for instance, to categorise the collocates of each search-word according to their position within the collocational span; thus, based on the example illustrated above, *electrical* can be classified as a left collocate of *light* as used in the science textbooks corpus.

Tables 3.7 and 3.8 below list the 20 most significant collocates of the ten most frequent content words in the twelve corpora. Information on statistical significance and frequency of occurrence has been removed from these tables to facilitate a more qualitative treatment of collocates. As in the frequency-based wordlists of the twelve corpora, several grammatical words appear as the top collocates of the most frequent content words of textbooks in particular; e.g., articles (*a, n, the*), prepositions (*in, on, at*) and pronouns (*you, I, he, she* etc.). These word linkages in which a dominant word 'fits together' with a grammatical word (typically a noun, verb or adjective followed by a preposition) result in grammatical collocations which are also referred to as colligation (Scott, 2004: 84; Hoey, 2005). From an examination of the full lists of collocates of these high-frequency content words, it can be seen that lexical collocates become predominant after the twentieth rank in most listings. Lexical collocations are understood as combinations of two basically 'equal' words such as noun + verb, verb + noun, and adjective + noun in which both words contribute to the meaning (Benson, 1985). The part of speech of the node word evidently influences the grammatical category of the collocates it attracts; thus, node words which are, for instance, nouns (e.g. *people*) collocate strongly with attributive adjectives (e.g. *young, homeless*) and verbs (e.g. *lived, had* etc.), while verbs which are transitive (e.g. *answer*) collocate with nouns (e.g. *questions*) and, less frequently, with adverbs (e.g. *then*). Beyond this, however, function words crop up at the top of all lists regardless of the node word.

Table 3.7 The top 20 collocates of the ten most frequent content words in the textbooks corpora ranked by descending order of significance based on G^2 scores (collocates are enclosed in diamond brackets; node = lexical word under analysis).

Subject corpus	Node-collocate pair
English	**story** <the, a, of, in, is, this, to, short, you, tells, or, tell, about, write, set, from, novel, that, read, your> **like** <a, the, you, to, I, what, would, that, it, of, is, this, or, in, was, just, do, dislike, be, she> **write** <a, the, you, to, about, your, paragraph, of, out, for, in, words, letter, story, down, article, an, two, sentences, own> **said** <he, she, I, the, to, mother, you, a, my, and, it, was, that, yes, uncle, me, father, in, as, Mr> **think** <you, do, I, the, is, about, what, of, that, to, would, it, best, this, why, they, which, a, be, are> **answer** <the, questions, your, following, reasons, follow, give, for, that, then, a, two, this, question, sample, to, of, read, support, poem> **poem** <a, in, the, this, of, is, read, you, or, to, about, questions, from, answer, with, carefully, following, that, line, reading> **words** <the, in, phrases, of, are, that, you, to, using, which, write, a, or, use, key, these, following, with, your, about> **people** <the, to, of, in, are, who, young, a, many, about, how, or, that, they, for, as, be, would, most, on> **time** <the, at, a, of, to, for, in, was, it, first, is, all, that, this, by, he, long, place, I, had>
Geography	**people** <in, of, to, the, live, who, many, for, have, million, young, are, from, work, their, a, migrate, that, areas, living> **map** <the, on, of, sketch, as, extract, ordnance, survey, draw, shown, OS [ordnance survey], area, evidence, page, figure, study, photograph, to, carefully, use> **population** <the, density, of, growth, densities, in, change, distribution, a, high, is, has, diversity, low, pyramid, very, world, to, increase, total> **areas** <in, of, are, rural, the, to, urban, lowland, inner, city, these, such, as, pressure, upland, from, people, low, high, some> **water** <the, of, supplies, is, clean, to, in, supply, a, fresh, from, cycle, for, air, as, into, by, it, condenses, rivers> **area** <the, of, in, a, an, map, shown, this, on, is, calculate, sketch, to, photograph, urban, pressure, for, one, that, around> **countries** <in, the, developing, of, rich, world, developed, poor, are, to, such, many, poorer, as, have, aid, south, poorest, third, these> **city** <the, inner, centre, of, in, a, to, areas, primate, is, suburbs, town, has, are, as, capital, housing, from, or, county> **river** <the, a, of, on, its, in, basin, valleys, flows, is, point, valley, flood, course, or, stream, which, bed, bridge, meander> **land** <the, of, is, use, in, flat, on, to, uses, agricultural, or, by quality, from, sloping, land, buildings, values, which, for>
History	**people** <the, to, of, in, were, many, lived, that, who, more, Irish, history, most, for, ordinary, their, about, a, had, poor> **war** <the, world, of, in, civil, cold, during, after, independence, to, ended, end, was, Korean, on, a, years, American, outbreak, when> **Irish** <the, to, of, in, volunteers, people, Gaelic, state, agreement, free, by, as, an, were, parliament, British, republic, a, language, government> **government** <to, of, British, in, a, Ireland, party, by, that, coalition, act, was, Irish, northern, Fianna, had, fail, set, formed, up> **century** <in, international, relations, change, early, social, political, year, of, during, end, first, beginning, decades, a, second, third, BC [Before Christ], AD [Anno Domini], years> **British** <government, to, army, of, forces, in, by, troops, Irish, prime, that, parliament, Americans, French, American, the, commonwealth, soldiers, monarch, minister>

	new <the, a, of, to, in, were, towns, world, called, built, as, party, lands, ideas, was, orders, constitution, routes, introduced, government> **world** <war, the, of, view, European, in, during, after, exploration, discovery, to, new, outbreak, was, ancient, a, on, parts, had, that> **became** <the, of, in, a, popular, he, more, as, known, leader, very, when, who, minister, so, many, famous, later, soon, they> **used** <to, the, were, was, for, they, as, be, a, by, could, in, steam, methods, tactics, which, of, that, describe, make>
CSPE	**people** <the, young, to, of, in, who, for, are, have, a, by, that, with, many, about, community, homeless, their, older, from> **action** <project, the, an, report, your, on, of, a, you, involvement, this, undertaken, in, part, community, to, for, marks, as, take> **community** <the, in, a, local, of, involvement, to, your, people, action, school, alert, neighbourhood, games, for, European, centre, as, members, groups> **rights** <human, the, of, responsibilities, declaration, to, universal, convention, on, child, abuses, are, for, in, UN [United Nations], dignity, a, that, protection, protect> **local** <the, authorities, authority, a, community, to, government, of, in, your, for, are, councillors, representatives, their, people, councillor, or, with, development> **world** <the, are, we, in, around, small, of, day, a, caring, developing, for, changing, to, third, crossword, war, countries, is, parts> **work** <the, to, of, course, in, module, their, a, for, assessment, on, with, book, who, about, hours, people, they, is, together> **school** <in, your, to, the, a, community, students, for of, or, go, class, my, on, at, children, number, from, day, an> **government** <the, of, local, to, is, a, in, by, on, policy, coalition, departments, head, for, opposition, Taoiseach, Irish, environment, ministers, has> **project**<action, the, report, on, an, your, a, of, this, doing, part, you, as, was, my, based, in, marks, did, title>
Mathematics	**find** <of, value, to, equation, values, image, a, slope, point, for, coordinates, area, solve, line, graph, if, hence, use, we, number> **cm[centimetre]** <of, is, radius, height, length, a, the, its, volume, diameter, by, cylinder, area, has, sphere, width, breadth, calculate, base, in> **number** <the, of, is, a, who, to, in, people, times, by, pupils, whole, larger, that, prime, total, on, per, less, second> **calculate** <the, of, its, value, area, to, radius, height, amount, mean, interest, total, length, cm, volume, tax, a, in, correct, earned> **line** <a, mirror, is, of, on, segment, equation, draw, to, through, point, symmetry, axis, slope, axial, perpendicular, find, parallel, meet, in> **area** <the, of, rectangle, is, triangle, calculate, find, volume, its, cm, base, circle, total, perimeter, parallelogram, length, height, width, maximum, in> **value** <the, of, find, calculate, minimum, maximum, to, estimate, corresponding, after, equation, is, write, or, down, hence, for, frequency, point, exact> **equation** <the, of, solve, an, to, find, in, line, write, this, given, down, form, quadratic, information, is, a, represent, we, value> **point** <the, of, on, a, is, intersection, cut, off, line, standard, find, rate, minimum, under, image, central, to, symmetry, coordinates, decimal> **example** <for, the, consider, of, is, in, solve, simplify, next, find, type, and, common, if, image, evaluate, sides, has, express, cm>
Science	**water** <the, distilled, of, in, to, a, is, with, boiling, vapour, beaker, from, dioxide, hard, into, salt, carbon, tap, test, that> **energy** <electrical, heat, to, the, of, is, food, into, chemical, light, in, stored, from, sound, form, potential, sources, by, source, sun> **used** <to, apparatus, is, chemicals, be, the, for, in, are, a, of, can, measure, test, it, an, small, two, separate, make> **light** <the, a, to, of, energy, rays, is, ray, white, from, bulb, that, it, in, travels, refraction, resistor, when, diode, towards> **air** <the, in, to, is, of, oxygen, water, from, dioxide, show, carbon, than, into, vapour, through, with, that, contains, glass, a> **food** <the, of, energy, to, chain, in, chains, is, for, make, web, plants, their, from, types, own, down, photosynthesis, digested, animals> **test** <tube, the, tubes, for, a, to, in, of, water, solution, into, each, used, with, chemicals, starch, presence, paper, stopper, apparatus>

heat <energy, the, of, to, a, is, in, loss, transfer, water, from, electrical, by, through, conductor, conduction, gently, show, conductors, chemical>
experiment <to, show, the, this, demonstrate, investigate, that, of, in, a, we, an, will, water, simple, find, examine, mandatory, describe, repeat>
called <is, the, a, are, this, of, which, tiny, they, in, an, to, particles, cells, process, by, these, because, sometimes, often>

Table 3.8 The top 20 collocates of the ten most frequent content words in the examinations corpora ranked by descending order of significance based on G^2 scores (collocates are enclosed in diamond brackets; node = lexical word under analysis).

English	**answer** <your, questions, question, the, two, reference, give, reasons, then, and, for, follow, support, following, either, one, which, to, sections, read> **page** <turn, over, on, paper, section, advertisement, picture, at, the, one, minutes, find, look, coimisiun, each, carefully, adapted, appears, writing, view> **section**<over, turn, read, of, writing, page, this, each, functional, fiction, minutes, poetry, personal, studies, media, drama, reading, carries, answer, write> **think** <you, do, a, about, is, the, what, this, why, or, I, would, that, of, poem, have, it, novel, good, poet> **poem** <this, you, the, poetry, studied, answer, name, choose, and, read, of, poet, have, or, which, a, from, questions, in, then> **write** <a, composition, the, personal, writing, to, one, you, of, letter, about, speech, on, for, any, section, conversation, prose, stated, in> **give** <reasons, your, for, answer, reason, the, you, title, must, a, of, two, poem, choice, name, play, based, would, text, opinion> **like** <you, to, a, would, I, it, dislike, or, the, what, and, this, that, did, she, was, looked, me, about, one> **questions** <the, answer, then, follow, and, following, which, two, carefully, piece, poem, of, answers, that, passage, extract, three, any, one, paper> **following** <the, of, answer, questions, one, two, edited, any, form, topics, extract, read, answers, drama, in, find, on, Shakespearean, titles, is>
Geography	**page** <over, box, tick, correct, the, either, answer, section, text, next, coimisiun, question, blank, warning, on, continues, in, and, to, of> **correct** <box, of, page, circle, each, statements, or, in, answer, are, option, numbered, therefore, a, north, terms, should, is, numbers, the> **box** <tick, correct, page, of, over, or, selection, ticking, warm, in, metres, statements, are, industrial, west, front, is, economic, north, south> **map** <the, survey, ordnance, on, sketch, shown, and, study, examine, shows, look, of, draw, at, supplied, area, key, legend, a, this> **tick** <page, or, a, activity, is, north, front, west, stack, east, isobars, rainfall, metres, statements, are, lines, linear, creep, delta, south> **answer** <either, questions, the, your, or, book, circle, of, three, follow, each, marks, return, correct, with, in, following, folder, geographical, mix> **population** <the, of, in, total, change, has, shows, structure, growth, pyramid, and, low, groups, a, age, males, which, between, largest, examine> **name** <one, the, and, of, two, feature, labelled, formed, country, river, given, features, in, explain, type, coastal, climate, way, erosion, you> **photograph** <aerial, study, this, of, shown, the, in, on, and, a, map, supplied, is, evidence, survey, using, ordnance, look, from, examine> **diagram** <a, the, shows, of, in, this, is, examine, shown, formed, feature, below, how, with, above, glacial, deposition, erosion, common, study>
History	**answer** <your, book, paper, following, this, and, a, in, the, marks, accompany, then, separate, history, over, with, include, return, examination, help> **picture** <shows, a, is, questions, of, page, following, the, pictures, from, source, and, evidence, in, an, answer, piece, artist, painting, give> **name** <one, the, two, of, leader, that, which, from, renaissance, and, hints, reformer, associated, tomb, stages, given, in, who, British, outside> **page** <of, turn, over, picture, source, document, answer, book, your, name, give, a, the, in> **following** <the, of, questions, answer, a, terms, explain, then, two, paper, relating, and, one, account, picture, marks, document, to, in, question> **give** <why, two, one, reason, reasons, evidence, the, of, piece, from, to, pieces, example, renaissance, century, your, picture, consequences, there, were> **people** <the, described, select, below, in, one, of, history, were, marks, to, many, answer, that, lived, who, dangers, and, why, saw> **write** <of, about, account, an, person, the, that, title, selected, down, one, at, two, need, since, fact, read, a, you, not> **source** <primary, of, picture, a, page, is, from, Duchas, Corbis, heritage, the, marks, service, secondary, national, which, study, what, by, extract>

	mention <two, one, in, of, the, change, or, by, major, apart, century, ways, taken, which, effects, consequences, changes, dangers, faced, have>
CSPE	**marks** <of, page, question, in, carries, two, section, questions, each, describe, the, total, answer, any, name, three, this, apart, from, one> **page** <the, a, printed, for, over, separate, you, has, section, answering, been, bonus, Comhairle, fairly, of, on, marks, top, will, total> **section**<question, answer, for, of, questions, in, all, marks, the, three, examiner, page, one, read, mark, leaflet, poster, total, answering, tick> **people** <homeless, of, young, to, the, group, with, in, who, encourage, disabilities, that, and, about, have, older, is, are, some, community> **name** <the, of, two, that, each, and, other, you, marks, three, one, actions, political, may, explain, a, organisation, party, only, write> **answer** <questions, the, section, this, in, any, studied, of, you, below, all, have, book, one, paper, three, back, extra, a, marks> **question** <section, for, carries, marks, each, please, superintendent, poster, extra, examiner, answer, the, ask, paper, examination, mark, any, tick, brochure, leaflet> **questions** <answer, numbered, below, this, studied, all, any, section, three, in, a, marks, of, what, poster, you, are, one, included, ask> **action** <second, first, project, explanation, marks, name, one, take, your, could, describe, this, to, EU, out, that, of, would, you, CSPE> **write** <a, the, your, spaces, answers, in, short, sure, each, to, make, you, down, for, number, speech, examination, marks, beside, would>
Mathematics	**page** <of, a, for, part, on, extra, work, grade, total, given, geometry, formulae, use, coimisiun, diagram, next, axis, reason, are, answer> **find** <of, value, the, area, to, volume, in, and, hence, cm, otherwise, exact, ordinates, cost, equation, slope, values, or, height, correct> **cm** <measures, and, of, height, the, wide, radius, is, has, find, text, by, diameter, in, length, with, block, box, sphere, cylinder> **box** <cm, Oideachais, Roinn, an, shown, in, the, is, of> **number** <the, of, whole, total, to, nearest, scruduithe, coimisiun, students, marks, who, days, per, mean, estimate, tiles, people, pupils, hours, in> **marks** <question, obtain, carries, full, level, each, morning, be, paper, shown, scruduithe, coimisiun, stait, state, to, ordinary, total, higher, other, questions> **calculate** <of, in, mean, area, length, value, interest, average, tax, speed, profit, interval, total, hence, as, amount, gross, for, the, percentage> **value** <find, when, exact, estimate, a, otherwise, and, or, diagram, calculate, down, to, write, in, of, cos [cosine], minimum, value, tan [tangent], the> **write** <down, the, of, in, a, an, as, expression, simplest, your, and, answer, its, equation, using, to, value, cos [cosine], form, length> **answer** <your, give, form, reason, in, to, correct, giving, simplest, a, the, nearest, for, its, decimal, express, write, places, and, two>
Science	**name** <the, a, of, part, that, one, labelled, two, letter, and, used, equipment, beside, piece, type, substance, device, in, found, which> **diagram** <shows, a, labelled, of, in, shown, aid, an, experiment, to, box, compare, text, circuit, show, apparatus, rounded, rectangle, question, human> **use** <for, one, give, of, question, the, number, page, name, a, mark, examination, you, this, in, headings, graph, biology, have, everyday> **give**<one, for, of, use, reason, a, two, the, your, function, and, advantage, why, this, way, everyday, an, example, examples, disadvantage> **water** <of, the, in, a, to, hard, and, is, hardness, oil, added, was, test, tube, boiled, salt, bath, that, level, solution> **page** <blank, examiner, question, for, section, only, over, turn, use, provided, extra, total, box, labelled, diagram, totals, end, grand, biology, back> **labelled** <diagram, the, aid, of, with, experiment, an, name, to, box, compare, part, text, show, in, section, parts, how, device, and> **section**<flowchart, marks, page, mark, question, this, questions, answer, of, biology, diagram, terminator, rounded, box, labelled, there, three, rectangle, text, are> **question** <a, page, of, examiner, use, only, answering, terminator, number, the, section, flowchart, biology, diagram, are, clearly, choose, you, mark, or>

shows <the, diagram, a, of, an, human, rectangle, rounded, system, apparatus, photograph, set, cell, three, circuit, structure, digestive, reproductive, experiment, up>

The lexical collocations which are prominently featured in the above tables convey information on the content of the text in which the node-collocate pair co-occurs because, as shown in the previous section, all content node-words are closely linked to the thematic needs of textbooks and examination papers. Thus, for example, the top ten content words of the CSPE textbooks corpus are all subject-specific, e.g. *people, action, community, rights, local, world, work, school government, project*, whereas those of the CSPE exams corpus can be characterised as exams-specific, e.g. *marks, page, section, name, answer, question, questions, write* (perhaps with the exception of *people* and *action*).

Comparing, therefore, the node-collocate pairs of textbooks corpora with those of exams corpora as a whole, a similar pattern emerges once again to that from the comparison of high-frequency content words in the two registers: in the textbooks corpora, significant word pairings are tied to subject-specific content while in the exams corpora, many node-collocate pairs are linked to the language of assessment and testing; for example, *homeless people, community involvement, human rights, third world, coalition government* etc. occur in the list of CSPE textbooks collocations whereas *marks total, page separate, answer section, question superintendent, write examinations* etc. can be found in the CSPE examinations corpus. Several collocations are evidently common across examinations corpora, e.g. *answer questions, section marks*, etc.

The dependency of the collocational behaviour of content words on the context of occurrence explains why the same node appears in different contexts with different sets of collocates every time (cf. section 3.3.1 on cross-curricular words). One example is the different collocates of *people* in the textbooks corpora of geography (e.g. *live, million, young, work, migrate*), history (e.g. *were, lived, Irish, ordinary*) and CSPE (e.g. *young, community, homeless, older*) corpora. It is worth pointing out here that 'there are entirely different sets of collocates for each sense of the node word' as well (Evert, 2008: 1219) but these distinctions are not made explicit in this research because separating collocations for different word senses automatically would require sense-tagged corpora.

The above observations clearly illustrate that words are idiosyncratic and that their full meaning can be determined only by considering their collocates. On these grounds it is argued that collocational knowledge should be part of word knowledge: 'Knowing a word involves knowing what words it typically occurs with' (Nation, 2001: 56; cf. Firth's (1957) theory of contextual meaning, Naggy's (1997) discussion of context-based meaning and the contribution of collocation, in psychological terms, to meaning emphasized by Aitchison (1994: 21): 'humans learn word-meaning from what occurs alongside'). The fact that meaning is context-based and that subject-specific words cannot be taught or learned divorced from their context

of use provides further support for the argument made at the end of section 3.2.1, that English language support must be subject-specific rather than general.

On the whole, collocational analysis in this section shows that lexical collocations and grammatical collocations which are formed on the basis of the most frequent content words in textbooks and examinations corpora are topic-specific, reflecting the information content of the corresponding corpus. The analysis of significant collocations as single units in the twelve corpora in the following section reveals that the same holds true even when a content word is not selected as a node word, showing therefore topic-specificity to be a characteristic property of all collocations.

3.2.2.3 The most significant collocations in the textbooks and examinations corpora

The analysis of the most significant collocations in the twelve corpora according to the G^2 ratio reveals similar but also some additional facts about the nature and properties of collocations in subject-specific texts. Tables 3.9 and 3.10 display the 20 most significant collocations in the six textbooks and the six examinations corpora respectively (the full lists of the top 100 collocations in the twelve corpora can be viewed in Appendices C3 and C4).

Table 3.9 The top 20 most significant collocations in the six subject textbooks corpora ranked by descending order of significance based on G^2 scores (frequency threshold F≥ 10, minimum range 75%).

Textbooks corpus	Collocations
English	at the, do you, from the, what is, do think, your answer, what do, give reasons, what you, will be, have been, answer questions, write, a, out of, would be, look at, had been, if you, going to, questions follow
Geography	such as, ordnance survey, per cent, have been, grid reference, more than, your answer, can be, what is, case study, may be, inner city, has been, acid rain, number of, they are, between and, these are, population density, sketch map
History	twentieth century, such as, set up, world war, De Valera, Northern Ireland, during the, Home Rule, international relations, Fianna Fail, more than, junior history, middle ages, prime minister, did not, relations century, how did, between and, political in, what was
CSPE	action project, human rights, United Nations, young people, find out, by the, has been, set up, there are, European Union, number of, answer questions, will be, do think, can be, right to, do you, more than, this is, asylum seekers
Mathematics	calculate the, value of, on the, how many, area of, both sides, write down, number of, each following, calculate of, correct to, standard rate, surface area, image under, under symmetry, at the, image of, Venn diagram, average speed, tax paid
Science	carbon dioxide, test tube, can be, Bunsen burner, experiment to, as shown, such as, set up, higher level, this is, what is, apparatus used, number of, used to, retort stand, living things, at the, result conclusion, centre gravity, through the

Table 3.10 The 20 most significant collocations in the six examinations corpora ranked by descending order of significance based on G^2 scores (frequency threshold F≥ 10, minimum range 70%).

Examinations corpus	Collocations
English	turn over, give reasons, do think, then questions, short story, do you, section read, reasons for, question two, give for, give your, reasons your, novel short, questions follow, which follow, turn section, either or, novel story, reasons answer, then answer
Geography	ordnance survey, either or, look at, survey map, ordnance map, aerial photograph, otherwise lost, return with, Junior Certificate, otherwise will, marks be, will lost, you have, will be, certificate examination, junior examination, examinations commission, scruduithe commission, stait commission, scruduithe examinations
History	turn over, your answer, paper answer, middle ages, twentieth century, account of, reasons why, your book, Junior Certificate, history level, accompany then, described below, coimisiun scruduithe, coimisiun stait, scruduithe stait, extract from, world war, scruduithe examinations, stait examinations, write about
CSPE	first second, first action, printed separate, human rights, will read, when studied, could take, young people, printed on, below carries, separate on, each carries, second marks, has been, studied answer, describe two, when have, action marks, top page, when you
Mathematics	your answer, write down, value of, give your, correct to, how many, Junior Certificate, supporting work, Junior examination, certificate examination, or otherwise, examination level, number of, examination mathematics, Junior mathematics, certificate mathematics, supporting be, space extra, mathematics level, obtained superintendent
Science	examiner only, examiner use, list right, list on, can be, describe aid, with aid, carbon dioxide, aid labelled, describe with, question a, experiment to, aid diagram, shown diagram, used to, shown in, piece equipment, answer questions, choose from, shown the

With specific regard to collocations in the six textbooks corpora, it can be seen from Table 3.9 and from a closer examination of the top 100 most significant collocations in the six textbooks corpora in Appendix C3 that many frequent content words that were ranked among the top 50 content words in the word frequency analysis of textbooks corpora (see section 3.2.1.2 and Appendix B4) are members of several collocations. The following table presents some of these content words and the collocations to which they belong.

Table 3.11 Examples of significant collocations that comprise frequent content words (marked in bold) in the six textbooks corpora (minimum range 75%).

Textbooks corpus	Collocations
English	**reasons** <for, your, answer>, **answer** <the, questions, follow, following>, **give**<reasons, for, your, answer>
Geography	**map**<draw, ordnance, survey, OS [ordnance survey], sketch, extract, page>, **people**<in, live, who, to>, **population**<density, densities, growth, distribution>
History	**century**<Ireland, change, nineteenth, twentieth, relations, change>, **change**<social, century, Ireland, in>, **war** <civil, cold, world, during>
CSPE	**action** <an, your, report>, **civic**<social, political, education>, **rights**<responsibilities, human>
Mathematics	**symmetry** <in, central, under>, **rate**<standard, cut, off>, **equation** <of, an, solve>
Science	**apparatus** <set, up, used, shown>, **shown**<that, in, the, diagram, experiment, apparatus, as>, **used**<to, be, apparatus, chemicals>

This pattern of the most frequent content words being embedded in the most frequent and significant collocations clearly demonstrates Sinclair's argument that 'the most common words are responsible for the most frequent word patterns' and it also provides evidence for the fact that the more frequent a word is, the more collocations it formulates. Thus, as shown in Table 3.11, the words *map*, *century*, and *shown* which are the second, fifth and thirtieth commonest words in the geography, history and science textbooks corpora respectively, belong to at least five different collocations. Exhaustive analysis of the collocations of the six textbooks corpora could usefully reveal words which appear with a large number of collocates and those with a small set of particular collocates (cf. Kilgarriff (2006) who used data from the British National Corpus to identify 100 nouns and verbs which show the 'most collocational' behaviour - for nouns, with respect to the verbs they are objects of, and for verbs, with respect to their object-nouns).

All textbooks-derived collocations in this research communicate the concerns of the subject corpus in which they occur. Some grammatical collocations, those with two grammatical words in particular, convey more subtle information: e.g. The fact that *such as* occurs so prominently in geography, history and science seems to indicate that exemplification is more important in these than in the other three subjects. Both the lexical and the grammatical collocations reflect concepts that are central to particular subjects, they reveal terms and also tools and instruments important to the study of the particular subjects, and they further mediate fragments of instructions that are typically used in the subject classroom. Accordingly, collocations in the English textbooks corpus refer mainly to asking and answering questions and expressing and justifying opinions (e.g. *give answer/reasons, reasons answer, questions follow, reasons for, what is, your opinion, point view, do think, what happened* etc.) as well as to writing tasks (e.g. *personal writing, words phrases, write about* etc.). The density of grammatical collocations which include pronouns in the English and CSPE textbooks conveys more subtle information about the style and tone of language that appears to be conversational, encouraging the active involvement of students and their subjective viewpoints. The CSPE textbooks corpus is further characterised by a wide range of collocations that refer to human, social and political entities (e.g. *Country Council, Defence Forces, Fine Gael, Amnesty International, United Nations, European Parliament, European Union, Bord Pleanala, asylum seekers, political parties, member states, local authorities, young people*). Collocations which mark the notion of time (e.g. *twentieth/nineteenth century, middle ages, Bronze Age* etc.), past events (e.g. *Industrial Revolution, Cold War*) and proper names of places and people (e.g. *Northern Ireland, Soviet Union, United States, Fianna Fail, De Valera, Cumann NGaedheal*) appear to be prominent in the history textbooks corpus. Most collocations of the geography textbooks corpus, on the other hand, are associated with people (e.g. *people live, population density, birth rate*) and places (e.g. *inner city, rural areas, West Ireland, Third World, developing countries*), geographical terms/phenomena (e.g. *fold mountains, mass movement, acid rain, shaping crust*), and resources and materials central to geography study (e.g. *OS [Ordnance Survey] map, aerial photographs, grid reference*).

Moving from the more theoretically-oriented to the more technical and activity-based subjects, the nature of collocations becomes more specialised. The most significant collocations in the mathematics corpus fall into two general categories; they either name mathematical concepts (e.g. *common factor, central symmetry, sine cosine, frequency distribution* etc.) or they are combinations of verbs and prepositions or nouns that denote mathematical processes (e.g. *multiple by, calculate cm, solve equation, correct to* etc.). In the same vein, most collocations in the science textbooks corpus either refer to important

scientific labels (e.g. *magnetic field, current flow, centre gravity, atmospheric pressure* etc.) and objects of scientific study such as chemical substances and materials (e.g. *carbon dioxide, cobalt chloride, iron filings, fossil fuels* etc.) and instruments used in experiments (e.g. *test tube, graduated cylinder, Bunsen burner, conical flask, bar magnet* etc.) or they can be linked to describing and carrying out experiments (e.g. *set apparatus, dilute acid, test for, what happens, using a, result conclusion, pass through, if you*).

Interestingly, although most textbooks-based collocations are subject-specific, as shown so far (with a few exceptions of some more neutral ones such as *give reasons, answer questions*), in the mathematics and science textbooks corpora certain highly specialised collocations stand out. Such instances are, for example, *under (a/the) translation* and *mean of* in mathematics and *gas jar* and *retort stand* in science. Although they comprise common everyday words, these word combinations acquire a different and more technical meaning in the discourse of the two subjects. These collocations are thus meaningful only within restricted contexts of use and seem odd outside their subject-specific domain. These instances attest to Bowker and Pearson's (2002) argument mentioned earlier (section 3.2.2) that LSP combines words in specialised ways that differ from the way they are used in LGP. These collocations are instances of Firth's 'more restricted technical ... collocations' (1957: 195), as opposed to 'general or more usual collocations'. In the same vein, Sinclair (1966) offers the examples of *vigorous depressions* and *dull highlights* which are respectively employed in the discourse of meteorology and photography. Moon, on the other hand, discusses the adjective *torrential* arguing that it must be learned as part of the lexical unit *torrential rain* (1997: 43). These common everyday English words which acquire extended meanings when used within restricted sets of lexical units in technical contexts pose difficulties to native and non-native speaking students alike (Tribble, 1985: 129-130; Nation, 2001: 191).

The peculiar nature of these collocations calls for an examination of their usage within subject-specific texts to provide an understanding of their meaning as this cannot be easily inferred from contextual cues. The following examples illustrate some of their typical contextual uses (collocations are indicated in italics):

Mathematics:

- Find the image of the point (2, 3) *under* the *translation* (1, 4) ~ (4, 2).
- *Under* a *translation*, the object moves along a given straight line.
- The shaded triangle is mapped onto B *under* the *translation* (3, 2) ~ (9, 4).

- The *mean of* a set of values is the sum of all the values divided by the number of values.
- Use the answer above to calculate the *mean* amount *of* money saved.

- Find the mode and the *mean of* each of the following frequency distributions.

Science:

- Suspend the magnet from the *retort stand* as shown.
- Place the thermometer, held in a *retort stand*, close to the tin can as shown in Fig. 15.
- Hang the metre stick from the *retort stand*, at the 50 cm mark until it is balanced.

- Pour a *gas jar* of carbon dioxide from one gas jar into an empty gas jar.
- Plunge the burning charcoal into a *gas jar* of oxygen.
- Invert the *gas jar*, filled with water, and place it on the beehive shelf.

On the basis of these examples it could be argued that it might not be possible for ESL students to derive the precise meaning of some of these restricted collocations (e.g. *under translation*) even if they are provided with contextual cues. ESL students would need to examine the visuals that accompany the sentences in which the aforementioned collocations occur, particularly when these describe technical equipment (e.g. *retort stand, gas jar*). Restricted collocations might be more easily deciphered, however, when they are explicitly defined (e.g. the first example of the collocation *mean of* above). The status of these collocations as specialised terminology unique to mathematics and science makes them pedagogically valuable features and necessitates excplicit instruction in the classroom.

With reference to the most significant collocations in the six examinations corpora, it can be once again observed that several frequent collocations include some of the 50 most frequent content words of the corresponding examinations corpora identified in their word frequency analysis (section 3.2.1.2). To illustrate this point, some examples are provided in Table 3.12 below.

With reference to the semantic value of examinations-specific collocations, it can be argued that, in their vast majority, they primarily reflect examinations instructions for candidates and, to a more limited extent, they convey thematic concerns of subject areas (it is interesting that the English collocations of both textbooks and exams corpora are similar cf. Tables 3.11 and 3.12, indicating how the English subject is conceptualised). Considering that the examination papers of the different subjects follow the same format and contain similar instructions and guidelines for candidates, some common collocations across examinations corpora can be detected. More specifically, what immediately stands out in all six examinations corpora is a set of all possible combinations of collocations based on the standard information that is provided (both in Irish and in English) on the first page of all examination papers, i.e. *Coimisiun na Scruduithe Stait / State Examinations Commission*, Junior

Certificate Examination, and the title of the subject (hence the collocations *certificate - examination, coimisiun - scruduithe, commission - examinations*, etc.).

Table 3.12 Examples of significant collocations that comprise frequent content words (marked in bold) in the six examinations corpora (minimum range 60%).

Examinations corpus	Collocations based on frequent content words
English	**reasons** <give, for, your, answer>, **answer** <give, question, reasons, support, then>, **give**<reasons, for, your, answer>
Geography	**North**<West, East, South>, **marks**<will, be, lost, questions, otherwise>, **folder**<your, with, return>
History	**paper**<which, accompany, answer, then, book >, **people** <described, select, below, about>, **write** <below, described, about, an, account, person>
CSPE	**actions** <could, your, describe>, **class**<to, decided, has, could>, **printed**<separate, on, a, page>
Mathematics	**work** <must, be, shown, extra, indicates, space, supporting, symbol>, **supporting**<symbol, indicates, work, must, be>, **question** <carries, marks, attempt, all>
Science	**choose** <a, word, from, the>, **describe** <with, the, aid, of>, **shown**<as, in, the, diagram>

Table 3.13 Collocations that are associated with subject-specific instructions for candidates and examination tasks in the six examinations corpora.

Examinations corpus	Collocations
English	read section, read carefully, support answer, write composition
Geography	match each, circle answer/correct, name one, describe two
History	write account, mention two, select one, evidence to
CSPE	tick box, describe actions, include slogan, study the
Mathematics	correct to/nearest, represent information, how many/much
Science	compare the, describe aid, choose from, list the

Another category of collocations that appear to be common comprises collocations which mediate instructions such as *please enclose, turn over, return folder*. There is, nevertheless, variation in the instructions-related collocations across exams-based corpora which results from the different types of tasks and exam questions of individual subject areas. Table 3.13 offers some illustrative examples of this variation. It may be added here that the top 100

collocation list of the CSPE exams corpus was the only one that had a limited number of verb-based collocations, as opposed to the top 100 collocations of the other five examinations corpora.

Although collocations that are related to examination instructions and guidelines prevail in the 100 most significant collocations of the six examinations corpora (and these typically include action verbs), there is a small number of lexical collocations which indicate subject-matter and differentiate the six collocation listings. These include the following:

- **English**: *short story, functional/personal writing, reference text*
- **Geography**: *ordnance survey, sketch map, aerial photograph, north east/west*
- **History**: *Middle Ages, world war, Industrial Revolution, twentieth century, international relations*
- **CSPE**: *first/second action, human rights, young people, United Nations, class decided, civic education*
- **Mathematics**: *value of, correct to/decimals, graph paper, simplest form, area of*
- **Science**: *carbon dioxide, with aid, piece equipment, test tube, switch closed*

Interestingly, no restricted collocations were found in any of the six examinations corpora. Taking into account the high frequency and wide range of occurrence of this small set of content-based collocations, it can be argued that these are very informative because they illuminate recurrent subject-specific features (concepts, issues, text types) that have been the object of assessment in Junior Certificate examination papers from 2001 to 2010.

3.2.2.4 Conclusions from the collocational analysis

This section described and discussed the most frequent collocations in the twelve subject-specific textbooks and examinations corpora from a double perspective, focusing first on the 20 collocates of the 15 most frequent content words in the twelve corpora and second, on the 100 most significant collocations in the twelve corpora (Appendix C). Three major conclusions are worth summarising here. One conclusion is that vocabulary cannot be conceptualised as single word items or simply as content words. Every single word is idiosyncratic which means that it has 'its own individual and unique pattern of behaviour' (Partington, 1998: 27), or a distinct group of 'lexical and grammatical friends' which contribute to its essence. To fully understand the meanings and usage of a word, it is essential to examine 'the company the word keeps' (Scott, 2004: 86). On these grounds, it was argued that collocational knowledge is

part of word knowledge (Nation, 2001: 321) and thus an integral aspect of vocabulary description and pedagogy (e.g. Lewis, 2000; McCarthy and O'Dell, 2005).

The second conclusion is related to the 'unrandomness' (Sinclair, 1991: 110) of language. Taking into account the nature of collocations that were identified in the subject-specific textbooks and examinations corpora in relation to their context of occurrence, it becomes clear that strong word co-occurrence patterns are not accidental but they are influenced by the communicative needs and information content of texts: 'Collocations are the surface, lexical evidence that words do not combine randomly but follow rules, principles, and real-world motivations' (Moon, 1998: 26). It was specifically shown in the present collocational analysis that the great majority of textbooks-based collocations reflect the topic concerns of subject discourse (e.g. *Industrial Revolution* in history, *frequency distribution* in mathematics, etc.) while collocations found in the examinations corpora are primarily linked to instructions for candidates (e.g. *circle answer* in geography, *write composition* in English, etc.). Lexical collocations, in particular, were found to have more transparent semantics, denoting context-specific concepts, notions, specialised terms, as opposed to grammatical collocations which often convey more subtle information. For example, the density of collocations with the preposition *of* in the science textbooks and examinations clusters, typically combined with nouns, represents the 'complex nominal nature of science writing' (Gledhill, 2000: 125). In addition, the collocational context of lexical words appears to be more restricted compared to grammatical words which collocate with a wider variety of words.

On the basis of the analysis of textbooks- and examinations-derived collocations, it can be argued that the collocational selectivity and normality of words depends on the genre, register, and style (Partington, 1998: 26). This explains the particular semantic value and functional role of collocations identified in the twelve corpora, as discussed above, and it also explains the emergence of more restricted collocations which are highly technical and specific to certain fields of knowledge (namely in mathematics and science textbooks corpora), even though they comprise everyday words. It is perhaps the collocations which involve terminology and 'reflect the recurrent semantics of the specialist domain' (Gledhill, 2000: 131) that are fundamental for constructing subject-specific meanings and probably the ones that have the highest pedagogic value in language support namely for subject-specific language learning.

Finally, the incremental effect of these regular word associations, which result from the repeated use of common words, is that: 'by far the majority of text is made of the occurrence of common words in common patterns or in slight variants of these patterns' (Sinclair, 1991: 108). This observation about the lexical patterning of texts challenges the compositional view of language and attests to Sinclair's idiom or collocational principle, that 'the language user has

available to him a large number of preconstructed or semi-constructed phrases that constitute single choices, even though they appear to be analysable into segments' (ibid.: 110). To explain this phenomenon, Sinclair remarks that: 'To some extent, this may reflect the recurrence of similar situations in human affairs; it may illustrate a natural tendency to economy of effort; or it may be in part due to the exigencies of real-time conversation'.

Corpus investigations have made it quite clear, however, that the collocational behaviour of words extends beyond the attraction of two words to the formation of multi-word combinations. Therefore the discussion proceeds to the analysis of longer word patterns that appear recurrently in the twelve corpora.

3.2.3 4-word clusters

The previous section examined the tendency of words to form statistically significant pairs, i.e. collocations. This section takes the vocabulary analysis of the twelve corpora a step further by adopting word clusters (Scott, 2004) as the primary unit of analysis and description of subject-specific language. Known also as lexical bundles (Biber et al. 1999) or multi-word expressions (Rayson, 2008), among many other labels (see Wray, 2002: 9 for an overview), word clusters can be regarded as 'extended collocations' (Biber et al., 1999: 989; Hyland, 2008: 4), specifically interpreted here as recurrent strings of more than two words that appear more frequently than expected by chance in a particular linguistic register.

The analysis of word clusters in the twelve subject-specific corpora of this research is motivated by their ubiquity in language and the instrumental role they play in academic discourse. Based on his observations in a series of papers about the lexical patterning of language, Sinclair (1996, 2005) was the first to uncover the phraseological nature of English, concluding that 'the normal carrier of meaning is the phrase' (2005) and that phraseology should be 'at the heart of linguistic description'. His view of language as a phraseological system (which is also captured by his idiom principle stated in the previous section), is supported by many other researchers (e.g. Partington, 1998; Hunston and Francis, 2000; Stubbs, 2001 etc.) who also reject the view of the single word as the primary unit of meaning.

An increasing number of recent studies focusing on word clusters in academic discourse, and in ESP in particular, document their importance. These studies specifically report that clusters function as fundamental building blocks of academic discourse (e.g. Biber et al., 1999; Biber et al., 2004; Biber, 2006) and contribute to the internal coherence of its structure

(Hyland, 2008: 4). They further shape context-specific meanings (ibid.: 5) which means that each linguistic register and genre can be characterized by a distinct set of recurrent word sequences (e.g. Biber et al., 2004; Biber, 2006; Scott and Tribble, 2006; Biber and Barbieri, 2007: 265; Hyland, 2008). It is therefore argued that the description of clusters can offer 'insights into important aspects of the phraseology used by writers in specific contexts' (Scott and Tribble, 2006: 132) which can be of great pedagogical utility. In addition, in psycholinguistic and language acquisition studies, clusters have been found to contribute to the fluency and naturalness of linguistic production by native and non-native speakers of a language (e.g. Pawley and Syder, 1983; Nattinger and DeCarrico, 1992; Howarth, 1998; Hyland, 2008). In the light of these findings, there are several reasons why the word clusters of subject textbooks and examinations corpora should be analysed in this research.

The study of word clusters has been approached from different theoretical perspectives (e.g. semantics-based, Makkai, 1972, Mitchell, 1971; syntax-based, Katz, 1973, Fraser, 1970; lexicographic research drawing on all previous approaches; functional methods drawing on pragmatics, Pawley and Syder, 1983, Sinclair, 1987, Nattinger and DeCarrico, 1992). In the present applied corpus linguistics research, a frequency-based approach is employed to the analysis of clusters. As already explained in Chapter 2, this approach is more robust because it does not rely on intuition and introspection and because high-frequency language features are pedagogically important. An additional strength of frequency information on clusters is that it reveals 'the extent to which a sequence of words is stored and used as a prefabricated chunk, with higher frequency sequences more likely to be stored [in the mental lexicon] as unanalyzed chunks than lower frequency sequences' (Biber et al., 2004: 376). This has direct implications for ESL teaching and learning, and specifically for the development of L2 fluency skills.

To narrow the scope of analysis, only 4-word clusters are computed here because these are 'far more common than 5-word strings' and usually have 'a clearer range of structures and functions than 3-word bundles' (Hyland, 2008: 8); furthermore, 4-word clusters 'hold three-word bundles in their structure (as in *as a result of* which contains *as a result*)' (Cortes, 2004: 401). The 4-word clusters of the twelve corpora are identified according to orthographic word units. The very nature of word clusters requires a more extended analysis than the one needed for words and collocations.Two questions guide the frequency analysis of 4-word clusters (adapted from Biber, 2006: 135):

a) Are there any 4-word clusters that are used repeatedly by writers in subject textbooks and examination papers?
b) What are the structural and functional characteristics of these word 4-clusters, and how can their repeated use in discourse be explained?

To address these questions, quantitative and qualitative perspectives are provided using WordSmith Tools (Scott, 2004). In the first stage of the research, the most frequent 4-word clusters are identified in the twelve corpora using the WordList tool. Biber's (2006: 134) conservative frequency threshold of 40 per million words was once again adjusted to the size of the corpora of this research (based on the formula: (40 x total number of tokens)/1,000,000) and frequency counts were normalized per mille words (‰). Manual editing of the results involves removing 4-word clusters which include symbols (#), numbers, as well as clusters with a range of occurrence in less than75% and 60% of the textbooks and examinations corpora respectively. The range threshold was essential to ensure that the idiosyncratic language use of individual textbooks authors does not influence the results.

Following the frequency analysis and manual editing of 4-word clusters, a more qualitative analysis was conducted. First, their structural correlates were analysed (in terms of grammatical types) and they were subsequently categorized according to major patterns, based on an adaptation of the structural classification of lexical bundles in academic prose by Biber et al., (1999: 997-1025; see section 3.2.3.2 below). Concordances of 4-word clusters were then closely examined to arrive at functional interpretations of their recurrent use in discourse. A close examination of the contextual environment of their occurrence allowed i) the investigation of their complementation patterns, i.e. the 'kind of [word] group or clause that may follow' (Hunston and Francis, 2000: 2) or precede each cluster, and ii) the description of their discourse meaning and functions. The results from the above analyses of textbooks- and examinations-derived clusters are discussed in detail below.The complete listings of the most frequent 4-word clusters in textbooks and examinations corpora can be viewed in Appendices D1 and D2 together with their structural patterns.

3.2.3.1 Density, forms and frequencies

The frequency analysis of 4-word clusters reveals significant differences in the total number of clusters that occur in the twelve corpora and meet the frequency and range thresholds, with examinations corpora exhibiting a denser use of clusters as a whole than textbooks corpora.

More specifically, the total numbers of clusters exceed the 2,000 in the six examinations corpora, except in the English examinations corpus in which 673 clusters meet the frequency threshold (40 per million words - adjusted to a minimum of four occurrences in this corpus). The CSPE examinations corpus contains by far the largest number of clusters with 7,125, followed by the mathematics corpus with 3,101 clusters. Similar numbers of 4-word clusters appear in the history, geography and science examinations corpora, with 2,494, 2,383 and 2,118 clusters respectively. With regard to textbooks corpora, the largest number of different 4-word clusters can be found in the mathematics corpus, with a total of 479 clusters. The science corpus includes 185 clusters and it is, unexpectedly, closely followed by the CSPE corpus which contains 184 word clusters. Slightly lower numbers of clusters appear in the geography and the history corpora, with 142 and 147 different 4-word clusters respectively. Finally, the English corpus appears to make a limited use of clusters with a total of only 68.

In view of the above aggregates and the total size of the corresponding corpora, as shown in Table 3.14 below, it follows that the overall density of 4-word clusters in a corpus does not correlate with its size. It would therefore be reasonable to argue that the observed variation in cluster density represents genuine differences among corpora. The results from the examinations corpora, however, should be treated with caution, as it is likely that their aggregates may not be entirely representative of the exact proportions of 4-word clusters in the six examinations corpora. This may be because the frequency cut-off points were quite low and perhaps not the most appropriate given the small size of the six corpora in question; these same frequency cut-off points as those adopted for the cluster analysis of textbooks corpora were, however, used in the interest of consistency. For this reason, it would make little sense to try to explain the reasons behind the above differences in the total numbers of examinations clusters. It is also important to point out that the above counts represent the original cluster lists that were derived from WordSmith prior to their being subjected to manual editing (i.e. they includen4-word clusters which meet the frequency but not the range threshold). The final lists of textbooks clusters that were analysed comprise the following numbers: 23 in English, 92 in geography, 86 in history, 117 in CSPE, 206 in mathematics, and 140 in science. The heavy load of examinations clusters renders their analysis a prohibitively labour-intensive task and, therefore, only the top 100 clusters of the six corpora are examined in the present research (see Appendix D2).

Table 3.14 Size (total number of types) and total number of 4-word clusters in the twelve corpora (prior to the manual editing of cluster listings, # - number).

Corpus	Corpus size	# of 4-word clusters
Textbooks		
English	451,784	68
Geography	330,257	142
History	367,708	147
CSPE	213,340	184
Mathematics	336,738	479
Science	328,426	185
Examinations		
English	93,898	673
Geography	48,178	2,383
History	46,351	2,494
CSPE	25,870	7,125
Science	68,617	2,118
Mathematics	55,476	3,101

Differences in the total numbers of clusters that are repeatedly used in the corpora of the two registers (textbooks and examinations) reveal the degree of formulaic language use and can be explained by considering the communicative needs of these text types or genres. The increased density of 4-word clusters in the examinations corpora in this research confirms the commonsense intuition that an important proportion of the discourse of examinations papers is made up of routinized, scripted language expressions. This stems from the assessment purpose of this genre which is linguistically enacted in concise instructions and questions formulated in conventionalized patterns that are specific to the different subjects and, more importantly, consistently used in examination papers over the years. The lower concentration of 4-word clusters in the subject textbooks corpora, on the other hand, could be a direct consequence of the informational focus of the genre of school textbooks. This evidently requires much more diverse language choices by textbook writers (Biber et al., 2004: 383) than pre-fabricated language expressions for the display and discussion of the wide range of topics covered in the different units of textbooks.

Similarly, comparisons of cluster density across the six subject textbooks corpora provide hints about the extent of language patterning in subject-specific textbooks. The extreme differences between the mathematics and science textbooks corpora on the one hand and English textbooks corpus on the other can be considered here as an example. The widespread use of 4-word clusters in the mathematics and science textbooks corpora in this research seems to indicate that technical subjects make a more extended use of routinized language. This can be understood by considering that the (spoken and written) texts of these

two subjects are constructed based on an amalgam of 'language, mathematical [and scientific] symbolism and visual display' (O'Halloran, 2004: 191), and, as a result of this 'multi-semiotic discourse' (ibid.), verbal language is primarily used to discuss information that is represented mainly through numerical and graphical means. Thus, formulating mathematical problems and providing solutions, explaining diagrams, describing processes in scientific experiments, etc. require the use of routinely patterned language, allowing little space for creativity. The low density of clusters in the corpus of English textbooks, on the other hand, reflects the creative and diverse linguistic expression that occurs in the wide range of literary genres (poetry, prose, plays, novels, short stories, media material) that are included in English textbooks. This is in line with the manifest aim of the Junior Cycle English syllabus to offer students 'a wide and varied linguistic experience' and to expose them to 'the cultural richness and diversity of English'. Interpreted in linguistic terms, these aims by definition, forbid an extensive use of (semi) fixed word patterns and require more 'liberal' language expression that does not heavily rely on the recurrent use of the same word sequences.

Although the 4-word clusters of the twelve corpora differ in respect of their total numbers, two similarities seem to emerge in terms of their forms. The first common pattern that can be readily observed from an examination of the 20 most frequent clusters in the twelve corpora (Tables 3.15 and 3.16) and the complete cluster listings in Appendices D1 and D2 is the repetition of slight variants of several 4-word clusters, differentiated by the addition or deletion of one or two items to the left or right. In connected discourse, these 4-word clusters are strung together and are repeated in overlapping sequence as longer clusters.

Table 3.15 The 20 most frequent 4-word clusters in the six textbooks corpora (italics = cluster with a range of occurrence in 100% texts of the corpus).

N	English	Geography	History	CSPE	Mathematics	Science
1	*give reasons for your*	each of the following	in the twentieth century	*has the right to*	*each of the following*	*shown in the diagram*
2	*reasons for your answer*	*draw a sketch map*	international relations in the	and answer the questions	find the value of	*as shown in the*
3	the questions that follow	the West of Ireland	*in Ireland in the*	*the work of the*	the values of x	*set up the apparatus*
4	answer the questions that	*a sketch map of*	write an account of	*why do you think*	values of x for	*the apparatus as shown*
5	and then answer the	*the ordnance survey map*	*the end of the*	*of the action project*	*the area of the*	*up the apparatus as*
6	*of the following questions*	*sketch map of the*	each of the following	answer the questions that	the value of x	*is made up of*
7	*what do you think*	*the area shown on*	answer the following questions	the questions that follow	*the equation of the*	*can be used to*
8	*the end of the*	to support your answer	*the Treaty of Versailles*	*social and political education*	*find the image of*	what is meant by
9	two of the following	*area shown on the*	the age of revolutions	*as part of the*	*at the end of*	*the bottom of the*
10	*do you think the*	*shown on the map*	*at the end of*	*a member of the*	*find the equation of*	*the centre of gravity*
11	*at the end of*	with the aid of	*the War of Independence*	*civic social and political*	*correct to the nearest*	*a form of energy*
12	*fill in the gaps*	per cent of the	the agricultural and industrial	of the European Union	divide both sides by	*for the presence of*
13	*one of the following*	study the ordnance survey	*the League of Nations*	declaration of human rights	of x for which	*a small amount of*
14	*the rest of the*	*in the West of*	agricultural and industrial revolutions	*what do you think*	standard rate cut off	*is a form of*
15	each of the following	tick the correct box	*of the twentieth century*	*the rights of the*	*the length of the*	apparatus as shown in *experiment to show*
16	*at the beginning of*	*map of the area*	during the middle ages	is made up of	*the radius of the*	*that*
17	*in the middle of*	*of the area shown*	per cent of the	*do you think the*	equation of the line	*the mass of the* carbon dioxide and
18	and answer the questions	is one of the	*in the middle ages*	*was set up in*	*axial symmetry in the*	water
19	*give a reason for*	the population of the	*the Home Rule Party*	*on the rights of*	*the end of the*	*the boiling point of*
20	this is a very	*using evidence from the*	*the Irish Free State*	*rights of the child*	*the volume of the*	*the density of a*

Table 3.16 The 20 most frequent 4-word clusters in the six examinations corpora (italics = cluster with a range of occurrence in 100% texts of the corpus).

N	English	Geography	History	CSPE	Mathematics	Science
1	*then answer the questions*	*tick the correct box*	*give two reasons why*	page you will read	*find the value of*	*for examiner use only*
2	*give reasons for your*	the correct box page	*two of the following*	printed on a separate	*the area of the*	*the list on the*
3	*reasons for your answer*	*the ordnance survey map*	*answer the following questions*	the page you will	space for extra work	*from the list on*
4	*and then answer the*	*in each of the*	*write an account of*	when you have studied	*be obtained from the*	*list on the right*
5	*novel or short story*	*marks will be lost*	give one reason why	of the page you	*obtained from the superintendent*	*shown in the diagram*
6	*one of the following*	*otherwise marks will be*	*one of the following*	on the top of	*correct to the nearest*	*the diagram shows a*
7	*the questions which follow*	circle the correct answer	*of the following terms*	the top of the	for extra work page	describe with the aid
8	*answer the questions which*	*the correct box or*	*and then answer the*	top of the page	*the length of the*	with the aid of
9	*two of the following*	the correct answer in	*paper and then answer*	answer the questions below	*the volume of the*	of a labelled diagram
10	with reference to the	correct answer in each	*then answer the following*	*of the questions numbered*	be shown to obtain	the aid of a
11	answer with reference to	which of the following	*write about that person*	you have studied this	indicates that supporting work	aid of a labelled
12	your answer with reference	of the statements below	during the middle ages	class has decided to	must be shown to	on the right to
13	*short story you have*	*with your answer book*	*of the people described*	your CSPE class could	shown to obtain full	*the diagram shows the*
14	*each question is worth*	answer in each of	*one of the people*	a blank page for	supporting work must be	give one use for
15	*of the following questions*	each of the statements	*select one of the*	blank page for the	symbol indicates that supporting	the diagram shows an
16	*story you have studied*	*of this paper is*	*the people described below*	of this answer book	that supporting work must	what happens to the
17	your answer by reference	ordnance survey map and	*accompany this paper and*	your CSPE class has	the symbol indicates that	word from the list
18	*or short story you*	a sketch map of	*this paper and then*	at the back of	to obtain full marks	a word from the
19	*the name of the*	draw a sketch map	*which accompany this paper*	been included at the	work must be shown	an experiment to show
20	*a novel or short*	sketch map of the	*with your answer book*	CSPE class has decided	*the value of x*	the piece of equipment

These slightly varied phraseological items specifically in the six textbooks corpora are characteristic of dominant instructions and activities in the different subject areas, as the following examples indicate:

English:

- *give reasons for your answer*
- *and then answer the questions that follow*
- *what do you think the*

Geography:

- *draw a sketch map of the area shown on the map*

History:

- *international relations in the twentieth century*

CSPE:

- *answer the questions that follow*
- *civic, social and political education*

Mathematics:

- *find the values of x for which*
- *find the equation of the line*

Science:

- and *set up the apparatus as shown in the diagram*

The same tendency of 4-word cluster variants to form longer patterns of repeated phrases can also be observed in the top 20 clusters in the examinations corpora (Table 3.16 and Appendix D2). As revealed by concordances, these phrases can often form long strings of discourse and only in very few cases do they exhibit internal variation with one or more words omitted or additional ones inserted. Here are some examples:

English:

- *and then answer the questions that follow*
- *give reasons for your answer, your answer with reference to the*
- *a novel or short story you have studied*

Geography:

- *you must return this paper with your answerbook otherwise marks will be lost*
- *circle the correct answer in each of the (following) statements below*
- *In the boxes provided, match each letter in Column X with the number of its pair in Column Y. One pair has been completed for you.*
- *draw a sketch map of the*

History:

- *and then answer the following*
- *Select one of the people described below. Write about that person. If you wish, you may use the hints to help you in your answer. Write the title selected at the top of your account.*
- *write an account of one/two of the following (topic/s)*
- *explain [one/two/three] of the following terms relating to*

CSPE:

- *the page you will read when you have studied this*
- *on the top of the page you will read*
- *printed on a separate page*
- *your CSPE class has decided to*
- *a blank page for the poster/invitation has been included at the back of this answerbook*
- *Put a tick in the box*
- *Write your answers in the spaces provided. Hand up this paper at the end of the examination.*
- *answer all of the questions in this section*
- *Answer one/any three of the questions numbered 1,2,3,4 below. Each question carries [...] marks. If you need extra paper to answer this question, please ask the Examination Superintendent for it.*
- *Make sure to write your examination number in the box above.*

Mathematics:

- *Attempt all questions. Each question carries [...] marks. Answers and supporting work should be written in the boxes provided. Extra paper and graph paper can be obtained from the Superintendent, if needed. The symbol indicates that supporting work must be shown to obtain full marks.*
- *be obtained from the superintendent*
- *space for extra work page*
- *symbol indicates that supporting work must be shown to obtain full marks*
- *what is/calculate the amount of investment at the end of [each/this/that/the first/second/third] year*
- *give your answer correct to [the nearest whole number/one/two decimal place(s)]*

Science:

- *In each case choose a word/choose the correct term from the list on the right to complete the sentences below.*
- *Describe with the aid of a labelled diagram*, a (laboratory) experiment to [compare/show/measure etc.]
- *Name the piece of equipment shown in the diagram.*

More phrases like these can be revealed by extending the size of clusters in WordSmith settings but only 4-word clusters are analysed here for the reasons explained at the beginning of section 3.2.3. As illustrated by the examples provided above, these (semi) fixed phrases are exam-oriented constructions which define assessment tasks or encapsulate instructions to

candidates. As such, it is not easy to identify the subject of the exam paper in which they occur. It should be further added that the examples listed above appear with a wide range of occurrence in the corresponding text collections (cf. Appendix D2). In general, the great reliance of a register on these longer fixed expressions attests to its highly formulaic nature which stems from the need for concise and consistent instructions to candidates.

Another noticeable pattern in the composition of the most frequent 4-word clusters is the recycling of common content words and significant collocations. Table 3.17 below lists some examples of highly-frequent words and collocations which can be found among many others to be nested in many 4-word clusters of textbooks and examinations corpora (these were identified based on Tables 3.15 and 3.16 and Appendices D1 and D2).

Table 3.17 Examples of highly-frequent content words and collocations embedded within common 4-word clusters in textbooks and examinations corpora.

English	
Textbooks corpus	**Examinations corpus**
• *think*	• *answer*
• *reasons-answer*	• *questions*
• *give-reasons*	• *story*
• *answer-questions*	• *reasons-answer*
• *do-think*	• *give-reasons*
• *questions-follow*	• *novel-story*
	• *reference-text*
	• *following-questions*

Geography	
Textbooks corpus	**Examinations corpus**
• *map*	• *map*
• *grid-reference*	• *page*
• *ordnance-survey*	• *tick*
• *ordnance-map*	• *ordnance survey*
• *survey-map*	• *survey-map*
• *west-Ireland*	• *sketch-map*
• *each-following*	• *aerial-photograph*
• *area-shown*	• *circle-correct*
	• *return-folder*

History	
Textbooks corpus	**Examinations corpus**
• *century*	• *answer*
• *international-relations*	• *write*
• *twentieth-century*	• *people*
• *industrial-revolution*	• *piece-evidence*
• *home-rule*	• *paper-answer*
• *middle-ages*	• *middle-ages*
• *reasons-why*	• *reason-why*
	• *international-relations*

CSPE	
Textbooks corpus	**Examinations corpus**
• *rights*	• *page*
• *action-project*	• *answer*
• *county-council*	• *questions*
• *European-Union*	• *all/any-questions*
• *human-rights*	• *top-page*
• *universal-declaration*	• *class-decide*
	• *describe-actions*
	• *tick-box*

Mathematics		Science	
Textbooks corpus	**Examinations corpus**	**Textbooks corpus**	**Examinations corpus**
• *value*	• *value*	• *energy*	• *diagram*
• *correct-decimal*	• *shown*	• *carbon-dioxide*	• *labeled*
• *gross-income*	• *work*	• *experiment-show*	• *shows*
• *standard-rate*	• *supporting-work*	• *centre-gravity*	• *examiner-use*
• *surface-area*	• *symbol-indicates*	• *gas-jar*	• *describe-aid*
• *average-speed*	• *obtained-superintended*	• *electrical-energy*	• *list-right*
• *central-symmetry*	• *correct-nearest*	• *test-tube*	• *aid-labelled*
	• *simplest-form*		• *diagram-shown*

It is particularly interesting that when these high-frequency words and collocations are not embedded within 4-word clusters, concordance lines show that they are manifested as their complementation items. Table 3.18 presents some instances of common content words (underlined and enclosed in parentheses) and collocations (enclosed in parentheses; intervening words are placed within square brackets) which follow, and less often precede, the commonest 4-word clusters in the twelve corpora.

Table 3.18 Examples of highly-frequent content words and collocations adjacent to common 4-word clusters in textbooks and examinations corpora.

English	
Textbooks corpus	**Examinations corpus**
• *the end/rest of the (story)*	• *with reference to the (text)*
• *this is a very good (answer)*	• *the name of the (poet/play/author)*
• *do you think the (poet)*	• *each question is worth* [...]*(marks)*
• *give a reason for (your answer)*	• [with] *(reference* [to] *the) novel or short story*
• *each of the following (words* [or] *phrases)*	• *(write* [a] *composition) on one of the* [following topics]
	• *read this piece and (then answer)* [the questions]

Geography	
Textbooks corpus	**Examinations corpus**
• *per cent of the (population)*	• *study the ordnance survey (map)*
• *using evidence from the (map)*	• *sketch map of the (area)*
• *the aid of a (diagram/reference grid)*	• *(match) each of the following*
• *the population of the (West* [of] *Ireland)*	• *using evidence from the (aerial photograph)*
• *in the case of (each* [of the] *following)*	• *(look at) the ordnance survey map*
• *calculate the area of* [*the*] *(map extract)*	• *(geographical mix:) answer any three of*

History	
Textbooks corpus	**Examinations corpus**
• *write an account of* [the] (*people*)	• (*explain*) [one/two/three] *of the following terms*
• *the end of the* (*war*)	• (*give*) *one piece of evidence*
• *per cent of the* (*land*)	• *during the age of* (*exploration*)
• *give two reasons why* (*Fianna Fail*)	• *is an extract from* [an/the] (*account of*)
• *the beginning of the* (*Civil/Cold War*)	• *from your study of*(*international relations*)
• *by the end of the/international relations in the* (*twentieth century*)	• (*international relations*) *in the twentieth century*

CSPE	
Textbooks corpus	**Examinations corpus**
• *the work of the* (*government*)	•*when you have studied* [this/the] (*poster*)
• (*organization*) *was set up in*	•*(answer)* [one/any three] *of the questions numbered*
• *the rights of the* (*child*)	•(*actions*) [your] *CSPE class could take*
• (*as part*) *of the action project*	•*you may use this* [title/name/concept/organization etc.] *only once*
• *a member of the* (*European Union/Parliament, Defence Forces*)	•*write your answers in [the] (spaces provided)*
• *write a letter to* [the/your] (*local authority*)	•(*would* [have] *to*) *set up in order*

Mathematics	
Textbooks corpus	**Examinations corpus**
• *the equation of the* (*line*)	• (*calculate/find*) *the area of the* (*triangle/parallelogram*)
• (*calculate*) *the area of the* (*triangle*)	• *draw the graph of* [the] (*function*)
• *find the image of* [the/each] (*point*)	• (*construct*) *the image of the*
• (*length of*) *the radius of the*	• (*write down*) *the length of the* (*side*)
• *as a percentage of* (*gross income/*[the] *selling price*)	• *give a reason for* (*your answer*)
• *find the area of* (*each* [of the] *following*)	• [write/give] (*your answer*) *in its simplest form*

Science	
Textbooks corpus	**Examinations corpus**
• *is made up of* (*water*)	• (*describe*) *with the aid of*
• *the density of a* (*liquid*)	• (*name*) [and] *give one use for*
• *the bottom of the* (*tube*)	• *the diagram shows an* (*experiment/apparatus*)
• (*test*) *for the presence of* (*starch/food/carbon dioxide*)	• *give a reason for* (*your answer*)
• (*current* [to] *flow*) *in one direction only*	• *describe an experiment to* (*compare the*) [absorbency/thermal conductivity/density/hardness/flexibility of]
• (*show that*)(*light*) *travels in straight lines*	• [a](*piece* [of] *equipment*) *that can be used* [to investigate]

The complementation patterns of 4-word clusters in all twelve corpora demonstrate, once again, Sinclair's argument that 'by far the majority of text is made of the occurrence of common words in common patterns or in slight variants of these patterns' (1991: 108) and

that words are in fact 'components of a *rich repertoire of multi-word patterns* that *make up a text*' (ibid.: emphasis added). By illuminating the lexical patterning that underpins the textbooks- and examinations-based discourse of the six subjects, the above findings are also a manifestation of the operation of Sinclair's idiom principle (ibid.: 110) mentioned earlier in the discussion (at the end of section 3.2.2.4). A general impression that can be imparted from the above findings is that texts can be visualized as interconnected webs of lexical phrases built upon frequent lexical words and collocations which convey the meanings that need to be constructed in subject-specific discourse.

With regard to the frequency of occurrence of 4-word clusters, only a few occur with very high frequencies in the twelve corpora (cf. Biber et al., 1999: 994). From their frequencies per mille words in Appendices D1 and D2 it can be seen that, for the most part, textbooks-derived clusters have quite low frequencies. More specifically, only the first three clusters in the English and geography textbooks corpora and the first four clusters in history and science occur with a frequency higher than 0.2‰ while all the clusters of the CSPE textbooks corpus appear with frequencies lower than 0.19‰, with the exception of the top cluster (*has the right to*, 0.33‰). The clusters of mathematics textbooks, by contrast, appear with considerably higher frequencies overall as the top 18 clusters do not fall below 0.2‰ and the normed frequencies of the top six clusters appear more than 0.37‰. Interestingly, most of the mathematics textbooks clusters also appear to have the widest distribution in the text collections compared to the clusters of the other corpora, with 83% of clusters occurring in 100% texts. The 4-word clusters with a range of 100% in the other five corpora represent percentages of 69.5% of the English list, 62.5% of the science list, 53.4% of the history list and 38.4% and 34.7% of the CSPE and geography listings of clusters respectively.

As regards the top 100 clusters in the examinations corpora (Appendix D2), they appear to have, as a whole, significantly higher frequencies than the top textbooks-derived clusters. Specifically, the three commonest clusters of all six corpora appear with normed frequencies higher than 2‰ and the proportional percentages of occurrence of all other clusters in the six lists do not fall below 0.2‰, except the clusters below the 72nd and 73rd rank positions in the English and science examinations corpora respectively. The science-specific clusters appear with the highest frequency percentages compared to the clusters of the other examinations corpora. Thus, the corpora of the two technical subjects of mathematics and science are the ones which stand out from the comparisons of cluster frequencies in the two registers of textbooks and examinations respectively. As regards the range of clusters in the exams corpora, the clusters with the highest consistency of occurrence (i.e. in 100% of text

collections) out of the top 100 are 67 in English, 28 in geography, 62 in history, 55 in CSPE, 23 in mathematics and 27 in science.

Although no important differences can be discerned among the frequencies of clusters in a single list, three exceptions are worth mentioning. Specifically in the examinations corpus for geography, there is a remarkable difference between the first cluster (*tick the correct box*), which has a frequency of 5.93‰, and all the rest that follow with frequencies lower than 1.78‰. This probably suggests that multiple-choice questions are a typical task in geography examinations. Important but less extreme differences also occur in the history textbooks corpus between the top two clusters (*in the twentieth century, international relations in the*) which occur 0.46 and 0.41 times per mille words and the other clusters in the same list whose frequency falls below 0.25‰. The large proportional use of these clusters clearly demonstrates their central importance in the discourse in which they recur. Thus, *international relations in the twentieth century* appears to be an important issue in historical discourse and multiple-choice exercises seem to be the commonest type of assessment tasks in Junior Certificate examinations in geography.

After these observations about the diversity, forms and frequenciesof 4-word clusters based on their general examination in this section, the discussion turns to a more detailed qualitative analysis of clusters in each of the twelve corpora. This brings to light the variation that is manifested in the structural patterns of the commonest clusters (section 3.2.3.2) as well as striking differences in the meanings and functions these clusters assume in subject-specific discourse (section 3.2.3.3).

3.2.3.2 Structural patterns

Following the frequency analysis of the commonest 4-word clusters in the twelve corpora, the investigation of their structural correlates (i.e. grammatical types of their constituents) is discussed in this section to illuminate the structures they form. Clusters are analysed and grouped according to a modified version of the structural classification developed by Biber et al. (1999: 997-1036) for the lexical bundles of the Longman Spoken and Written English (LSWE) corpus (40 million words of British and American English in 37,000 texts). More specifically, the structural categories used to characterise the bundles of the academic prose component of the LSWE corpus (representing 5.3 million words in the corpus; ibid.: 1014-1036) are employed but some additional categories are necessarily inserted to account for certain types of structure

that emerged in this research (see below for details). The analysis reveals quantitative and qualitative differences in the total number and types of principal structures of clusters across the twelve corpora.

Table 3.19 lists the main structural categories identified in the clusters of the six textbooks corpora and Table 3.20 presents the top 20 structures of the main categories that characterise the top 100 clusters of the examinations corpora.The lists are sorted by the total number of different clusters (specified in parentheses) that assume each type of structure. Although clusters typically assume a single structure, there are two instances (*what happens to the* and *would you expect to* respectively in the science textbooks and exams corpora) which act both as WH-clause fragments and as WH-question fragments in discourse. All instances are assigned to the category to which the majority belongs. Thus, *what happens to the,* which occurs 15 times as a WH-question fragment and only eight as a WH-question fragment, is classified as a WH-question fragment, and *would you expect to* acts more often as a WH-clause fragment (eight times) than as a WH-question fragment (two times).

Table 3.19 The main structures of the 4-word clusters in the six textbooks corpora (the number of clusters that manifest each structure is specified within parentheses).

N	English	Geography	History	CSPE	Mathematics	Science
1	Transitive V+NP fragment (6)	NP+of fragment (30)	NP+of fragment(36)	NP+offragment (32)	NP+offragment (73)	NP+offragment (43)
2	NP+of fragment (5)	Other PP fragment(14)	Other PP fragment(21)	Other PP fragment(18)	Transitive V+NP fragment (28)	Other PP fragment(15)
3	PP with embedded of-phrase fragment(3)	Transitive V+NP fragment (11)	NP (11)	NP+Post-nominal clause fragment (10)	Other PP fragment(25)	Transitive V+NP fragment (13)
4	yes-noquestionfragment(3)	Passive verb+PP fragment(9)	PP with embedded of-phrase fragment(8)	Transitive V+NP fragment (9)	NP+PP fragment clause (23)	(V/Adj+) to-clause fragment (12)
5	WH-questionfragment(2)	NP (8)	Transitive V+NP fragment (4)	NP+PP fragmentclause (8)	NP(10)	Passive verb+PP fragment(11)
6	3rd person pronoun+VP fragment (1)	PP with embedded of-phrase fragment (8)	Copulabe + AdjP(2)	Passive verb+PP fragment(6)	PP with embedded of-phrase fragment(7)	Intransitive V+PP fragment (9)
7	NP + post-modifierfragment(1)	NP + Post-nominal clause fragment (4)	Adj+Adjfragment(1)	(V/Adj+) to-clause fragment (5)	Passive verb+PP fragment(6)	PP with embedded of-phrase fragment(7)
8	NP+that-clausefragment(1)	WH-questionfragment(3)	NP+post-modifierfragment(1)	NP (5)	PP+copulabe(6)	Copulabe + NP(5)
9	Other PP fragment(1)	Copulabe + AdjP(2)	Passiveverb + PP fragment(1)	PP with embedded of-phrase fragment(4)	Transitive V+NP (5)	NP + Post-nominal clause fragment(5)
10		either-orfragment(1)	Transitive V + PP fragment(1)	Transitive V + NP(4)	Intransitive V+PP fragment (5)	NP + PP fragmentclause(5)
11		NP + PP fragmentclause(1)		yes-noquestionfragment(4)	Copula be+NP(4)	WH-questionfragment(3)
12		to-clausefragment(1)		WH-questionfragment(3)	Copulabe + AdjP(3)	NP(3)
13				NP fragment(3)	Copulabe + AdjP(2)	NP+passive VP(2)
14				NP+verb(1)	N+VP fragment(2)	WH-clausefragment(2)
15				(connector+) Transitive	(V/Adj +) to-clause	Copulabe + AdjP(1)

	V+NP (1)	fragment (1)	
16	VP fragment(1)	Anticipatoryit + AdjP(1)	N or N(1)
17	Copulabe + AdjP(1)	Comparativeexpression(1)	NP+verb(1)
18	N+passive VP(1)	copulabe+PP(1)	NP+VP fragment(1)
19	N+VP(1)	If-clausefragment(1)	PP+copulabe(1)
20		NP + Post-nominal clause fragment(1)	that-clausefragment(1)
21		NP+copulabe+Adj(1)	
22		NP+to-clausefragment(1)	
23		NP+verb(1)	

Table 3.20 The top 20 structures of the 100 commonest 4-word clusters in the six examinations corpora (the number of clusters that manifest each structure is specified within parentheses).

N	English	Geography	History	CSPE	Mathematics	Science
1	NP+PP fragmentclause(19)	NP+offragment(12)	NP+offragment(19)	NP+post-modifierfragment(20)	NP+offragment(15)	NP+post-modifierfragment(18)
2	transitive V+NP fragment(14)	PP fragment(12)	transitive V+NP fragment(14)	NP+offragment(14)	transitive V+NP fragment(12)	NP+offragment(15)
3	NP + offragment(8)	passive VP fragment(8)	NP+post-modifierfragment(11)	transitive V+NP fragment(13)	NP+post-modifierfragment(8)	transitive V+NP fragment(11)
4	VP fragment(7)	NP+post-modifierfragment(6)	PP(11)	PP fragment(9)	PP fragment(8)	PP fragment(10)
5	PP fragment(8)	transitive VP+NP fragment(6)	PP fragment(8)	passive VP fragment(5)	PP(6)	(noun+) to-clausefragment(5)
6	NP fragment(5)	NP fragment(5)	transitive V+NP(5)	transitive V+NP(5)	N+modalV+passive VP fragment(4)	NP+VP fragment(5)
7	2nd person pronoun+VP fragment(4)	(verb+) to-clausefragment(4)	modal V+NP(3)	NP fragment(3)	NP+modal VP fragment(4)	WH-questionfragment(5)
8	N+VP(4)	NP(4)	(noun+) to-clausefragment(3)	2nd person pronoun+VP fragment(3)	passive VP fragment(4)	(noun+) to-clausefragment(4)
9	NP+VP(4)	V+NP(3)	2nd person pronoun+VP fragment(2)	NP+modal VP fragment(2)	modal V+passive VP fragment(3)	transitive V+NP(4)
10	VP withpassiveverb(4)	VP+NP(3)	NP+PP fragment(2)	NP+VP fragment(2)	VP fragment(3)	N+VP fragment(3)
11	yes-noquestionfragment(3)	VP+NP fragment(3)	NP(2)	PP (2)	comparisonexpression(2)	NP fragment(3)
12	N or N fragment(3)	N+passive VP(2)	(connector) transitive V+NP fragment(1)	that-clausefragment(2)	N+NP(2)	(verb+) to-clausefragment(2)
13	intransitive V+PP fragment(2)	passive V+PP fragment(2)	(connector+) VP fragment(1)	to-clausefragment(2)	(connector+) NP fragment+modal verb(2)	NP(2)

14	VP+NP fragment(2)	PP+copulabe(2)	3rd person pronoun+VP fragment(1)	WH-clausefragment(2)	N+offragment(2)	passive V+PP fragment(2)
15	VP+NP(2)	V+NP fragment(2)	copulabe+NPfragment(1)	(noun+) to-clausefragment(1)	NP(2)	passive VP fragment(2)
16	VP+NP(1)	VP fragment(2)	If-clausefragment(1)	(verb+) to-clausefragment(1)	passive V+PP fragment(2)	that-clausefragment(2)
17	(connector+) VP fragment(1)	WH-questionfragment(2)	N+2nd person pronoun+VP fragment(1)	anticipatoryit+AdjP(1)	passive VP fragment+to-clause fragment(2)	V+NP fragment(2)
18	(NP+) that-clausefragment(1)	(adverb+) N+VP fragment(1)	N+post-modifierfragment(1)	If-clausefragment(1)	transitive V+NP(2)	WH-clausefragment(2)
19	2nd personpronoun+VP(1)	2nd person pronoun+VP fragment(1)	N+PP(1)	N+2nd person pronoun+VP fragment(1)	(connector+) NP fragment(1)	2nd person pronoun+VP fragment(1)
20	modal VP+NP(1)	copulabe+NPfragment(1)	N+transitive V+NP(1)	N+modal VP fragment(1)	(connector+) transitive V+NP(1)	copulabe+NP fragment (1)

The major structures that can be found in the clusters of individual corpora and the differences that emerge are summarised in Table 3.21 below:

Table 3.21 Total number of 4-word clusters and their structural types across the twelve corpora (# signifies number).

Corpus	# of clusters	# of structures
Textbooks		
English	23	9
Geography	92	12
History	86	10
CSPE	117	19
Mathematics	206	23
Science	140	20
Exams		
English	100	26
Geography	100	39
History	100	31
CSPE	100	24
Mathematics	100	35
Science	100	25

Although there is a common core of structural patterns across the twelve corpora, Tables 3.19 and 3.20 show that there are important differences in the total number of clusters that fall into each category. The variation that exists becomes clear also from the fact that only four types of structure are shared by the clusters of the six textbooks corpora (i.e. noun + *of* fragment, other prepositional phrase fragment, prepositional phrase with embedded *of*-phrase fragment, transitive verb + noun phrase fragment) while only one (prepositional phrase fragment) appears to be common to the lists of all six examinations corpora. It cannot be argued with certainty, however, that this variation is statistically significant and that differences are genuine. It is also difficult to account for these quantitative differences, unless the actual types of structures that the above numbers represent are examined in relation to the nature of discourse in which they occur.

Looking at all the different structures collectively and from a more qualitative perspective, four major categories can be distinguished: i) structural types which incorporate verb phrase fragments, ii) structures which comprise verb phrases as part of dependent clause fragments, iii) clusters which include noun phrase or prepositional phrase fragments (Biber, 2006: 136), and iv) clusters which do not contain a fragment but stand as complete structural

units that comprise either a verb or a noun or a prepositional phrase. The third structural category (noun and prepositional phrase fragments) is prominently featured in all lists, in various combinations, characterising most 4-word clusters of the twelve corpora. This finding is similar to Biber et al.'s conclusion, based on the analysis of the LSWE corpus, that 'In academic prose, over 60% of all lexical bundles are parts of noun phrases or prepositional phrases' (1999: 995).

More specifically, the structure that occupies the top position in most lists presented in Tables 3.19 and 3.20 is the noun phrase + *of* fragment (e.g., *a sketch map of, the end of the, a small amount of* etc.). Even when this specific structural pattern is not the top one, the top ranks are still occupied by noun phrase structures in different variants such as noun phrase + post-modifier clause fragments like *the list on the, a word from the, reasons for your answer*, etc. In the English textbooks and the CSPE and science exams lists, for instance, the noun phrase + *of* fragment is the second most frequent structure and in the English exams list it is the third commonest structure. The frequency of this structure highlights the nominalised nature of the written academic discourse of subject textbooks and examination papers and it also overlaps with the extensive use of noun phrases found in the academic prose of university textbooks by Biber (2006: 138) who explains that 'the productive use of complex noun phrase structures of academic writing' is exploited for information packaging purposes. Nominalised 4-word clusters express many different meanings in subject-specific discourse: i) physical description (e.g. *the surface of the, the slope of the*), ii) place (e.g. *north and south of, the location of the*), iii) size (e.g. *the length/height/mass/volume of the*), and iv) amount (e.g. *the maximum value of, a tax credit of, the cost of the, a few drops of*)(Biber et al., 1999: 1015).The next most frequent structure is the 'other prepositional phrase fragment' (e.g., *at the end of, in each of the, during the age of*, etc.). The clusters which incorporate a prepositional phrase + *of* fragment describe logical relations (e.g. *as a result of*), temporal relation (e.g. *by the end of*) or refer to time periods (e.g. *at the age of*) (ibid.). After the different nominal and prepositional types of structure, the patterns that follow are manifested by fewer clusters.

Another general conclusion that can be drawn regarding the structures of textbooks and examinations 4-word clusters reflects once again Biber et al.'s finding that the vast majority of clusters 'span two structural units' (ibid.: 999; e.g. noun phrase + beginning of prepositional phrase) and that 'functions words are by far most common as the ending word of four-word lexical bundles' (ibid.: 995). The function words in this role are typically articles (*a, an, the*), prepositions (*of, in, to, at* etc.), and complementizers (*that*). A large number of these clusters, and particularly the ones that are composed of nominal or prepositional elements, act as 'extremely productive frames' (ibid.: 1000) with a single or more variant slots, such as *'the ... of*

...', 'the ... of the', 'at the ... of', '... the ... of', filled by a range of words resulting in different 4-word clusters (cf. 'collocational frames' by Renouf and Sinclair, 1988). These structural frames function as 'a kind of discourse anchor for the new information in the slot' (Biber, 2006: 172). Several examples can be found in all twelve corpora, as can be seen in Table 3.22.

Similar frames which are verb-based have also been detected in this research. These often contain high-frequency verbs and a single variable slot which is typically filled by frequent content nouns (indicated in the parenthesis), as illustrated in the examples below:

Mathematics textbooks clusters:

- *calculate the ... of* (*area, height, length, radius, cost, amount, rate*)
- *find the ... of* (*value, equation, area, number, values, image, slope*)

Science textbooks clusters:

- *find the ... of* (*density, mass, volume*)

The above examples demonstrate how word clusters 'form recurrent discourse building blocks, with the following slot being used to express the content specific to each individual situation' (Biber et al., 1999: 991). On these grounds, 4-word clusters are useful devices for the comprehension and construction of subject-specific texts (e.g. Biber and Barbieri, 2007: 284).

Table 3.22 Examples of 4-word clusters which are underpinned by structural frames.

Noun phrase fragments			Prepositional fragments	
the ... of the		**... of the ...**	**at the ... of**	**... the ... of**
the amount of the	*the name of the*	*area of the triangle/rectangle*	*at the age of*	*in the case of*
the area of the	*the part(s) of the*	*centre of the circle*	*at the back of*	*in the form of*
the base of the	*the people of the*	*each of the following*	*at the base of*	*in the west of*
the beginning of the	*the perimeter of the*	*end of the first/second/third*	*at the battle of*	*on the outskirts of*
the bottom of the	*the population of the*	*equation of the axis/line*	*at the beginning of*	*to the rest of*
the centre of the	*the position of the*	*head of the church*	*at the bottom of*	*with the aid of*
the cost of the	*the power of the*	*law of the lever*	*at the end of*	*with the help of*
the direction of the	*the presidency of the*	*one of the most*	*at the rate of*	
the edge of the	*the radius of the*	*one of the world's*	*at the top of*	
the end of the	*the resistance of the*	*radius of the circle*		
the equation of the	*the rest of the*	*rights of the child*		
the function of the	*the results of the*	*slope of the line*		
the image of the	*the rights of the*	*temperature of the water*		
the law of the	*the sign of the*	*volume of the box/cylinder*		
the leader(s) of the	*the size of the*			
the length of the	*the slope of the*			
the location of the	*the subject of the*			
the mass of the	*the sum of the*			
the mean of the	*the surface of the*			
the measure of the	*the volume of the*			
the middle of the	*the walls of the*			
the mouth of the	*the weight of the*			
the name of the	*the temperature of the*			

Attention should also be drawn to the fourth structural category of clusters, that is, structurally complete units. Biber et al. found that '[i]n academic prose, less than 5% of the lexical bundles [in the LSWE corpus] represent complete structural units' (1999: 995) and that '[w]hen a lexical bundle is structurally complete in academic prose, it is typically a prepositional phrase that functions as a discourse signaling device' e.g. *for the first time, in the same way, in the first place, on the other hand* etc. (ibid.: 999). These structurally complete units are not classified as lexical bundles by Biber et al. and are thus not included in the major structures of their taxonomy. In this research, however, the structural analysis of 4-word clusters reveals a considerable number of structurally complete units with variable meanings in all corpora which are realised as follows:

noun phrase:
- *the ordnance survey map*
- *an either or choice*
- *the Irish Free State*
- *the human digestive system*
- *a six grid reference*
- *a frequency distribution table*
- *the highest common factor*
- *the League of Nations*
- *blue chloride cobalt paper,* etc.

(transitive) verb (phrase) + noun (phrase):
- *tick the correct box*
- *answer the following questions*
- *circle the correct answer*
- *read carefully the following*
- *draw up a list*
- *draw a Venn diagram*
- *complete the table below* etc.

prepositional phrase:
- *during the Industrial Revolution*
- *in twentieth century Ireland*
- *during the Middle Ages*
- *in a separate answer*
- *the age of exploration*

Structural frames have also been found to underpin a large number of structurally complete units in this research. These are typically noun phrases and less often prepositional phrases as demonstrated by the following examples:

Noun phrases:
the ... of ...
- *the act of union*
- *the age of exploration*
- *the age of revolutions*
- *the amounts of money*
- *the axis of symmetry*
- *the back of this*
- *the battle of Britain*
- *the centre of gravity*
- *the Council of Ministers*
- *the Council of Trent*
- *the density of water*
- *the end of year*
- *he government of Ireland*
- *the league of nations*
- *the name of one*
- *the number of people/pupils*
- *the oath of allegiance*
- *the piece of equipment*
- *the plantation of Ulster*
- *the point of intersection*
- *the presence of starch*
- *the rate of interest*
- *the reign of terror*
- *the Republic of Ireland*
- *the rights of others*
- *the transfer of heat*
- *the Treaty of Rome*
- *the Treaty of Versailles*
- *the war of independence*
- *the west of Ireland*

Prepositional phrases:

of the following ...

- *questions*
- *topics*
- *statements*
- *terms*

of the ... ...

- *action project*
- *Defence Forces*
- *European Parliament*
- *European Union*
- *following grid*
- *Periodic Table*
- *test tube*
- *United Nations*

Structurally complete clusters, and particularly those realised as: (transitive) verb (phrase) + noun (phrase), are often part of the long fixed expressions that were discussed in the previous section. They are taken into account in this analysis because they have important functional uses, as demonstrated in the discussion of the pragmatic meanings and functions of clusters in the next section (3.2.2.3), and they are thus important elements of the target communicative repertoire of ESL students.

Finally, it is also important to note that certain structures appear to be unique to the commonest clusters of particular corpora. Characteristic instances include the following:

Mathematics textbooks corpus:

- comparative expression (e.g., [Band A gets] *twice as much as* [Band B])
- *If*-clause fragment (e.g., *If the area of* [the triangle, rectangle, parallelogram])
- Noun phrase + copula be + Adjective (e.g., [How can we prove/What does it mean if] *two triangles are congruent*)
- Noun phrase + to-clause fragment (e.g., *use your graph to* [find/estimate the values])

Mathematics examinations corpus:

- comparative expression (e.g., [Note: 10 – 20 means 10 or] *more but less than* [20])
- modal verb + passive verb phrase fragment + *to*-clause fragment (e.g., [supporting work] *must be shown to* [obtain full marks])
- noun + modal verb + passive verb phrase fragment (e.g., [extra/graph] *paper must/can/may be obtained* [from the Superintendent], [supporting] *work should be written* [into the boxes provided])

English textbooks corpus:

- 3rd person pronoun + verb phrase fragment (e.g., *this is a very* [good response/answer])

English examinations corpus:

- 2nd person pronoun + verb phrase (e.g., *you must give the* [title of the poem/text/name of the play], *you must also answer* [any two/three of sections], *you must attempt all* [three sections], [In answering,] *you may not use* [the poem given on this paper/the extract given above])

Geography examinations corpus:

- noun + passive verb (phrase) (e.g., [a sketch map of the] *area shown on the* [photograph/Ordnance Survey map])

Possible explanations for the above could be found in the functional utility of these structures in discourse. Thus, the comparative structures of mathematics clusters serve to quantify and compare numerical results (e.g., [Band A gets] *twice as much as* [Band B]). The *If*-clause fragment (e.g., *If the area of* [the triangle, rectangle, parallelogram]) is essential in hypothesis testing that is found in theorems and mathematical problems. The noun phrase + copula be + Adjective structure often appears in geometry propositions to express that something is true (e.g., [How can we prove/What does it mean if] *two triangles are congruent*). The noun phrase + to-clause fragment, on the other hand, can be linked to imperative statements which instruct students to perform an action (e.g., *use your graph to* [find/estimate the values]). The 2nd person personal pronoun (e.g. *you must give*) of the English examinations clusters is indicative of the more conversational tone of language which addresses the reader directly. This can be contrasted with the impersonal structure manifested by the examination clusters of mathematics (e.g. *work should be written*). The frequent occurrence of passive verb structures in geography examinations clusters can be linked to multiple references to graphic displays of data (e.g. *as shown in the*). Although it may not be argued with certainty that the above structures are characteristic or truly representative of the specific registers in general, the fact that they appear to be manifested by the commonest clusters of the above individual corpora and not of others indicates their importance in these text collections. These structural patterns reveal characteristics of the types of discourse that is used in particular subjects and they can be brought to ESL students' attention with explanations of their functions.

On the whole, the interpretations and conclusions that emerge from the structural analysis of 4-word clusters should be accepted with caution. Due to the conservative cut-off points of frequency and range of occurrence and the limited data analysed, the caveat that comes with the above observations is, first, that the structural patterns presented here might not be representative of all the clusters of exams corpora and the particular subject-specific registers beyond the corpora of the present study (but they are still representative of the language of the Junior Cycle textbooks and examinations). Second, the manifested variation that can be observed in the structures of clusters of different corpora and in the number of clusters realised in each structural category may not represent genuine differences.

3.2.3.3 Pragmatic meanings and functions in subject-specific texts

Considering the differences in the structural correlates of 4-word clusters across the twelve corpora discussed in the previous section, it will come as no surprise that the typical meanings and functions of these clusters in discourse differ as well. The discussion of the semantic relations of 4-word clusters with the content of the corresponding textbooks and examinations corpora (sections 3.2.3.3.1 and 3.2.3.3.2 respectively) indeed reveals that recurrent 4-word patterns are influenced by the semantic (information content) and functional (communicative) requirements of particular subjects and registers.

3.2.3.3.1 Textbooks-derived 4-word clusters

From a brief glance at Table 3.15, the topic-specificity of textbooks-derived 4-word clusters becomes apparent. This is not surprising because, as mentioned earlier,the commonest words are responsible for the commonest patterns or slight variants of these patterns (Sinclair, 1987: 208) and, as shown in section 3.2.3.1, several subject-specific content words and collocations are recycled within 4-word clusters or appear as their complementation patterns. The 4-word clusters of the six textbooks corpora (Appendix D1) are discussed separately in this section in an attempt to isolate the influence of the subject area.

Beginning with the English textbooks corpus, it can be clearly observed that its limited number of clusters (Table 3.14) revolve around questions, answers and reasons and subjective viewpoints (similarly with the significant collocations of this corpus). The small number of the remaining clusters specify discourse structure (e.g. *the end/rest of the, at the beginning/end of, in the middle of*). The focus of clusters on asking and answering questions and on expressing personal opinions reflects the emphasis of the English syllabus on helping students to develop their communication skills and their thinking capacity as part of their 'personal growth through English' (draft English syllabus; NCCA, 2008: 3).

Moving to the geography textbooks corpus, it appears that the majority of 4-word clusters either denote areas or phenomena of geographical significance (e.g. *the West of Ireland, Pacific Ring of Fire, the Mid-Atlantic Ridge, of the earth's surface* etc.), or they express instructions from activities in which students engage in their geography textbooks (e.g. *draw a sketch map, tick the correct box, to support your answer, why do you think* etc.). A smaller number of clusters describe graphical means of representing information (e.g. *sketch map of*

the, the ordnance survey map, evidence from the photograph, six figure grid reference) or refer to information represented in these visual aids (e.g. *evidence from the map/photograph, of/in the area shown, shown on the map/photograph* etc). It can also be noticed that the language of geography does not contain a lot of technical terms. This could be explained by considering that the discourse of geography is primarily 'visual', i.e. the language of maps, globes, photographs, grids, bar graphs, pie charts, diagrams, pictorial models, statistics, etc., and a great deal of technicality is manifested in the 'language of images' (Martin et al., 1988: 150-155) in geography textbooks to encourage students to 'look through a geographer's eyes' (van Leeuwen and Humphrey, 1996: 30).

From an examination of the 4-word clusters of the history textbooks corpus, two types of clusters clearly stand out. First, a significant number of clusters refer to events (e.g. *American War of Independence*), issues (e.g. *the Treaty of Versailles*) and people (e.g. *the leaders of the*) from the past. Secondly, a large number of clusters mark the notion of time which is fundamental to the study of history and can be linguistically constructed in many different ways: 'Any historical writing is likely to draw on a wide range of linguistic expressions for construing time' (Coffin, 2006: 10). There are, therefore, clusters which mark particular historical eras (e.g. *the age of revolutions, during the middle ages, the early twentieth century*etc.) and others which mediate the chronological ordering of events (e.g. *at the beginning/end of, by the end of* etc.). Indirect reference to the past is also made by certain evaluative clusters which contain simple past verbs (and simultaneously signal the author's stance), namely, *was the most important, was one of the,* and *became known as the*. Only eight out of the 88 clusters in the history textbooks corpus can be directly linked to instructions. Ordered by frequency of occurrence, these are: *write an account of, an account of the, answer the following questions, give two reasons why, find out about the, out about the past, give one reason why, with the help of*.

For the most part, 4-word clusters in the CSPE textbooks corpus are nominal phrases which act as referents of civic, social and political entities and organisations from the local level of school and students' community, Ireland and Europe, as well as titles of official documents and agreements of political importance. The following are only few among many others:

- *of the European Union/European Parliament/United Nations*
- *Minister for the Environment*
- *South Dublin County Council*
- *The European Court of*

- *of the Defence Forces*
- *(Universal) Declaration of Human Rights*
- *The UN Convention on*

The concept of *rights*, which is fundamental to the study of CSPE, is also present in several clusters of this corpus:

- *the rights of others*
- *rights of the child*
- *convention on the rights*
- *right to freedom of*
- *the right to be*
- *has the right to*
- *everyone has the right*

It is interesting that the 4-word clusters which formulate questions in the CSPE textbooks corpus overlap with those that have the same functional use in the English textbooks corpus, and make use of the following four lexical words: *think, answer, questions, reasons*:

- *and answer the questions*
- *answer the following questions*
- *answer the questions that/below*
- *give one reason why*
- *give reasons for your*
- *reasons for your answer*
- *do you think is/that*
- *how do you think*
- *what/why do you think*

These clusters, in combination with the repeated use of directive clusters that contain the personal pronoun *you* and directly address the reader (e.g., *can you think of, do you think that, why/what/how do you think*), aim to encourage the active involvement and participation of students and help them develop their critical thinking skills which can indirectly contribute to the goal of CSPE, that is, to help students discover the importance of active, participatory citizenship, as part of their personal development. In addition, there are some clusters which indicate the real-life oriented tasks that are assigned to students, e.g. *to take part in, report on*

an action, carry out a survey, write a letter to, to protect the environment, to raise awareness of.

It is not accidental that the widest range of different 4-word clusters can be found in the mathematics textbooks corpus. The eight distinct thematic areas of mathematical study encompassed in Junior Cycle textbooks (listed below) create a need for varied linguistic resources which - in turn - results in a dearth of 4-word clusters. Direct links can therefore be established between most 4-word clusters and these eight thematic areas such as the following:

i) **sets**: *the point of intersection*
ii) **number systems**: *the total number of, the highest common factor, as a percentage of*
iii) **applied arithmetic and measure**: *calculate the cost of, amount of tax paid, volume of the box, an average speed of*
iv) **algebra**: *find the equation of*
v) **statistics**: *a frequency distribution table, the mean of*
vi) **geometry**: *the perpendicular sector of, two triangles are congruent, an axial symmetry in, the point of intersection, the slope of the*
vii) **trigonometry**: *a right-angled triangle, angle at the centre*
viii) **functions and graphs**: *your graph to estimate, maximum/minimum value*

In addition to these topic-specific (nominal) clusters, there are also clusters which are fragmentary (typically verb-based) instructions and descriptions of mathematical procedures to be undertaken by students such as:

- *write an equation in*
- *draw the graph of*
- *draw a Venn diagram*
- *use your/the graph to*
- *asked to find the*
- *we are given the*
- *divide both sides by*
- *construct the image of*
- *correct to one decimal/the nearest*
- *factorise each of the*

As mentioned earlier in the discussion, some of the commonest lexical verbs of mathematics textbooks are incorporated in several of these clusters, the two conspicuous ones being: *calculate the area/height/length/volume/radius/cost/amount/rate/of, calculate the mean mark*, and *find the value(s)/image/equation/area/slope/number of,* etc. (cf. the structural

frames in the previous section). Clusters which can be related to the reporting of findings and results are also present in the list of mathematics textbooks clusters:

- *are equal in measure*
- *is equal to the*
- *is the same as*
- *difference of two squares*
- *twice as much as*
- *has a radius/standard rate/volume of*
- *the results are shown*

Taken together, the mathematics textbooks clusters directly reflect the key practices of this subject: counting, measurement, pattern, and geometry. Similarly to the specialised nature of mathematics collocations discussed in section 3.2.2.3 (e.g. *under translation*), the 4-word clusters of mathematics textbooks corpora are also highly technical and quite distinct from the clusters found in other textbooks corpora. They might seem odd outside mathematical discourse not only because of the technical terms they contain (e.g. *bisector, perpendicular, equation, intersection, radius, factorise, standard rate cut off* etc.) but also because many of the common words they include acquire very precise meanings and unfamiliar usage in the world of mathematics (e.g. Pimm, 1987; 1995), e.g. *slope, image, mean, translation, problem, value, area, prime* etc. (cf. Skemp's 1976 *faux amis*). Even the use of a simple preposition like *by* can take on the unfamiliar meaning of *according to* (Bailey, in press, cited in Bailey, 2007: 13) in a phrase like *divide both sides by*.

As a result of these two facts, and also because of the less common syntactical patterns (e.g. *correct to the nearest whole number*), the meanings of 4-word clusters such as *under a central symmetry, an axial symmetry in, the highest common factor, slope of the line, calculate the mean mark*, etc. can be understood only 'in a strictly mathematical sense' (Ongstad, 2007: 2; cf. the seven types of structures which appear to be exclusively manifested by the commonest clusters of mathematics textbooks and examinations corpora, revealed though the structural analysis discussed in section 3.2.3.2). Taking account of the heavy load of terminology in mathematical discourse, which becomes evident from the above discussion of 4-word clusters, it is not surprising that the use of a mathematics glossary is recommended in the guidelines for post-primary mathematics teachers: 'a lexicon of mathematical terms can be built up gradually in a natural way for students'.

As regards, finally, the science textbooks corpus, its dense use of 4-word clusters reflects the routinized wordings that are used in conducting experimental work and in reporting

observations, results and conclusions. It should be noted that the 4-word clusters that occur in this corpus represent the topic concerns of the three sub-disciplines of scientific study, i.e. biology, chemistry and physics. Thus, the following links between thematic areas of the three sub-disciplines, as listed in the science syllabus (www.ncca.ie), and 4-word clusters can be established:

Biology:

a) the theme of energy conversion

- *a form of energy*
- *energy to electrical energy*
- *energy to heat energy*
- *one form to another to*

b) the study of plants and micro-organisms

- *make their own food*
- *energy from the sun*

Chemistry:

a) substances and materials

- *carbon dioxide and water*
- *acid or a base*
- *oxygen in the air*
- *blue cobalt chloride paper*

b) the processes by which materials can change and be changed

- *expand when heated and*
- *and contract when cooled*
- *heated and contract when*
- *the heating effect of*
- *is caused by the*

Physics:

4-word clusters which reflect the three major topics specified in the syllabus:

i) force and energy:

- *the force of gravity*
- *the density of water*
- *find the density/mass/volume of*
- *Law of the Lever*
- *boiling point of water*

ii) heat, light and sound:

- *the transfer of heat*
- *the heating effect of*
- *temperature of the water*
- *the boiling point of*

- *light travels in straight*

iii) magnetism, electricity and electronics:

- *the magnetic field around*
- *current flow in*
- *flow in one direction*
- *set up the circuit*

A lot of 4-word clusters in science textbooks are parts of phrasings which describe activities performed and investigations and experiments conducted in the context of the above three areas of study such as, for instance:

- *set up the apparatus/circuit*
- *experiment to show that/the*
- *find the density/mass/volume of*
- *test for the presence*
- *are made up of*
- *is measured in*
- *is caused by the*
- *is connected to the*
- *demonstrate the effect of*
- *describe with the aid*
- *add a few drops*
- *describe an experiment to*
- *flow in one direction*
- *heated and contract when*
- *what happens to the*

Pieces of equipment used in scientific study are also described by several 4-word clusters:

- *the piece of equipment*
- *a beaker of water*
- *gas jar of oxygen*
- *the mouth of the*
- *the north pole of*
- *the walls of the*
- *in a test tube*
- *set up the apparatus*

Measurement units and various quantities are expressed by some clusters:

- *the density/mass/volume of a*
- *the force of gravity*
- *the resistance/size/temperature/weight of the*

- *a few drops of*
- *a small amount/piece of*
- *for a few days etc.*

Finally, reference is made to graphic/visual representations of information and displays of data e.g. *shown in the diagram, of the Periodic Table, plot a graph of, make a list of,* etc. Technicality is inherent in the investigative and experimental study of science since the aim for students in the Junior Cycle science classroom is 'to understand the ''world'' by looking at it through a technical framework: turning commonsense understandings into technical understandings' (Martin and Veel, 1998: 298). This technicality is manifested in most science-specific 4-word clusters that denote substances and materials (e.g. *blue chloride cobalt paper*), equipment (e.g. *in a test tube*), experimental processes (e.g. *set up the circuit*) and scientific phenomena (e.g. *the force of gravity).*

From the discussion of textbooks-based clusters it becomes clear that their meanings and functions are directly associated with the informational content of subject textbooks and are thus markedly different across the six corpora. On these grounds, in addition to lexical words and collocations, 4-word clusters can also be a reliable indicator of disciplinary variation (e.g. Biber and Conrad, 1999; Biber, 2006; Biber and Barbieri, 2007; Hyland, 2008). In pedagogical terms, this means that an ESP perspective, considering each subject area on its own terms, is required in language support to help students appreciate the distinct multi-word patterns that are favoured by the discourse of different subjects and learn how to use these appropriately.

3.2.3.3.2 Examinations-derived 4-word clusters

By contrast with the diverse and topic-specific functional associations of textbooks-based clusters which mediate the subject-specific content of textbook units, most clusters in examinations corpora (Appendix D2) appear to share the same 'functional language use' (Schmitt, 2000: 101). This meets the rhetorical needs of examination paper writers: to formulate concise instructions to candidates and questions in conventionalized patterns in the different subjects. This is a direct consequence of the assessment focus of Junior Certificate examination papers and their uniformity, owing to their standardized format which is maintained over the years, and their restricted subject-specific factual content. It is therefore possible to identify three broad categories of 4-word clusters in examinations corpora: i)

clusters which convey instructions to candidates for the completion and submission of examination papers, most of which are quite similar across the six corpora, ii) clusters which mediate subject-specific questions and activities for assessment purposes and are evidently characteristic of individual corpora, and iii) clusters which signal the nature of specific subject areas by describing major concepts, themes, characteristic genres, instruments, materials, resources, graphic-visual means, etc. These three semantic and functional categories are the points of reference for the discussion of examinations clusters in this section. Only some illustrative examples are offered throughout the discussion rather than an exhaustive list of all the clusters that fall into each category.

With regard to the first category, there are small sets of 4-word clusters in the top 100 lists of the six corpora that provide instructions to candidates on how to complete and submit the examination paper. As can be seen in the examples listed below, these clusters refer specifically to the structure and to particular parts of the examination paper, to what must be submitted to examiners and to the marks assigned to each section or question:

English:

- *you must attempt all*
- *you must answer sections*
- *must also answer any*
- *sections on this paper*
- *minutes on this section*
- *each question is worth*

Geography:

- *please remember to return*
- *return this folder/paper with*
- *folder with your answerbook*
- *separate folder where spaces*
- *spaces are provided for*
- *otherwise marks will be (lost)*
- *write* your examination number (here)

History:

- *you must return this paper*
- *a separate answer book*
- *please enclose this paper*
- *not include these pages*
- *paper/pages for your answer*
- *centre stamp examination number*

CSPE:

- *make sure to write*

- *extra paper to answer*
- *you may use each*
- *hand up this paper*
- *one question in each section*
- *blank page for the*

Mathematics:

- *space for extra work*
- *extra paper can be*
- *obtained from the Superintendent*
- *the symbol indicates that*
- *work must be shown*
- *to obtain full marks*

Science:

- *for examiner use only*
- *indicate clearly the number*
- *if you require extra*
- *back of this booklet*
- *write your examination number*
- *number in the box*

Naturally, none of these clusters appears in textbooks corpora, as opposed to clusters from the other two categories mentioned above. Concordancing reveals that the clusters listed above are, in fact, parts of longer formulaic word sequences (see the long fixed expressions in examination papers described in section 3.2.3.1) which convey the standard information that is provided at the beginning of every examination paper. Their consistent occurrence increases their added value for pedagogical purposes in the language support classroom since they can facilitate ESL students' understanding of standard examinations instructions.

As regards the second functional category of examinations clusters, i.e. those which serve to formulate subject-specific questions and assessment tasks, there are striking differences across the six corpora, as can be seen in the examples below. The typical complementation patterns of these clusters (enclosed in square brackets) demonstrate their surrounding contexts of use.

English:

- *you must give the* [title/name of the text/poem you choose and the name of the author/poet]
- *read this poem* [and then answer the questions]
- *read the extract carefully* [and then answer the questions which follow]
- *read this piece and* [and then answer the questions]
- [in answering] *you may not use* [the poem given on this paper/ the extract given above as the basis for your answer]
- [explain (any two of) the following/tell the story of the poem] *in your own words*

- *from your reading of* [the extract/passage/poem/short story ...]
- [write a composition on one/about one page on one] *of the following topics*
- *pick only one topic*
- [support your answer with] *reference to the text/extract/poem*
- *will be rewarded for* [well-structured answers, clarity of expression, an appropriate tone, good grammar, spelling and punctuation]
- [write a speech] *for or against the* [motion]
- [you need to decide whether you are] *for or against the* [motion]
- [which character/teenager/part] *would you like to* [play?]
- *would you like to* [read more of this novel?]
- *what do you think* [the author/poet means/ the author's message is]
- *you have been asked* [to design a poster encouraging people not to waste water etc.]

Geography:

- *tick the correct box*
- [using the ordnance map provided/name/describe and] *explain two reasons why ...*
- *circle the correct answer* [in each of the (following) statements (below)]
- *draw a sketch map* [of the area shown on the Ordnance Survey map]
- [in the boxes provided] *match each of the* [letters/grid references in column X with the number of its pair in column Y]
- [one pair has] been completed for you
- [You have] *an either or choice* [within 4 questions]
- [Circle] *the correct answer in* [each of the following statements/sentences]
- [Look at/examine/study the] *ordnance survey map supplied* [with this paper]
- *the correct statements are* [numbered]

History:

- *write an account of* [one/two of the following]
- *write about that person*
- *write the title selected* [at the top of your account]
- *why do you think* [that/the ...]
- *do you think that*
- [If you wish, you may use the] *hints to help you* [in your answer]
- *this is an extract* [from an (oral/eye-witness) account by/of ...]
- [Write down] *the name of one* [government leader/civilization/country etc.]
- [Select one of] *the people described below*
- [Explain two of] *the following terms relating* [to ancient Ireland/a medieval manor/the Renaissance/the Middle Ages, etc.]
- *from your study of* [International Relations in the twentieth century/social change in twentieth-century Ireland/World War II/Irish history etc.]
- *give two reasons why* [the Renaissance began in Italy/Ireland stayed neutral in World War II/the 1916 Rising was a failure etc.]
- *explain one of the* [words below/following terms from the Age of Revolutions/Middle Ages/relating to the Irish Famine of the 1840s, etc.]

CSPE:

- *put a tick in the box* [opposite the correct answer]
- *please tick the question*
- *describe two actions that* [a school/your class/community/theDáil/government could take]
- *you may use each* [title/name and position/issue or topic/concept/political party only once]
- [briefly explain one reason/why you think] *it is important for* [young people to be properly informed/know about/Irish citizens to vote/people to respect the law]

Mathematics:

- [A parallelogram has dimensions/A triangle has measurements] *as shown in the* [diagram]
- *calculate the value of* [x in the diagram/triangle]
- *complete the table below*
- [Give your answer] *correct to one decimal* [place]
- [Give your answer] *correct to the nearest* [whole number/degree/minute/metre/km etc.]
- *draw the graph of* [the function of f]
- [Find/factorise fully/write] *each of the following*
- [use the graph/your answers to] *estimate the value of*
- *find the area of* [the triangle/rectangle/circle/square/garden/flowerbeds etc.]
- *find the equation of* [the line]
- [using calculator, or otherwise] *find the exact value* [of]
- *find the volume of* [the cylinder/cone/sphere/tank/block/box/this cube etc.]
- *is a point on* [the circle/line/graph]
- *list the elements of*
- *write down the value* [of cos., tan., sin.]

Science:

- [Describe, with the aid of a labelled diagram,] *an experiment to show* [that oil can reduce friction/the release of energy from food etc.]
- [Name one method/a substance that] *can be used to* [prevent the rusting of iron/put out fires etc.]
- *describe with the aid* [of a labelled diagram]
- [What] *does this tell us* [about the speed of light/light and sound etc.]
- *draw a graph of* [voltage against current/solubility against temperature/velocity against time etc.]
- [Describe, with the aid of a labelled diagram, an] *experiment to investigate the* [effect of grain direction on the strength of wood]
- *give one function of* [teeth/the stomach/a root/leaf etc.]
- [Name and] *give one use for* [this piece of equipment/the following two devices/the alloy/metal you have named]
- [what] *is the function of* [fat in the diet/the resistor in the circuit]
- *name the part labelled* [A/B/C]
- *name the piece of* [equipment labelled A/shown/apparatus]
- *name the type of* [switch shown in the circuit/bonding in sodium chloride/joint shown in the diagram etc.]
- [name] *the following pieces of* [equipment]
- [What happens to the aluminium strip when] *the switch is closed* [?]
- *use the graph to* [estimate the distance/solubility/melting point/weight etc.]
- [Which piece of equipment A or B is usually] *used to measure the* [hydrochloric acid during this experiment?]
- *what is meant by* [the safety symbol shown/flame-proofing a fabric/electroplating?]
- *what is the name* [given to the separation technique shown in the diagram?]
- [Choose a] *word from the list* [on the right to complete the (following) sentences (below)]
- [what] *would you expect to* [notice/observe/see/happen]
- [name one food that is] *a good source of* [fat/starch/protein/fibre]

Although the 4-word clusters presented above are a limited sample of clusters selected from only the top 100 clusters of the six corpora (Appendix D2), they are sufficient to demonstrate the wide variation that exists in the construction of examination questions and tasks across the

six subject areas. By describing subject-specific assessment questions and exercises, these 4-word clusters also indicate aspects of knowledge and skills that appear to be central to Junior Certificate examinations.

A sub-category that can be detected in the same group of clusters, however, comprises a set of semantically neutral clusters that all serve the function of phrasing questions, requesting answers and/or explanations. It is the items that complement these clusters to each side that add the subject-specificity dimension to examination instructions, as shown below:

English:

- [read this piece/poem/the extract carefully and] *then answer the questions*
- [give] *reasons for your answer*
- *support/explain your answer with* [reference to the poem/text/extract]
- [base your answer/give reasons for your answer based] *on evidence from the* [poem/text/extract]
- [you may not use the extract given above/choose either of the scenes quoted on this examination paper as the] *basis for your answer*

Geography:

- [Geographical mix:] *answer any three of* [the questions below]
- [Examine the weather map/ Study the pie charts/cartoon below/the diagram below] *and answer the questions* [that follow/below]
- [all] *questions to be answered*
- *explain two reasons why*

History:

- *answer the following questions [in a separate answerbook]*
- *[study the documents which accompany this paper] and then answer the [following questions]*
- *[answer ten/two] of the following questions*

CSPE:

- [When you have studied this poster/brochure/information leaflet/postcard/web article etc.] *answer the questions below*
- [If you need extra paper to] *answer this question please* [ask the Examination Superintended for it]
- *answer all/one/two/three of the* [questions in this section]
- [Write your] *answers in the spaces* [provided]

Mathematics:

- [give] *a reason for your* [answer]
- [Give your] *answer in the form*
- [Give your answer] *in its simplest form*

Science:

- *give a reason for* [your answer]

- *give one reason why* [dieting can be dangerous/such a heater should be earthed etc.]
- *answer the questions in* [the spaces provided in the booklet]
- *give a reason why* [the treatment that you have named is carried out/this is the case etc.]

Tied to this same sub-category, clusters which request evidence to support one's answer can be found in the following three corpora:

English:

- [base/basing your answer/give reasons for your answer based] *on evidence from the* [poem/text/extract]

Geography:

- *using evidence from the* [(Ordnance Survey) map/aerial photograph]

History:

- [give one piece of] *evidence to show that*
- [give one piece of] *of evidence from the* [account/picture/extract to show/support your answer]

The above clusters can also be found in the lists of the commonest textbooks clusters as their functional role is fundamental in formal education in general. They can therefore usefully provide a basis for a common core of academic word clusters that have added pedagogical value as they are important in both registers (textbooks and examinations) and across subject areas.

Notable differences are also identified in the final category of clusters which mark the 'subject-specific identity' of individual examinations corpora. As the following examples demonstrate, clusters in this category reveal aspects that are important to the discourse of the six subject areas and thus different across the six corpora:

Genres in English:

- [Think about/name/select/describe the opening of a etc.] *novel or short story* [you have studied]
- *a poem/film/play you have studied*
- *extract in edited form*
- *the advertisement on page*

Visual aids in geography:

- *the Ordnance Survey map*
- [Draw] *a sketch map of* [the area shown on the photograph/Ordnance Survey map]
- *survey map and legend key*
- [Study] *the aerial photograph of* [Wicklow/Galway/Killarney/Mullingar etc.]

- [The feature/instrument/process/type of rainfall] *shown in the diagram*

History:
Genres

- *an account of one*
- *write an account of*

Time

- *the age of exploration*
- *during the middle ages*
- *during the Industrial Revolution*
- *in the twentieth century*

Themes

- *international relations in the*
- *in twentieth century Ireland*

Text types in CSPE:

- [a blank] *page for the poster* [has been included at the back of this answer book]
- [Write] *a short article for* [your school magazine explaining ...]

Mathematics:
Time

- *at the end of* [this/the first/second etc. year]

Comparison

- *more but less than*
- *or more but less*

Mathematical concepts and terms

- *the area of the* [triangle/rectangle/parallelogram/lawn/garden/figure etc.]
- *the length of the* [side opposite to the angle/hypotenuse/radius/perimeter etc.]
- *the volume of the* [cylinder/cone/sphere/container/block/cube/capsule etc.]
- *the image of the* [point/triangle/rectangle/figure/letter etc.]
- *amount of the investment*
- *the height of the* [tree/cliff/office block/mast/cylinder etc.]
- *the right angled triangle*
- *the equation of the* [line]

Science:
Equipment and visual aids

- *the piece of equipment*
- *a three pin plug*
- *the apparatus shown in*
- *of a labelled diagram*

with particular emphasis on ***diagrams*** and their depicted content:

- *diagram shows part of* [the human digestive system/human skeleton/human breathing system etc.]
- *diagram shows the structure* [of a flower/tooth/human lung etc.]
- *the diagram shows a* [ray of light/circuit/flowering plant/food pyramid/beaker etc.]
- *the diagram shows an* [apparatus/experiment/instrument/atom of lithium etc.]
- [the diagram shows] *the human digestive system*
- [name/identify] *the part of the* [cells/atom/plant/flower]

Measurement unit

- *the volume of the* [stone/water/acid/base]
- [friction/starch is] *an example of a* [force/carbohydrate]

Location

- *in a school laboratory*

On the whole, the discussion of examinations clusters in this section is likely to refute any expectations about examinations language being more neutral than subject-specific. On the contrary, the analysis shows that, with the exception of the small number of generic clusters which are linked to posing questions and inviting answers, like the ones listed above, the functional associations and pragmatic meanings of examinations clusters are directly rooted in subject-specific concerns. Although it is not the 4-word clusters that are characterised by subject-specificity in most cases but rather their complementation patterns that mark their content area, there are still important differences in the actual clusters that are preferred by individual subjects to formulate assessment questions. It appears that the communicative functions of clusters reflect the distinct types and purposes of assessment tasks of the different subjects. Thus, the clusters of science examinations, for instance, serve the purposes of defining, classifying, describing and explaining (Martin and Veel, 1998: 298) while the English examinations clusters are concerned with formulating and justifying different points of view, the assumption being that one or another is more preferable.

To conclude, it can be argued that the 4-word clusters of examinations corpora discussed in this section probably represent some of the most fundamental and institutionalised linguistic means employed for assessment purposes by the six curriculum subjects because of their repeated occurrence and subject-specific functional language uses and meanings. Thus, their status makes them undoubtedly an important and very useful phraseological resource for ESL students.

3.2.3.4 Conclusions from the 4-word cluster analysis

This extended section of Chapter 3 described and discussed the 4-word clusters that recur in the textbooks and examinations corpora of the six subjects in relation to their density, forms, frequencies, complementation patterns, and their structural and functional characteristics. At all levels of analysis, important differences were revealed across corpora, indicating that the nature and contextual use of 4-word clusters are conditioned by the requirements of subject-specific texts.

Striking differences were first observed in the density of 4-word clusters between textbooks and examinations corpora as a whole and among the six subject corpora in each register, providing hints to the extent of formulaic language used in individual corpora. More specifically, examinations corpora exhibit a larger number of different clusters occurring with higher frequencies as a whole than the clusters of textbooks corpora. Looking at individual subject corpora, the textbooks and examinations corpora of mathematics and science appear to make more extended use of 4-word clusters than those of geography and history, while the two corpora of English manifest a limited number of different 4-word clusters.

To explain the differences in the density of textbooks and examinations clusters, it was argued that the need for concise and consistent phrasing of examination questions, regardless of the subject area, is addressed by the use of routinised, fixed language expressions manifested in 4-word (and longer) clusters, as opposed to the more diverse language choices made by textbook writers to mediate the informational load of subject knowledge. The higher concentration of 4-word clusters in the corpora of the two more technical subjects of mathematics and science can be attributed to the fact that conventionalised language is employed to discuss numbers, symbols and graphic displays of data, which make up a large proportion of the mathematical and scientific discourse. These findings are in line with conclusions drawn from other corpus studies of disciplinary variation of word clusters (e.g. Conrad, 1996; 2004; Biber and Barbieri, 2007; Hyland, 2008, etc.). For instance, Biber (2006: 164) reports that university textbooks of social science, humanities and natural science occupy opposite ends of the continuum in cluster density. His explanation becomes relevant here as he identifies technical content and stylistic preferences to be the two factors responsible for this difference:

> Natural science textbooks convey dense technical content, and thus use specific terms and expressions to refer to that content, aiming to achieve an explicit conveyance of meaning. It is likely that this reliance on a specific set of technical terms contributes to the dense

> use of lexical bundles. In contrast, humanities textbooks are more concerned with the critical discussion of ideas and interpretations. In addition, humanities authors value stylistic variation, often expressing the same idea in multiple ways for the same of a more highly varied style. Such stylistic variation would contribute to the overall lesser reliance on lexical bundles.
>
> (ibid.)

What came as a surprise was the heavy reliance of CSPE textbooks and examinations corpora on a large stock of different 4-word clusters. Although it is difficult to account for the inflated cluster numbers in CSPE texts, a possible explanation could be found in the pragmatic meanings of these clusters. An examination of these reveals that several clusters denote labels, names and titles of civic, social and political entities (organisations, political parties and documents, etc.) which are realised in four words (e.g. *The Council of Ministers, South Dublin County Council, The Council of Europe*, etc.).

As regards the forms of the commonest 4-word clusters in the twelve corpora, two readily observable patterns emerge. Several textbooks and examinations clusters appear to be repetitions of the same but longer cluster, with one or more words inserted or deleted. Concordance lines of 4-word clusters specifically reveal that these variants can often extend from five up to twenty words, forming (semi)fixed sentences strung together in a series and recur particularly in examinations corpora, describing subject-specific tasks and instructions. Another trend is the recycling of high-frequency lexical words and significant collocations within the commonest 4-word clusters or in adjacent positions as their complementation patterns. The range of examples that were provided to demonstrate these two tendencies constitute a clear manifestation of the operation of Sinclair's (1991) idiom principle.

The structural analysis of 4-word clusters in the twelve corpora highlights their nominalised and structurally incomplete nature, supporting the findings of Biber et al. regarding word clusters in academic prose (1999: 998 - 1036). Most 4-word clusters constitute noun and prepositional phrases which incorporate *of* + fragments (e.g. *a sketch map of, in the form of*, etc.). The vast majority of 4-word clusters also bridge two structural units (e.g. noun phrase and the beginning of a prepositional phrase such as *the top of the,* or prepositional phrase + of fragment such as *with the aid of*, etc.). A number of clusters in all twelve corpora, however, were found to be structurally complete units (e.g., *the government of Ireland, tick the correct box*, etc.). It was further demonstrated that many fragmentary and structurally complete clusters are underpinned by productive structural frames such as *the ... of the, at the ... of*, etc. (cf. Sinclair and Renouf, 1988), with missing slots filled by content-specific items. The relevant examples that were provided demonstrate the role of 4-word clusters as building

blocks of subject-specific discourse, 'reflecting routinized ways of unfolding and presenting information in continuous discourse' (Altenberg, 2007: 243).

The qualitative analysis of the discourse meanings and functions of 4-word clusters revealed that, similarly with lexical words and collocations, recurrent 4-word combinations are also influenced by the semantic and functional requirements of subject-specific texts. On the one hand, textbooks clusters directly reflect the informational content of subject textbooks and can be thus used as a reliable indicator of linguistic variation across the six subject areas. Examinations clusters, on the other hand, also reflect disciplinary variation but mostly when combined with their complementation patterns and few of them do so in their own right. A set of semantically neutral 4-word clusters common to the six corpora was further identified. These serve to instruct students to answer questions and provide evidence and explanations in their answers (e.g. *answer the questions below, reason for your answer*, etc.), and it is again their complementation patterns that add subject-specific substance.

On the whole, the findings from the detailed analysis of 4-word clusters that recur in the twelve corpora reveal that, as with lexical words and collocations, recurrent word clusters are also influenced by the needs of individual subject-specific texts, thus adding further evidence to the existing research that demonstrates that each register employs a distinct set of lexical bundles (e.g. Biber and Barbieri, 2007: 265). These contribute to the discourse structure and meanings of subject-specific texts and their semantics and functional utility make them valuable for accessing textbooks and examination papers and for constructing appropriate texts.

3.3 Common lexical features across corpora: cross-curricular words and 4-word clusters

Having drawn frequency profiles of the twelve subject-specific corpora, comprising the commonest i) words (and their collocates), ii) significant collocations, and iii) 4-word clusters, this section proceeds to corpus comparisons in order to detect commonly shared language features and investigate their uses in subject-specific discourse. Cross-curricular language features are pedagogically important because 'they are required to understand written texts from different registers and different academic disciplines' (Biber, 2006: 34; i.e. post-primary subject textbooks and Junior Certificate examination papers of the six curriculum subjects

analysed in the present research). They, therefore, have an added value compared to certain subject-specific language features that are likely to be less useful outside the discipline to which their use is restricted. In post-primary language support, lexical linkages among subjects can promote a multidisciplinary approach to vocabulary teaching and learning with considerable pedagogic benefits for ESL students that can be derived from exposure to the same lexical items in multiple thematic contexts. This section establishes lexical links across the six textbooks corpora and the six examinations corpora specifically at the level of content words and 4-word clusters, and compares and contrasts the typical uses of these features in subject-specific contexts.

In the first stage, the Detailed Consistency analysis function of WordSmith 4 (Scott, 2004: 138) is employed to identify the content words and 4-word clusters which meet the 40 per million words frequency threshold and recur consistently a) across textbooks corpora and b) across examinations corpora. This function displays the results (cross-curricular words in Appendices E1 and E4 and cross-curricular clusters in F1, F3) as follows:

N	Word	Total	Texts	English	Geography	History	CSPE	Mathematics	Science
1	WATER	4291	24	0.27	3.05	0.46	0.53	0.44	9.65

Next to each word (or cluster), the *Total* column shows how many instances the word occurs overall (raw frequency count), Texts shows how many text-files it was found in. Then information in provided on the frequency of the word in each corpus in two columns; the first column lists the raw frequencies that emerge from the Detailed Consistency Analysis functionality of WordSmith and the second column displays the corresponding normalized frequencies (‰) which were manually calculated to facilitate direct comparisons of the results. In the present chapter, as shown in the above table, only normalised frequencies are provided to enable comparisons of the quantitative occurrences of cross-curricular features across corpora. Accordingly, in the above example, which is based on the comparison of the textbooks-based content words, the word *water* occurs in all 24 texts of the six textbooks corpora (recall that each textbooks corpus is composed of four different textbooks), with a raw frequency of occurrence of 4,291 in all, and it is most frequent in the science textbooks corpus with a normed frequency of 9.65‰. Manual editing of the detailed consistency analysis lists involved the deletion of numbers and single letters.

As a second step, the usage of these cross-curricular words and clusters is examined at the concordance level, focusing on the most significant collocates (computed using the G^2 ratio) of words and the complementation patterns of clusters. A wide range of authentic

examples of contextual use are provided throughout the discussion (i.e. concordance lines) for a qualitative treatment of the above items in order to illustrate their variable uses across subject textbooks and examinations corpora.

3.3.1 Vocabulary use: cross-curricular content words

Comparisons of content words across subject corpora bring to light general academic words that uniformly appear in a wide range of contexts and which are used (in)variably across subject areas (cf. the General Service List by West, 1953; the University Word List by Xue and Nation, 1984; and the Academic Word List by Coxhead, 2000). Function words are excluded from the comparison because their pervasiveness in all corpora and their importance to effective language use in all communicative contexts is taken for granted. The comparison of content words in the six corpora of each register that meet the 40 per million words frequency restriction reveals 203 content words to be common to the six textbooks corpora (Appendix E1) and only 34 content words to recur in the six examinations corpora (Appendix E3) respectively. From these content words, only those that recur consistently in every single text-file of the text collections that comprise the six corpora of each register (i.e. in all 24 files of the six textbooks corpora and in all 60 files of the six examinations corpora) are described and discussed in detail in the following sections (3.3.1.1 and 3.3.1.2). The complete listings of cross-curricular words in textbooks and examinations corpora are presented in Appendices E1 and E4 respectively. The most significant collocates (G^2 score ≥ 15.13) of some of the top cross-curricular content words in the twelve corpora are displayed in Appendices E2 and E5. Concordance lines demonstrating the uses of these cross-curricular words together with their collocates across subject-specific texts are provided in Appendices E3 and E5.

3.3.1.1 Cross-curricular words in textbooks corpora, their central collocates and contextual use

From the 203 content words that are found to be common among the textbooks corpora, 95 recur in all 24 text-files of the six corpora, the top 20 of which are presented in this section (Table 3.23). The rest are listed in Appendix E1 (53 words occur in 23 text-files, 35 words in 22 text-files, 16 words in 21 text-files and four words in 20 text-files). As can be seen in Table 3.23

and in Appendix E1, most of these 95 content words are highly frequent English words of general use, featured in West's (1953) General Service List of English words which comprises the 2,000 words selected to be of the greatest 'general service' to English language learners (e.g. *find, explain, make, answer, give, example, following,* etc.). Not surprisingly, only a small number of content words could be characterised as subject-specific (e.g. *water, speed, travel, red*) rather than of general applicability.

Table 3.23 The top 20 cross-curricular content words in the six textbooks corpora.

N	Word	Texts	Frequency ‰					
			English	Geography	History	CSPE	Maths	Science
1	WATER	24	0.27	3.05	0.46	0.53	0.44	9.65
2	FIND	24	0.58	0.33	0.31	1.23	6.85	1.15
3	FOLLOWING	24	0.98	1.93	1.19	0.97	2.39	0.73
4	USED	24	0.67	1.12	1.57	0.68	0.4	2.9
5	AREA	24	0.12	2.82	0.27	1.29	3.6	0.43
6	NUMBER	24	0.15	0.85	0.38	0.88	5.38	0.97
7	USE	24	1.09	1.51	0.67	0.92	1.28	1.58
8	TIME	24	1.36	0.98	1.19	0.92	1.28	1.08
9	NEW	24	1.08	1.6	1.85	1.23	0.25	0.42
10	MAKE	24	1.3	0.61	0.64	1.76	0.35	1.45
11	LIKE	24	2.19	0.42	0.47	1.26	0.42	0.7
12	WORK	24	0.49	0.98	1.12	2.26	0.24	0.69
13	ANSWER	24	1.67	0.79	0.47	0.79	0.9	0.2
14	PLACE	24	0.65	0.88	0.49	0.85	0.41	1.86
15	EXAMPLE	24	0.53	0.64	0.36	0.79	2.67	0.73
16	GIVE	24	1.15	0.86	0.57	1.02	0.35	0.7
17	SEE	24	1.15	0.95	0.31	0.81	0.22	1.08
18	LINE	24	0.51	0.58	0.15	0.13	3.65	0.36
19	POINT	24	0.55	0.6	0.12	0.19	2.84	0.64
20	DIFFERENT	24	0.59	0.73	0.26	1.14	0.23	1.43

Examining the 95 common words in respect of their lexical categories, the majority appear to be nouns, followed by verbs and fewer adjectives and many fewer adverbs. Cross-curricular nouns fall into many different semantic categories (adapted from Biber, 2006: 248; specifically, the labels for the semantic classification of verbs and adjectives and adverbs are borrowed from Biber et al.'s analysis of the LGSWE corpus, 1999: 361-371, and the semantic categories for nouns are based on the analysis of the T2K-SWAL corpus by Biber, 2006: 248-250):

- cognitive nouns, i.e. 'mental/cognitive processes or perceptions', e.g. *information, question*
- concrete nouns, i.e. 'inanimate objects that can be touched', e.g. *box, parts, water*
- technical/concrete nouns, i.e. 'tangible objects that are not normally perceived and/or cannot be normally touched', e.g. *line, point, centre, end, list*
- quantity nouns, i.e. 'nouns specifying a quantity, amount, or duration', e.g. *number, time, times, piece, group, type(s)*
- place nouns, i.e. 'places, areas, or objects in a fixed location', e.g. *area, place, school, class, ground, space, top*
- abstract/processes, i.e. 'intangible, abstract concepts or processes', e.g. *form, use, example, speed, way, side, order, range*

The verbs that consistently appear in the six textbooks corpora can be classified as follows (Biber et al., 1999: 361-371):

- activity verbs which 'primarily denote actions and events that could be associated with choice, and so take a subject with the semantic role of an agent' (LGSWE, pp. 361-362, 367-368, 370), e.g. *make(s), give/given, take/taken, put, used/using, show(s)/shown, work, go, travel, turn, bring, change, clear, cut, try, join, look, left*
- mental verbs which 'denote a wide range of activities and states experienced by humans', e.g. *know/known, find/found, read, study, like, see*
- communication verbs, 'a special sub-category of activity verbs that involve communication activities (speaking, writing) (LSGWE, pp. 362, 368, 370), e.g. *say/said, answer, explain, state*
- aspectual verbs which 'characterize the stage of progress of some other event or activity, typically reported in a complement clause following the verb phrase' (LGSWE, pp. 364, 369, 371), e.g. *complete, keep*

Commonly shared attributive adjectives denote size (*large, long, higher*), time (e.g. *new*), colour (e.g. *red*), evaluation (e.g. *important, main, same, different, right*), and they can be also relational (e.g. *following, middle*) (Biber, 2006: 245). Finally, adverbs refer mainly to place (e.g. *inside, outside, away, across*).

Considering that these words are consistently used in all text collections of the six textbooks corpora, it would be useful to compare and contrast their central collocates within their wider context of use in order to make explicit their (in)variable uses in the discourse of the six subjects. Table 3.24 below displays the 15 most central collocates of the top ten cross-curricular words (within a 5-word span to either side) according to the G^2 ratio (sorted by descending significance value scores). To minimise the load of data presented here, collocates

with a G^2 score greater than or equal to 15.13 (G^2 score ≥ 15.13) are presented.[2] For some words, less than 15 collocates are provided because the rest appear with lower G^2 scores and are thus not viewed as statistically significant. Complete lists with all the collocates of the top ten cross-curricular textbooks words and detailed information on G^2 scores, total frequency and frequency in each position can be found in Appendix E2.

Table 3.24 The top 15 collocates of the top ten cross-curricular content words in the six textbooks corpora ranked by descending order of significance based on G^2 scores (≥ 15.13).

Textbooks corpus	Cross-curricular word and collocates
	water
English	bottle, like
Geography	supplies, clean, supply, fresh, cycle, air, condenses, river, atmosphere, shortage, resource, surface, mark, available, irrigation
History	running, supply, steam, pump, toilets, frame, powered, spinning, tap, pumps, supplies, drinking, filled, engine, power
CSPE	clean, supply, pollution, sewerage, air, chemicals, access, food, day, safe, waste
Mathematics	level, cm, tank, container, cylindrical, spheres, volume, height, filled, litres, difference, cylinder, calculate
Science	boiling, vapour, beaker, dioxide, hard, salt, carbon, tap, test, cold, oxygen, air, tube, hot, soft
	find
English	evidence, answers, words, following, passage, interesting, extract, (you'll), silently, reading, story, useful, read, book, people, information
Geography	directions, area, difficult, grid, references, map, help, following, countries
History	past, year, routes, explorers, historians, new, North, archaeologists, wanted
CSPE	names, summarise, survey, chapter, information, newspaper, work, research, helpful, discovery, students, words, activity, following, people
Mathematics	value, equation, values, image, slope, point, coordinates, area, solve, line, graph, use, number, corresponding, range
Science	density, mass, volume, experiment, stone, gravity, graph, use, centre, let's, weight, want, ph, number, liquid
	following
English	answer, questions, read, carefully, words, meanings, poem, extract, answers, complete, extract, find, write, passage, using
Geography	grid, answer, questions, mark, features, references, identify, statements, name, terms, headings, CD, locations, reference, show
History	questions, answer, page, explain, terms, account, hints, headings, using, years, relating, name, author, war, lines
CSPE	words, questions, read, answer, statements, sentences, ticked, missing, indicate, true, false, decide, boxes, whether, help
Mathematics	table, steps, complete, frequency, distribution, express, form, simplify, factorise,

[2] G^2 scores can be interpreted as follows: A G^2 of 3.8 or higher is significant at the level of $p < 0.05$ and a G^2 of 6.6 or higher is significant at $p < 0.01$.

- 95th percentile; 5% level; $p < 0.05$; critical value = 3.84
- 99th percentile; 1% level; $p < 0.01$; critical value = 6.63
- 99.9th percentile; 0.1% level; $p < 0.001$; critical value = 10.83
- 99.99th percentile; 0.01% level; $p < 0.0001$; critical value = 15.13 (http://lingo.lancs.ac.uk/llwizard.html)

	evaluation, fraction, equations, shows, diagrams, write
Science	blanks, fill, rewrite, copy, science, circuits, discovering, complete, answer, table, set, questions, use, symbols, apparatus
	used
English	words, techniques, language, English, kind, effectively, describe, image, phrases, comment, effect, advertisements, quotation, poem, create,
Geography	make, peat, generate, transport, measure, instrument, power, energy, humidity, concrete, offices, relative, resource, electricity, fuel
History	steam, methods, tactics, describe, make, build, stone, propaganda, people, weapons, storing, power, bronze, tools, cooking
CSPE	skills, skill, land, action, school
Mathematics	link, form, method, given, equation, case, letters, problems, cm
Science	apparatus, chemicals, measure, test, small, separate, make, beaker, widely, commonly, power, energy, tubes, water, units
	area
English	people, local, community, development, population, joined, neighbourhood, litter, youth, cities, facilities, young, Dublin, own, problems
Geography	map, shown, calculate, sketch, photograph, urban, pressure, Dublin, surrounding, woodland, harbour, Dungarvan, low, large, people
History	pale, called, small
CSPE	people, local, development, population, cities, famous neighbourhood, youth, language, own, currency, joined, particular, facilities, young
Mathematics	rectangle, triangle, calculate, find, volume, cm, base, circle, total, perimeter, parallelogram, length, height, width, maximum
Science	pressure, force, surface, square, unit, per, measuring, shaped, well, low, large, measured, small, object, find
	number
English	points, makes, question
Geography	column, squares, per, pair, count, people, average, children, grid, smaller, shopping, rate, factors, mother, ways
History	people, reasons, total, voters, small, increased, Irish
CSPE	people, votes, children, school, increase, needed, flags, decide, street, name, countries
Mathematics	people, times, pupils, whole, larger, prime, total, per, smaller, let, line, added, students, factors, find
Science	atomic, protons, electrons, mass, atom, neutrons, count, same, units, per, times, hours, shell, total, atoms
	use
English	language, words, quotations, poets, writers, made, imagery, dialogue, phrases, good, reference, make, support, quotes, evidence
Geography	land, evidence, map, headings, atlas, intensity, photograph, support, pencil, disuse, identify, help, USAID, references, values
History	language, sources, land, public, made, methods, vernacular, made, wanted, violence, propaganda, against, able, own, archaeologists
CSPE	energy, bags, products, people, plastic, information, questions, like, school
Mathematics	graph, information, estimate, find, form, calculator, formula, curve, solve, ratio, represent, protractor, adjacent, fact, graphs
Science	make, electricity, find, experiment, test, energy, graph, plants, apparatus, show, key, name, tongs, careful, instead
	time
English	long, place, same, spend, part, year, different, waste, take, setting, setting, day, Shakespeare's, story, came, (didn't), happy
Geography	long, change, full, year, changes, period, part, population, changed, same, function, men, women, functions, grow
History	same, part, spend, full, leisure, come, invented, short, soldiers, long, militia, historians, studied, money, little

CSPE	full, date, spend, part, election, years, people, school
Mathematics	taken, distance, speed, total, journeys, minutes, kilometres, hours, years, journey, average, axis, temperature, latest, calculate
Science	distance, taken, velocity, graph, putting, long, horizontal, against, takes, seconds, same, axis, change, speed, short
	new
English	York, blouse, line, land, paragraph, year, light, students, words, look, school, world, see, own, story
Geography	towns, town, Tallaght, suburbs, built, housing, York, city, industries, Shannon, houses, estates, developments, created, planning
History	towns, world, called, built, York, party, lands, ideas, orders, constitution, routes, introduced, government, industries, brought
CSPE	laws, subject, York, skills, school, proposals, friends, teachers, EU, commission, building, lob, law, UN, member
Mathematics	situation, principal, price, level, total, paid, money, year
Science	form, substance, change, form, seed, plant, growth, physical, called, plants, substances, read, volume, properties, take
	make
English	sure, points, costume, think, sense, reader, list, speech, hair, story, point, want, feel, use, clear
Geography	used, soil, sure, fertile, help, use, droplets, people, living, products, difficult, cloud, area, together, building
History	laws, used, tools, own, weapons, moulds, easier, Ireland, sure, bread, iron, barley, use, wanted, parliament
CSPE	sure, difference, list, history, laws, posters, poverty, team, changes, better, information, decisions, people, leaflets, power
Mathematics	sure, true, coefficients, term, values, sides
Science	sure, own, food, list, used, plants, use, collection, simple, things, sketch, photosynthesis, chart, different, bacteria

Taking account of the findings from the collocational analysis of textbooks corpora (sections 3.2.2.2 and 3.2.2.3), the topic-specificity of collocates in Table 3.24 and Appendix E2 comes as no surprise; cross-curricular content words embedded in subject-specific contexts reflect each time the information content of the corresponding texts.

To paint a clear picture of striking differences, and of any existing similarities, in the use of these common items across subject areas, the typical occurrences of some of them (namely of the nouns *water, time, area,* and verbs *find, use*, *make* listed above) within subject-specific texts are described here. The focus of the qualitative analysis is primarily semantic (rather than grammatical or syntactical), placing emphasis on the core meanings cross-curricular words acquire in different contexts and their thematic relation to the broader discourse. The examples provided are based on concordance lines of only some of the top collocates of these content words that were manually selected with a pedagogical intent in mind (cross-curricular words are marked in bold and their collocates are indicated in italics).

An examination of the word that is, surprisingly, found to be the most frequent and most widely distributed content word across the six textbooks corpora, *water*, yields interesting insights into its discursive treatment in the texts of different subjects. In its five

occurrences in the English textbooks corpus, it is twice used with the collocate *bottle*: e.g. 'I threw a **water** *bottle* at Connor', and it appears three times with *like* forming similes that are characteristic of literary texts: e.g. '**Water** that glistened *like* silver', '... slipping away under the **water** *like* dark-brown dreams'.

In geography, *water* is discussed in different thematic areas, as the following examples demonstrate:

Water vapour in the atmosphere and the water cycle:

- Uncontaminated **water** *vapour* in the atmosphere has a pH of about 5.6.
- The amount of **water** *vapour* in the atmosphere is expressed in terms of relative humidity.
- These gases combine with **water** *vapour* in the atmosphere and fall as acid rain.
- Nature constantly cleans and renews the earth's **water** *supply* by a process called the water cycle.

Water supply systems today and in the past:

- The following case study tells the story of South Dublin's main **water** *supply*.
- The steep slopes and **water** *supply* in cirques has been used to generate hydro-electricity.
- South Co. Dublin gets much of its **water** *supply* from the River Liffey.
- Most of the earth's **water** *supply* exists in the form of salt water in the oceans or as ice in the polar ice caps.
- In past times, rivers were used for **water** *supply* and transport.

The shortage of clean water in Third World countries and provision of international aid:

- Fifty years ago, less than one tenth of Third World people had access to reliable and *clean* **water** *supplies*.
- The development of *clean* **water** systems for all, vaccinations of the children of the developing world and the education of mothers are the ways by which population growth can be controlled.
- Unlike Calcutta, Hong Kong's public **water** *supply* is *clean* and drinkable.

In relation to the technical term of irrigation:

- Name one river from which **water** is drawn for *irrigation* in the South of France.
- So much **water** is removed for *irrigation* and **water** *supply* that the river has dried up long before it reaches its mouth.
- The Nile provides **water** for *irrigation* in the vicinity of the river.

The reference to water supply systems in the past in geography directly reflects the discussion of water in history which also focuses on its supply and uses in the past:

- Describe how Rome obtained its **water** *supply*.
- The aqueducts: these were used to *supply* **water** to the fountains.
- There was no proper *supply* of *clean* **water** in the medieval cities and there was no way to dispose of rubbish or sewage.
- Thomas Newcomen invented the *steam* engine, which could *pump* **water** out from the mines.
- During the Industrial Revolution *steam* power replaced muscle and **water** *power*.
- Electricity also enabled *running* **water** to be pumped into houses.

In CSPE, water is featured in two thematic areas: it is, on the one hand, discussed as a human right, in relation to its shortage in Third World countries, a motif which can also be found in geography, as mentioned above:

- This has led to shanty towns (known as favelas in South America), which lack even the most basic services like *clean* **water**, sewerage systems, rubbish and electricity.
- Children have the right to enough food and *clean* **water**.
- Millions of people do not have basic needs like *access* to *clean* **water**.

and in relation to green issues, on the other hand:

- Tell your local authority about suspected air or **water** *pollution*.
- Passing laws, e.g. the **Water** *Pollution* Act 1997.
- *Clean* **water** is not in endless *supply*.
- The use of pesticides and fertilisers in *chemicals* into the air or **water**.

Considering that the aim of science education is 'to understand the "world" by looking at it through a technical framework: turning commonsense understandings into technical understandings' (Martin and Veel, 1998: 298), as explained in section 3.2.3.3.1, *water* is examined from technical perspectives in the science corpus, as opposed to the other five subject textbooks corpora. As a cross-curricular word, it appears most frequently in the science corpus with the largest number of collocates compared to the number of collocates it attracts in the other five corpora (i.e. 481 collocates with G^2 scores higher than 15.3) probably because it is one of the key concepts examined in this curriculum subject (and mostly in chemistry combined with *air, oxygen*, and *carbon dioxide*). Together with some of its top collocates, *water* appears in several topics of the three sub-disciplines (chemistry, biology, physics) (cf. *water vapour* in geography):

The respiratory system in the study of human biology:

- The lungs excrete **water** *vapour* and *carbon dioxide*.

- The word equation for respiration is: glucose + oxygen =>*carbon dioxide* + **water** + energy.

The study of plants in biology:

- **Water** *vapour* evaporates from cells in the leaves of plants and exits the leaves by way of tiny pores.

As a substance used in experiments:

- Place the test tube into a *beaker* of cold **water**.
- Describe an experiment to show that *carbon dioxide* and **water** are formed when fossil fuels are burned.
- Making invisible ink and cobalt chloride paper to *test* for **water**.
- Showing the presence of **water** *vapour* and *carbon dioxide* in the *air*.
- To investigate the effects of decreased pressure on the *boiling* point of **water**.

Or to describe results from experimental work where *water* is used in various forms:

- As it rises, the **water** *vapour* cools and condenses to form clouds of tiny droplets.
- The cobalt chloride paper is pink because **water** *vapour* is present in the *air*.
- *Hard* **water** has a much better taste than *soft* **water**.
- To examine the difference between *hard* and *soft* **water**.

Finally, in mathematics, *water* is mentioned in mathematical problems that are concerned with measuring quantities (e.g. *litres*) in various containers (e.g. *tank*) and calculating dimensions (e.g. *volume*), as shown below:

- If 2.2. *litres* of the **water** are removed, calculate (i) the drop in the *level* of the **water** and (ii) the percentage of **water** remaining in the *cylinder*.
- If the radius of the cylinder is 8 cm, by how much will the *level* of the **water** drop if the *spheres* are removed from the cylinder?
- *Calculate* the *height* of **water** in the *tank*.
- *Calculate* the percentage of **water** in the *container*.
- What is the total *volume* of **water** put into the fish *tank*?
- Calculate (to the nearest *cm*) the rise in **water** in the *tank*.

It is further noteworthy that, based on a frequency analysis of all content words in a single subject textbooks corpus, that was created by merging the six textbooks corpora in one, *water* appears as the second most frequent content word (with a frequency of 2,1‰, after *people* with a proportional use of 2,4‰; this corpus is not included in the thesis appendix).

As regards the word *area,* its scarce use in English (0.12‰) with only one significant collocate (*school*) indicates its minor importance in the thematic content of this corpus (e.g.,

'Other *school* libraries in the **area** were visited to inspect their facilities'). By contrast, *area* is a fundamental concept in the geography corpus since this subject is by definition oriented to the discussion of places and locations. This is confirmed by numbers: with a frequency of 2.82‰, *area* is the cross-curricular word that is the second most frequent in the corpus in question (followed by its highest frequency in mathematics 3.6‰), and it is the sixth most frequent and widely distributed (100% range) content word in the same corpus. It consequently attracts a large number of collocates as well (131 in total) with high significance (G^2) values (see Appendix E2). Concordances show that *area*, together with several of its collocates, typically refers to actual geographical areas that are represented in visual aids:

- *Calculate* the **area** of the *map* extract *shown.*
- Draw a *sketchmap* of the **area** *shown* on the *photograph.*
- Ireland has a small population but *Dublin* and its *surrounding* **area**.
- 62% of the nation's toxic waste is produced in the Cork *Harbour* **area**.
- On it [sketch map] mark and name the following: one **area** of evergreen *woodland*, one **area** of deciduous *woodland* and one **area** of mixed *woodland.*

Another use of the word refers to the area in the air that is created by atmospheric pressure:

- Warm air is lighter and rises, creating an **area** of *low pressure.*
- Air moves from a *high-pressure* **area** to a *low-pressure* **area**.
- A depression is an **area** with a low atmospheric *pressure* at the centre.

As in the English textbooks corpus, the word *area* is not very frequent in history (0.27‰), appearing with only three significant collocates *pale, called, small.* Concordances reveal that the collocate *pale,* which immediately strikes one as odd is, in fact, a proper name that denotes an actual area:

- The *Pale* was a*small* **area** around Dublin, stretching from Dundalk in the north to the Wicklow mountains in the south.

When used with *called*, the word *area* in history does not refer to actual geographical locations but it is rather used as a generic term to define the terms used to name a space within a building or a space within a region:

- This led to a rectangular **area** *called* the atrium.
- The **area** around the fireplace, *called* the hearth, had stone flags.
- The bishops ruled an **area** *called* a diocese.

In CSPE the word *area* is used to denote location in general. Combined with its various collocates, it is used to refer to people (and particularly the young) living in an area:

- Find out if any *young people* in your **area** have got involved in any campaign.
- There is a need for more *youth facilities* in this **area**.
- At weekends most TDs meet *people* of the **area** they have been elected to represent.
- Constituencies are smaller than counties and are based on the number of *people* living in the **area**.
- We decided to help older *people* in our **area** by organising a Christmas party for them.
- Let each group identify some *local* problems in your **area**.
- Everyone has the right to a say in the future planning and *development* of their **area**.
- Design a poster that would encourage people to attend a meeting to set up a *Neighbourhood* Watch scheme in your **area**.

The frequent occurrence of *area* in the mathematics textbooks corpus (3.6‰), on the other hand, acquires a different technical sense and it is always used with reference to specific geometrical shapes and dimensions:

- The **area** of a *rectangle* is equal to the frequency it represents.
- *Calculate* the **area** of the *triangle* in the diagram.
- *Find* the **area** of the shaded region.
- *Find* the **area** of the *parallelogram*.
- Find the *volume* and the surface **area** of each of the following solids.
- *Calculate* the circumference and **area** of a *circle*.
- *Calculate* the *perimeter* and the **area** of the figure shown.
- Find the *length* and *width* corresponding to the *maximum* **area**.
- **Area** of *triangle*: ~ (*base*) x perpendicular height.

Finally, in science, *area* is also used in relation to shapes and objects (as in mathematics):

- The **area** of an *object* is the amount of surface it has.
- Every *square* has an **area** of 1 mm.
- *Measuring* the **area** of regularly/irregularly-*shaped* objects.

as well as in the discussion of atmospheric pressure (as in geography):

- Pressure is the *forceper unit* **area**.
- The atmospheric *pressure* over an **area** has a huge effect on the weather that area will have.
- *Low pressure* over an **area** gives rain, strong winds and generally unsettled, poor weather.
- *High pressure* over an **area** gives sunny days, light winds and clear, blue skies.
- The smaller the **area**, the greater the *pressure*.

Comparing and contrasting the word *time* across subject corpora, no important differences can be discerned in its meanings and uses. It is most frequent in the English textbooks corpus (1.36‰) and the history textbooks corpus (1.19‰). In the humanities and social science subjects of English, geography, history and CSPE, *time* is used to express duration, to mark particular eras, or as a component of the adjective *part/full-time* and the phrase *at the same time*. Some examples of the word's use with the top collocates in individual textbooks corpora include the following:

English:

- Point out three details that tell you this *story* is *set* a *long* **time** ago.
- It covers an unusually long period of **time** for a short *story*.
- In bed that night, I lay awake a *long* **time**, thinking about it all.
- In this first extract, we see Billy arriving at his *part*-**time** job.
- Work out the **time** you will *spend* on each section before the exam.
- *Spend* some **time** planning you points carefully, making sure that you address the question directly.
- They *spend* their **time** 'waiting for Godot' who, of course, never comes.
- You are often asked to write about '*setting*' (**time** and place) in the examination.
- Some poems are *set* at a particular **time** of day or night.
- Before *Shakespeare's* **time**, drama took the form of short, religious plays which were presented in the church to teach people the stories of the Bible.
- In *Shakespeare's* **time** Elizabethan audiences loved puns.
- Name a novel or a short story you have studied where the writer describes one of the following: a sad time, a *happy* **time**, a worrying time.

Geography:

- The British iron and steel industry provides a good example of *change* over **time** in industrial location.
- Explain how social and historical events resulted in population *changes* over **time** in the West of Ireland.
- Many Georgian buildings have changed in *function* over **time**.
- Soil takes a *long* **time** to form.
- The world's *population* has increased unevenly over **time**.
- People who study population agree that the world's population will continue to *grow* for some **time**.

History:

- At the *same* **time**, a civil war had started in England between the Royalists and the Parliamentarians.
- The knights *spent* much of their **time** hunting.

- The city of Paris formed its own *militia* (*part*-**time** voluntary *soldiers*) called the National Guard.
- Paper, made from rags, was also *invented* at this **time**.
- *Historians* usually measure **time** in years and in centuries.
- Primary sources are those which come directly from the **time** that is being *studied*.

CSPE:

- Choose a *date*, **time** and venue for the event.
- Councillors *spend* a great deal of **time** away from their homes.
- When *election* **time** comes around, candidates send out leaflets to encourage people to vote for them.

In the textbooks corpora of the two more technical subjects of mathematics and science, *time* is used in its core sense but within the context of mathematical problems and scientific exercises, in conjunction with other units of measurement. The word therefore appears with several same collocates in the two textbooks corpora (e.g. *distance, speed, calculate*), as can be seen below:

Mathematics:

- Express *distance* in terms of **time** and *averagespeed*.
- Calculate, in hours and *minutes*, the **time** *taken* for a man to run a race of 51 km in length at an average speed of 12km/hr.
- The *total* **time** for both *journeys* was 7 hours.
- The x-*axis* represents **time**.
- *Calculate* the **time** *taken* for the first *journey*.
- Use the graph to find the **time** when the *temperature* was lowest.

Science:

- The velocity of an object is the *distance* it travels per unit **time**, in a given direction.
- The acceleration of a body is the change in its *velocity* in unit **time**.
- Plot a *graph* of temperature *against* **time**, *putting* **time** on the *horizontalaxis*.
- Measure the **time** difference in *seconds*.
- Speed is the rate of *change* of *distance* with **time**.

Having examined the thematic manifestation of the nouns *water, area* and *time*, it would also be interesting to explore the uses of the basic English verbs *find, use* and *make*, in the six subject textbooks corpora. Information can obviously be derived from the nouns that accompany these transitive activity verbs but also from the agents who (are asked to) perform the actions. Answers to 'who' *finds/uses/makes* 'what' and 'for what purpose' in the six textbooks corpora provide revealing information about the type of knowledge and skills that are considered important in the six areas of curriculum study. For example, on account of the preoccupation of English textbooks with communication skills and language, the verb *find* and

its surrounding environment are tied to providing and justifying answers and expressing specific opinions:

- **Find** *evidence* in the *extract* to support your answer.
- What *evidence* can you **find** in the above scene to support this view?
- **Find** *answers* to the following questions.
- Choose one character that you **find** *interesting* or striking for some reason and say what kind of person you think she will be in the story.
- What do you **find** *interesting* about the person who is speaking in this poem?
- Most *people* seem to **find** personal writing easiest.

very often with a focus on language use and text types and materials studied:

- Read the story silently first and **find** *words* that have the following meanings.
- **Find** the key *words* in the question and underline or highlight them, as in this sample question.
- As you read the *story*, **find** *words* which have the following meanings.
- Skim: read through the *passage* quickly to **find** out what it is about.
- **Find** illustrations, photographs and *book* covers.
- At the beginning of a *book* you will **find** the publisher's name and the date of first publication.

In the geography textbooks corpus, *find* seems to be a useful device for directly engaging students in the study of geographical areas because, in all its occurrences, it invites students to identify areas illustrated in visual aids:

- **Find** the *area* of the OS map extract.
- **Find** Eritrea on the *map* in figure 1.
- Six-*grid references* are used to **find** the exact location of a specific feature or place.

The verb *find* is equally important in the discourse of history as it reflects the exploratory nature of historical study, and it is central to the major thematic unit of the first year of Junior Cycle history syllabus, i.e. 'How we find about the past'. In the history textbooks corpus, the verb refers to what different people try to find out about the past or to what people tried to find in past eras:

- The travels [of Marco Polo] encouraged *explorers* to **find** *new routes* to the great wealth of the East.
- Some *explorers* wanted to **find** a *north*-west passage.

- Using sources *historians* search for evidence to **find** out about our *past*.
- Suggest how artefacts help *archaeologists* **find** out about the *past*.
- *Archaeologists* try to **find** out about prehistoric people by examining materials excavated from the ground.

In the CSPE textbooks corpus *find* is used to encourage students to identify specific types of information in various sources, reflecting the research and survey-based tasks and activities that are characteristic of the CSPE course:

- **Find** out the *names* of the four people who represent Ireland in the Council's Parliamentary Assembly.
- **Find** the *names* of five Senators and say whether they belong to political parties.
- **Find** out the *names* of your local representatives.
- Research: **Find** out more *information* about the United Nations.
- **Find** *information* on two projects carried out by the Leader Programme in any part of the country.
- **Find** out more *information* that you could include in the poster/leaflet.
- In *Chapter* 5, you will **find** advice and helpful hints on how to make posters and leaflets.
- Fill in the blanks and then **find** the *words* in the Word Search.
- **Find** and *summarise* newspaper and television reports of political parties during the next week.
- Carry out a *survey* in your area to **find** out about attitudes towards unemployment or poverty.

Find is the commonest content word (with a normed frequency of 6.85‰) in the mathematics textbooks corpus, by far exceeding its frequency of occurrence in all other textbooks corpora (0.58‰ in English, 0.33‰ in geography, 0.31‰ in history, 1.23‰ in CSPE, 1.15‰ in science). It is recurrently used in mathematical exercises, asking students to find a number of different mathematical concepts:

- **Find** the *value* of x when the diver's speed is greatest.
- **Find** the *equation* of line L.
- Solve the *equation* to **find** the unknown *number*.
- **Find** the *image* of the point (3, 5) under the axial symmetry in the line y=2.16.
- To **find** the *slope* of a line when given its equation, do the following ...
- **Find** the *point* of intersection of the following pairs of lines using two methods.
- Use your *graph* to **find** the maximum *point* and the maximum value of f(x).

In the same vein, in science textbooks, students are instructed to find a number of quantities, dimensions and measurement units associated with basic scientific concepts:

- To **find** the *density* of a liquid/an irregular shaped object/a solid which floats, etc.

- **Find** the *mass* of the liquid/stone/dry clean beaker/water/flask and attachments, etc.
- **Find** the *volume* of a small stone/regular shaped object/cork/the substance, etc.
- To **find** the centre of *gravity* of an irregular shaped piece of cardboard.
- To **find** the force of *gravity* of different objects.
- Use the *graph* to **find** the acceleration by taking any two points/the athlete's velocity/the time taken to travel 27 m.

With regard to the verb *use*, straightforward conclusions about its subject-specific use in individual corpora cannot be easily drawn due to its semantic versatility. However, an attempt is made here to identify some patterns of similar usage across particular corpora. The verb, combined with its top collocates, is almost exclusively used in discussions of language use in the English textbooks corpus, which reflects the fact that language constitutes the fundamental tool for the study of English. The following are some examples that illustrate this point:

Definition of language use:

- **Use** of *language* means word choices, emotive words and phrases to affect the audience, description, persuasion phrases, imagery, alliteration etc.

Multiple references to the use of appropriate language in different text types for different purposes and relevant advice directed at students:

- When you write about poetry you must closely examine the *language* and **use** various poetic terms.
- The [ballads] **use** simple *language* and tell the tale through a mixture of description, action and dialogue.
- In looking closely at travel writing, facts, opinions and the writer's **use** of *language* are important when examining the text.
- When writing to advise, always **use** *language* in an appropriate way to communicate ideas to your audience.
- If you are writing about school rules for a class of very young children, the *language* you will **use** will be clear and simple.
- As with all types of writing, the examiner will be looking at the way you **use** *language*, e.g. spelling, punctuation, sentences and paragraphs.
- Always **use** *quotations* from the passage to support your answers.
- Remember to **use** brief *quotations* as evidence for your points and work these into your sentences.
- Underline *quotes* which you can **use** to explain your answer.
- **Use** *quotes* to *support* your opinions.
- Don't **use** too much *dialogue* in the story.
- **Use** *words* and *dialogue* from the poem.
- Underline words and *phrases* that you can **use** in your answers.
- What *words* or *phrases* would you **use** to describe the rhythm you are making?
- Address all questions directly and **use** relevant *reference* to support your point.

- **Use** close *reference* to the text in your answer.
- When writing about characters **use** *evidence* from the text to back up your statement.
- To make a good case, think about the issue carefully and **use** *evidence* and examples to support what you are saying.

Inviting students to comment on authors' use of language:

- How does the poet **use** *language* to create this setting?
- What kind of *language* does the writer **use**?
- Discuss important features of an author's style and **use** of *language*.
- Comment on the **use** of violent *language* in the poem.

Expressing specific critical comments on the use of language and literary devices by poets, writers and others:

- *Poets* frequently **use** figurative (metaphorical) *language*, such as metaphors, similes, personification and symbols, which makes their writing more vivid and interesting.
- Both *poets* **use** *language* that is elegant, respectful and delicate.
- Why do *poets* **use** sound effects?
- Many *poets* **use** comparisons to help us more clearly understand what is being said.
- Some *poets* **use** straightforward vocabulary and can be taken literally.
- *Poets* **use** rhyme to emphasise and draw attention to certain words.
- Sometimes *poets* **use** rhythm to imitate the motion of something.
- *Poets* **use** *language* in a special way.
- *Writers* also **use** sound effects in their writing to draw attention to a particular word or phrase.
- *Writers* who **use** a friendly, conversational style usually seem sincere.
- *Writers* **use** cliffhangers to hold their readers.
- *Writers* often **use** similes, metaphors and personification to create vivid images.
- Dickens' **use** of colour *imagery*, especially red and black, emphasizes the unnatural atmosphere of the dreary industrial town.
- *Writers* **use** *dialogue* to help us to know characters in stories.
- They [tabloid newspapers] tend to **use** emotive *language* and sensationalise stories.
- Have you noticed how advertisers **use** *words* to persuade us to buy things?
- A very fine answer, focusing clearly on the **use** of violent *language* in creating an unsettling atmosphere.

It is important to note that, beyond the usage of *use* and its collocates, the above examples illustrate much of the terminology employed in the written academic discourse of Junior Cycle English. Concordance lines of *use* and its collocates bring to the surface a wide range of literary terms that accompany a variety of genres and which are required of students for appropriate literary expression in the English classroom.

Moving on to the geography textbooks corpus, most occurrences of *use* and its top collocates emphasize:

a) the need to provide evidence on the basis of a source to justify one's answer:
   - **Use** *evidence* from the *map* to support/justify/explain your answer.
   - **Use** *evidence* from the *photograph* to support your decision.
   - **Use** the following *headings* in your description/to explain your answer/to develop three points fully.
   - **Use** *grid references* to justify your choice
   - **Use** *map evidence* to *support* your answer.

b) the use of instruments to perform tasks and activities:
   - **Use** your *atlas* to *identify* the towns of Ireland's most important roads.
   - **Use** a sharp *pencil* to mark the position and height of the end points.

Another type of occurrence of *use* in the same corpus is in the compound noun *land-use*:

- This photograph shows that *land*-**use** in the town centre is more intensive than in the outskirts.
- Both land *values* and *intensity* of *land*-**use** tend to increase towards a city centre.

As regards the history textbooks corpus, it is difficult to pin down similar thematic instances of *use* and detect patterns, as the variable examples indicate below:

- By the late nineteenth century, **use** of the Irish *language* was in decline in Ireland.
- The planters would **use** English law, *language* and customs.
- *Vernacular* writing humanists encouraged scholars to **use** Greek and Latin but ordinary people continued to use their vernacular (everyday) languages and many writers used them to produce great works of literature.
- The knights kept some of the *land* for their own **use** (their demesne) and rented the rest to peasants.
- The **use** of *violence* against socialists.
- Do you think the Conservative party would support unionist **use** of *violence* to resist Home Rule?
- Hitler censored the media but made clever **use** of *propaganda* to build up popular support.
- Fascists and Nazis made clever **use** of *propaganda* to portray themselves as mighty leaders of great nations.

Perhaps one identifiable thematic context in which the verb *use* recurrently appears concerns the means and tools historians and archaeologists use for the study of the past:

- *Archaeologists* **use** a technique called dendrochronology to discover the age of wooden objects.
- *Archaeologists* can therefore **use** pollen to build a picture of the way people used the land at various times in the past.
- The job of the historian is to **use** the *sources* to reconstruct the past.

- Good historians are very careful in their **use** of *sources* and they like to use several different sources when studying the past.
- We can also **use** pictorial *sources* to compare the past with the present.

In CSPE, on the other hand, it can be easily observed from concordance lines that the verb *use* appears, once again, mainly in the discussion of green issues:

- **Use** renewable *energy* sources *like* wind, solar and wave power to make electricity
- Could you cut down on your **use** of *plastic bags* and reuse them when you shop?
- They started to **use** environmentally-friendly *products* and they closed down the old town dump.

In the mathematics and science textbooks corpora, the verb *use* is used to instruct students to use something to perform an action, and that is why several collocates of this cross-curricular word are verbs, in particular high-frequency verbs of the mathematics and science corpora.

Mathematics:

- **Use** your graph to *estimate* the median mark.
- **Use** Pythagoras's theorem to *find* the length of the sides indicated by a letter in each of the following diagrams.
- We will **use** two methods to *solve* this quadratic equation.
- **Use** your *calculator* to *find* the angle.
- **Use** your *protractor* to draw an angle of 70°.
- **Use** your *graph* to find the minimum point and the minimum value.
- **Use** this *information* to *form* an equation.
- Write down an equation in x to *represent* the information and **use** it to calculate x.

Science:

- **Use** litmus to *test* whether the solution is acidic or basic.
- **Use** iodine solution to *test* B for starch.
- **Use** the graph to *find* the weight of an object.
- *Name* a chemical you would **use** to test for the presence of starch in a leaf.
- Solar panels on buildings **use** the sun's *energy* to heat water.
- In this experiment, we will **use** a *battery* and some conducting wire to show the heating effect of a current.
- **Use** the *tongs* to hold the piece of magnesium ribbon in the Bunsen burner flame.
- When most of the water has been evaporated, **use** a *tongs* to place the evaporating basin over a beaker of water, as shown in the diagram.
- The *plants* **use** the sun's energy to make food in their leaves through photosynthesis.
- Draw a diagram of the *apparatus* you would **use** to copper plate an iron key.

In addition, in science, the verb also refers to the use of instruments and forms of energy:

- Always **use** a *test* tube holder.
- Name the pieces of *apparatus* you would **use** in the following activities.
- Very often plants develop at different times to *make* **use** of light while it is available.
- The power rating of any appliance is a measure of how much *electricity* it will **use**.
- The ESB meter in your home measures the number of units of *electricity* you **use**.

With reference to the last cross-curricular verb examined here, *make*, together with its distinct collocates in the six corpora, refers i) to the different things that students, or other agents, create or produce in the six subjects ('someone makes something'), or ii) to the use of an instrument, material etc. in order to construct or put together something (often through the phrase *make use of*), iii) to cause change/ to develop into, and iv) as part of the idiom *make sure*. Concordances of *make* and its distinct collocates in all the above instances reveal the distinct preoccupations of the six curriculum subject areas. Thus, looking at the first core meaning of the verb in the six corpora, information is obtained on what students, or other agents, produce in individual subjects:

English:

- **Make** a *list* of the adjectives in the following passage.
- **Make** a *list* of words that you have difficulty with and learn them.
- **Make** a *list* of three good qualities in this student's story.
- Write out the *speech* you would **make** for or against this motion.
- Write out the text of a *speech* you would **make** to inform your school year group about a deserving charity that you support.
- Support the *points* you **make** by detailed and relevant reference.
- **Make** *clear* your *point* of view.
- Short sentences can really **make** a *point*.
- If taking notes from a book, skim the relevant section first in order to **make** *sense* of the whole piece.
- The following jumble of words will **make** *sense* if you organise it into a grammatical sentence.
- Address questions directly and develop the main *points* you **make** by using relevant reference.

History:

- The British Parliament could **make** *laws* for Ireland.
- When did the Celts learn to **make** iron *weapons*?

CSPE:

- **Make** a *list* of the basic human rights that people who are bullied are being deprived of.
- **Make** a suitable *list* of questions and choose a selection of people to survey.
- **Make** a *list* of the different ways you use water in a day.
- One of the government's jobs is to **make** *laws*.
- The Parliament will have more power to **make** *laws* with the Council of Ministers.
- You could **make** out some *poster*/information *leaflets* or make announcements at Assembly or over the school public address system.
- Name three *changes* a person can **make** to their lifestyle that will help the environment.

- Members of the local authority meet each month to **make** *decisions* about their local area and how money is to be spent.
- By making small changes in our lifestyle, we can **make** a big *difference* in the environment.

Science:

- **Make** a *list* of the changes man has made in nature.
- **Make** a *list* of the different animals that you observe in the habitat.
- **Make** a *list* of the amount of food proteins, fats etc. per 100 g. of the food.
- Producers are *plants* that **make** food by photosynthesis.
- **Make** a *collection* of well-known wild flowers/the fungi to be found in the woodland/insects/seashells etc.
- To **make** a *simple* map of the habitat/electromagnet/transformer/periscope using two plain mirrors.
- **Make** a *sketch* of the onion cells/a spider's web/the plants you have identified.
- **Make** a bar *chart* of your results.
- **Make** an energy *chart*.

In a similar vein, the following examples demonstrate the second pattern of *make use of* something for a purpose or make use of something to produce something else:

English:

- If you wish, you can **make** *use* of some of the phrases you wrote down or you can write entirely from your own imagination.
- **Make** *use* of the following words or phrases when you write about an advertisement.

Geography:

- Materials can be *used* to **make** lots of different *things*.
- Limestone is *used* to **make** monuments and headstones.
- The rock is crushed and *used* to **make** concrete blocks and for road chippings.
- Sheffield steel is *used* to **make** specialised, high quality expensive products such as surgical instruments and cutlery.
- Light industries *use* light resource materials to **make** relatively small *products*.

History:

- Wheat and *barley* were *used* to **make** *bread* and porridge.
- The press, radio, television and posters can be *used* to **make** propaganda.
- What *use* did the Celts **make** of the new *weapons*?
- The bronze was poured into *moulds* to **make** *tools* and *weapons.*

Science:

- Fuses, circuit breakers and plugs are all *used* to **make** electricity safer for use in the home.
- Nylon can be formed into fibre *used* to **make** stockings, carpets and ropes.
- Chloroplasts contain chlorophyll, which plants *used* to **make** food cell wall.
- Using *bacteria* to **make** cheese.
- Thermometers usually **make** *use* of the expansion of liquids.

The third sense of *make*, that is, to transform or cause a change (realised by a complement or a verb) is displayed through the examples below:

English:

- Longer sentences **make** the *story* flow more smoothly.
- The setting can **make** the *story* seem sadder, more scary etc.
- They [poems] will **make** you *think* about things and some of them might even affect your feelings.
- An opening must **make** the *reader* curious.
- A sequence of events must be set in motion which will catch our interest and **make** us *want* to read on.
- The last three lines are intended to avoid panic and **make** listeners *feel* reassured and calm.

Geography:

- How can earthworm *help* to **make** *soil* fertile?

History:

- The Fianna Fail government wanted to **make** *Ireland* self-sufficient in farm products.

CSPE:

- Urban planners try to **make** an *area* more pleasant and convenient for residents to live in.

Mathematics:

- Solving an inequality means findings the *values* of x that **make** the inequality *true*.
- **Make** *coefficients* of y the same.

Science:

- A force can **make** *things* move or stop when moving.

To conclude, the idiom *make sure* recurs consistently in the six corpora, drawing students' attention to important instructions and guidelines about completing subject-specific activities and exercises in the subject classroom and in examinations:

English:

- **Make** *sure* you understand what you have to read before you begin to summarise.
- **Make** *sure* you know how exam papers are organised.
- **Make** *sure* you stick to this schedule during the examination.
- Start your answer with a plan and **make** *sure* that this plan is legible.
- **Make** *sure* to get the layout right.
- Always **make** *sure* that what you write is relevant.
- Get straight into your answer with only a brief introduction, remembering to *use* Standard English, that is, **make** *sure* your writing is formal.

Geography:

- **Make** *sure* you have at least three pieces of relevant detail in each point.
- **Make** *sure* you to include a frame, title or heading and direction arrow on your sketch map.
- Always give at least three reasons just to **make** *sure* you have enough facts written down.

CPSE:

- **Make** *sure* to have prepared a *list* of the important questions that you want answered.
- **Make** *sure* you go back to your diary or log to remind yourself of all the activities and information you discovered and were involved in.
- **Make** *sure* you use these questions as part of your revision for the exam, as the same style of questioning is used.
- **Make** *sure* that the *information* is easy to show on a graph.

Mathematics:

- **Make** *sure* both *sides* are measured in the same units.
- **Make** *sure* all the bars are the same width.
- **Make** *sure* you know which angle you are using and which sides are the opposite and adjacent.
- **Make** *sure* both *sides* are measured in the same unit.

Science:

- **Make** *sure* the moss peat does not dry out.
- **Make** *sure* there is no air bubble trapped underneath the plasticine.
- **Make** *sure* the thermometer is held steadily and does not touch the test tube.

Exceptionally, in the history textbooks corpus the idiom appears in contexts like the following:

History:

- The Supreme Court's main task is to **make** *sure* that the *laws* passed by Congress and the actions of the President are constitutional.
- It [the guild] set examinations to **make** *sure* that its craftsmen were skilled at their trade.

Overall, the discussion of cross-curricular content words of the six textbooks corpora and their uses in subject-specific contexts shows that, with a few exceptions, the same lexical words attract different collocates and acquire different meanings, depending on the thematic needs of the context. Concordance lines show that it is not the actual cross-curricular words that have specialised meanings but the angle from which they are discussed becomes more technical especially in mathematical and scientific discourse (see, for example, the subject-specific occurrences of *water*). This emphasises the importance of context and underlines that meaning is context-based. From a pedagogical perspective, the above observations imply the need to expose ESL students to the range of senses and uses of cross-curricular content words

(through a multi-disciplinary approach) in multiple contexts in order to maximise the depth of their vocabulary knowledge rather than focusing only on increasing vocabulary breadth.

3.3.1.2 Cross-curricular words in examinations corpora, their central collocates and contextual use

The detailed consistency analysis of the content words (frequency ≥ 40 per million words) of examinations corpora shows that 34 words recur consistently in the six examinations corpora. The ten most frequent items which occur across all 60 text-files are displayed in Table 3.25 below. The words which appear in a smaller number of text-files that comprise the six corpora can be viewed in Appendix E4 (specifically, one word appears in 59 texts, three words in 58 texts and one word in 57 texts, and three or two words appear in 44-56 text-files).

Table 3.25 The top ten cross-curricular content words in the six examinations corpora.

N	Word	Texts	Frequency ‰					
			English	Geography	History	CSPE	Maths	Science
1	ANSWER	60	7.33	5.92	8.15	7.28	4.07	4.07
2	WRITE	60	2.84	0.81	3.91	3.62	4.35	4.35
3	QUESTIONS	60	2.44	2.99	3.16	5.32	1.39	1.39
4	USE	60	0.8	1.84	0.83	1.49	2.56	2.56
5	FOLLOWING	60	2.39	2.72	5.33	1.28	1.84	1.84
6	PAPER	60	2.2	1.69	2.69	1.23	2.7	2.7
7	EXAMINATION	60	0.8	1.47	2.03	2.21	1.24	1.24
8	CERTIFICATE	60	0.51	0.93	0.93	0.47	1.27	1.27
9	JUNE	60	0.51	1.1	0.59	0.6	2.27	0.45
10	JUNIOR	60	0.5	0.93	0.91	0.47	1.24	0.43

Perhaps the most interesting cross-curricular words in Table 3.25 above are the two communication verbs *answer, write,* the activity verb *use* and the noun *questions*, all of which are clearly fundamental terms in the formal assessment discourse of Junior Certificate examinations. It is therefore worth investigating the collocates and typical patterns of use of these items in the examinations corpora of the six subjects to identify any existing differences and/or similarities (see Table 3.26).

Table 3.26 The top 15 collocates of four cross-curricular content words in the six examinations corpora ranked by descending order of significance based on G^2 scores (≥ 15.3).

Examinations corpus	Cross-curricular word and collocates
	answer
English	questions, question, reference, give, reasons, follow, support, following, sections, read, text, functional, extract, poem, explain
Geography	questions, book, circle, follow, marks, return, correct, following, folder, geographical, mix, section, statements, page, otherwise
History	book, paper, following, marks, accompany, separate, history, include, return, examination, help, pages, turn, support, centre
CSPE	questions, section, studied, book, paper, extra, marks, question, need, please, ask, numbered, approach, expected, follow
Mathematics	give, form, reason, correct, giving, simplest, nearest, decimal, express, write, places, part, whole, place, metre
Science	questions, following, book, spaces, marks, reason, parts, section, question, sections, follow, give, study, core, enclose
	write
English	composition, personal, writing, letter, speech, section, conversation, prose, stated, report, free, rest, wish, form, inspired
Geography	examination, number, coimisiun
History	account, person, title, selected, need, fact, read, name, changes
CSPE	spaces, answers, short, sure, make, number, speech, examination, marks, name, article, letter, provided, questions
Mathematics	expression, simplest, answer, equation, using, value, cos [cosine], form, length, fraction, calculator, single, equations, sin [sine], mode
Science	letter, table, examination, number, opposite, equation, right
	questions
English	answer, follow, following, carefully, piece, poem, answers, passage, extract, paper
Geography	answer, marks, answered, follow, choice, section, equal, following, name
History	following, picture, short, stamp, document, marks, number, examination, level, history
CSPE	answer, numbered, studied, section, marks, poster, included, ask, follow, answering, paper, information, write, page
Mathematics	attempt, time, hours, marks
Science	answer, marks, spaces, section, provided, follow, equal, carry, following, textiles, plastics, diagram, using, table
	use
English	answering, given, correct, capitals, poem, spelling, punctuation, grammar, extract
Geography	land, terms, explain, photograph, answer, diagram, people, map, correct
History	wish, help, hints
CSPE	poster, name, people
Mathematics	stamp, mark, graph, used, examiner, centre, question, superintendent, calculator, estimate, draw, wish, theorem, Pythagoras, drawn
Science	give, question, number, page, name, mark, examination, headings, graph, biology, everyday, estimate, section, piece, equipment

Unlike their counterparts in textbooks corpora, the cross-curricular words of examinations corpora share many common lexical collocates, e.g. *answer(s), question(s), follow, following, marks, section(s), give, reason(s), number, page, examinations*, and behave in similar ways

within context. Little variation can be noticed in the interpretation and uses of *answer* as it is evidently used in all examination papers mainly in two ways: i) as a verb to ask candidates to provide answers (with supporting evidence, 'reasons' in English, mathematics and science) or to select the correct answer from given options and ii) as a noun in the compound *answer book*:

English:

- **Answer** the *questions* which/that *follow*.
- *Explain* your **answer** with *reference* to the *text*/extract/story.
- *Support* your **answer** by *reference* to the text.
- *Give reasons* for your **answer**.
- *Read* this *poem* and then **answer** the *questions*.
- You must also **answer** *sections* 1, 2, 3.

Geography:

- **Answer** the *questions* which/that *follow*.
- Use the pie charts to **answer** the *questions* below.
- You must return this paper with your **answer** *book*.
- *Circle* the *correct* **answer** in each of the *statements* below.
- Please remember to *return* your *folder* with your **answer** *book*.

History:

- Answer the following questions in a separate **answer** *book*.
- Study the two documents/pictures which *accompany* this *paper* and then **answer** the following questions.
- Do not *include* these pages with your **answer** *book*.
- You must *return* this paper with your **answer** *book*.
- If you wish, you may use the *hints* to help you in your **answer**.

CSPE:

- **Answer** all of the *questions* in this section.
- **Answer** any three *questions* in *section* 2.
- When you have *studied* [this information brochure/leaflet/newspaper article/postcard/extract etc.] **answer** the *questions* below.
- If you *need extra* paper to **answer** this *question*, *please* ask the Examination Superintendent for it.
- When you are answering the questions on this paper, you are *expected* to **answer** from the human rights *approach* of the CSPE course.

Mathematics:

- *Give* your **answer** *correct* to two *decimal places*.
- *Give* your **answer** in its *simplest form*.
- *Give* your **answer** *correct* to the *nearest decimal*.
- *Give* a *reason* for your **answer**.

Science:

- *Study* the diagram and **answer** the *questions* that *follow*.

- You must return this paper with your **answer** *book* otherwise marks will be lost.
- **Answer** any three *sections* from sections B, C, D, E.
- *Give* a *reason* for your **answer**.
- Return this *section* of the examination paper. *Enclose* it in the **answer** *book* you use in *answering* the other sections.

The use of the key word *write* and its collocates in the English and CSPE examinations corpora reveals the text types students have to produce for assessment purposes:

English:

- *Section* 2: *Personal writing*: **Write** a *composition* on one of the following topics.
- *Section*2: *Personal writing*: **Write** a *prose composition* on any one of the following titles.
- **Write** a *letter* of complaint to one of these shops.
- **Write** a *letter* to the editor of a newspaper stating your point of view.
- **Write** a *speech* for or against the motion: 'All teenagers should have to participate in sport'.
- **Write** out a *conversation* between the two animals in Photo 1 of Paper X.
- You have been asked to **write** a *report*/review about your favourite TV programme.
- Look at the picture on Page 1 of Paper X which accompanies this paper and **write** a *composition inspired* by it.

CSPE:

- **Write** a *short speech* for a school assembly explaining why it is important to celebrate International Women's day.
- **Write** out the welcome *speech* you would make to introduce the Garda to your class.
- **Write** an *article* for your school magazine explaining why Earth Day is important.
- **Write** a *letter* to the Courts Service Information Office asking for a guided tour of the Four Courts for your CSPE class.
- In the space *provided*, **write** down the *name* of the politician which matches each picture.
- **Write** three *questions* that you would ask in the survey.

In geography examinations, the verb *write* does not appear in the formulation of exam questions but only in the following instance: '**Write** your *examination number* here' (which appears also in other exams corpora, e.g. in science: '**Write** your *examination number* in the box provided on this page' and in CSPE: 'Make sure to **write** your *examination number* in the box above'). This may be because students are not asked to produce a piece of writing in geography examination papers but they are instead required to *tick, answer, name, explain, describe, examine, look, use, study, state, identify, match, show, draw,* etc., as can be seen in the frequency list of geography content words (Appendix B5).

In the history examinations corpus, *write* and its collocates express i) the text type students are typically asked to construct, i.e. account, ii) different historical figures and iii) past events they have to discuss in writing:

- **Write** an *account* of the growth of the European Union between 1973 and 1992.
- **Write** an *account* of the life of a Serf on a medieval manor.
- **Write** about that *person*.
- **Write** the *title selected* at the top of your account.
- **Write** down one *fact* about the War of Independence in Ireland, 1919-1921/about the 1916 Easter Rising etc.
- **Write** down the *name* of an important Protestant reformer whom you have studied/one political figure in Ireland after 1945 etc.
- **Write** about *changes* in two of the following areas since the 1940s: a) Sport and leisure, b) Transport, c) Women's lives.

Although in mathematics students do not have to write (short or long) texts, as in geography, the verb *write* is frequently used to ask them to write numbers or symbols that express mathematical values (e.g. *value, length* etc.) or concepts (e.g. *equation*):

- **Write** down an *expression* in x to represent the cost of 1 gram of the powder.
- **Write** down an *expression* in x for the width of the garden.
- **Write** your *answer* in its *simplest* form.
- **Write** in its *simplest form* 3x2 . 2x+6 . x(2x . 3).
- **Write** an *equation* in x to represent this information.
- **Write** down the *value* of *cos*. A as a fraction.
- **Write** down the *length* of the side adjacent to the angle A.
- Without *using* a *calculator* **write** + as a *single fraction*.
- **Write** down the *mode* of the following numbers.

Students are also asked to *write* symbols and numbers (*letter, equation*) in science examination papers:

- **Write** the *letter* Z *opposite* the function of the fuse in a plug.
- **Write** the *letter* F beside the *name* of the process by which plants make food.
- In the *table* on the *right* **write** the *letter* A beside the name of each of the two alloys listed.
- **Write** a chemical *equation* for the reaction between the acid and the base.
- **Write** a balanced *equation* for this reaction.

The cross-curricular noun *questions* evidently appears in sentences which invite students to answer exam questions or in sentences that specify the marks assigned to each question. The only differentiation in contextual use can be noticed in important instructions and advice that accompany the prompt to answer questions:

English:

- Read this *piece/poem* and then *answer* the **questions**.

- Read the *extractcarefully* and then *answer* the **questions** which *follow*.
- Read this *passage carefully* and *answer* the **questions** which *follow* it.
- *Answer* two of the *following* **questions**.

Geography:

- Examine the Geological map of Ireland and *answer* the **questions** that follow.
- *Section*1 (60 *marks*) All **questions** to be *answered*.
- You have an either / or *choice* within three **questions**.
- All **questions** carry *equal marks*.
- *Answer* the **questions** which *follow*: *Name* one city/the country/the river/four features of erosion etc.
- *Answer* three of the *following* **questions**.

History:

- Read documents 1 and 2, which accompany this paper and then *answer* the *following* **questions**: *Document* 1: This is an extract from the Proclamation of the Irish Republic issued by the leaders of the 1916 Rising.
- Study the pictures A, B, C which accompany this paper and then *answer* the *following* **questions**: *Picture* A shows the St. Patrick's Rock, Cashel, Co. Tipperary. What name is given to the type of building marked X?
- *Short-answer* **questions** (20 *marks*)

CSPE:

- *Answer* all of the **questions** in this *section*.
- *Answer* any three of the **questions** *numbered* 1, 2, 3 and 4 below.
- When you have *studied* this *information/poster*, *answer* the **questions** below.
- A blank *page* for the *poster***questions** has been *included* at the back of this answer book.
- *Write* three **questions** that you would *ask* in this survey.
- Study the *information* carefully and answer the **questions** that *follow*.
- When you are *answering* the **questions** on this paper, you are expected to answer from the human rights approach of your CSPE class.

Mathematics:

- *Time* 2 *hours*: *Attempt* all **questions**.
- Attempt Question 1 (100 *marks*) and four other **questions** (50 *marks* each).

Science:

- *Answer* the/all **questions** in the *spaces provided*.
- Applied Science/Chemistry/Biology/Physics (72 *marks*) There are three **questions** in this *Section*.
- Study the *diagram* and *answer* the **questions** that *follow*.
- *Answer* the *following* **questions** *using* the *table*/on the urinary system/about photosynthesis/about the electrolysis of water etc.
- All **questions** *carry equal marks*.
- *Answer* any one of the **questions**: A (*Plastics*), B (*Textiles*), C (METALS), D (TIMBER) which are on the following two pages.

Looking, finally, at the cross-curricular verb *use*, the following differences can be noticed in the types of things that are used by students to answer exam questions in the different examinations corpora:

Literary texts and language in **English**:

- N.B. In *answering*, you may not **use** the *poem given* on this paper/the *extract given* above as the basis for your answer.
- You will be rewarded for: Well-structured answers, Clarity of expression, An appropriate tone, Good *grammar, spelling, punctuation* and *correct* **use** of *capitals*.

Visual aids and *land use* in **geography**:

- **Use** the *diagram* to *explain* how technology has been used to exploit the bogs.
- *Explain* how you would **use** any two of the instruments to record the weather.
- **Use** the *map* to *answer* the following questions.
- **Use** the Ordnance Survey *map* and/or the aerial *photograph* provided to support your answer.

Hints in **history**:

- If you *wish*, you may **use** the *hints* to *help* you in your answer.

Posters, labels and slogans in **CSPE**:

- Design a *poster* that you would **use** to encourage young *people* in your school to vote in the mock general election.
- Draw a sketch of a *poster* that you would **use** to raise awareness about Earth Day.
- You may **use** the *name* and title/activity/position etc. only once.
- Write a slogan which you would **use** to encourage *people* to get involved in this campaign.

Graphic aids and theorems in **mathematics**:

- **Use** your *graph* to *estimate* the median/inter-quartile range/maximum height reached by the ball.
- If you *wish* to *draw* a diagram, **use** the next page.
- **Use** the theorem of *Pythagoras* to find the length of the side marked y in the right-angled triangle.

Equipment, substances and visual aids in **science**:

- *Give* one **use** for this *piece* of *equipment* in the laboratory/this mineral in the human body/this device.
- *Give* one *everyday* **use** of a magnet/periscope/LDR etc.
- *Examination number* - For examiner **use** only.
- *Name* a liquid suitable for **use** in thermometers.
- *Name* a substance you could **use** to show that this liquid was water.
- **Use** the *graph* to calculate the resistance of resistor R shown in the diagram/*estimate* the solubility at 60°C/*estimate* the melting point of lauric acid etc.
- **Use** the *headings* below.

The examination of collocates and contextual uses of this small sample of content words that uniformly appear in the six examinations corpora reveals that some cross-curricular words, such as *write* and *use*, have more variable uses than others, e.g. *answer, questions*, which tend to appear in fixed sentences (e.g. *Answer the questions that follow*) in examination papers across subjects. Nevertheless, they are all equally important, even words which are used in different contexts with similar meanings and little variation in their phrasing. In other words, ESL students need to be familiar with different possible formulations of the instruction to answer exam questions to be able to cope with the linguistic challenges of assessment tasks: e.g. *answer (one/two/any of) the following questions*, *answer the questions that follow*, *answer the questions below*, *answer all of the questions in this section*, etc.

3.3.2 Recurrent phraseology: cross-curricular 4-word clusters, complementation patterns and context-based meanings and functions

The purpose of comparing the 4-word clusters of textbooks and examinations corpora is to detect any lexical phrases that are uniformly used across subject corpora and to investigate whether they are used with the same or different pragmatic meanings and functional roles. This information can provide insights into patterns of recurrent phraseology favoured by different subjects. Contrary to content words, however, the Detailed Consistency Analysis of 4-word clusters that meet the 40 per million words frequency threshold (adjusted to the size of the corpora analysed in the present research) does not yield any commonly shared clusters across the six corpora of textbooks or across examinations corpora, probably because the frequency limit is too strict given the small size of text corpora. For this reason, frequency lists of 4-word clusters with a minimum frequency of five occurrences in individual textbooks corpora and one mention in individual examinations corpora are computed and compared in this section. Their comparison reveals a small number of lexical phrases that occur in six, five, and four corpora (see Appendices F1 and F3) and these are briefly discussed below.

3.3.2.1 Cross-curricular clusters in textbooks corpora, their complementation patterns and contextual use

Three 4-word clusters appear to recur in all six textbooks corpora and five different clusters in five textbooks corpora (none of which can be found in all 24 text-files however). Only the raw frequency counts of the occurrence of these clusters in total and in individual corpora are provided here (and in Appendix F3), as these emerged from the Detailed Consistency analysis. It would be of little value to normalize these frequencies as they are too low to be discussed meaningfully (that is why only raw frequencies are displayed in Tables 3.27 and 3.28). More common clusters may be found in four textbooks corpora and even more in three and two corpora but these are not discussed in this section. As can be seen in Tables 3.24 and 3.25, the 4-word clusters which are uniformly used across six and five textbooks corpora manifest the two most frequent structural patterns, as these emerged from the structural analysis of textbooks clusters in section 3.2.3.2. More specifically, from the ten 4-word clusters listed in Tables 3.27 and 3.28 below, half are composed of a noun phrase + *of* fragment and the other half comprise a prepositional phrase with an embedded *of* phrase fragment, except *at the same time* which is classified as 'other prepositional phrase fragment'.

Table 3.27 Cross-curricular4-word clusters consistently used across the six textbooks corpora (with information on raw frequency).

N	Cluster	Total	Texts	English	Geography	History	CSPE	Maths	Science
1	**at the end of**	206	23	28	7	39	14	95	23
2	**the end of the**	201	23	37	11	64	9	63	17
3	**each of the following**	482	19	26	115	59	8	248	26

Table 3.28 Cross-curricular4-word clusters consistently used across five textbooks corpora (with information on raw frequency).

N	Cluster	Total	Texts	English	Geography	History	CSPE	Maths	Science
1	**the centre of the**	116	18	9	23	19	0	46	19
2	**the top of the**	83	18	16	10	0	9	17	31
3	**the rest of the**	83	17	27	5	20	12	0	19
4	**at the same time**	67	16	13	8	22	0	10	14
5	**in each of the**	72	15	0	10	7	5	40	10
6	**at the centre of**	34	12	5	7	8	0	8	6
7	**of each of the**	114	12	11	29	23	0	41	10

As explained in section 3.2.3.2, nominalised 4-word clusters express many different meanings, particularly when the same ones are used in different subject-specific contexts:

noun phrase fragments can denote i) physical description, ii) place, iii) size, iv) amount (Biber et al., 1999: 1015) and clusters which incorporate a prepositional phrase + *of* fragment describe i) logical relations, ii) temporal relation, or iii) refer to time periods (ibid.). As regards the clusters listed in the above tables, these cannot be characterised as subject-specific with transparent context-based meanings, in the sense that they do not incorporate content words that are characteristic of particular subject areas. This probably explains their uniform use and wide distribution across the text collections of six or five corpora. What was argued in section 3.2.3.3.2 about examinations clusters appears to be true for the cross-curricular clusters presented here: although they are not composed of subject-specific words, their distinct complementation patterns in different subject textbooks corpora differentiate their pragmatic meanings and functional roles in ways that reflect the thematic and communicative concerns of subject-specific texts. To provide evidence for this claim, the complementation patterns of the above 4-word clusters which cut across six and five textbooks corpora are presented in Table 3.29 below. Their context-based use in subject-specific texts is discussed with reference to the concordance lines which can be viewed in Appendix F2.

Table 3.29 Complementation patterns of 4-word clusters which cut across six and five textbooks corpora ((in)definite articles are omitted).

Textbooks corpus	4-word clusters and complementation patterns
	at the end of
English	poem, story, extract, book, lines, sentence, stanzas, words, chapter, paragraph
Geography	the Ice Age, the eighth century, passageways
History	World War I/II, Civil War, nineteenth century, Middle Ages, passage, hall
CSPE	chapter, book, unit, letter, talk
Mathematics	two/three years, a (second) year, one year
Science	experiment, processes, formula, food chain, tube, rods
	each of the following
English	quotations, statements, words, phrases, questions, paragraphs, types of functional writing
Geography	locations, grid references, counties, directions, names on the map, placenames, terms [e.g. colonialism, empire, plantation], statements, questions, features
History	areas, political developments, events, the Berlin Blockade, the Korean War, the Cuban Missile Crisis, political parties, King Edward VI, Queen Elizabeth, Queen Mary, pastors, deacons
CSPE	sentences, headings, [terms such as:] voting machine, ballot paper, polling station, quota, etc., situations, areas of activity
Mathematics	arrays of numbers, equations, sets, grouped frequency distributions, diagrams, lines, angles, circles, triangles, statements
Science	circuits, groups of animals, [propositions such as:] 'light can be used to power calculators'
	the centre of the
English	stage, set, bench, road, picture
Geography	Solar System, photograph, photo
History	universe, city, empire, empire, trade routes

CSPE	room
Mathematics	circle
Science	flask, glass bowl, ice, root, iris, cards
	the top of the
English	steps, high cliff, dunes, letter, article, front page
Geography	photograph, map, waves, shore, bog
History	tower, keep, theatre, cross
CSPE	page, sheet, chart
Mathematics	vessel, cliff, antenna, fraction, vertical axis
Science	test/boiling tube, beaker, bottle, jar, chromatography paper
	the rest of the
English	poem, story, article, week, day
Geography	world, country, class
History	world, country, crew, nineteenth century
CSPE	population, class, community
Mathematics	lead, race
Science	plant, body

The first 4-word cluster, *at the end of*, typically refers to time, place, and physical description. On the basis of the complementation patterns it acquires in the six textbooks corpora (Table 3.29 above) and the listings of concordance lines displaying authentic examples of contextual use (Appendix F2), the following distinctions in its semantic and functional role across corpora can be drawn:

- to signal the end of literary genres (*story, poem* etc.) and other units of text (*chapter, paragraph*) in **English** and the end of spoken and written texts (*talk, chapter*) in **CSPE**
- to refer to time periods (*the Ice Age*) and, in fewer instances, to indicate physical description of areas depicted in visual aids (*passage ways*) in **geography**
- to mark the end of past time periods (*nineteenth century*) and historical events (*World War I*) and, less often, in the physical description of locations in buildings (*passage, hall*) in **history**
- to specify time periods of years in the description of problems in **mathematics**
- to mark the completion of experiments and processes and the position of pieces of equipment (*test tube holder, wire* etc.) in **science**

The cluster *the end of the* is not included in the above table because it is incorporated in the cluster *at the end of* and it is thus realised in the same ways in all textbooks corpora except in the science textbooks corpus where it marks the end point of objects which are used as pieces of equipment in experimental work: e.g. *the end of the diode, glass tube, magnet, rods* etc. (see Appendix F2).

Each of the following is generally used to express amount, quantity and for specification purposes (see Appendix F2). In the English textbooks corpus, it refers to written texts (*paragraphs, types of functional writing*) and language features (*words, phrases, quotations*).

In geography the cluster refers students to particular visual aids and locations of places in these (*grid references, counties*) and to terms and statements they are asked to explain and discuss (*placenames, colonialism, empire, plantation* etc.). Similarly in CSPE, the adjective *following* mainly specifies terms and concepts students have to define (*voting machine, vote, quota* etc.) or use in fill-in-the-gap exercises. In the history textbooks corpus, on the other hand, *following* is used to list areas, historical figures (*Queen Mary*) or groups of people (political parties e.g. *Home Rule, the Liberals, the Unionists*), historical events (e.g. *the Korean War*), topic-specific terms (linked, for example, to Renaissance art: *fresco, portrait, perspective*). In mathematics, the cluster appears as a referent of numbers and mathematical relations between numbers (*sets, equations, grouped frequency distributions*), shapes (*triangles*) and diagrams. Finally, in science, *each of the following* refers to instruments used in scientific study (*circuits*), organisms (*animals*) and natural phenomena (*light energy*) examined.

The next two word clusters, *the centre of the* and *the top of the* are both used for physical description purposes (see Appendix F2). Thus, in the English textbooks corpus they denote locations or places where stories take place (*road, hall, dunes*) but they can also refer to written texts studied by students (e.g. *picture, letter, article*). In geography, the two clusters specify the location of places of geographical interest (e.g. *bog, shore*) or indicate specific points in visual aids (e.g. *map, photograph*). In history, they normally describe the location of places (*city, empire*), buildings (*tower*), objects (*cross*) etc. The two clusters appear with few occurrences in the CSPE textbooks corpus where *the centre of the* refers to the centre of the classroom and *the top of the* directs students to particular points in written sources of information (*chart, sheet, page*). In mathematics, *the centre of the* is always used with reference to a circle while *the top of the* is used with close reference to mathematical entities (*vertical axis*) and, further, to describe parts of objects or places mentioned in mathematical problems (e.g. *vessel, cliff, antenna*). The recurrent use of the clusters in question in science textbooks facilitates the precise and detailed physical description of pieces of equipment (*test tube, beaker, glass bowl of water*) and materials (*ice, chromatography paper*) that is important when conducting experiments.

The last cluster listed in Table 3.29 above, *the rest of the*, clearly signifies the remaining amount of different entities (Appendix F2). In the English textbooks corpus it specifically indicates the remainder of literary texts (*poem, story, article*) and, in fewer instances, it marks temporal relation (*day, week*). In the three corpora of geography, history and CSPE textbooks the cluster is complemented by words that denote (groups of) people (e.g. *class, community, country, population, world*). In addition, in history, *the rest of the* also functions as a marker of time when followed by lexical items such as *the nineteenth century*. In its few instances in

mathematics and science textbooks corpora, the cluster is complemented by various terms (*lead, race* and *plant, body* respectively).

3.3.2.2 Cross-curricular clusters in examinations corpora, their complementation patterns and contextual use

The Detailed Consistency analysis of 4-word clusters of examinations corpora (with a minimum frequency of one occurrence) reveals the three clusters listed in Table 3.30 to be common to all six corpora (but not to all 60 text-files that comprise them). There appears to be no important reason to normalise frequencies again here due to the quite low numbers of occurrence. These clusters represent some of the most frequent structures identified in the structural analysis of examinations clusters in section 3.2.3.2.: noun phrase + post-modifier fragment (*reason for your answer*), prepositional phrase with embedded *of* fragment (*of one of the*), and WH-question fragment (*what is the name*). Even if all three clusters occur with quite low frequencies, they are worthy of examination because they are the only ones that uniformly appear in the six examinations corpora.

To shed light on their use in subject-specific exam papers, their adjacent items are presented in Table 3.31 and the corresponding concordance lines are included in Appendix F4 and discussed in this section.

Table 3.30 Cross-curricular 4-word clusters consistently used across the six examinations corpora (with information on raw frequency).

N	Cluster	Total	Texts	English	Geography	History	CSPE	Maths	Science
1	**reason for your answer**	82	33	20	3	2	2	26	29
2	**of one of the**	30	18	2	1	19	3	2	3
3	**what is the name**	22	18	5	3	1	3	1	9

Table 3.31 Complementation patterns preceding and following the 4-word clusters used across the six examinations corpora.

Examinations corpus	4-word clusters and complementation patterns
	reason for your answer
English	give a
Geography	explain/give one
History	give a/one
CSPE	give one
Mathematics	give/giving a

Science	give a
	of one of the
English	a modern version – following, the behaviour – characters
Geography	in the case – named poorer regions
History	an account – following topics/military operations, the name – plantations
CSPE	the Ambassador – countries, the work – organisations, leadership – political parties
Mathematics	length – semi-circular ends
Science	circle – meter symbols, the presence – products, the function – tissues
	what is the name
English	of the person/book/magazine/island/most famous painting
Geography	given to the slabs of limestone/a diagram
History	of the well-known politician
CSPE	of the organisation, given to a local authority
Mathematics	given to a line
Science	of the group/piece of equipment, given to the separation technique/this type of reaction

As regards the cluster *reason for your answer*, no differences can be discerned in its complementation patterns and use across the six corpora. It consistently occurs in the sentence *Give a reason for your answer* with little internal variation. This phrase is a fundamental part of examination questions as it functions as the examiner's defence against yes/no answers to yes/no questions. In all instances, its functional role is to request supporting evidence. Differences evidently exist in the issue or question for which students have to provide an answer and justify it, as these are rooted in subject-specific content matter. Concordance lines (Appendix F4) reveal, for instance, that these questions can be related to points of view (e.g. *Do you agree with the poet's view in the above two lines?*) and literary genres in English (e.g. *Do you think it is a good poem?*), to the specification of numerical values in mathematical problems (e.g. *Name a triangle congruent to the triangle acd*), and to explanations of results from experiments in science (e.g. *Is sulphur oxidised or reduced in this reaction?*).

The cluster *of one of the* is used to provide students with options in answering examination questions: e.g. *Write an account of one of the following topics* (history examinations corpus), or as a numerical device: e.g. *You are surprised or upset by the behaviour of one of the characters in a novel or short story you have studied* (English examinations corpus). Concordance lines (Appendix F4) indicate that the cluster refers to i) stories and characters in English, ii) regions/places in geography, iii) historical issues/topics and past events in history, iv) organisations and political parties in CSPE, and v) in various entities/technical terms in the mathematics and science examinations corpora (e.g. *semi-circular ends* and *tissues*, respectively).

From Table 3.31 and the few concordance lines of *what is the name* in Appendix F4, it appears that students are asked to name the following in the six examinations corpora:

- books, magazines, paintings etc., which are probably presented in source texts, in **English**
- materials in **geography**
- political figures in **history**
- social - political entities in **CSPE**
- technical mathematical concepts and scientific terms and equipment in **mathematics** and **science**, respectively.

Table 3.32 Cross-curricular 4-word clusters consistently used across five examinations corpora (with information on raw frequency).

N	Cluster	Total	Texts	English	Geography	History	CSPE	Maths	Science
1	**give a reason for**	80	31	22	0	1	4	22	31
2	**two of the following**	126	31	52	7	54	1	0	12
3	**at the end of**	56	30	6	0	5	10	33	2
4	**each of the following**	77	30	0	23	3	5	13	33
5	**one of the following**	123	30	58	13	47	1	0	4
6	**the end of the**	39	28	9	0	1	10	17	2
7	**the name of the**	90	28	33	3	4	17	0	33
8	**give two reasons why**	79	26	6	3	59	5	0	6
9	**answer the following questions**	77	25	4	6	51	0	3	13
10	**and answer the questions**	65	24	25	18	2	5	0	15
11	**answer the questions that**	53	22	23	12	1	7	0	10
12	**the questions that follow**	53	22	23	12	1	7	0	10
13	**in your answer book**	45	18	1	2	31	0	1	10
14	**the top of the**	41	18	1	0	4	32	3	1

Looking at the 4-word clusters which recur in five examinations corpora in Table 3.32 above, it can be observed that they have higher frequencies than the ones which cut across six examinations corpora and that there are more of them. Their examination at concordance level, however, (see Appendix F4) reveals that many of these are, in fact, duplicates of the same but longer (5/6-word) clusters; e.g. *give a reason for your answer, at the end of the, and answer the following questions, answer the questions that follow, what is the name of the*. Based on the concordance lines of the rest of the above clusters, i.e. *two of the following, at the end of, give two reasons why*, which are indicatively listed in Appendix F4, some remarks can be briefly noted here. The cluster *two of the following* specifies the number of questions to be addressed by examination candidates. It is most often preceded by the communication verbs *answer* and *explain* and it is followed by the cognitive nouns *questions* and *terms*. At *the*

end of signifies finality and, more specifically, signals the end of the following in the five examinations corpora:

- literary texts (*play, extract, story*) in **English**
- time periods and years (*World War I, 1940*) in **history**
- the examination in **CSPE**
- year(s) linked to calculating hypothetical investment amounts in mathematical problems in **mathematics**
- experiments in **science**

Lastly, little can be said about the cluster *give two reasons why* since, similarly with *reason for your answer*, it clearly indicates the need for explanation in answering examination questions. According to the relevant concordance lines, it is normally found at the beginning of a sentence and its typical complementation patterns are propositions which mediate facts for which students have to provide a rationale drawing on their subject knowledge to explain, for instance, why they liked a poem in English, the high life expectancy in First World countries in geography, the defeat of Germany in World War II in history, the famine in Africa in CSPE, and why Earth can support life in science.

3.3.3 Conclusions from the comparisons of content words and 4-word clusters

It has been the purpose of this section to identify language features that are uniformly used across textbooks corpora and across examinations corpora and, further, to compare and contrast their typical uses in subject-specific contexts. The Detailed Consistency analysis function of WordSmith 4 (Scott, 2004) revealed 203 lexical words (frequency ≥40 per million words) and three 4-word clusters (frequency ≥5 in each corpus) to be common to the six textbooks corpora and showed that 34 words (frequency ≥40 per million words) and three 4-word clusters are commonly shared by the six examinations corpora. More words and clusters were found to recur in five corpora (see Appendices F1 and F3). The manifestations of these cross-curricular features and the items typically found in their immediate environment in subject-specific texts were described on the basis of concordance lines (listed in Appendices F2 and F4).

As regards the semantic nature of these items, the analysis shows that all the lexical words and 4-word clusters which were found to cut across the six corpora of curriculum subject textbooks and Junior Certificate examinations, can be characterised as neutral rather

than subject-specific, that is, they are not charged with subject-specific meanings in their own right (perhaps with the exceptions of the less generic words *water, speed, travel, red*). Concordance lines reveal that it is the topic-specificity of the significant collocates of cross-curricular words and the complementation patterns of cross-curricular 4-word clusters that determine and differentiate the semantic and functional associations of these features across subject corpora. These are, in turn, influenced by the broader thematic preoccupations of the six subject areas and the communicative requirements of the two registers of textbooks and examinations.

Comparing the cross-curricular words and clusters in textbooks corpora with those in examinations corpora as a whole, it can be argued that the former have more variable meanings and uses across subject corpora, most probably due to the variability of the information content of the register they represent. The common words of examinations corpora, by contrast, have several identical collocates, e.g. *answer(s), question(s), follow, following, marks, section(s), give, reason(s), number, page, examinations*, and 4-word clusters have the same functional roles in context across the six examinations corpora. These similarities in the uses of cross-curricular words and clusters in the six examinations corpora evidently stem from the assessment focus of language in this particular register.

It is important to underline, however, that the findings from the comparisons of words and clusters presented in section 3.3 should be interpreted with caution because of the limitations involved in their analysis. The results may not be statistically significant and they almost definitely do not represent all the possible and genuine lexical linkages across subjects because the Detailed Consistency analysis functionality of WordSmith 4 (Scott, 2004) is not underpinned by a rigorous statistical measure of significance. They nevertheless represent some lexical links that can be established among the corpora analysed in the present research at a lexical level and, by extension, at a thematic and communicative level, thus facilitating a multi-disciplinary approach to language learning in the language support class.

3.4 Conclusion

In response to the lack of a concrete understanding of the nature of the language ESL students encounter in the post-primary subject classroom, it was the purpose of this chapter to provide detailed, empirically-based descriptions of the most important lexical characteristics of six Junior Cycle curriculum subjects (English, geography, history, CSPE, mathematics, and science).

By analysing subject textbooks and Junior Certificate examination papers of the aforementioned subjects, the following lexical features were identified for individual subject corpora in each of the two registers: i) function and content words, ii) collocates of the commonest content words, iii) significant collocations, and iv) 4-word clusters (together with their structural and complementation patterns, and their semantic and functional characteristics). Concordance lines displaying authentic use in subject-specific contexts were also provided for several of these items. In addition, features (i.e. content words and 4-word clusters) which are frequently and consistently used across subject corpora were detected and their variable semantic and functional associations in different contexts of use were compared and contrasted. From the analysis and interpretation of the above corpus data, several important conclusions can be drawn about the language of the above Junior Cycle curriculum subjects.

One finding that consistently recurs throughout the discussion is the unrandomness of subject-specific language and the linguistic variation that exists across subjects and registers, which is extensively documented in the relevant literature (e.g. Biber, 1988; Halliday and Martin, 1993; Biber et al., 1999; Hyland, 2008, etc.). Language features do not randomly occur in subject-specific texts but they are selected to fulfil particular functions (Conrad, 1996: 300). Regardless of the unit of lexical analysis (i.e. words, collocations, 4-word clusters) and the subject or register (i.e. textbooks, examinations) of the corpus under examination in this research, the results, in all instances, reflect the communicative and thematic concerns of the corresponding texts. More specifically, the choice, frequencies, meanings and uses of the lexical features of textbooks corpora reflect the distinct topics, themes, concepts, tasks and practices of the different subjects, whereas the lexical features in the examinations corpora are primarily linked to the discourse of assessment and testing (which also varies across corpora but to a lesser extent). As a result, the interrelation of lexical features and subject-specific content appears to be stronger in the case of the six textbooks corpora than in examinations corpora because of the heavy informational load of the register of subject textbooks, as opposed to the less diverse language of assessment in the six examinations corpora. Even in the case of cross-curricular words and clusters, the selection of co-text, i.e. significant collocates of content words and complementation patterns of clusters, and their meanings and uses are not free but conditioned by meanings that need to be constructed in subject-specific contexts. On these grounds, it becomes clear that the language of Junior Cycle curriculum is not an undifferentiated and monolithic register but it manifests diverse linguistic varieties. This calls for an ESP perspective (Hutchinson and Waters, 1987) in language support

which takes account of the different varieties of curriculum language, as argued in Chapter 4 (section 4.2.1.3).

Another major aspect of subject-specific language that is illuminated by corpus analysis in this research is its pre-patterned nature, which becomes particularly noticeable in examinations corpora. The ubiquitousness of the commonest lexical words in the top collocations and of the commonest lexical words and collocations in the most frequent 4-word clusters observed in the twelve corpora results in the recycling of the commonest language items. This reveals the lexical patterning that underpins subject-specific texts and illustrates that 'By far the majority of text is made of the occurrence of common words in common patterns, or in slight variants of those common patterns' (Sinclair, 1991: 108). This, in turn, demonstrates the operation of the idiom principle, according to which 'a language user has available to him or her a large number of semi-preconstructed phrases that constitute single choices, even though they might appear to be analyzable into segments' (ibid.: 110). These observations highlight the phraseological nature of subject-specific language and the pedagogical value of lexical phrases. They simultaneously emphasize the value of frequency words because, looking at language from the angle of single words, the above arguments also project a view of language as a web of interconnected networks that are built on high frequency lexical words. From a pedagogical perspective, it is thus possible to describe words in subject-specific texts in terms of their preferred phraseologies (Hunston, 2002: 143) and their grammatical properties (cf. 'local grammars', Hunston and Sinclair, 2000). This emphasizes the importance of textual context for effective language description and it, unavoidably, leads to arguments for a lexical approach to pedagogical grammar and pedagogy in general (Willis, 1990; Lewis, 1993; 2002; Little, 1994: 99-122; see Chapter 4, section 4.2.1.2).

Tied to the previous point, the findings in this chapter further underline that the vocabulary of the different curriculum subjects should not be viewed, and thus taught and learned, simply as single word items (cf. Nattinger and DeCarrico, 1992; Lewis, 1993). It was illustrated through the contextual usage of words and the distinct collocates of cross-curricular words across corpora that the full meanings of words depend on their significant collocates and broader context of use. In addition, the semantic and functional roles of collocations and 4-word clusters show that subject-specific meaning can be found in units longer than single words (cf. Sinclair's (1996) model of extended units of meaning). These facts make it necessary to take account of all three units of analysis (words, collocations, clusters) for a more complete picture of vocabulary use in the six subjects. As Hunston notes: 'One description does not necessarily supersede or replace another but reminds us that language has many more dimensions than a single model of description can comfortably encompass' (2002: 169). The

word-based, collocational and 4-word cluster analyses of the twelve corpora accordingly provide complementary and equally useful perspectives on the lexical dimension of the textbooks and examination papers of the six curriculum subjects.

Finally, it is important to stress the indispensable role of discoursal context for the purposes of analysing and describing corpus-derived data and for making qualitative interpretations of their usage. In this research, the importance of textual context is emphasised through i) the examination of the immediate environment of use of lexical words and collocations and the description of the complementation patterns of word clusters, ii) the concordance lines provided, and iii) the emphasis on the functional utility of language features in subject-specific discourse. In pedagogical terms, constant reference to context implies a discourse-based view of language (cf. McCarthy and Carter, 1994) and a discourse-based approach to ESL teaching and learning, according to which the target language features are considered from a language-in-use (and not decontextualised) perspective, that is, as essential devices to perform communicative acts (see Chapter 4, section 4.2.1.3).

To conclude, the corpus data provided here are not meant to be an exhaustive inventory of the language features ESL students need to learn. They represent, however, pedagogically important information which could not have been identified on a purely intuitive basis and which should be viewed as an integral part of ESL students' target linguistic repertoire. The conclusions drawn from corpus analysis carry important implications for the provision of post-primary language support and could undoubtedly have multiple applications in the ESL curriculum, pedagogy and assessment. These are discussed in the chapter that follows.

4 Pedagogical implications of corpus-based findings and applications in post-primary language support

4.1 Introduction

This thesis has so far i) described the gap that exists between post-primary language support practices and the linguistic demands of the mainstream curriculum (Chapter 1), ii) proposed a corpus linguistics methodology and tools to analyse the language of Junior Cycle subject textbooks and Junior Certificate examination papers in order to address this gap (Chapter 2), and iii) documented many important features of subject-specific language that were revealed by the empirical analysis (Chapter 3). Taking into account the empirical findings reported in the previous chapter, the present discussion aims to address the second question of the study: How could English language support make use of the insights and descriptions of language which emerge from this empirical research?

The response to this question focuses on the following areas: i) ESL teaching and learning, ii) curriculum reform and the design of a lexical pedagogic agenda, iii) materials development, and iv) language assessment. The conclusion comments on the overall contribution of research findings to post-primary language support and on the strengths of a corpus-informed language pedagogy in general.

4.2 Summary of empirical findings

Aiming to shed light on the language of six Junior Cycle curriculum subjects (English, geography, history, CSPE, mathematics, science), corpus analysis in Chapter 3 revealed a wide range of language features characteristic of subject textbooks and Junior Certificate examinations. Further, the qualitative interpretation of these features yielded several descriptive insights into subject-specific language as a whole. This chapter begins with a summary of these empirical findings, touching on their relevance to post-primary language support pedagogy, and it then proceeds to their specific applications in different areas of language support.

All the linguistic features that were identified through corpus analysis, and were subsequently analysed in this research, can be immediately useful in the provision of English language support. Function words were shown to be essential for effective language use regardless of the context. Most lexical words appeared to mediate the topic vocabulary of the different subject areas and to reflect key thematic concepts and terms. Collocations, in their dual manifestation (i.e. collocates of lexical words and strongly associated word pairs viewed as single units), were argued to be essential for word-sense disambiguation, i.e. understanding the different senses of individual words, particularly in subject-specific language texts where several lexical items acquire technical, specialised meanings that can be inferred only from context (cf. restricted collocations, e.g. *retort stand* in the science textbooks corpus, *under translation* in the mathematics textbooks corpus, etc.). The pedagogical value of statistically-significant word pairings, or collocations, was further underlined explaining that they demonstrate appropriate word usage and that they can therefore contribute to the naturalness of ESL students' language use. As Nation (2001) observes, inappropriate word combinations mark non-native language speakers/users. Finally, 4-word clusters, described and analysed in respect of their forms, structure, and complementation patterns, were also presented as pedagogically important features, mainly owing to their functional utility. The discussion illustrated their role as useful devices for the comprehension and construction of discourse and for the internal coherence of texts. It was further argued that their use, together with their complementation patterns, can shape subject-specific meanings and that they can be a reliable indicator of linguistic variation across subjects and registers. It was also noted that both collocations and clusters are essential for real-time communication (Skehan, 1992: 186) because they minimise the amount of cognitive processing that is required in language comprehension and production and they can contribute to the naturalness and fluency of language use. In short, it can be concluded that all the features that emerged from the empirical analysis are fundamentally relevant to language support pedagogy.

The same is true for the conclusions drawn about the nature of subject-specific language in textbooks and examination papers as a whole. Three major findings are worth repeating here. Perhaps the most conspicuous is the strong interdependence between lexical choices and the information content and communicative requirements of a particular text, which results in marked variation across texts of different subjects and registers. This language variation offers substantial evidence to support the repeated calls for ESP and content-based language teaching and learning (section 4.2.1.4). Another observation that is important to repeat here is the recurrence of the commonest topic-based (lexical) words in the commonest collocations and 4-word clusters, as well as their tendency to appear within the

complementation patterns of clusters. This suggests that common lexical words can be held responsible for the commonest meanings (Willis, 1990) in subject-specific texts and that lexical items could be usefully employed as the point of departure in language teaching and learning (cf. section 4.3.2 on the value of a lexical approach). The repeated use of the commonest features (words and collocations) in the commonest patterns (clusters) simultaneously illuminates the lexical patterning that underpins language use in subject-specific domains and attests to its pre-patterned nature. This also reinforces the need for a lexical approach which places emphasis on 'pedagogical chunking' (Lewis, 1993: 120) for teachers and students alike (cf. section 4.3.2). Finally, the dependence of the full meanings and appropriate usage of lexical items (words, collocations, clusters) on the broader textual context revealed by concordances highlights the importance of co-text and discourse (cf. section 4.3.3) and undermines the assumption that words are independent entities. As Sinclair remarks: 'the words of a language [should be viewed] as acting more like the letters of an alphabet – each one contributing to the recognizable shape of the higher unit, but not necessarily adding a clearly defined meaning' (1997: 35).

Overall, the above summary shows that both the corpus data and the descriptive qualities of subject language revealed by this empirical research are pedagogically relevant to the post-primary language support classroom. The remainder of the chapter exemplifies their implications and applications in relation to pedagogy, curriculum, materials and assessment. It should be noted that an inevitable consequence of the focus of research is that the empirical findings have direct implications for ESL students' reading and writing skills and become less relevant to speaking and listening.

What is important to underline here is that, although the present empirical research is the first of its kind in the Irish post-primary context, and it was conducted with a concern for ESL students, its findings carry implications for native English speaking students in the post-primary classroom. Considering that such an analysis has not been undertaken before for language across the Irish post-primary curriculum, it is argued throughout this chapter that the corpus-based insights into subject language demands can help teachers support all students who face difficulties gaining control of the academic register (and have a 'restricted code' Bernstein, 1971) and enhance their academic achievement. This argument is based on the role of language as a medium of acquiring knowledge in school education and the importance of the explicit teaching of language in classroom instruction to promote advanced literacy (e.g. Colombi and Schleppegrell, 2002: 5).

4.3 ESL learning and teaching

The insights into the lexical characteristics and overall nature of subject-specific language in textbooks and Junior Certificate examinations that emerge from this empirical analysis have a number of implications for ESL learning and teaching in the language support classroom. They can generally facilitate informed decisions about what needs to be taught (language content) and how (teaching approach). Specific pedagogical implications are discussed in this section in relation to: i) prioritising the language to be taught, ii) a lexical approach to ESL teaching and learning, iii) the importance of context and discourse considerations, iv) ESP and content-based language learning, v) raising the awareness of subject and language support teachers and encouraging teacher cooperation and whole-school approaches to language learning across the curriculum.

4.3.1 Giving priority to the commonest lexical features

It was argued in Chapter 1 that, due to its limited duration (two years maximum), language support needs to be focused on the most important language ESL students need to learn. The frequency-based approach to language analysis and the restriction of range of occurrence in this research help to address this need by revealing the most important lexical features of the six curriculum subjects, thus helping teachers to prioritise the language items to be taught. As explained in Chapter 3 (section 3.2), the value of language features with high frequency and wide range lies in the fact that they are encountered many times across different texts and normally have large text coverage (i.e. a large proportional use in texts) and can be learned more easily (frequency correlates with learnability; e.g. Milton, 2009). They are therefore essential for a wide range of uses of the target language and consequently very useful for ESL students.

In addition, frequency information is important for proficient word use because the overuse of certain relatively infrequent words and the underuse of certain relatively frequent words may be one reason why L2 learners do not sound native (Schmitt, 2000: 76). Arguably, the same applies to combining words in pairs (collocations) and longer sequences (clusters). The strength of the above arguments can be further supported by the demonstration of the enormous power of common words in English, which is considered a lexical language (Willis,

1990: 22-24; on the evidence that its use entails the repeated use of common words and phrasal verbs), as revealed by the analysis of the COBUILD corpus (cited in Willis, 1990: 46):

> The most frequent 700 words of English constitute 70% of English text.
> The most frequent 1,500 words of English constitute 76% of text.
> The most frequent 2,500 words of English constitute 80% of text.

As Willis concludes, 'in many texts, even if they are highly specialised, the incidence of words outside the 2,500 frequency band is increasingly low' (ibid.).

For language support teachers, on the other hand, information about the most frequently and consistently used language features of the subject textbooks and Junior Certificate examinations corpora can inform the design of the vocabulary component of language support and shape decisions about content selection and sequencing. Teachers can, for instance, set (short- and long-term) vocabulary teaching goals based on word frequency bands which can be established for gradation purposes and as benchmarks of levels of vocabulary knowledge (O'Keeffe et al., 2007: 31). Frequency information on the different meanings of word forms and clusters might also prove to be useful for the above purpose (Flowerdew, 2009: 330) - the editing function of the concordancer in WordSmith 4 (Scott, 2004) allows the grouping of items according to their different meanings. In this way, teachers can ensure students' principled exposure to carefully-selected language.

Frequency data should not be accepted uncritically however, and low frequency, and even certain hapax legomena, should not be excluded from the ESL teaching agenda. Nation (2001) reports that it is, in fact, low frequency words that pose difficulties to students and suggests that teachers should rather train learners in the use of strategies to deal with such vocabulary (e.g. guessing from context clues, using word parts to help remember words, using vocabulary cards and dictionaries, etc.) than spend time teaching them (ibid.: 20-21). This is an important argument for language support, considering that it can be provided only for two years, and it could be operationalised via the ELP (autonomous vocabulary learning). Additional criteria are very often needed to evaluate the importance of linguistic features in order to determine the lexical content of teaching, such as saliency (Hunston, 2002: 193), communicative usefulness, the specific needs of individual students, etc.

A debatable issue relates to the extent to which input from post-primary subject teachers could also play a role in language selection, particularly when some of the corpus data appear to be counterintuitive. This issue can be linked to arguments in favour of or against corpus-informed, or corpus-based, pedagogies as opposed to corpus-driven pedagogies. The first use corpora to 'expound, test or exemplify theories and descriptions that

were formulated before large corpora became available to inform language study' (Tognini-Bonelli, 2001: 65), and therefore opponents of this argument discard inconvenient corpus data i.e. 'data not fitting the pre-corpus theory' (McEnery et al., 2006: 8). Corpus-driven pedagogies, in contrast, are committed to 'the integrity of the data as a whole' (ibid.: 84) and willingly accept them. Taking account of this opposition, teachers' contributions to filtering the empirical data of this research would be welcome in the first case but not in the second. Considering that the present research was undertaken to inform subject specialists who have not received any formal linguistic training, encouraging their manipulation of corpus analysis would distort the linguistic reality of empirical findings, and would thus defeat the original purpose of this research. This argument could be further supported by considering that teachers' impressionistic judgements about what language is important, has very often resulted in a number of distortions and serious omissions in the language classroom (cf. Sinclair and Renouf, 1988: 152-153; Willis, 1990: 51; Römer, 2008 etc.). These failings may be partly attributed to the fact that 'Native speakers often notice the marked, or unusual, rather than the unmarked, or typical uses of language' (Reppen, 2010: 4) and thus 'it is in this [latter] area that corpus linguistics can make the greatest contribution to language teaching' (ibid.). It cannot be denied that a corpus-based language study, like the present research, can provide information that even the most sophisticated native speaker would not be able to offer based on intuition.

Overall, it may be argued that, classroom time spent on the features analysed in the present research could be 'well justified by their frequency, coverage and range' (Nation, 2001: 16) and also by the fact that they represent lexical characteristics of core curriculum subjects (English, geography, history, CSPE, mathematics, science) in the two most important educational registers (subject textbooks and Junior Certificate examinations papers).

4.3.2 Adopting a lexical approach to ESL teaching

The focus of the present research on the lexical dimension of corpora and the lexical descriptions of subject- and register-specific language in the present research clearly promote a lexical approach to ESL teaching and learning. Put forward by Lewis (1993), this language teaching method is based on the increased understanding of the nature of lexis in naturally occurring language. It takes as its starting point the lexical item, arguing that this can lead to powerful generalisations about language use which are more meaningful for learners than

structures (Lewis, 1993; 2000; 2002). Particular emphasis is therefore placed on collocations and chunks, i.e. lexical items longer than single words. More specifically, the collocational range of words and restrictions on that range are viewed as a major element of procedural vocabulary knowledge (Lewis, 1993: 119), that is, the ability to apply knowledge in practice. The value of training in pedagogical chunking, i.e. 'breaking continuous text into useful component "bits"' (ibid.: 120), is also underlined on the evidence that 'It is the gradual perception of pattern which underlies the development of competence, and what is usually thought of as "grammar"' (ibid.). Grammar is not neglected but it is examined in terms of individual words, i.e. their principal collocates and institutionalised sentences which contain them (ibid.: 115) rather than structures, supporting the view that 'language consists of grammaticalised lexis, not lexicalised grammar' (Lewis, 2000: 137; see also Little, 1994 for arguments for a lexical approach to pedagogical grammar).

The present research is consistent with the above tenets of a lexical approach. It employs the lexical item as the primary unit of linguistic analysis and recognises the importance of different types of lexical items (words, collocations and word clusters). It also provides information on the collocational range of frequent lexical words and on collocations restricted to particular subjects. Further, it acknowledges the importance of 4-word clusters and their functional utility in subject-specific contexts and it could thus facilitate pedagogical chunking. It could therefore be argued that the subject-specific and cross-curricular lexical inventories offered in this research could provide the basis for a lexical approach to ESL teaching and learning; they could inform activities for language 'noticing' (Schmidt, 1990), consciousness raising, to convert input into intake; and could facilitate a principled organisation of vocabulary (Lewis, 1993: 118; e.g. in terms of the semantic links between words, i.e. co-ordination, collocation, superordination, synonymy; Aitchison, 1987: 74f.), all of which are viewed as important for the successful implementation of a lexical approach. Converting the Benchmarks (IILT, 2003a) into a lexically-based pedagogic agenda is another application of empirical findings that could clearly contribute to the implementation of a lexical approach, as argued later in this discussion (section 4.4.2).

4.3.3 Encouraging content- and discourse-based teaching and learning

Although in this research the primary focus of corpus analysis is on the examination of different types of lexical items, several perspectives are also provided on co-text, i.e. linguistic

environment and the use of extended text or discourse (Lewis, 1993: 105). These recognise the importance of textual context and imply the need for a discourse-based approach to ESL teaching and learning.

Contextual and discoursal considerations form an integral part of communicative language teaching and learning. Contextualised language learning provides authenticity because it 'places words in contexts of use, so that the conditions of learning closely resemble the conditions under which the words will need to be used' (Nation, 2001: 300). At the same time, students become familiar with the range of contexts in which a particular vocabulary item occurs; in this case, it can be the context of individual subjects or of different themes of a single subject, contexts across different subjects and of different school registers. As Willis argues, 'only by drawing students' attention to occurrences in text can learners begin to build up a picture of (these forms) in use' (Willis, 1990, cited in Lewis, 1993: 140).

In this research, the importance of co-text is acknowledged in two ways. First, through the context-based examinations of several lexical features using concordancing and second, through the constant references to the thematic and communicative concerns of texts and registers in which features under examination can be found. Both underline the importance of co-text and imply the need to teach target language items embedded in their 'natural habitat' rather than decontextualised. In the language support classroom, this can be achieved by making explicit use of two types of information from this research.

First, teachers can draw upon the interrelations between the semantic and functional interpretations of subject-specific language features and the thematic and communicative requirements of subject texts and registers described in Chapter 3. This information could motivate teachers and students to take account of the broader discourse context and pay attention to the functional utility these features have in different contexts. By making this information explicit to ESL students, teachers can highlight the discourse dimension of vocabulary and reinforce a view of lexis as a communicative resource (i.e. from a language-in-use perspective) rather than a sterile list of lexical items. Secondly, the concordance lines provided for several lexical features in this research can be exploited in numerous ways to introduce ESL students to the importance of co-text and to expose them to authentic language use. Using edited versions of concordances in materials development (see section 4.5.2.1) for direct learning and for language focused instruction etc., or using unedited concordances for data-driven learning to promote incidental, discovery learning (see section 4.5.2.2) are two among many possibilities.

It is, however, important to note here that a broader understanding of context, beyond co-text, and wider discourse dimensions should also be taken into consideration in language

support. As explained in Chapter 2 (section 2.3.1), the examination of subject-specific discourse can focus on i) discourse as text, ii) discourse as genre, encompassing structure and organization, and iii) the wider socio-cultural context of language use. Although the primary focus of this study was the textual dimension of written academic discourse, the importance of the other two dimensions should not be ignored. With specific regard to information on subject discourse as genre, future research is needed that examines organisational and structural patterns of subject- and register-specific texts adopting, for instance, text linguistics and genre analysis.

As regards the sociocultural context of language use, language support teachers should provide relevant information to ESL students. As Nagy (1997: 83) remarks: 'the notion of context cannot be restricted to the textual neighbourhood of a word'. Apart from linguistic meaning, texts provide the opportunity to deal with affective meaning, i.e., attitudinal and emotional factors conveyed through words (for instance, 'single woman' as opposed to 'spinster'; Gairns and Redman, 1986: 18) and with the sociocultural connotations of words (which are universal but also context-specific). All these types of meaning are essential for appropriate vocabulary use and they become increasingly important in the pedagogical context in question, considering the diverse sociocultural profile of ESL students.

On the whole, it is only through the combination of corpus linguistics techniques and other discourse analytic methods, e.g. genre-based approaches, text-linguistics, and Systemic Functional Linguistics, that a holistic examination of curriculum subject discourse as text, genre and social practice (Bhatia, 2004: 18) becomes possible (see the prospects for future research in Chapter 5, section 5.3). The added value of using combined methods lies also in its potential to address criticisms levelled against corpus data, e.g. decontextualisation of language information (Widdowson, 1998; 2002) and a lack of focus on large stretches of discourse, as is the case in genre analysis (Swales, 2002).

4.3.4 Facilitating ESP and content-based language instruction and learning

The lexical variation across corpora of different subjects and registers described in Chapter 3 carries direct implications for the type of ESL learning and teaching methodology that needs to be employed in the language support class. More specifically, the striking differences in the lexical realisation of the six content areas in textbooks and examination papers evidently call for an ESP perspective (e.g. Hutchinson and Waters, 1987) in the language support classroom

that distinguishes the language varieties of different subject areas and makes explicit the linguistic conventions of different registers. This implies the need for content-based ESL teaching and learning. Although the importance of content-based language instruction to facilitate subject learning was emphasised in Chapter 1 (section 1.2.3.1), the empirical findings of this research provide concrete, quantifiable evidence to legitimately argue for ESP rooted in the demands of the different curriculum subjects and can be used in actual practice to this end.

The empirical data of the research can be used in multiple ways to support content-based language instruction in the classroom. The corpus-based lexical profiles of the six Junior Cycle subjects enacted in textbooks and Junior Certificate examination papers can, first of all, provide teachers and students with a clear and concrete understanding of the preferred lexical patterns of the six subject areas. Both quantitative and qualitative findings can be used to mediate this understanding through all areas of pedagogy; e.g. the content of instruction, the ESL curriculum, the design of a lexical syllabus, classroom tasks and activities, materials development, and the content of assessment instruments. In this way, teachers can help ESL students to i) appreciate the lexical variation that exists in subject-specific language varieties, ii) gain an understanding of some of the key features of the sub-languages that different subjects operate, and iii) actively engage in English language learning embedded within the communicative context of content areas. In addition, the use of concordance lines, displaying the use of key lexical features in multiple subject contexts, can effectively facilitate students' learning of new language items against the conceptual background of subject-specific themes and topics. Introducing students to disciplinary modes of language use entails, by extension, their introduction to patterns of communication and thinking within the framework of individual disciplines. In this way, students can be gradually drawn into the communicative roles they have to perform within the discourse communities of subject areas.

Inter-disciplinary perspectives could also be usefully incorporated in content-based ESL instruction on the basis of the cross-curricular features discussed in this research. Contextual encounters of commonly used vocabulary items across subjects can consolidate students' language learning and increase the depth of their vocabulary knowledge, i.e. 'the knowledge of various aspects of use of a word, including beyond its formal properties, its collocations, its sub-senses, and its semantic prosody' (O'Keeffe et al., 2007: 54). This knowledge ultimately contributes to the learner's ability to create associations between words and to place them meaningfully within various networks in relation to other words (Meara, 1996; Haastrup and Henriksen, 2000). Lexical associations are likely to promote linkages at a thematic level which could ultimately promote organic rather than fragmentary lexical and disciplinary knowledge.

From the above it follows that the present research can promote the integration of language and content in language support giving rise to many benefits for ESL students. Some of these benefits are worth considering here.

The integration of subject-matter content and English language in language support is, first of all, in line with communicative approaches to language learning which emphasize the meaningful use of language in appropriate contexts (Brinton et al., 1989; Snow et al., 1989). As Davison and Williams remark: 'the subject matter of academic learning is adopted as a meaningful basis for the selection of language for teaching and the meanings of integrating various language features and skills to be taught' (Davison and Williams, 2001: 53). Supporting students' learning and use of academic English language within the context of post-primary curriculum subjects can also speed up the learning process: 'The provision of concurrent instruction in both subject-matter and language may speed up the process of achieving subject-based learning as language development and content go hand-in-hand' (ibid.: 2001). This is not to argue, however, that language support teachers are expected to teach the curriculum.

Content-and-language integration in the language support classroom is further likely to stimulate higher order thinking skills which can, in turn, contribute to the development of advanced levels of English language proficiency (CALP; Cummins, 1984; cf. Chapter 1, section 1.2.3.1). This, however, 'will obtain to the extent that higher order thinking skills require more complex or elaborate language skills in more cognitively demanding tasks' (Snow, Met and Genesee, 1989: 215). It can be argued, moreover, that ESL learning can become more effective if the emphasis of instruction is placed on the communicative information of curriculum subjects rather than on decontextualized language: 'the student can most effectively acquire a second language when the task of learning becomes incidental to the task of communicating with someone [...] about some topic [...] which is inherently interesting to the students' (Tucker and D'Anglejan, 1975). This can have a motivational effect for ESL students who all share the common goal of accessing subject-specific language learning: 'the use of informational content which is relevant to learners [i.e. subject-matter] can increase motivation for learning' (Brinton, Snow and Wesche, 2003: 3).

This research provides, overall, strong empirical evidence in support of an ESP perspective in language support. More importantly, it offers a substantial bank of subject-specific language data that can be used to facilitate content-based language instruction and learning; as such, the research could contribute to supporting the third and most challenging phase of ESL instruction, i.e. subject-specific language learning (IILT, 2003b; cf. Chapter 1,

section 1.2.3.1 on the phases of ESL instruction) and directly contribute to a principled approach to the planning, design and delivery of content-based language instruction.

It could be further argued here that the empirical insights into curriculum language can also benefit native English speaking students in the subject classroom who are less familiar with the 'dialects' of post-primary education. This can be supported by recognising that 'developing knowledge and understanding in school subject areas and developing control of the linguistic resources that construct and communicate that knowledge and understanding are essentially the same thing' (Hasan, 1996). Informed by the current empirical findings, subject teachers can differentiate their teaching methods to make the language demands of curriculum subjects explicit in the classroom which can benefit native and non native speaking students alike.

4.3.5 Stimulating awareness raising, teacher cooperation and whole-school approaches

The empirical findings of this research might also prove to be useful for raising the awareness of post-primary subject specialists and language support teachers in three ways: in terms of i) their personal responsibility and role in facilitating ESL students' linguistic integration, ii) the diverse nature of curriculum language and its role as a medium of content learning, and iii) the need for whole-school approaches towards promoting language learning across the curriculum for all post-primary students.

The subject- and register-specific nature of linguistic features and the resulting striking differences across corpora that were revealed through corpus analysis automatically dispel the assumption held by some language support practitioners that there is a single uniform and monolithic English language that can be taught to ESL students (cf. Chapter 1, section 1.3). In addition, the analysis of the heavy lexical load of curriculum subject textbooks and examination papers in this study can be used to sensitise mainstream subject teachers to the linguistic dimension of the content areas they teach and help all teachers to appreciate the challenge involved in ESL students' endeavours to access subject textbooks and examination papers. At a more practical level, teachers might also realise the importance of subject textbooks and examination papers as authentic linguistic resources and find ways to exploit them for pedagogical purposes in language support (cf. Chapter 1, section 1.3; Lyons and Little, 2009). This increased awareness could encourage subject teachers in particular to accept their

responsibility in facilitating ESL students' linguistic integration, which, although it is clearly stated in the official DES Circular - 'The EAL [English as an Additional Language] pupil remains the responsibility of [...] the subject specialist teachers at post primary level who will work closely with the EAL support teachers' (2009: 2), has not yet been universally embraced, as mentioned in Chapter 1 (section, 1.3; Lyons and Little, 2009).

By demonstrating that each curriculum subject operates its own language variety, this research highlights the instrumental role of language as a medium of subject learning for native and non-native students alike. It illustrates that the lexical preferences of individual subjects can have a gate-keeping function for accessing content-based meanings. Accordingly, it can be argued that there is a reciprocal relationship between ESL learning in the language support class and content learning in the mainstream subject class. This relationship is summarised by Mohan (1986: 18) thus:

> Language learning in the communicative environment of the content classroom furthers the goals of language teaching by offering a context for language. It provides language use in a context of communication about important subject matter. Language ceases to be taught in isolation. At the same time, language learning in the language classroom can further the goals of content teaching by offering learners help with the language of the thinking process and the structure or shape of content.

It could be inferred on the basis of Mohan's argument that ESL learning in the language support class and content-based learning in the mainstream subject class are two sides of the same coin for ESL students. This suggests that it is only through joint efforts for a systematic whole-school approach to language learning across the curriculum that ESL students' integration can be successful; but this depends on the close cooperation of subject and language support teachers.

While the need for language support teachers to 'share their expertise with mainstream class teachers and assist in developing and disseminating good practice' is clearly stated in the DES policy document (2009: 2), it has been reported that this appears not to happen on a large scale in actual practice (cf. Chapter 1, section 1.3). It is hoped that the descriptive insights into curriculum subject language offered by the present research could have a role to play in this direction. By promoting a visible pedagogy (Bernstein, 1990), the empirical findings of this study can lay the foundations for an explicit agenda of pedagogical actions and intervention that could act as a common platform of cooperation between teachers. This cooperation is essential to promote whole-school approaches to learning across the curriculum. The Toolkit for diversity in post-primary schools (IILT, 2009), could also be fundamental to this end since it

was designed 'to help schools to create and sustain a welcoming and inclusive environment for students and parents from all backgrounds and ensure equal access for all'. This can be accessed on the ELSP website (www.elsp.ie).

Whole-school approaches to cross-curriculum language learning would have a positive impact on native speaking students too, particularly on students whose 'gap between their own language and the textbook is so great that the textbook becomes mere noise' (Rosen, 1972: 123). This 'linguistic-intellectual bafflement' that besets secondary school students in subject learning (ibid.: 119) stems from the role of language as a medium of learning, discussed in Chapter 1 (section 1.2.3). Subject teachers need to recognise that 'language permeates the whole curriculum both as a vehicle for learning and as an object of study in its own right' (Bearne, 1998: 11) and that 'language is so deeply embedded in many subjects of the secondary curriculum that it is sometimes difficult to separate learning the concepts and processes of a subject from learning to use language to represent and use these concepts and processes' (Barnes, 1972: 113). On this evidence, the pedagogical methods that are employed to support ESL students' L2 learning embedded in the post-primary curriculum (e.g. providing definitions, explanations, reformulating language in the mainstream class and recycling new language, etc.) may well facilitate native speaking students' mastery of the academic register. The empirical findings of the present research could make a contribution to this end.

Overall, the discussion in this section has shown how the present research can become immediately useful to ESL teaching and learning in post-primary language support. The empirical findings from corpus analysis were proposed as a basis for a lexical and discourse-based approach to teaching and learning the most important (i.e. most frequent and consistently used) lexical features of the two fundamental registers (subject textbooks and Junior Certificate examinations) of six curriculum subjects (English, geography, history, CSPE, mathematics, science). It was further argued that, promoting a visible pedagogy in the Irish post-primary context taking account of the present empirical findings can also benefit native English speaking students with learning difficulties and low academic achievement.

The discussion that follows puts forward additional ways of using the empirical insights of the research to establish direct links between language support and the mainstream classroom; namely, by incorporating corpus-based findings into the existing ESL curriculum framework and the design of a lexically-based pedagogic agenda (section 4.4), materials development (section 4.5) and assessment procedures (section 4.6).

4.4 Informing curriculum reform: enhancing the Benchmarks and ELP

It was established at the beginning of this thesis (Chapter 1) that language support aims to give students access to English-medium education. In line with this aim, it was shown how the curriculum framework of Benchmarks (IILT, 2003a) and ELP (IILT, 2004) that is currently in place reflects target learning outcomes expressed in 'can do' statements that are rooted in the communicative demands and major themes and topics of the post-primary curriculum. It was explained, however, that, the Benchmarks and ELP were not developed as a curriculum that specifies the language to be taught in language support. They rather represent an action-oriented curriculum which describes the developmental path ESL students have to follow to become able to fully operate in the mainstream subject classroom. Taking into account language support teachers' critique of the Benchmarks as abstract (Lyons and Little, 2009; Fionda, 2011) and the need for concrete language specifications, this section discusses how the empirical findings from corpus analysis could be used to enhance the effectiveness of the Benchmarks and ELP for teachers and ESL students respectively. It is specifically argued that corpus-based language descriptions with subject-specific substance can be used i) to supplement and elaborate the content of the subject-specific scales of the Benchmarks, ii) to convert these scales into a lexically-based pedagogic agenda for the language support classroom, and iii) to enrich the ELP language biography and dossier.

4.4.1 Enhancing the subject-specific scales of the Benchmarks

The empirical findings of the present research can be used to enhance the subject-specific scales of the Benchmarks by making them more concrete and by expanding their content. More specifically, the lexical profiles of the textbooks and examination papers of the six curriculum subjects offered here can add further specificity to the subject-specific scales of the Benchmarks (Part IV; Appendix A1) which capture: 'the varieties of communication the language support student must master in order to become fully integrated in the mainstream' (IILT, 2003a: 15). An exception would evidently be the Physical education and sports scale, as this was not included in this empirical analysis. This suggestion could be implemented in different ways; two are proposed here.

The simplest way would be to supplement each subject-specific scale of the Benchmarks with comprehensive datasets which would comprise all the corpus-derived lexical

characteristics of the corresponding subject areas and which would be separately provided. Thus, the set of communicative descriptors for mathematics, for instance, can be accompanied by a database with i) mathematical vocabulary items, ii) collocates of the commonest mathematical words and significant collocations, and iii) 4-word clusters characteristic of mathematical discourse. Concordance lines displaying authentic use should also form an integral part of each database. To facilitate their use by language support teachers, the information of each dataset could be categorised in ways that are meaningful for pedagogical purposes. For instance, lexical words could be grouped thematically, to reflect the units of study of individual curriculum subjects, while 4-word clusters might be more effectively categorised according to their functional role in discourse etc.

A different possibility might be to match corpus-derived lexical features to individual 'can do' descriptors (perhaps in the form of hyperlinks) based on the criterion of semantic and functional relevance. More specifically, the lexical words and collocations revealed from corpus analysis can be matched to the content of the abstract terms and concepts included in these descriptors, while 4-word clusters (together with their complementation patterns) can be used as concrete examples of the language functions (speech acts) that are described by 'can do' statements (i.e. correspondence between the subject-specific notions and functions of 'can do' statements and corpus-derived lexical items which mediate these). To provide an example, drawing again on the mathematics scale, the 'can do' descriptors which make reference to students' ability to recognise and say 'the words for shapes' can be accompanied by a lexical inventory which includes corpus-derived words associated with this particular semantic field such as the following:

> *line, angle, circle, cylinder, triangle, rectangle, sphere, cone, parallelogram, perpendicular, square, rectangular, curved, cylindrical, curve, shape, horizontal, arc, straight, vertical, squares, cube, right-angled, semicircle, cubic, diagonals, geometry, shaped, diagonal, circular, equiangular, quadrilateral, sharp, vertices, arcs, 3d, semi-circular, corners, semicircles* etc.

A further categorisation that might be deemed pedagogically useful would involve grouping the above words according to the four main word classes (nouns, verbs, adjectives, adverbs).

Similarly, the following list of 4-word clusters could be provided to illustrate part of the language knowledge that is involved in the ability to 'read and understand most textbook tasks and problems' in mathematics (Reading, B1 Level):

- *calculate the area/height/length/volume/radius/cost/amount/rate/of/the mean mark/an average speed of*
- *find the value(s)/image/equation/area/slope/number of*
- *write an equation in*
- *draw the graph of*
- *draw a Venn diagram*
- *use your/the graph to*
- *divide both sides by*
- *construct the image of*
- *correct to one decimal/the nearest*
- *factorise each of the, etc.*

Descriptions of the complementation patterns of the above listed clusters (e.g. *use your graph to estimate the maximum value, draw the graph of the function of f, construct the image of abed under an axial symmetry in line L*, etc.), as well as concordance lines displaying their manifestations in full sentences, like the ones below, could further contribute to making language functions described in the 'can do' descriptors more concrete.

- *Calculate the values of the letters representing the angles in each of the following diagrams. Find the area of each of the following triangles, where the lengths of the sides are in cm.*
- *Draw a histogram to represent each of the following grouped frequency distributions.*
- *Name the shaded region in each of the following Venn diagrams.*

Regardless of the way in which corpus findings become incorporated in subject scales, the use of functional descriptions of target communicative behaviour in tandem with corpus-based language specifications for the different subjects could considerably improve teachers' understanding of subject-specific communicative and linguistic demands, and could contribute to an explicit teaching agenda.

In addition to enriching the subject-specific scales of the Benchmarks, it could be further argued that the findings of this research carry some indirect implications for revising the content of subject-specific scales. More specifically, it could be suggested that the content of the scales can be elaborated and expanded to account for certain aspects of subject-specific language that were revealed by corpus analysis but have not been captured by the present scales. Vocabulary items associated with the semantic field of substances and materials in science is one such example; although substances and materials constitute an integral part of scientific knowledge, no relevant reference is made in the 'can do' descriptors of the Benchmarks scale for science subjects.

Other possible changes that could enhance the quality of subject-specific scales involve distinguishing between geography and history descriptors. Although there are some overlaps between the two subject areas in terms of language (e.g. proper names of people and places and geographical areas), corpus analysis reveals that their overall lexical preferences are quite distinct. This can be attributed to the different topics and communication patterns of the two areas of study. Therefore, it might be useful to designate different scales for the two subjects. The addition of a scale for the subject of CSPE is also recommended here, as this is one of the core subjects of the post-primary curriculum, and its content could be informed by the results from the analysis of the CSPE textbooks and examinations corpora in the present research.

Although the discussion in Chapter 3 makes reference to distinctions of linguistic features among the sub-disciplines of mathematics (algebra, trigonometry, geometry) and science (biology, chemistry, physics), it does not seem appropriate to recommend introducing these distinctions in the 'can do' descriptors of the corresponding Benchmarks scales. In the same vein, distinctions between the language of textbooks and examinations of individual subjects are not considered useful either. It would be essential, however, to make the distinct lexical demands of the two registers explicit in the language support classroom through instruction and appropriate language activities in order to make students aware of the distinct conventions and norms of subject registers.

Having explained how the subject-specific scales of the Benchmarks can benefit from the empirical findings of this research, it is important to consider also the ways in which these findings might be viewed as somewhat limited in their present form to effectively complement subject-specific scales. Considering that corpus-based language descriptions provided in this study primarily represent the written language of textbooks and examination papers, they might appear to be immediately relevant only to descriptors for reading and writing skills. It could be argued, however, that these corpus-based descriptions of subject textbooks and examinations language have implications for listening skills and skills of spoken production and spoken interaction because, as explained previously (Chapter 2, section 2.2), the language of textbooks and examinations should, at least in principle, underpin the language of instruction and the spoken production and interaction that is required of students in the subject classroom.

An important limitation of the findings of this research however is that they have been derived from a linguistic analysis of Junior Cycle subjects while the subject-specific scales of the Benchmarks mediate the communicative demands of both the Junior and Senior Cycle of post-primary education. It would therefore be necessary to extend the empirical analysis to Senior Cycle textbooks and Leaving Certificate examinations in order to draw a complete

profile of the language of post-primary education (cf. prospects for further research in Chapter 5, section 5.3).

The discussion in this section has overall argued that the subject-specific findings of the present research can be used to enhance the subject-specific scales of the Benchmarks in two ways. First, they can enrich the present 'can do' descriptors by adding further specificity to the linguistic dimension of the learning outcomes they describe. Second, they can inform an elaboration and expansion of the content of the present scales to account for aspects of content-based lexical knowledge and communicative functions which can be inferred from the empirical findings of the research.

The discussion that follows extends the argument made in this section arguing for converting the contents of these scales into a lexically-based pedagogic agenda for the language support classroom.

4.4.2 Converting the Benchmarks into a lexically-based pedagogic agenda

The corpus-based lexical profiles of curriculum subject textbooks and examination papers could be profitably channelled into the design of a free-standing lexically-based teaching agenda which would be firmly embedded in the subject-specific scales of the Benchmarks. In this way, the Benchmarks could inform teachers about the functional development of ESL students in different content areas and teachers could use this lexically-based 'pedagogic agenda' (Widdowson, 1990: 127) to inform their teaching on a daily basis. The envisaged pedagogic agenda could take the form of a lexical syllabus which would serve as a most effective tool for promoting a lexical approach to ESL teaching and learning in language support (cf. section 4.3.2).

As its name suggests, a lexical syllabus (Sinclair and Renouf, 1988; Willis, 1990) is underpinned by a lexical description of language, its underlying principle being frequency (Flowerdew, 2009: 338). According to the basic tenets of a lexical approach (cf. section 4.3.2), the specification of the lexical syllabus content involves the description of the commonest lexical items, their central patterns of use, and a consideration of the natural environment of these patterns (Sinclair and Renouf, 1988: 148). All of this information is provided by the present research in relation to post-primary curriculum subjects. Instead of incorporating a grammar component, the realisation of a lexical syllabus in the classroom helps students to acquire grammatical knowledge in a bottom-up fashion, by examining the context-based

manifestations of lexical items; as Willis observes: 'listing words and their behaviour [...] generates automatically the structural environments and the words which are likely to occur with them' (1990: 52).

A paradox that might be noted here is that, although the process of lexical syllabus design entails the itemisation of language in discrete units, the methodology that is associated with its implementation in the classroom promotes holistic rather than piecemeal vocabulary acquisition (Sinclair and Renouf, 1988: 155). In his detailed discussion of a lexical syllabus, Willis (1990) explains how the lexical content of this type of syllabus can be mediated through authentic texts which exemplify the target language features. He describes how these texts should, in turn, lead to task-based classroom activities, which will essentially engage students in *meaningful use* of *real language* (ibid.: 124) for *true communication* (ibid.: 5; emphasis added). In the language support class, the above recommendations accordingly imply i) the use of authentic texts from subject textbooks and examination papers, ii) task-based activities which are firmly embedded in curriculum learning and which are described by the Benchmarks, and iii) a clear awareness of target language features which can be facilitated by the corpus analysis findings of this research.

Based on the above requirements of a lexical syllabus, it can be argued that the lexical descriptions of curriculum subjects, the information on their authentic contexts of use and their pragmatic meanings and functions in subject-specific texts provided by this research could be used to turn the Benchmarks into a lexically-based pedagogic agenda for language support teachers. The resulting lexical agenda could be further enhanced by i) a principled organisation of subject-specific lexis in semantic fields (to facilitate thematic vocabulary teaching and learning) and ii) information on the grammatical and syntactic environments of lexical features. These could be achieved by conducting semantic and grammatical analysis of subject-specific corpora using Wmatrix2 (Rayson, 2008) as this software offers automatic semantic (USAS; Rayson et al., 2004) and grammatical taggers (CLAWS7; Garside and Smith, 1997).

4.4.3 Enhancing the ELP language biography and dossier

Revising the subject-specific scales of the Benchmarks, automatically points to the need to update their mediation tool to ESL students, i.e. ELP (IILT, 2004) in order to provide students

with a more visible learning agenda. Accordingly, this section puts forward ways of exploiting the corpus-based lexical profiles of curriculum subjects to enhance the contents of the ELP.

Taking account of the ELP structure (see Appendix A2), corpus-based information on subject-specific language is relevant to the aims and contents of the language biography and the dossier sections (ibid.: 25-28). With regard to the language biography, corpus-based findings become relevant, first of all, to the section which invites students to record their learning targets (*What I need to learn in English - my learning targets*; ibid.: 19-20). The tasks of recording language that has been acquired and setting new language learning goals require conscious reflection on and principled organisation of linguistic knowledge. Providing ESL students with direct access to appropriately modified and user-friendly versions of the corpus-based language databases described in the previous section could considerably facilitate this task. The thematic, or topic-based, categorisation of lexical features in each database would be necessary to help students set content-based language learning targets.

Access to these same subject-specific language databases can be particularly useful for ESL students when they engage with the subject-specific 'I can' checklists of the biography (ibid.: 25-28). As with the 'can do' descriptors of the Benchmarks subject scales, the subject-specific 'I can' checklists in the language Biography describe communicative tasks that learners are expected to be able to perform in different content areas using the target language, e.g. *I can understand a short text if I know the main words and ideas* (English, Listening A2); *I can understand labels on scientific diagrams and equipment* (Science subjects, Reading A2), etc. (see Appendix A2). Contrary to what was proposed for the 'can do' descriptors of the Benchmarks scales, it might be of little value to attach corpus-based lexical descriptions to each of these 'I can' checklists. This is because in their present form, these functional statements can effectively serve their intended purpose to encourage learners to reflect upon and evaluate their learning outcomes and to set new language learning targets. As Little (2005) points out, 'can do' descriptors render the act of self-assessment feasible for all learners as even those with low levels of L2 proficiency are likely to know what they can or cannot do in the target language. Although there appears to be no need to attach corpus-derived lexical information to individual 'I can' statements, students could benefit considerably if they were presented with information on language features which are typically associated with these tasks. This subject-specific information could increase the transparency of language learning content and facilitate students' assessment of their own language proficiency. There is a growing body of research data in support of learners' ability to produce accurate and reliable judgements of language proficiency when they are provided with explicit language criteria (e.g. Oscarson, 1978; 1984; 1989; LeBlanc and Painchaud, 1985; Bachman and Palmer, 1986;

Blanche, 1990; Ross, 1998). On these grounds, it would be useful to insert appropriately modified and user-friendly versions of the aforementioned language databases for each subject in the biography section or to offer them as a separate booklet.

Corpus-derived language profiles of the different subjects can also inform the contents of the ELP dossier (Appendix A2). This section is intended to function as a personalised support for language learning in the content areas where ESL students record their vocabulary learning and organise lexical context in principled ways (e.g. by subject and theme) and in various formats (e.g. table, vocabulary tree, grid, wordlist, personal dictionary with words put into sample sentences, spider grams, etc.). Drawing upon the detailed language descriptions that have been made available by the present research, students could be encouraged to take control of the lexical content of their dossier on the basis of a tangible language learning agenda. If the empirical findings are offered to students in a meaningfully organised fashion, as suggested previously in the discussion (e.g. by theme, topic, functional utility, semantic networks among words, lexical linkages across different subject areas, etc.), students could become aware of different possible ways of grouping and presenting new knowledge and identifying relationships between lexical items.

For instance, the dossier could be divided into different sections devoted to curriculum subject areas and learners could compile an alphabetic index of subject-specific vocabulary in the corresponding sections. Sorting subject-specific vocabulary items by topics or semantic fields is another possibility whose usefulness lies in its relevance to a real world context (Lewis, 1993: 118). These collections of subject-based word items can be the object of comparison with peers, spelling tests in pairs offering opportunities for peer assessment, etc. Translations of terms in students' L1 might also be encouraged; considering that ESL students are teenagers, it follows that they probably attended school in their country of origin for many years, and thus finding L1 equivalents for words and concepts should help them to draw on knowledge they acquired in their first language. Drawing upon subject-specific language descriptions might also allow students to more easily detect and record lexical features and patterns they find particularly challenging and decide upon strategies to overcome vocabulary learning difficulties. Concordance lines could be usefully included in the dossier to demonstrate examples of authentic use for subject-specific and cross-curricular features. The inclusion of short text extracts (e.g. from subject textbooks) should also be encouraged to afford students a way of understanding the contextual uses of lexical features and 'the central function of lexis in structuring discourse' (Willis, 1990: 124).

In short, corpus-derived language descriptions provided to ESL students in an accessible form can be employed as a reference tool when they set their personal vocabulary learning

goals in the ELP biography (IILT, 2004: 10-28); when they assess language learning outcomes in the thematic 'I can' checklists; and when they record their lexical knowledge in the dossier (ibid.: 29-31). Being equipped with empirically-based and detailed language specifications, ESL students should more easily gain control over the lexical content of their target lexical repertoire and their ELP can become a personalised lexical syllabus for each student, supporting differentiated and individualised vocabulary development even beyond the language support classroom and independently of the teacher (i.e. autonomous vocabulary acquisition).

To conclude, it was the purpose of this section to discuss the potential that the corpus-based findings of this research have to enhance the effectiveness of the current ESL curriculum framework for post-primary language support. It was specifically argued that, mediating the empirical insights into subject language through the subject scales of the Benchmarks and even converting these scales into an explicit lexically-based pedagogic agenda could help teachers to make principled decisions. With regard to the ELP, it was proposed that, providing students with direct access to the subject-specific language descriptions of the present research when they engage with the ELP biography and dossier could have an empowering effect. This empowerment of students stems from developing a clear understanding of the lexical knowledge they need to acquire for content learning in the different subject areas. On the basis of this increased language awareness, students could more easily gain control of the lexical content of their language learning and consequently set learning targets more effectively and develop personalised strategies for vocabulary learning (see Little, 1997 for a discussion of language awareness and the autonomous language learner).

4.5 Contribution to materials development

An area where the empirical findings of this research can have direct application, and potentially a major impact, is the design and development of pedagogical materials for post-primary language support. In response to the repeated calls for context-specific and appropriate materials (cf. Chapter 1, section 1.3), the corpus-based linguistic findings of the present research can serve as a substantial basis for the development of resources and supports that are tailored to the language learning needs of post-primary ESL students.

Although materials can be developed in a number of forms, the focus here is primarily on written materials, which can be 'either specifically designed for language learning and

teaching (e.g. textbooks, worksheets, etc.); authentic materials [...] that have been specifically selected and exploited for teaching purposes by the classroom teacher; teacher-written materials; and learner-generated materials' (McGrath, 2002: 7). The discussion that follows refers first to the corpus-informed materials that were developed for post-primary language support by the ELSP project and then to a range of other materials which could be developed in the future based on the findings of the present research.

4.5.1 Corpus-informed materials developed by the ELSP project

One of the key deliverables of the ELSP of the TII was a substantial bank of pedagogical materials, responding to the lack of context-specific and appropriate resources for post-primary language support (cf. Chapter 1, section 1.3). The ELSP developed a website providing over 10,000 pages of corpus-informed language activities, firmly embedded in post-primary curriculum subjects and examinations (www.elsp.ie). These can be freely accessed by language support teachers in post-primary schools across the country.

Although the website was originally created for language support teachers to help them foster ESL students' language development, it soon proved to be useful for mainstream subject teachers for two purposes: i) to facilitate ESL students' curriculum access within the subject classroom and ii) to scaffold the subject-specific language learning of native speaking students with learning difficulties. The materials that support the development of the English language that is required for academic achievement in the different subject areas of the curriculum are the Language Support Activity Units.

The Activity Units are worksheets with language activities and exercises informed by the corpus analysis of Junior and Senior Cycle subject textbooks and extensive consultation with subject teachers. They are categorised according to curriculum subject and topic; a number of Activity Units are provided for each curriculum subject, focusing on the key topics studied in that subject. Thus, for Junior Cycle mathematics, for example, Activity Units are provided based on the following thematic strands: *arithmetic, algebra, set, indices and surds, area and volume, coordinate geometry, functions and quadratic graphs, graphic inequalities*, and *statistics*. The selected topics correspond to the units of subject textbooks; this means that the Activity Units reflect the content of the official subject syllabuses that post-primary subject teachers are expected to follow. Activity Units that focus exclusively on the language and requirements of examinations are also provided for each curriculum subject.

All the Activity Units follow the same format to provide teachers and students with a sense of structure and consistency. In the first page of each Unit, information is provided on i) the theme, ii) the level, iii) the language focus, iv) the learning focus, v) activity types, and vi) an acknowledgement of the source (i.e. subject textbook) of the worksheet. Next, a list of topic-based keywords is presented (divided into the four word classes of nouns, verbs, adjectives and adverbs), and a set of language exercises follows based on the given keywords. Exercises can be, for instance, gap-filling, matching, providing definitions, reading comprehension questions, labelling, as well exercises that require the use of subject textbooks used in the mainstream classroom, etc. As explained on the website, language exercises in these Units fall into three categories: i) simple ones which require letter and word manipulation, ii) some more in-depth exercises related to grammar and writing, and iii) others which have an element of fun (e.g. crosswords, snap, word chains, etc.). These exercises are scaled according to the first three CEFR levels (A1-B1) and several can be adapted to more than one levels. Many of the activities in these units are suitable as homework tasks, for self-study, or for use in the subject classroom with the agreement of the subject teacher. The answer key is provided at the end of each worksheet.

Additional materials which accompany the Activity Units include notes and some sample lesson plans for teachers (specifically for the Junior Certificate Activity Units of particular curriculum subject areas), with information on how the Units can best be used. For ESL students, the use of a Keyword list is recommended for reference when working on individual Units, and Learning and Vocabulary Files are provided where students can respectively record their learning progress and list newly acquired vocabulary items. These may well be incorporated into the student's ELP.

The website offers additional language activities in relation to the following six areas: i) learning to learn, ii) whole-book activities, iii) listening, iv) speaking, v) reading, and vi) writing, as well as some activities which are 'short, fun and interactive while remaining language and curricular content-focussed' (e.g. quizzes, crosswords, word chains, etc.; www.elsp.ie). Information is also offered on key words and phrases linked to safety rules (e.g. safety in the science classroom and the Wood-Metalworking rooms) which is derived from a corpus-linguistics analysis of various textbooks, classroom safety manuals and official guidance/directives.

The ELSP pedagogical resources in general, and the Language Support Activity Units in particular, were designed to help students 'penetrate and access textbooks and classroom activities' and empower them 'to build new knowledge on that which is already clear, comprehensible, and familiar to them in: learning support, language support, subject

classroom' (www.elsp.ie). The Activity Units constitute a clear manifestation of the invaluable role that corpus analysis can play in the development of materials which i) are context-specific, i.e. accurately mediate the most important lexical features of the major topics across the subjects of the Irish post-primary curriculum and ii) bring authenticity to the classroom, on the evidence that the language that is presented, the texts that are used and the tasks and activities they contain are characteristic of the subject classroom. Further, the reported usefulness of these materials for all post-primary students provides supporting evidence for the value of scaffolding the subject-specific language learning of native English speaking students underlined in the previous chapter (section 4.3.5). For its original contribution to post-primary language support, the ELSP website received the 2009 European Award for Languages.

The website materials were developed on the basis of the initial corpus analysis whereas the purpose of the present research was to take the analysis much further. A wide range of pedagogical materials could be developed for post-primary language support capitalizing on the empirical findings of the present research and which would be in line with the existing curriculum framework of the Benchmarks and ELP. These could include, on the one hand, materials whose design structure and content are underpinned by corpus findings and, on the other hand, resources which would engage students in hands-on language development work with 'raw' corpus data for data-driven learning (e.g. Johns, 1991). Some concrete suggestions are provided for both possibilities in sections 4.5.2 and 4.5.3 respectively.

4.5.2 Contribution to the development of additional context-specific and authentic materials

4.5.2.1 Ready-made corpus-based materials for the language support classroom

By creating new knowledge about the commonest patterns of lexis and its pragmatic meanings and functional characteristics in the textual contexts of two registers in each of six curriculum subjects, this research can, firstly, inform the design of ESL textbooks which can be developed to support ESL students' content-and-language-integrated learning. Corpus-informed ESL/EFL textbooks have relatively recently started to become available with the pioneering example being *Touchstone* (McCarthy, McCarten and Sandiford, 2004; see also Carter, Hughes and

McCarthy, 2000). This four-level textbook series for English language instruction includes real-life dialogues encompassing all the key features of real conversations (e.g. pauses, fillers, false starts, etc.), information on the frequency of language forms, collocations, and grammar structures and on the ways these are used in different contexts. Other examples can be found in ESL/EFL textbooks for university students that use the Academic Word List (AWL; Coxhead, 2000) in ready-made, corpus-informed material for vocabulary instruction (e.g. Schmitt and Schmitt, 2005). However, to the best of my knowledge, no corpus-informed textbooks for ESL/EFL for primary and secondary level education have been cited, and undoubtedly no corpus-based textbooks have been developed in the Irish primary and secondary context. It has been reported, in fact, that many of post-primary language support teachers use textbooks for general EFL purposes (Lyons and Little, 2009). On this evidence, the empirical perspectives on language across the post-primary curriculum provided by this research constitute an innovation as they offer the way forward to ESL textbooks development for post-primary language support.

Contrary to the newly emerged corpus-informed textbooks, corpus-informed dictionaries have been long established (e.g. Collins COBUILD English Language Dictionary; Sinclair, 1987; Longman Dictionary of Contemporary English, Oxford's Advanced Learner's Dictionary; MacMillan English Dictionary for Advanced Learners). Recognising the distinct nature of the academic and specialised vocabulary of the different curriculum subjects and its gate-keeping role to students' accessing subject-specific texts, it is argued that the corpus findings of this study could instigate the creation of subject-specific glossaries and dictionaries rooted in the Irish post-primary curriculum. These could contain information on all aspects of curriculum-specific vocabulary knowledge, i.e. form, meaning, use (Nation, 2001); e.g. word senses, pronunciation, collocations, grammatical properties, examples of authentic use etc., and they could be used as reference materials both by native and non native English speaking students. These reference tools would provide an obvious bridge between textbooks and the Benchmarks/ELP and ESL students in particular could consult them to inform their personalised dictionaries in the ELP dossier (cf. section 4.4.3; see also Chapter 3, section 3.2.3.3.1 where it was mentioned that a mathematics glossary for all students is recommended in post-primary teacher guidelines). Given their pervasiveness and pedagogical value, word clusters and their complementation patterns should be also included in the glossaries in question (cf. Sinclair's call for the compilation of a dictionary which captures the phraseology of language: 'A dictionary containing all the lexical items of a language, each one in its canonical form with a list of possible variations, would be the ultimate dictionary'; Sinclair in Sinclair et al., 2004: xxiv; Similarly, Henry and Roseberry (2001: 121) propose the compilation

of 'Language Pattern Dictionaries', i.e. genre-specific glossaries, based on their specialised corpus (i.e. English letters of application)).

Corpus-informed activities and exercises (free-standing or as part of ESL textbooks) specifically designed to support ESL students' vocabulary development across the curriculum could, undoubtedly, be one of the areas where this research could have a powerful impact. The richness of empirically-based information on the subject-specific vocabulary and lexis of subject textbooks and examinations papers made available by this research could be channelled into a wide range of vocabulary activities to promote direct and incidental learning. Space restrictions, however, forbid a detailed discussion of concrete suggestions (examples abound in the voluminous literature; e.g. Nation's work on vocabulary acquisition and learning); therefore only some fundamental considerations for developing vocabulary exercises that emerge from the present research are briefly considered here.

This research, first of all, points to the need for vocabulary tasks and activities which are underpinned by a comprehensive view of vocabulary, in the sense that they incorporate different types of lexical features instead of being confined to single words. For this purpose, subject-specific lexical words, collocations, clusters, as well as cross-curricular words and clusters should all be included in vocabulary exercises, to help students understand the role that these features play in language learning and use. Vocabulary exercises should be further built upon an understanding of the preferred lexical patterns of the different curriculum subjects and registers, as these have been made explicit in this research, to help students appreciate lexical variation and the role of context in making appropriate language choices. As Conrad (2004: 67) argues: 'ignoring this variation [in the language classroom] has undermined the effectiveness of teaching materials' in general.

Although it might appear that the aforementioned pedagogical resources do not differ greatly from some that may have been already used in language support, there would be significant differences in the content and the order of information presentation (Reppen, 2010: 14-15), and these could enhance the quality of language support. In short, these materials would essentially encompass the characteristics of corpus findings; that is, they would provide coverage of the most frequent language items, they would be lexically-oriented and content-based, while also taking account of subject and register requirements.

4.5.2.2 Working with 'raw' corpus data for data-driven learning

In addition to offering language support teachers ready-made corpus-based materials, the linguistic findings of the present research could be further exploited in more direct and novel ways in the language support classroom. More specifically, 'raw' corpus data could be made directly available to ESL students to be used mainly in two ways.

On the one hand, students can treat corpus data (e.g. frequency wordlists, node-collocate pairs and collocations, 4-word clusters, and concordance lines) as reference tools, to obtain 'examples and therefore clarify doubts on particular problems which had arisen in other language activities' (Aston, 1997). On the other hand, corpus data can be used as teaching materials; students can interact with corpus data for hands-on language development work which should be 'determined by a preselected goal' and 'progressive discoveries [will] occur on a negotiated step-by-step basis' (ibid.). For this second use of corpus data as a source of language activities, a number of suggestions have been put forward by the proponents of corpus-aided language pedagogy (e.g. Johns, 1988, 1991, 1994; Tribble and Jones, 1990; Cobb, 1997; Aston, 2001; Gavioli, 2001; Hunston, 2002; Reppen, 2010). Most of these include the use of frequency wordlists and concordances and could be introduced to the language support classroom (see also Thurston and Candlin's (1993) commercially produced workbook based on concordance print-outs).

Subject-specific frequency wordlists could be provided to ESL students to examine differences in the proportional use of lexical and function words, the role of prefixes and suffixes (word formation) in relation to word meaning (Reppen, 2010: 8), and the lexical variation across subject textbooks and similarities across examination papers. Printed versions of concordances can be exploited in numerous ways. They can inform activities such as deducing the meaning of lexical items from context, observing phenomena of collocation and colligation, studying homonyms and synonyms and word senses, idiomatic, metaphorical and technical uses, stylistic features, connotations (Tribble and Jones, 1990: 36-37; Johns, 1994). Concordance lines may well be used for gapfill exercises, matching exercises, in which the left or right contexts are jumbled and have to be re-instated, and remedial exercises based on learners' own writing (Tribble and Jones, 1990: 55). An additional possibility would be to use the concordance lines of cross-curricular lexical words and 4-word clusters provided in this research (cf. Chapter 3, section 3.3) to explore the multiple meanings of these words and the functions of clusters in different subject areas. To examine the discourse dimension of vocabulary, the role of words and clusters in text structure and organisation (Hoey, 1991;

McCarthy, 1991; cf. Nation, 2001: 206-216) could also be studied at the level of concordancing. Further, a lexical approach to pedagogical grammar teaching could be facilitated by concordances which, according to Bernardini, may even be, in some cases, more effective than grammars 'whose level of abstraction often works against their effectiveness' (2004: 17; see Carter and McCarthy's (1995) illustration - interaction - induction approach). In general, a wide range of activities are possible, located along a cline ranging from teacher-led to learner-led (Gilquin and Granger, 2010: 362-363).

It is important to note that, regardless of the type of activity, the language support teacher should be in control of the data students examine to ensure that the focus is placed on the intended target language feature for noticing (Schmidt, 1990) and consciousness-raising (Sharwood-Smith, 1981). Selecting data tailored to students' level and needs may be possible in other L2 contexts to ensure that students are provided with comprehensible input (Krashen, 1988) when they engage in 'raw' corpus work. In the pedagogical setting discussed here, however, there is no such possibility. It is the aim of language support to turn the language of the mainstream classroom (textbook, teacher talk, learning activity) into comprehensible input as ESL students have to grapple with it from the first instance. Every effort should thus be made, to ensure that the corpus-based language activities that are introduced reflect 'as far as possible the full range of linguistic and communicative features of the raw data' (Johns, 1994: 298).

Although corpus interrogation by students has also been proposed as a classroom activity (e.g. Dodd, 1997; Bernardini, 2000; see some possible activities in Hunston, 2002: 172-177), this might not be feasible for all post-primary ESL students. The feasibility and effectiveness of this pedagogical venture is questioned here on the grounds that it might prove to be too challenging both for several students and schools. As Hunston argues, this appears to be 'most suitable for more advanced learners who are filling in gaps in their knowledge rather than laying down the foundations' (2002: 171), as is the case with ESL students in the language support classroom. In addition, helping students (and teachers) to develop the knowledge and skills that are required for conducting corpus investigations effectively would require a lengthy and gradual apprenticeship over many years, which is not a realistic goal to achieve (with ESL students) since language support is provided for two years maximum. Beyond the training, insufficient computer access in Irish post-primary schools and teachers' inability to have consultations with individual students might be considered additional barriers to encouraging ESL students' corpus investigations (Hunston, 2002: 171). Despite these challenges, it should be nevertheless pointed out that, after ESL students have exhausted their allocation of

language support, they might still find it useful to be able to access a curriculum-focused corpus.

The preparation of corpus-based self-access materials, on the other hand, requires slightly less technological equipment but the downside is that this can be time-consuming for teachers (Gilquin and Granger, 2010: 336) and these materials may be viewed as peripheral to the main activities of the classroom (Hunston, 2002: 171). Nevertheless, developing and using these materials based on the corpus data of this research is strongly recommended as it can give rise to several benefits for ESL students.

One obvious advantage of corpus-based self-access resources is the authenticity they bring in the classroom in terms of the language and texts they contain (e.g. Gilquin and Granger, 2010: 359-360). Perhaps more importantly, hands-on corpus work promotes data-driven learning (DDL; Johns, 1986; 1988; 1991) or discovery learning. This helps ESL students develop valuable skills which can be transferred to other educational activities (e.g. subject learning in the mainstream); namely, hypothesis formation and testing and perceiving similarities and differences (Johns, 1991: 30-37), identifying patterns, drawing conclusions based on evidence and making valid generalisations. By examining real language data, students assume the role of researchers (Johns, 1991) and thus become actively involved in the learning process.

In terms of language development, DDL through corpus work has been reported to promote students' vocabulary expansion (e.g. Stevens, 1991; Cobb, 1997) and to improve their writing skills by exposing them to 'natural contextual learning' (Cobb, 1997: 314) and by facilitating error correction based on native speaker language use (Bernardini, 2004; Gaskell and Cobb, 2004; Chambers, 2005). In language support, the language of textbooks and examinations could model the language students need to use in classroom interaction, learning activities and examinations. These benefits appear to arise from the 'condensed language experience' (Gabrielatos, 2005) that corpus data can offer students; i.e. multiple encounters of the commonest lexical items in different contexts presented in a concentrated form can consolidate students' vocabulary knowledge. The aforementioned benefits for students' learning skills and language growth indicate that the use of corpus-based self-access materials can, in short, give rise to increased learning and language awareness (Bernardini, 2004: 31); both of which foster learner empowerment.

On these grounds, hands-on corpus activities seem to be accommodated more easily within a learner-centred classroom which is oriented to ESL students' language learner autonomy. According to Little (2007), this entails the simultaneous growth of target language proficiency and language learning skills. To achieve this growth, three principles need to be

operationalized in the classroom: active learning participation, critical reflection (on the content and process of learning), and appropriate target language use (ibid.: 1991). The analytical approach of corpus linguistics is an obvious way to stimulate students' incidental and conscious reflection on linguistic form. All the materials and corpus-based activities suggested here combined with the use of the ELP, enriched by the empirical findings of the present research could enhance ESL students' language awareness through critical reflection on and appropriate use of the target language.

Essential prerequisites for the successful implementation of the aforementioned corpus-based activities include teacher and learner training. Teachers need to adopt a descriptive rather than a prescriptive approach to language teaching when they engage students in work with corpus data. New skills and knowledge are also required to tackle the tasks of data interpretation and classification. Teachers and students need to use knowledge, reasoning and intuition about the language in order to 'move from an examination of the physical evidence to some generalisations that group the data into sets which seem to show identical or sufficiently similar patterning in their environments' (Sinclair, 2005: xv). This presupposes students who are active agents, and teachers who become facilitators, scaffolding students' hands-on corpus activities.

To conclude, it should be pointed out that the effectiveness of working with 'raw' corpus data and DDL in the language classroom has not been sufficiently-tested (Hunston, 2002: 170; Gilquin and Granger, 2010: 365). As mentioned above, pedagogical benefits have been reported by some small-scale studies (e.g. Stevens, 1991a; 1991b; Cobb, 1997; Cobb and Horst, 2001) but further empirical research is needed to validate this approach. This is another area where the present research has the potential to contribute. Implementing the above suggested corpus-based activities in the language support classroom with ESL students and evaluating their effectiveness could provide informative insights into their potential benefits and challenges. In any event, corpus-based activities are not put forward here as replacements for but rather as an enhancement of text-based work (Chambers, 2010: 353), or as an added value offered by corpus-aided discovery learning (Bernardini, 2004: 32-33).

4.6 Applications to language assessment for specific purposes

The corpus-based language descriptions of this research could have useful applications in the domain of assessment in post-primary language support. Corpora have been, in general, less

visible in language assessment, as opposed to other domains of L2 education. Alderson (1996) was the first to identify their potential uses to the different stages in the lifecycle of a test in theory (test construction, compilation and selection; test presentation; response capture; test scoring, and calculation and delivery of results). Today the major test services draw upon language corpora for a number of purposes; e.g. to enable item writers to base their examination tasks on real texts (Ball, 2001: 7) and to develop domain-specific wordlists for use in test materials (Ball, 2002; Horner and Strutt, 2004), among many others.

With specific regard to post-primary language support, corpus-based assessment instruments could be developed tailored to the needs of ESL students. The empirical findings of the present research could inform the construction of language tests for specific purposes (ESP testing), which would be in line with ESP instruction and learning. Douglas (2000: 19, emphasis added) defines a specific purpose test thus:

> one in which test *content* and *methods* are derived from an analysis of a *specific purpose target language use situation*, so that *test tasks* and *content* are *authentically representative* of tasks in the target situation, allowing for *an interaction between the test taker's language ability and specific purpose content knowledge*, on the one hand, *and the test tasks* on the other.

According to the above definition, ESP testing has two distinguishing features: first, test tasks and content directly reflect the nature and uses of language in the target domain, and second, students' language ability is viewed as interwoven with the content knowledge of the target language use situation. Although the first condition applies to language testing in general, the interaction between language and background, or topical (Bachman and Palmer, 1996), knowledge does not. Douglas explains that in general testing background knowledge is considered 'a confounding variable, contributing to measurement error and to be minimized as much as possible' whereas in ESP testing, 'background knowledge is a necessary, integral part of the concept of specific purpose language ability' (Douglas, 2000: 2).

Applied to language assessment in the context of post-primary language support, the above considerations point to the need for language tests that contain authentic curriculum-specific language and content-based language tasks that are typical of the subject classroom. Both subject-specific and cross-curricular lexical information can serve as a basis for selecting language items to be tested (i.e. test items) and authentic language to be used as test input. The selection of test tasks and exercises could draw upon the subject-specific scales of the Benchmarks (IILT, 2003a) as the 'can do' descriptors suggest communicative tasks that are typical of the subject classroom. Informed by the 'can do' descriptors of the Benchmarks and

corpus analysis results, the envisaged tests would contain tasks and language that would be 'authentically representative' of the content-based language use and activities of the target situation since both the Benchmarks and corpus-based findings derive from an 'analysis of [the] specific purpose target language use situation'. This implies that 'the topicality of the input' (in this case, subject-specific knowledge) will also be 'derived from the specific purpose language use situation' (Douglas, 2000: 60).

Taking account of the lexical nature of corpus-based data provided in this research study, it follows that these could effectively inform the content of ESP tests that focus specifically on ESL students' lexical knowledge across the six curriculum subjects analysed here and the two registers of Junior Cycle textbooks and Junior Certificate examination papers. ESP vocabulary tests could be constructed which would encompass all types of lexical items (words, collocates, collocations, clusters) and engage all aspects of vocabulary knowledge, that is, form, meaning and use (Nation, 2001). A wide variety of exercises can be designed inviting ESL students to identify, for instance, lexical features of different subjects, to explain the differences in the meaning of cross-curricular lexical words, to identify (in)appropriate word combinations, i.e. collocations (e.g. *mighty/feeble tea*instead of *strong/weak tea*), to select the appropriate complementation patterns of cross-curricular word clusters depending on the given subject-specific context, etc.

It is worth adding that collaboration between test developers and practitioners in the specialist area is generally recommended for specific purpose test development. Rea-Dickins, in fact, argues that this collaboration 'would seem to be a pre-requisite for the design of a "special purposes" test as the domains incorporated within the specialist area go beyond those in which the linguist - independently - is competent to make judgements' (1987: 196). In making this recommendation, however, she refers to English for doctors which can be highly specialised. This recommendation may not apply to language support teachers and subject teachers in post-primary education as the level of linguistic specialization is much lower.

It follows from the discussion in this section that the empirical findings of this research could inform the design and development of language tests tailored to the needs of post-primary language support. ESP tests that focus on ESL students' lexical knowledge across the curriculum may be developed for different purposes (e.g. diagnostic, achievement, proficiency, etc.). Underpinned by empirical findings, corpus-based tests would ensure the authenticity of language and tasks and their enhanced content validity. It is important to underline here that piloting would be essential to test the face validity of these tests. Corpus-informed language tests could add a further dimension to the Post-Primary Assessment Kit (IILT, 2009; downloadable from www.ncca.ie); a suite of Benchmarks-based language tests administered

mainly for placement purposes and for the measurement of proficiency (especially at the end of a student's two-year allocation of language support). To conclude, it is worth recalling that corpus analysis insights have the potential to enhance learner self-assessment of language achievement and proficiency based on the ELP biography (cf. section 4.4.3). It may therefore be argued here that the kind of corpus-informed language tests described above should be combined with ELP learner self-assessment of progress and proficiency for a more complete and holistic assessment in language support.

4.7 Conclusion

It has been the purpose of this chapter to discuss the pedagogical implications and applications of the empirical findings of the present research. The discussion showed that corpus-based information on curriculum languagecould be profitably incorporated into: i) ESL teaching and learning, ii) the Benchmarks and ELP, iii) materials development, and iv) language assessment.

More specifically, the results from subject-specific corpus analysis could inform the content of an explicit ESL teaching agenda (i.e. information on the commonest subject-specific and cross-curricular lexical features and their context-specific uses) and further facilitate an appropriate teaching methodology for language support (i.e. a lexical and discourse-based approach with a dual focus on content and language). Corpus evidence about the lexical variation that exists across the curriculum might also have a role to play in raising teacher awareness of the need for a whole-school approach to helping all students, native speakers as well as non-native speakers, to gain control of the subject-specific language varieties or 'dialects' of school education. As regards the current curriculum framework for post-primary language support (i.e. Benchmarks and ELP; IILT, 2003a; 2004), this could also be enriched by the linguistic findings of corpus analysis. Subject-specific lexical profiles could add further specificity to the functional descriptions of English language proficiency of the Benchmarks and ELP. In addition, they could be used to convert the Benchmarks into a rigorously-defined and lexically-based pedagogic agenda to promote a linguistically-explicit and principled approach to ESL teaching.

The possible applications of corpus findings to materials development were argued to be many and varied. The discussion first described the corpus-informed materials which were developed by the ELSP. It then suggested a number of other materials which could be designed based on the corpus-based language specifications of this research, and which could be easily

accommodated into the existing ESL curriculum framework of the Benchmarks and ELP (i.e. ESL textbooks, glossaries and dictionaries, and vocabulary activities). The use of raw corpus data (wordlists, collocations, clusters and concordance lines) as direct sources of language activities in the classroom was also recommended as a means of promoting DDL and learner empowerment. Finally, the design of ESP tests would be another domain that could greatly benefit from the lexical inventories made available by the present research. These tests would focus on ESL students' subject-specific lexical knowledge and could be administered for different purposes (e.g. diagnostic testing, achievement tests, proficiency tests, etc.).

To conclude, this chapter has addressed the second aim of the present research since numerous possible ways have been identified for post-primary language support to capitalize on corpus-based lexical profiles and descriptive insights into the language of different curriculum subjects. The discussion accordingly demonstrates the value of language corpora as useful sources for language description and for deriving pedagogically important information in any other context.

5 Conclusion

5.1 Aims of the research

Embedded within the ELSP of TII, the present research was conducted to contribute to the project's general aim, that is, to enhance the quality of English language support for immigrant students in Irish post-primary schools. This thesis sought to address a twofold aim: first, to analyse the language of Junior Cycle curriculum subjects, as this manifests itself in textbooks and Junior Certificate examination papers, and second, to discuss the pedagogical implications and possible applications of findings from this analysis in post-primary language support.

Following a critical overview of the language support policy in Irish post-primary education and the existing needs and challenges in Chapter 1, the first aim of the research was addressed in Chapters 2 and 3. Chapter 2 argued in favour of an applied corpus linguistics methodology for analysing the most frequent and consistently used language features of subject textbooks and examination papers on the grounds of its objectivity, accuracy and reliability. The quantitative and qualitative analysis of twelve subject-specific corpora of textbooks and examination papers, built specifically for the present research, was proposed using the lexical analysis software WordSmith Tools 4 (Scott, 2004). The design rationale and milestones in corpus development as well as the stages of the methodological procedure were discussed in detail.

The findings from the proposed empirical analysis were presented and analysed in Chapter 3. These findings paint a colourful picture of subject-specific language. In the first part of the discussion, detailed lexical profiles of the twelve corpora were drawn comprising the commonest words, the most significant collocations, and the most frequent 4-word clusters (together with a classification of cluster structures, their complementation patterns, and their meanings and functions). The second part of the discussion focused on lexical words and 4-word clusters which appear to recur across (six and five) textbooks and examinations corpora. Drawing on the collocates and complementation patterns of these features respectively, as well as on concordance lines, their semantic and functional characteristics in different contexts were compared and contrasted. Taken together, these findings demonstrate the lexical variation that is manifested across curriculum subjects and registers and the reciprocal relation of language use and the thematic and communicative priorities of context. In pedagogical

terms, corpus data represent important features of ESL students' target repertoire and they indicate the need for them to process and use fundamentally different sets of lexical features across the six subjects and in the two registers of textbooks and examinations.

Drawing on the corpus-derived language descriptions, the second aim of the research was addressed in Chapter 4 which explored their pedagogical implications and applications in relation to language support. In particular it was shown that empirical findings could contribute to the following aims:

- To inform principled decisions about the content of ESL learning and appropriate teaching approaches (frequency-based language learning and teaching, ESP with a dual focus on curriculum content and language, lexical and discourse-based approaches);
- To raise teachers' awareness of curriculum language, the demands it makes of students, its gate-keeping function as regards access to the curriculum, and encourage them to promote cross-curriculum language learning for all students;
- To enrich the existing ESL curriculum framework (IILT's Benchmarks (2003a) and ELP (2004)) and facilitate its conversion into an explicit pedagogic agenda formulated in a lexically-based form;
- To develop a wide range of context-specific and authentic materials, firmly embedded within Junior Cycle curriculum subjects and in line with the Benchmarks; and
- To inform the content of language tests for specific purposes and ensure the content validity and authenticity of these tests in terms of language, texts, and tasks.

5.2 Contribution

5.2.1 Providing an empirical understanding of curriculum language

The present research represents the first empirical analysis undertaken to inform post-primary language support for ESL students in Ireland, and it also constitutes the first cross-curriculum language analysis in the context of Irish post-primary education. One of its major contributions is, accordingly, the explicit understanding it provides of important language features and conventions in six Junior Cycle curriculum subjects. The value of the corpus-based language

descriptions that are provided in this thesis stems from the fact that i) they offer information that is empirically-derived and not based on intuition, ii) they identify language that is frequency-driven and filtered by range restrictions, iii) they characterise language from multiple perspectives (subject-specific words, collocations, clusters in each of the twelve corpora, and cross-curricular words and clusters), iv) they are representative of curriculum subjects that are important in post-primary education (English, geography, history, CSPE, mathematics, science), and v) they are, finally, representative of two fundamental school registers (textbooks and Junior Certificate examinations) which play an instrumental role in students' access to subject knowledge and academic achievement.

By offering the aforementioned curriculum language descriptions, this research can be viewed as a first step in addressing the lack of comprehensive information on the nature of language that students encounter in the Irish post-primary classroom. On these grounds, the research also responds to the repeated calls for a visible pedagogy in school education and the need to appreciate and document the fundamental role that curriculum language has in all students' academic achievement (e.g. the Bullock Report, 1975; Bernstein, 1971; 1990;Cope and Kalantzis, 1993; Halliday, 1993; Cummins, 1984; Schleppegrell, 2001; 2004; Bailey, 2007, etc.). By laying the foundations for a visible pedagogy, this empirical study should motivate further research in Irish post-primary education that also seeks to bridge the gap between rhetoric and practice and raise all students' standards of academic language use.

5.2.2 Enhancing post-primary language support and addressing challenges

On the basis of the empirical understanding of the language of curriculum subject textbooks and examinations it provides, this research can further contribute to addressing some of the major deficits, challenges and needs of language support provision. Corpus evidence serves to disprove misguided views of curriculum language as a monolithic, uniform register of general English. It further increases the transparency of some of the language demands that are imposed on ESL students by the different curriculum subjects, revealing important language features of students' target repertoire.

At a more practical level, if the linguistic data and descriptive insights that emerge from corpus analysis are incorporated into the different components of language support, as proposed in Chapter 4, the findings of this research could be used to address i) the lack of a rigorously defined teaching agenda, ii) the need for context-specific materials beyond those

developed by the ELSP, and iii) the lack of appropriate language tests for specific purposes tailored to ESL students' needs. As regards the implications of this research for the existing language support curriculum of the Benchmarks (IILT, 2003) and ELP (IITL, 2004), a more extended empirical research would be required to make any substantial recommendations beyond the suggestions made in Chapter 4 (section 4.4.1). More specifically, it would be necessary to extend the scope of corpus research to encompass a larger sample of texts from both Junior and Senior Cycle education as well as additional genres/text types both of the written and spoken medium.

A corpus-based language support pedagogy in which ESL teaching/learning, the Benchmarks and the ELP, materials and assessment instruments are all underpinned by the same corpus-based language specifications would result in a coherent and visible pedagogy. In other words, the empirical findings of the research could promote a linguistically-explicit, organised and principled language pedagogy, provided that they inform all aspects of language support provision.

5.2.3 Informing teacher education programmes

The present research might also have the potential to make a contribution to an area which has been very little affected by corpus linguistics, that is, teacher education. Both the corpus findings and the subject-specific corpora could be exploited as tools for professional development in training programmes for language support teachers and subject specialists alike, to promote teachers' 'pedagogic and linguistic awareness' (Farr, 2010: 621). Exploration by teachers of the subject-specific corpora generated by this research could inform them about the demands of curriculum language and raise their awareness of lexical variation across curriculum subjects. This increased awareness of curriculum language and its implications for teaching could help teachers improve their classroom practice and would point newly appointed teachers in the right direction. It might also convince subject teachers of the need to help all students, including native speakers, to gain control of the 'dialects' of the different curriculum subjects and develop an ability to move freely between them.

The findings of the present research could, moreover, help teachers develop awareness of the language they use for instructional purposes. Through conscious reflection on their personal use of language in the classroom stimulated by the analysis of teacher corpora, teachers could examine the extent to which they make use of the fundamentally different

lexical features and patterns of subject areas which are enacted in school textbooks. This suggestion is predicated on the assumption that teaching should be underpinned by the language of textbooks (cf. Chapter 2, section 2.2). Given that students need to access and use the academic language of textbooks, teacher talk could be informed by corpus evidence regarding the specialised lexical characteristics of different subjects. This might be one of many possible ways to enhance high quality oral input from the teacher and foster a lexically-rich classroom environment which is considered fundamental to students' learning of different registers (Gibbons, 2001; see also Hunston, 1995; Coniam, 1997; O'Keeffe and Farr, 2003; Chambers, 2005; Farr, 2008 on the use of corpora in teacher education contexts to raise language awareness).

The present research could further have a role to play in the development of teachers' pedagogic awareness through professional training programmes. This type of awareness can be associated with the knowledge and skills that are required in order to provide scaffolded mediation to students' learning and to engage them in language development work with corpus data. To this end, teachers could be first introduced to using corpus-based materials, like those described in Chapter 4 (section 4.5), and could be then shown how to manipulate the corpus data and conduct investigations of the corpora offered by the present research. This apprenticeship could help teachers develop the pedagogic knowledge they need in order to support their students' engagement with the same type of learning tasks.

The above recommendations might help language support teachers overcome their reported lack of knowledge and skills and respond to classroom reality, adopting differentiated methods, mediation skills and flexibility to support the needs of diverse students. In short, introducing corpora to teachers encourages them to become reflective practitioners and action researchers (e.g. Schön, 1983; Wallace, 1998); both are qualities which characterise autonomous language teachers who are capable of nurturing autonomous language learners (Little, 1995). A positive side-effect of this professional enhancement would be the improved classroom practices with native speaker students as well. Whole-school approaches could be promoted aiming to help all students learn how to: i) use language in specific contexts, ii) use language to learn through the curriculum, and iii) talk about, reflect on, and evaluate language (Bearne, 1999).

5.2.4 Contribution to applied corpus linguistics research and the current literature

At the level of research, the findings from the present study contribute to the current literature on applied corpus linguistics, ESP, and the status of vocabulary and lexis. The current findings add to a growing body of literature on the value of corpora as tools for domain-specific (especially academic) language descriptions and for deriving pedagogically-useful information (e.g. Johns, 1991; Conrad, 1996; Coxhead, 2000; Tognini-Bonelli, 2001; Bernardini, 2004; Biber, 2006; Hyland, 2008 etc.). As such, this research responds to the calls for i) more applied corpus linguistics studies motivated by a concern with language pedagogy (e.g. Flowerdew, 2010), ii) corpus research specifically into the language of secondary education to support students in their school-based studies and in preparation for further studies (e.g. Coxhead, 2010: 466; Flowerdew, 2010: 345), iii) localised ESP corpora to examine additional language varieties. By illustrating how corpus analysis could make a contribution to the needs of post-primary language support in Ireland, the present research echos Bernardini's suggestion that we need to 'rethink language pedagogy [and] language description [...] from a corpus perspective' (2004: 16). In addition, although the study is context-specific, tied to Irish post-primary education, the research design and methodology it employs should be viewed as immediately useful and applicable to any other context of migrant education with similar pedagogical needs.

As regards its contribution to the current literature on ESP and vocabulary and lexis studies, the research confirms previous findings supporting i) the interdependence of language use and the thematic and communicative priorities of particular contexts (e.g. Hutchinson and Waters, 1987; Naggy, 1997 etc.) and ii) the variation of language across disciplines and registers (e.g. Conrad, 1996; Biber et al., 2004; Hyland, 2008; 2010, etc.), demonstrating that each curriculum subject operates its own 'sub-language' (Rosen, 1972). The results from the present corpus analysis also contribute additional evidence about the manifestation of the idiom principle (Sinclair, 1991), specifically in subject-specific texts used in Irish post-primary education. The operation of the idiom principle is illustrated by the recycling of the commonest lexical features (words and collocations) within or adjacent to the commonest word patterns (4-word clusters), revealing the lexical patterning that underpins subject textbooks and particularly examination papers. The research also corroborates the findings of studies which reveal particular facts about the status of different vocabulary items: i) the topic-specificity of lexical (as opposed to function) words (e.g. Corson, 1985; Nation, 2001), ii) the contextual variation of significant collocations (e.g. Gledhill, 2000), and iii) the role of 4-word

clusters as reliable indicators of language variation (e.g. Biber and Barbieri, 2007) and as fundamental building discourse blocks (e.g. Biber et al., 1999).

5.3 Limitations and scope for further research

The research undertaken for the present study could be extended in several ways to overcome its limitations, some of which are inherent in all applied corpus linguistics studies. First of all, the scope of corpus analysis could be profitably extended in different ways by increasing, for example, i) the number of corpora (more corpora for more post-primary subjects from Junior Cycle and Senior Cycles, including corpora for Leaving Certificate examinations), ii) the size of subject-specific corpora (including more texts in each corpus), and iii) the sample of the target language (including additional genres used in the subject classroom, such as workbooks, classroom handouts, etc.).

The scope of empirical analysis should also be extended from frequency to keyness analysis and grammatical/syntactic and semantic analysis of tagged corpora using linguistic annotation and additional software programs (e.g. Wmatrix 2; Rayson, 2008 and the Sketch Engine; Kilgarriff et al., 2004). Future analysis of tagged corpora could also facilitate the word sense and part-of-speech disambiguation which was not possible in this research. The examination of additional aspects of subject-specific language would evidently result in comprehensive and more holistic descriptions of curriculum language. More concordances would be further needed for all the features that are analysed. In addition, the use of corpus linguistics techniques and tools should be combined with discourse analytic approaches to ESP analysis, capitalising on the fruitful interface of corpus linguistics with SFL, text linguistics, and discourse analysis which has been so far exploited to a limited extent (e.g. Thompson and Hunston, 2005; Biber, Connor and Upton, 2007; Charles, Pecorari and Hunston, 2009). This integration could provide complementary linguistic perspectives and could address the criticisms that have been levelled against corpus linguistics, e.g. decontextualisation of language (Widdowson, 1998; 2002) and a lack of focus on large pieces of texts (Swales, 2002).

Information from the analysis of learner and teacher corpora could also be of great value for language support. More specifically, learner corpora, comprising ESL students' written and spoken language, would be a useful source of information on the interlanguage of different students to inform the design of tailor-made pedagogical materials which would

benefit students with the same L1 (Granger, 2009: 20). Similarly, multi-lingual corpora could be built 'to provide insights into how different languages express similar academic concepts in different ways' and 'how [subject areas] employ words differently' (Coxhead, 2010: 467). Teacher corpora, comprising teacher talk and oral interaction in the classroom, could serve as a springboard for reflection in teacher education programmers (cf. section 5.2.3), helping teachers develop critical language awareness skills and modify oral input for students. The construction of multimodal corpora (e.g. Knight et al., 2009), which would include annotated videos of classroom interaction, might be another possibility in order to analyse the communicative and linguistic demands of the subject classroom and to overcome the decontextualisation of written text language.

Another limitation of this research that is inherent in all corpus studies concerns the generalisability of findings. The corpus analysis results documented in this thesis cannot be applied to the language under analysis as a whole; results are rather representative of the particular corpora that were examined here. As Sinclair argues: 'no corpus, no matter how large, how carefully designed, can have exactly the same characteristics as the language itself' (2005: 2). A final limitation can be found in the identification of cross-curricular words and clusters which was facilitated by the Detailed Consistency Analysis of WordSmith 4 in this research study. As explained in Chapter 3 (section 3.3), this functionality of WordSmith 4 is not underpinned by any test of statistical significance; thus the results may not be entirely representative of genuine similarities in words and clusters - future research might employ a more reliable methodology for corpus comparison. Notwithstanding these limitations, the empirical findings of this research consistute a substantial amount of pedagogically useful information that could be utilised in numerous ways to establish direct links between English language support and the mainstream subject classroom.

5.4 Concluding remarks

Although many questions about the language of Irish post-primary curriculum remain, it is hoped that this research has provided valid and pedagogically-useful insights into the language of Junior Cycle textbooks and Junior Certificate examination papers. It is also hoped that this thesis will inspire further empirical research designed to characterise curriculum language and extend the current methodological framework of corpus analysis. Moving beyond the Irish context and considering that migrant education is, and will remain, a common concern across

Europe, the present research may be immediately relevant and applicable to other educational contexts with similar pedagogical needs.

To conclude, it should be pointed out that, in addition to its methodological rigour and its contribution to academic research and pedagogical practice, the value of an applied corpus linguistics research study should be evaluated 'ultimately on the basis of its potential for positive impact on societal and educational problems' (Ortega, 2005: 430, cited in Dornyei, 2007: 277) since research in applied linguistics is inherently a social activity. It is therefore my hope that the present research could contribute to addressing both educational and societal challenges in the post-primary language support classroom, and make a difference by promoting teacher empowerment, learner empowerment and equality of education for all.

References

Adolphs, S. (2006). *Introducing electronic text analysis: a practical guide for language and literary studies*. Abingdon: Routledge.

Aitchison, J. (1987). *Words in the mind: an introduction to the mental lexicon.* Oxford: Basil Blackwell.

Aitchison, J. (1994). *Words in the mind*. (2nd Ed.). Oxford: Blackwell.

Alderson, J. C. (1996). Do corpora have a role in language assessment?. In J. A. Thomas & M. H. Short (Eds.), *Using Corpora for Language Research*. London: Longman.

Allwood, J. (2008). Multimodal Corpora. In A. Lüdeling & M. Kytö (Eds.), *Corpus linguistics: an international handbook.* (Vol. 1). Berlin: Mouton de Gruyter.

Aston, G. (1997). Enriching the learning environment: corpora in ELT. In A. Wichmann, S. Fingelstone, T. McEnery & G. Knowles (Eds.), *Teaching and language corpora*. London: Longman.

Aston, G. (Ed.) (2001). *Learning with corpora*. Houston, TX: Athelstan.

Bachman, L. F., & Palmer, A. S. (1996). *Language testing in practice*. Oxford: Oxford University Press.

Bailey, A. (Ed.) (2007). *The language demands of school: putting academic English to the test*. New Haven, CT: Yale University Press.

Baker, P. (2006). *Using corpora in discourse analysis*. London: Continuum.

Baker, P., Hardie, A., & McEnery, T. (2006). *A glossary of corpus linguistics*. Edinburgh: Edinburgh University Press.

Ball, F. (2001). Using corpora in language testing. *Research notes*, *6*, 6-8.

Ball, F. (2002). Developing wordlists for BEC. *Research Notes*, *8,* 10-13.

Barber, C. L. (1962). Some measurable characteristics of modern scientific prose. In J. Swales (Ed.) (1988), *Episodes in ESP*. Cambridge: Cambridge University Press.

Barker, F. (2004). Using corpora in language testing: research and validation of language. *Modern English Teacher*, *13* (2), 63–67.

Barnbrook, G. (1996). *Language and computers: a practical introduction to the computer analysis of language*. Edinburgh: Edinburgh University Press.

Barnes, D. (1972). Language and learning in the classroom. In *Language in education: a source book*. London: Kegan Paul in association with The Open University Press.

Barrett, J., & Richardson, F. (2000). *Impact!: Civil, social and political education for Junior Certificate*. Dublin: Gill and Macmillan.

Bartsch, S. (2004). *Structural and functional properties of collocations in English*. Tübingen: Narr.

Bazerman, C. (1994). Systems of genres and the enactment of social intentions. In A.Freedman & P. Medway (Eds.), *Genre and the new rhetoric*. London:Taylor & Francis.

Bearne, E. (Ed.) (1998). *Use of language across the primary curriculum*. London: Routledge.

Bearne, E. (Ed.) (1999). *Use of language across the secondary curriculum*. London: Routledge.

Benson, M. (1985). Collocations and idioms. In R. Ilson (Ed.), *Dictionaries, lexicography and language learning*. Oxford: Pergamon.

Berkenkotter, C., & Huckin, T. N. (1993). Rethinking genre from a sociocognitive perspective. *Written Communication*, 4, 475-509.

Bernardini, S. (2000). *Competence, capacity, corpora: a study in corpus-aided language learning.* Bologna: CLUEB.

Bernardini, S. (2004). Corpora in the classroom: an overview of some reflections of future developments. In J. Sinclair (Ed.), *How to use corpora in language teaching*. Amsterdam: J. Benjamins.

Bernstein, B. (1971). *Class, codes and control.* (Vol. 1). London: Routledge and Kegan Paul.

Bernstein, B. (1973). *Class, codes and control: applied studies towards a sociology of language.* (Vol. 2). London: Routledge and Kegan Paul.

Bernstein, B. (1977). *Class, codes and control: towards a theory of educational transmissions.* (Vol. 3). London: Routledge and Kegan Paul.

Bernstein, B. (1990). *The structuring of pedagogic discourse*. London: Routledge.

Bernstein, B. (1996). *Pedagogy, symbolic control, and identity: theory, research, critique*. London: Taylor & Francis.

Bernstein, B. (1999). Horizontal and vertical discourse: an essay. *British Journal of Sociology of Education, 20* (2), 157-173.

Berry-Rogghe, G. L. M. (1973). The computation of collocations and their relevance to lexical studies. In A. Aiken, R. Bailey & N. Hamilton-Smith (Eds.), *The computer and literary studies*. Edinburgh: Edinburgh University Press.

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

Bhatia, V. K. (2004). *Worlds of written discourse*. London: Continuum.

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

Biber, D. (1993). Representativeness in corpus design. *Literary and Linguistic Computing, 8*, 1-15.

Biber, D. (2006). *University language: a corpus-based study of spoken and written registers*. Amsterdam; Philadelphia: J. Benjamins.

Biber, D., & Barbieri, F. (2007). Lexical bundles in university spoken and written registers. *English for Specific Purposes, 26*, 263-286.

Biber, D., & Conrad, S. (1999). Lexical bundles in conversation and academic prose. In H. Hasselgard & S. Oksefjell (Eds.), *Out of corpora: Studies in honour of Stig Johansson*. Amsterdam: Rodopi.

Biber, D., & Finegan, E. (Eds.) (1994). *Sociolinguistic perspectives on register*. New York; Oxford: Oxford University Press.

Biber, D., Connor, U., & Upton, T. (2007). *Discourse on the move: using corpus analysis to describe discourse structure*. Amsterdam; Philadelphia: John Benjamins.

Biber, D., Conrad, S., & Cortes, V. (2004). If you look at... : Lexical bundles in university teaching and textbooks. *Applied Linguistics, 25*, 371-405.

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

Biber, D., Johansson, S., Leech, G., Conrad, S., & Finegan, E. (1999). *Longman grammar of spoken and written English*. London: Longman.

Blanche, P. (1990). Using standardized achievement and oral proficiency tests for self-assessment purposes: The DLIFLC study. *Language Testing, 7* (2), 202-229.

Bowker, L. (1996). Towards a corpus-based approach to terminography. *Terminology, 3* (1), 27-52.

Bowker, L., & Pearson, J. (2002). *Working with specialized language: a practical guide to using corpora*. London: Routledge.

Brinton, D. M., Snow, M. A., & Wesche, M. B. (1989). *Content-based second language instruction*. New York: Harper/Newbury.

Brinton, D., Snow, M. A., & Wesche, M. (2003). *Content-based second language instruction.* (Revised Edition). Michigan: University of Michigan Press.

Carter, R., Hughes, R. and McCarthy, M. (2000). *Exploring grammar in context: grammar reference and practice*. Cambridge: Cambridge University Press.

Cassidy, C., & Kingston, P. (2004). *We are the world: Civil, social and political education*. Dublin: Mentor.

Celce-Murcia, M., & Olshtain, E. (2000). *Discourse and context in language teaching: a guide for language teachers*. Cambridge: Cambridge University Press.

Chambers, A. (2005). Integrating corpus consultation in language studies. *Language learning and Technology, 9* (2), 111-125.

Chambers, A. (2010). What is data-driven learning? In A. O'Keeffe & M. McCarthy (Eds.), *The Routledge handbook of corpus linguistics*. London: Routledge.

Chamot, A. U., & O'Malley, J. M. (1994). *The CALLA Handbook: how to implement the cognitive academic language learning approach.* Reading, MA: Addison Wesley.

Chamot, A. U., &O'Malley, J. M. (1987). A cognitive academic language learning approach: a bridge to the mainstream. *TESOL Quarterly, 21*, 227-49.

Charles, M., Pecorari, D. and Hunston, S. (Eds.) (2009). *Academic writing: at the interface of corpus and discourse*. London: Continuum.

Christie, F. (1999). *Pedagogy and the shaping of consciousness: linguistic and social processes*. London: Cassell.

Christie, F. (2002). *Classroom discourse analysis: a functional perspective*. London: Continuum.

Christie, F., & Misson, R. (1998). *Literacy and schooling*. London: Routledge.

Christie, F., J. R. Martin & J. Rothery (1991). Teaching functional grammar. In *Teaching English Literacy: a project of national significance on the preservice preparation of teachers for teaching English literacy* Volume 2: Papers.

Church, K. W., & Hanks, P. (1990). Word association norms, mutual information, and lexicography. *Computational Linguistics, 16* (1), 22-29.

Church, K. W., Gale, W., Hanks, P., & Hindle, D. (1991). Using statistics in lexical analysis. In U. Zernik (Ed.), *Lexical acquisition: using on-line resources to build a lexicon*. Hillsdale, NJ: Lawrence Erlbaum Associates.

Clear, J. (1993). From Firth principles: computational tools for the study of collocation. In M. Baker, G. Francis & E. Tognini-Bonelli (Eds.), *Text and technology: in honour of John Sinclair*. Amsterdam: Benjamins.

Cobb, T. (1997). Is there any measurable learning from hands-on concordancing? *System, 25* (3), 301-315.

Cobb, T., & Horst, M. (2001). Reading academic English: carrying learners across the lexical threshold. In J. Flowerdew & M. Peacock (Eds.) *The English for academic purposes curriculum*. Cambridge: Cambridge University Press.

Coffin, C. (2006). *Historical discourse: the language of time, cause and evaluation*. London: Continuum.

Coleman, K. (2007). *Shortcuts to success: Science experiments for Junior Certificate*. Dublin: Gill and Macmillan.

Collier, V. (1987). Age and rate of acquisition of second language for academic purposes. *TESOL Quarterly, 21*, 617-641.

Collier, V. (1989). How long? A synthesis of research on academic achievement in a second language. *TESOL Quarterly*, *23* (3), 509-531.

Coniam, D. (1997). A practical introduction to corpora in a teacher training language awareness programme. *Language Awareness*, *6*, 199-207.

Conrad, S. M. (1996). Investigating academic texts with corpus-based techniques: An example from biology. *Linguistics and Education*, *8*, 299-326.

Conrad, S. M. (2004).Corpus linguistics, language variation, and language teaching. In J. M. Sinclair (Ed.), *How to use corpora in language teaching*. Amsterdam: J. Benjamins (Studies in corpus linguistics 12).

Cope, B. & M. Kalantzis (Eds.) (1993). *The powers of literacy: α genre approach to teaching writing.*London: Falmer Press.

Corson, D. (1985). *The lexical bar*. Oxford: Pergamon.

Cortes, V. (2004). Lexical bundles in published and student disciplinary writing: examples from history and biology. *English for Specific Purposes*, *23*, 397-423.

Cotter, L. (2007). *Less stress, more success: English revision for Junior Certificate - Higher level.* Dublin: Gill and Macmillan.

Council of Europe. (2001). *Common European Framework of Reference for Languages: Learning, teaching, assessment.* Cambridge: Cambridge University Press.

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

Coxhead, A. (2010). What can corpora tell us about English for Academic Purposes? In M. McCarthy & A. O'Keeffe (Eds.), *The Routledge handbook of corpus linguistics*. London: Routledge.

Crystal, D. (2003). *The Cambridge encyclopedia of language* (2nd Ed.). Cambridge: Cambridge University Press.

Cullen, G. (2003). *Discovering science*. Dublin: Mentor.

Cummins, J. (1979). Cognitive academic language proficiency, linguistic interdependence, the optimum age questions and some other matters. *Working Papers on Bilingualism, 19*, 121-129.

Cummins, J. (1981). Age on arrival and immigrant second language learning in Canada: a reassessment. *Applied Linguistics*, *2*, 131-149.

Cummins, J. (1984). Wanted: a theoretical framework for relating language proficiency to academic achievement among bilingual students. In C. Rivera (Ed.), *Language proficiency and academic achievement*. Clevedon: Multilingual Matters.

Cummins, J. (2000). *Language, power and pedagogy: bilingual children in the crossfire*. Clevedon: Multilingual Matters.

Cummins, J., & Swain, M. (1986). *Bilingualism in education: aspects of theory, research and practice*. London: Longman.

Darmody, M. (2007). Strengthening the school social climate. In P. Downes & A. L. Gilligan (Eds.), *Beyond educational disadvantage*. Dublin: Institute of Public Administration.

Davison, C., & Williams, A. (2001). Integrating language and content: unresolved issues. In B. Mohan, C. Leung, & C. Davison (Eds.), *English as a second language in the mainstream*. New York: Longman-Pearson.

Dennis, S. F. (1965). The construction of a thesaurus automatically from a sample of text. In M. E. G. Stevens, V. E. Heilprin & B. Lawrence (Eds.), *Proceedings of the Symposium on statistical association methods for mechanized documentation*. Washington: National Bureau of Standards (National Bureau of Standards Miscellaneous Publication 269).

DES. (2006). *Looking at English: teaching and learning English in post-primary schools*. Dublin: Stationery Office.

DES. (2008). *Language education policy profile: country report - Ireland.* Strasbourg: Council of Europe.

DES. (2009). *Meeting the needs of pupils for whom English is a second language* (Circular 0015/2009). Dublin: DES.

Devine, D. (2005). 'Welcome to the Celtic Tiger?' - Teacher responses to immigration and increasing ethnic diversity in Irish schools. *International Studies in Sociology of Education, 15*, 49-71.

Devine, D., & Kelly, M. (2006). 'I just don't want to get picked on by anybody': dynamics of inclusion and exclusion in a newly multi-ethnic Irish primary school. *Children and Society, 20*, 128-139.

Devine, D., Kenny, M., & McNeela, E. (2002). *Ethnicity and schooling: a study of ethnic diversity in selected Irish primary and post-primary schools. Report for the Department of Education and Science*. Dublin: University College Dublin.

Devine, D., Kenny, M., & McNeela, E. (2004). Experiencing racism in the primary school - children's perspectives. In J. Deegan, D. Devine & A. Lodge (Eds.), *Primary voices - equality, diversity and childhood in Irish primary schools*. Dublin: Institute of Public Administration.

Devine, D., Kenny, M., & McNeela, E. (2008). *Naming the 'other' – Children's construction and experience of racisms in Irish primary schools. Race Ethnicity and Education, 11*(4), 369-385.

Dodd, B. (1997). Exploiting a corpus of written German for advanced language learning. In A. Wichmann, S. Fligelstone, T. McEnery, & G. Knowles (Eds.), *Teaching and language corpora*. London: Longman.

Douglas, D. (2000). *Assessing language for specific purposes*. Cambridge: Cambridge University Press.

Dunning, T. (1993). Accurate methods for the statistics of surprise and coincidence. *Computational Linguistics, 19* (1), 61-74.

Eurydice. (2004). *Integrating immigrant children into schools in Europe*. Brussels: Eurydice.

Evert, S. (2005). *The statistics of word co-occurrences: word pairs and collocations.* Ph.D. Thesis, University of Stuttgart.

Evert, S. (2008). Corpora and collocations. In A. Lüdeling & M. Kytö (Eds.), *Corpus linguistics: an international handbook.* (Vol. 2). Berlin: Mouton de Gruyter.

Farr, F. (2008). Evaluating the use of corpus-based instruction in a language teacher education context: perspectives from the users. *Language Awareness, 17* (1), 25-45.

Farr, F. (2010). How can corpora be used in teacher education. In M. McCarthy & A. K'Keeffe (Eds.), *The Routledge handbook of corpus linguistics*. London: Routledge.

Fionda, R. (2011). *English language support in Irish post-primary schools: an empirical exploration of policy and practice*. Unpublished Ph.D. Thesis, University of Dublin, Trinity College.

Firth, J. R. (1957). *A synopsis of linguistic theory, 1930-1955. Studies in linguistic analysis: special volume of the Philological Society*. Oxford: Blackwell.

Fitzharris, T. (2008). *Junior Certificate rapid revision geography*. Dublin: Folens.

Flowerdew, J. (1993). Concordancing as a tool in course design. *System, 21* (2), 231-244.

Flowerdew, J. (Ed.) (2002). *Academic discourse*. Harlow: Longman.

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.

Flowerdew, L. (2009). Applying corpus linguistics to pedagogy: a critical evaluation. *International Journal of Corpus Linguistics, 14* (3), 393-341.

Flowerdew, L. (2010). Using corpora for language instruction. In M. McCarthy & A. O'Keeffe (Eds.), *The Routledge handbook of corpus linguistics*. London: Routledge.

Foucault, M. (1972). *The archaeology of knowledge*. London: Tavistock.

Francis, W. N., & Kučera, H. (1982). *Frequency analysis of English usage*. Boston: Houghton Mifflin Company.

Fraser, B. (1970). Idioms within transformational grammar. *Foundations of Language, 6* (1), 22-42.

Gabrielatos, C. (2005). Corpora and language teaching: just a fling, or wedding bells? *TESL-EJ, 8* (4), 1-37.

Gairns, R., & Redman, S. (1986). *Working with words: a guide to teaching and learning vocabulary*. Cambridge: Cambridge University Press.

Garside, R., & Smith, N. (1997). A hybrid grammatical tagger: CLAWS4. In R. Garside, G. Leech & A. McEnery (Eds.), *Corpus annotation: linguistic information from computer text corpora*. London: Longman.

Gaskell, D. and Cobb, T. (2004). Can learners use concordance feedback for writing errors? *System 32*, 301-319.

Gavioli, L. (2001). The learner as researcher: Introducing corpus concordancing in the classroom. In G. Aston (Ed.), *Learning with corpora*. Houston, TX: Athelstan.

Gibbons, P. (2001). Learning a new register in a second language. In C. Candlin & N. Mercer (Eds.), *English language teaching in its social context: a reader*. London: Routledge.

Gilquin, G., & Granger, S. (2010). How can data-driven learning be used in language teaching? In A. O' Keeffe & M. McCarthy (Eds.), *The Routledge handbook of corpus linguistics*. London: Routledge.

Gledhill, C. (2000). The discourse function of collocation in research article introductions. *English for Specific Purposes, 19*, 115-135.

Grabe, W., & Kaplan, R. B. (1996). *Theory and practice of writing: an applied linguistic perspective*. New York: Longman.

Granger, S. (Ed.) (1998). *Learner English on the computer*. London; New York: Addison Wesley.

Granger, S. (2009). The contribution of learner corpora to second language acquisition and foreign language teaching: a critical evaluation. In K. Aijmer (Ed.), *Corpora and language teaching*. Amsterdam & Philadelphia: John Benjamins.

Haastrup, K., & Henriksen, B. (2000). Voluntary acquisition: acquiring depth of knowledge through network building. *International Journal of Applied Linguistics, 10* (2), 221-240.

Halliday, M. A. K. (1988). On the language of physical science. In M. Ghadessy (Ed.), *Registers of written English: Situational factors and linguistic features*. London: Pinter, Open Linguistics Series.

Halliday, M. A. K. (1993). *Language in a changing world*. ACT: Applied Linguistics Association of Australia (Occasional Paper 13).

Halliday, M. A. K. (1994). *An introduction to functional grammar*. (2nd Ed.). London: Edward Arnold.

Halliday, M. A. K., & Hasan, R. (1989). *Language, context, and text: aspects of language in a social-semiotic perspective*. (2nd Ed.). Oxford: Oxford University Press.

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

Halliday, M. A. K., McIntosh, A., & Strevens, P. (1964). *The linguistic sciences and language teaching*. London: Longman.

Halliday, M.A.K. (1988). On the language of physical science. In M. Ghadessy (Ed.), *Registers of written English: situational factors and linguistic features*. London: Pinter, Open Linguistics Series.

Halpin, M. (2005). *Shortcuts to success: Maths - Junior Certificate Higher Level*. Dublin: Gill and Macmillan.

Hannan, D. F., Smyth, E., McCoullagh, J., O'Leary, R., & McMahon, D. (1996). *Coeducation and gender equality: exam performance, stress and personal development*. Dublin: Oak Tree Press / ESRI.

Harrison, C., & Wilson, M. (2001). *Make a difference!: Civil, social and political education for Junior Certificate*. Dublin: Folens.

Hasan, R. (1996). *Ways of saying: ways of meaning: selected papers of Ruqaiya Hasan*. London: Cassell.

Hasan, R., & Williams, G. (Eds.) (1996). *Literacy in society*. London: Longman.

Hayes, C. (2006). *Shortcuts to success: History Junior Certificate*. Dublin: Gill and Macmillan.

Henry, A. and R. L. Roseberry (2001). Using a small corpus to obtain data for teaching a genre. In Ghadessy, M., Henry, A. and Roseberry, R. L. (Eds.), *Small corpus studies and ELT: theory and practice*. Amsterdam: J. Benjamins.

Hoey, M. (1991). *Patterns of lexis in texts*. Oxford: Oxford University Press.

Hoey, M. (2005). *Lexical priming: a new theory of words and language*. Abingdon; New York: Routledge.

Horner, D., & Strutt, P. (2004). Analyzing domain-specific lexical categories: evidence from the BEC written corpus. *Research Notes, 15,* 6-8.

Howarth, P. (1998). Phraseology and second language proficiency. *Applied Linguistics, 19* (1), 24-44.

Humphrey, G. (2001). *New Concise Maths 1*. Dublin: Gill and Macmillan.

Humphrey, G. (2001). *New Concise Maths 2*. Dublin: Gill and Macmillan.

Hunston, S. (1995). Grammar in teacher education: the role of a corpus. *Language Awareness, 4* (1), 15-31.

Hunston, S. (2002). *Corpora in applied linguistics*. Cambridge: Cambridge University Press.

Hunston, S., & Francis, G. (2000). *Pattern grammar: a corpus-driven approach to the lexical grammar of English*. Amsterdam: John Benjamins.

Hunston, S., & Sinclair, J. (2000). A local grammar of evaluation. In S. Hunston & G. Thompson (Eds.), *Evaluation in text: authorial stance and the construction of discourse*. Oxford: Oxford University Press.

Hutchinson, T., & Waters, A. (1987). *English for specific purposes: a learning-centred approach*. Cambridge: Cambridge University Press.

Hyland, K. & Tse, P. (2009). Academic lexis and disciplinary practice: corpus evidence for specificity. *International Journal of English Studies, 9* (2), 111-130.

Hyland, K. (2008). As can be seen: lexical bundles and disciplinary variation. *English for Specific Purposes, 27* (1), 4-21.

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

IILT. (2003a). *English Language Proficiency Benchmarks for non-English-speaking students at post-primary level*. Dublin: IILT.

IILT. (2003b). *Immersion programme for non-English speaking students entering post-primary school: A framework for school implementation*. Dublin: IILT.

IILT. (2003c). *Integrating non-English-speaking students into the school and curriculum: handbook for schools*. Dublin: IILT.

IILT. (2004). *European Language Portfolio for post-primary learners*. Dublin: IILT.

IILT. (2007a). *Integrate Ireland Language and Training: annual activities report 2007*. Dublin: IILT.

IILT. (2007b). *A resource book for language in post-primary schools*. Dublin: IILT.

IILT. (2009). *EAL post-primary assessment kit*. Dublin: IILT.

INTO. (1998). *The challenge of diversity*. Dublin: INTO.

INTO. (2004). *Intercultural education in the primary school - an INTO discussion document*. Dublin: INTO.

Johansson, S., & Hofland, K. (1989). *Frequency analysis of English vocabulary and grammar.* (Vol. 1 and 2). Oxford: Clarendon Press.

Johns, T. (1986). Microconcord: a language-learner's research tool. *System, 14* (2), 151-162.

Johns, T. (1988). Whence and whither classroom concordancing? In T. Bongaerts et al. (Eds.), *Computer applications in language learning*. Dordrecht: Foris.

Johns, T. (1991). From printout to handout: grammar and vocabulary learning in the context of data-driven learning. *English Language Research Journal, 4*, 27-45.

Johns, T. (1994). From printout to handout: grammar and vocabulary teaching in the context of data-driven learning. In T. Odlin (Ed.), *Perspectives on pedagogical grammar*. Cambridge: Cambridge University Press.

Jones, S., & Sinclair, J. (1974). English lexical collocations - a study in computational linguistics. *Cahiers in Lexicologie, 24*, 15-61.

Katz, J. (1973). Compositionality, idiomaticity, and lexical substitution. In S. Anderson & P. Kiparsky (Eds.), *A festschrift for Morris Halle*. New York: Holt.

Kearney, F. (2008). *Inclusion or invasion? How post-primary teachers view newcomer students in the mainstream classroom*. Unpublished Master Thesis, University of Dublin, Trinity College.

Kelly, P. (2003). *New Frontiers in first year English (Junior Certificate)*. Dublin: Folens.

Kelly, P. (2004). *Junior Certificate English: New frontiers for second and third year*. Dublin: Folens.

Kennedy, G. D. (1998). *An introduction to corpus linguistics*. London: Longman.

Keogh, T., & Whyte, S. (2003). *Getting on, the experiences of ethnic minority second-level students*. Dublin: Trinity College.

Kieran, M. (2004). *Exam start: Higher level Junior Certificate English*. Dublin: Gill and Macmillan.

Kilgarriff, A. (2006). Collocationality and how to measure it. In E. Corino, C. Marello Carla and C. Onesti (Eds.) *Proceedings XII Euralex International Congress*, Turin, Italy, Alessandria Edizioni dell'Orso.

Kilgarriff, A., Rychly, P., Smrz, P., & Tugwell, D. (2004). The Sketch Engine. In G. Williams & S. Vessier (Eds.), Proceedings of EURALEX 2004. France: Université de Bretagne-Sud.

Kjellmer, G. (1990). Patterns of collocability. In J. Aarts & W. Meijs (Eds.), *Theory and practice in corpus linguistics*. Amsterdam and Atlanta: Rodopi.

Knight, D., Evans, D., Carter, R., & Adolphs, S. (2009). HeadTalk, handTalk and the corpus: towards a framework for multi-modal, multi-media corpus development. *Corpora, 4* (1), 1-32.

Kostopoulou, S. 2009. Democratising and enhancing the quality of institutionalised language assessment through the European Language Portfolio. In Taylor, L. and Weir, C. J. (Eds.) *Language testing matters: investigating the wider social and educational impact of assessment – Studies in Language Testing 31*. Cambridge: Cambridge University Press and Cambridge ESOL.

Krashen, S. (1988). *Second language acquisition and second language learning*. NY: Prentice Hall.

Lazenby-Simpson, B. (2006). *Teacher support programme for post-primary schools 2006-2007.* Paper presented to the English inspectors at the Department of Education and Science, Tallaght, Dublin, September.

LeBlanc, R., & Painchaud, G. (1985). Self-assessment as a second-language placement instrument. *TESOL Quarterly, 19* (4), 673-686.

Lee, D. Y. W. (2000). *Modelling variation in spoken and written language: the multidimensional approach revisited.* Unpublished PhD thesis, Linguistics Department, Lancaster University.

Leech, G. (1992). Corpus linguistics and theories of linguistic performance. In J. Svartvik (Ed.) *Directions in corpus linguistics: Proceedings of Nobel Symposium 82*, Stockholm, 4-8 August, Berlin: Mouton de Gruyter.

Leech, G. (2001). The role of frequency in ELT: new corpus evidence brings a re-appraisal. *Foreign Language Teaching and Research* 33 (5), 328-339.

Leech, G., Rayson, P., & Wilson, A. (2001). *Word frequencies in written and spoken English.* Harlow: Longman.

Lemke, J. (1990). *Talking science: language, learning and values.* Norwood, NJ: Ablex.

Leung. C. (2003). Integrating school-aged ESL learners into the mainstream curriculum. Working Paper 21. King's College London Working Papers on Urban Language and Literacies. Retrieved from http://www.kcl.ac.uk/content/1/c6/01/42/29/paper21.pdf

Lewis, D. (1990). *The lexical syllabus: a new approach to language teaching.* London: Collins, ELT.

Lewis, M. (1993). The lexical approach. Hove: Teacher Training Publications.

Lewis, M. (2000). Language in the lexical approach. In M. Lewis (Ed.), *Teaching collocation: Further developments in the lexical approach.* Hove: Language Teaching Publications.

Lewis, M. (2002). *The lexical approach: the state of ELT and a way forward.* Boston, Mass.: Thomson Heinle.

Lindquist, H. (2009). *Corpus linguistics and the description of English.* Edinburgh: Edinburgh University Press.

Little, D. (1991). *Learner autonomy 1: Definitions, issues and problems.* Dublin: Authentik.

Little, D. (1994). Words and their properties: arguments for a lexical approach to pedagogical grammar. In Odlin, T. (Ed.), *Perspectives on pedagogical grammar.* Cambridge: Cambridge University Press.

Little, D. (1995). Learning as dialogue: the dependence of learner autonomy on teacher autonomy. *System* 23 (2), 175-181.

Little, D. (1997). Language awareness and the autonomous language learners. *Language Awareness* 6, 93-104.

Little, D. (2002). The European Language Portfolio: structure, origins, implementation and challenges. *Language Teaching*, *35*, 182-189.

Little, D. (2005). The Common European Framework of Reference for Languages and the European Language Portfolio: involving learners and their judgements in the assessment process. *Language Testing*, *22*, 321-336.

Little, D. (2007). Language learner autonomy: some fundamental considerations revisited. *Innovation in Language Learning and Teaching*, *1* (1), 14-29.

Little, D., & Lazenby-Simpson, B. (2003). *European Language Portfolio: the intercultural components and learning to learn*. Strasbourg: Council of Europe.

Little, D., & Lazenby-Simpson, B. (2004). Using the CEF to develop an ESL curriculum for newcomer pupils in Irish primary schools. In K. Morrow (Ed.), *Insights from the Common European Framework*. Oxford: Oxford University Press.

Little, D., & Perclovà, R. (2001). *European Language Portfolio: a guide for teachers and teacher trainers*. Strasbourg: Council of Europe.

Louw, W. E. (1993). Irony in the text or insincerity in the writer? The diagnostic potential of semantic prosodies. In M. Baker, G. Francis, & E. Tognini-Bonelli (Eds.), *Text and technology: In Honour of John Sinclair*. Amsterdam: J. Benjamins.

Lucey, D. (2002). *The past today: Complete Junior Certificate History*. Dublin: Gill and Macmillan.

Lyons, Z., & Little, D. (2009). *English language support in post-primary schools in Ireland: Provision, challenges and deficits*. Dublin: Trinity Immigration Initiative, English Language Support Programme.

Makkai, A. (1972). *Idiom structure in English*. The Hague: Mouton.

Marco, L. M. J. (2000). Collocational frameworks in medical research papers: a genre-based study. *English for Specific Purposes, 19,* 63-86.

Martin, J. R. (1992). *English text: system and structure*. Amsterdam: Benjamins.

Martin, J. R., & Veel, R. (1998). *Reading science: critical and functional perspectives on discourses of science*. London: Routledge.

Martin, J. R., Christie, F., & Rothery, J. (1987). Social processes in education: a reply to Sawyer and Watson (and others). In I. Reid (Ed.), *The place of genre in learning: Current debates*. Geelong, Australia: Deakin University, Centre for Studies in Literary Education.

Martin, J. R., Wignell, P., Eggins, S. and Rothery, J. (1988). Secret English: discourse technology in a junior secondary school. In Gerot, L., Oldenburg, J. and van Leeuwen, T. (Eds.) *Language and socialization: home and school*. Proceedings from the Working Conference on Language in Education. Sydney: Macquarie University.

Matthiessen, C. (1995). *Lexicogrammatical cartography: English systems*. Tokyo: International Language Sciences.

McCarthy, M. (1991). *Discourse analysis for language teachers*. Cambridge: Cambridge University Press.

McCarthy, M., & Carter, R. (1994). *Language as discourse: perspectives for language teaching*. London: Longman.

McCarthy, M., & Carter, R. (1995). Spoken grammar: what is it and how do we teach it? *ELT Journal, 49* (3), 207-218.

McCarthy, M., & O'Dell, F. (2005). *English collocations in use*. Cambridge: Cambridge University Press.

McCarthy, M., McCarten, J., & Sandiford, H. (2004). *Touchstone*. Cambridge: Cambridge University Press.

McEnery, T., & Wilson, A. (1996). *Corpus linguistics*. Edinburgh: Edinburgh University Press.

McEnery, T., & Wilson, A. (2001). *Corpus linguistics: an introduction.* (2nd Ed.). Edinburgh: Edinburgh University Press.

McEnery, T., Xiao, R., & Tono, Y. (2006). *Corpus-based language studies: an advanced resource book*. London: Routledge.

McGrath, I. (2002). *Materials Evaluation and design for language teaching.* Edinburgh: Edinburgh University Press Ltd.

Meara, P. (1996). The dimension of lexical competence. In G. Brown, K. Malmkjaer & J. Williams (Eds.), *Performance and competence in second language acquisition*. Cambridge: Cambridge University Press.

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

Miller, C. R. (1994). Rhetorical community: The cultural basis of genre. In A. Freedman & P. Medway (Eds.), *Genre and the new rhetoric*. London: Taylor & Francis.

Milton, J. (2009). *Measuring second language vocabulary acquisition*. Bristol: Multilingual Matters.

Mitchell, T. F. (1971). Linguistic 'goings on': Collocations and other Lexical matter arising on the syntactic records. *Archirum Linguisticum*, *2*, 35-69.

Mohan, B. (1986). *Language and content*. Reading, MA: Addison Wesley

Molcho, M., Kelly, C., Gavin, A., & Nic Gabhainn, S. (2008). *Inequalities in health among school-aged children in Ireland*. Galway: HBSC Ireland.

Moon, R. (1997). Vocabulary connections: multi-word items in English. In N. Schmitt & M. McCarthy (Eds.), *Vocabulary: description, acquisition and pedagogy*. Cambridge: Cambridge University Press.

Moon, R. (1998). *Fixed expressions and idioms in English: a corpus-based approach*. Oxford: Clarendon Press.

Mullally, S. (2007). *Science revision for Junior Certificate.* (4th Ed.). Dublin: Gill and Macmillan.

Nagy, W. (1997). On the role of context in first- and second-language vocabulary learning. In N. Schmitt & M. McCarthy (Eds.), *Vocabulary: description, acquisition and pedagogy*. Cambridge: Cambridge University Press.

Nation, I. S. P. (1990). *Teaching and learning vocabulary*. Boston, MA: Heinle and Heinle.

Nation, I. S. P. (2001). *Learning vocabulary in another language.* Cambridge: Cambridge University Press.

Nattinger, J., & DeCarrico, J. (1992). *Lexical phrases and language teaching*. Oxford: Oxford University Press.

NCCA. (2006). *Intercultural education in the post-primary school*. Dublin: NCCA.

NCCA. (2008). *Junior Certificate English – Draft syllabus for consultation*. Dublin: NCCA

Nesselhauf, N. (2004). Learner corpora and their potential for language teaching. In J. Sinclair (Ed.), *How to use corpora in language teaching*. Amsterdam: J. Benjamins.

North, B. (2000). *The development of a common framework scale of language proficiency*. New York: Peter Lang.

O'Halloran, K. (2004). Discourses in secondary school mathematics classrooms according to social class and gender. In J. Foley (Ed.), *Language, education and discourse: functional approaches*. London: Continuum.

O'Keeffe, A., & Farr, F. (2003). Using language corpora in language teacher education: pedagogic, linguistic and cultural insights. *TESOL Quarterly, 37* (3), 389-418.

O'Keeffe, A., McCarthy, M., & Carter, R. (2007). *From corpus to classroom: Language use and language teaching*. Cambridge: Cambridge University Press.

Oakes, M. (1998). *Statistics for corpus linguistics*. Edinburgh: Edinburgh University Press.

Oakey, D. (2002). Formulaic language in English academic writing. In R. Reppen, S. M. Fitzmaurice & D. Biber (Eds.), *Using corpora to explore linguistic variation*. Amsterdam; Philadelphia: J. Benjamins.

O'Dywer, P. (2002). *The human planet: Geography for Junior Certificate*. Dublin: Gill and Macmillan.

OECD. (2003). *Where immigrant students succeed: a comparative review of performance and engagement in PISA 2003*. Paris: OECD.

OECD. (2006). *Where immigrant students succeed: a comparative review of performance and engagement in PISA 2003*. Paris: OECD.

OECD. (2007). *Science competencies for tomorrow's world: Results from PISA 2006*. Paris: OECD.

OECD. (2009). *OECD reviews of migrant education: Ireland*. Paris: OECD.

O'Leary, D. (2007). *History revision for Junior Certificate.* (3rd Ed.). Dublin: Gill and Macmillan.

Ongstad, S. (2007). *Language and communication in Norwegian curricula for mathematics*. Strasbourg: Council of Europe. Retrieved from http://www.coe.int/lang

Ortega, L. (2005). For what and for whom is our research? The ethical as transformative lens in instructed SLA. *The Modern Language Journal*, 89, 427-443.

Oskarsson, M. (1978). *Approaches to self-Assessment in foreign language learning.* Council of Europe: Pergamon Press.

Oskarsson, M. (1984). *Self-assessment of foreign language skills: a survey of research and development work.* Strasbourg: Council of Europe.

Oskarsson, M. (1989). Self-assessment of language proficiency: rationale and applications. *Language Testing, 6*, 1-13.

Östman, J. & Virtanen, T. (1995). Discourse analysis. In J. Verschueren, J. Östman & J. Blommaert (Eds.), *Handbook of pragmatics*. Amsterdam: J. Benjamins.

Painter, F. (1999). Preparing for school. In Christie, F. (Ed.), *Pedagogy and the shaping of consciousness*. London: Continuum.

Palmer, H. E. (1933). *Second interim report on English collocations*. Tokyo: Kaikatusha.

Paltridge, B. (2006). *Discourse analysis: an introduction*. London: Continuum.

Partington, A. (1998). *Patterns and meaning: using corpora for English language research and teaching*. Amsterdam: J. Benjamins.

Pawley, A. K., & Syder, F. H. (1983). Two puzzles for linguistic theory: Native like selection and nativelike fluency. In J. C. Richards & R. W. Schmidt (Eds.), *Language and communication*. London and New York: Longman.

Pecina, P. (2005). An extensive empirical study of collocation extraction methods. In Proceedings of the ACL Student Research Workshop (pp. 13–18). Association for Computational Linguistics.

Pimm, D. (1987). *Learning Mathematics: issues, theory and classroom practice*. London: Cassell Education.

Pimm, D. (1995). *Symbols and meanings in school mathematics*. London: Routledge.

Quinn, R. (1998). *Taking action: Civil, social and political education for Junior Certificate*. Dublin: C. J. Fallon.

Quinn, R., & O'Leary, D. (2002). *Junior Certificate History: Door to the past*. Dublin: Folens.

Ravid, D. & L. Tolchinsky, L. (2002). Developing linguistic literacy: a comprehensive model. *Journal of Child Language* 29, 419-448.

Rayson, P. (2008). From key words to key semantic domains. *International Journal of Corpus Linguistics, 13* (4), 519-549.

Rayson, P., Archer, D., Piao, S., & McEnery, T., (2004). The UCREL semantic analysis system. Proceedings of the LREC-04 Workshop, beyond Named Entity Recognition Semantic Labelling for NLP Tasks (pp. 7-12). Paris: European Language Resources Association.

Rea-Dickins, P. (1987). *The relationship between grammatical abilities and aspects of communicative competence: with special reference to the testing of grammar.* Unpublished PhD Thesis, University of Lancaster, UK.

Reppen, R. (2010). *Using corpora in the language classroom*. Cambridge: Cambridge University Press.

Reppen, R., Fitzmaurice, S. M., & Biber, D. (Eds.) *Using corpora to explore linguistic variation*. Amsterdam; Philadelphia: J. Benjamins.

Rocks, F. (2004). *Exam start: Ordinary level Junior Certificate English*. Dublin: Gill and Macmillan.

Römer, U. (2008). Corpora and language teaching. In A. Lüdeling & K. Merja (Eds.), *Corpus Linguistics. An International Handbook.* (Vol. 1). Berlin: Mouton de Gruyter.

Rosen, H. (1972). The language of textbooks. In Language and Learning Course Team at The Open University (Eds.), *Language in education: a resource book.* London: Kegan Paul in association with The Open University Press.

Ross, S. (1998). Self-assessment in second language testing: a meta-analysis and analysis of experiential factors. *Language Testing, 15*, 1-20.

Rumberger, R. & Scarcella, R. (2000). Academic English. *Linguistic Minority Research Institute Newsletter* 1, 1.

Scarcella, R. (2003). *Academic English: a conceptual Framework*. (Technical Report 2003-1). Irvine, CA: University of California Santa Barbara, University of California Linguistic Minority Research Institute.

Schleppegrell, M. (2001). Linguistic features of the language of schooling. *Linguistics and Education, 12* (4), 431-459.

Schleppegrell, M. (2004). *The language of schooling: a functional linguistics perspective.* Mahwah, New Jersey: Lawrence Erlbaum Associates.

Schleppegrell, M., & Colombi, C. (2002). *Developing advanced literacy in first and second languages: meaning with power*. Mahwah, N.J.; London: Lawrence Erlbaum Associates.

Schmidt, R.W. (1990). The role of consciousness in second language learning. *Applied Linguistics, 11* (2), 129-158.

Schmitt, D., & Scmitt, N. (2005). *Focus on vocabulary: mastering the Academic Word List*. Longman: White Plains, NY.

Schmitt, N. (2000). *Vocabulary in language teaching*. Cambridge: Cambridge University Press.

Schneider, G., & Lenz, P. (2001). *European Language Portfolio: guide for developers*. Strasbourg: Council of Europe.

Schön, D. (1983). *The reflective practitioner*. London: Basic Books.

Scott, M. (1999). *WordSmith Tools users help file*. Oxford: Oxford University Press.

Scott, M. (2004). *WordSmith Tools version 4*. Oxford: Oxford University Press.

Scott, M., & Tribble, C. (2006). *Textual patterns: keyword and corpus analysis in language education*. Amsterdam: John Benjamins.

Sharwood Smith, M. (1981). Consciousness-raising and the second language learner. *Applied Linguistics, 2* (2), 159 - 168.

Sinclair, J. (1966). Indescribable English. Inaugural lecture, University of Birmingham (Abstract in Sinclair, J. and R.M. Coulthard (1975) *Towards an analysis of discourse*. Oxford: Oxford University Press).

Sinclair, J. (1987). *Collins COBUILD English language dictionary*. London: Collins.

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

Sinclair, J. (1992). Trust the text. In M. Davies and Ravelli, L. (Eds.), Advances *in systemic linguistics*. London: Pinter.

Sinclair, J. (1995). *Introduction to the Collins Cobuild English Dictionary.* (2nd Ed.). London: HarperCollins.

Sinclair, J. (1996). The search for units of meaning. *Textus, 9* (1), 75-106.

Sinclair, J. (1997). Corpus evidence in language description. In A. Wichmann, S. Fligelstone, T. McEnery & G. Knowles (Eds.), *Teaching and language corpora*. London: Longman.

Sinclair, J. (1998). The lexical item. In E. Weigand (Ed.) *Contrastive lexical semantics*. Amsterdam: J. Benjamins.

Sinclair, J. (2003). *Reading concordances: an introduction*. London: Longman.

Sinclair, J. (2004). Interview with John Sinclair conducted by Wolfgang Teubert. In J. M. Sinclair, S. Jones, R. Daley & R. Krishnamurthy (Eds.), *English collocation studies: the OSTI report*. London: Continuum.

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

Sinclair, J. (2005). Corpus and text - basic principles. In M. Wynne (Ed.), *Developing linguistic corpora: a guide to good practice*. Oxford: Oxbow Books.

Sinclair, J., & Renouf, A. (1988). A lexical syllabus for language learning. In R. Carter & M. McCarthy (Eds.), *Vocabulary and language teaching.* Harlow: Longman.

Sinclair, J., Jones, S. & Daley, R. (1969). *English collocation studies: the OSTI report*. Reprinted in Sinclair, J., Jones, S., Daley, R., & Krishnamurthy, R. (Eds.), *English collocation studies: the OSTI report*. London: Continuum.

Skehan, P. (1992). Strategies in second language acquisition. *Working Papers in English Language Teaching, No. 1*, Thames Valley University.

Skemp, R. R. (1976). Relational understanding and instrumental understanding. *Mathematics Teaching, 77*, 20-26.

Smyth, E., Darmody, M., McGinnity, F., & Byrne, D. (2009). *Adapting to diversity: Irish schools and newcomer students*. Dublin: The Economic and Social Research Institute.

Smyth, E., McCoy, S., & Darmody, M. (2004). *Moving up: the experiences of first-year students on Irish post-primary education*. Dublin: Liffey Press.

Snow, C. E., & Uccelli, P. (2009). The challenge of academic language. In D. R. Olson & N. Torrance (Eds.), *The Cambridge handbook of literacy*. Cambridge, New York: Cambridge University Press.

Snow, M. A., Met, M., & Genesee, F. (1989). A conceptual framework for the integration of language and content in second/foreign language programs. *TESOL Quarterly, 23* (2), 201-217.

Stevens, V. (1991a). Classroom concordancing: vocabulary materials derived from relevant, authentic text. *English for Specific Purposes, 10*, 35-46.

Stevens, V. (1991b). Concordance-based vocabulary exercises: a viable alternative to gap-filling. *English Language Research Journal, 4*, 47-61.

Stubbs, M. (1996). *Text and corpus analysis: computer-assisted studies of language and culture*. Oxford: Blackwell.

Stubbs, M. (2001). *Words and phrases: corpus studies of lexical semantics*. Oxford: Blackwell.

Swales, J. (2002). Integrated and fragmented worlds: EAP materials and corpus linguistics. In Flowerdew, J. (Ed.), *Academic discourse*. London: Longman, Pearson Education.

Swales, J. M. (1981). *Aspects of article introductions* (Aston ESP Reports No. 1).The Language Studies Unit: The University of Aston in Birmingham.

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

Taylor, L. and Barker, F. (2008). Using corpora for language assessment. In E. Shohamy and N. H. Hornberger (Eds.), *Encyclopedia of language and education* (2nd Ed.) Vol. 7, Language testing and assessment. New York: Springer.

Thompson, G. and S. Hunston (2005). *System and corpus: exploring connections*. London: Equinox.

Thurston, J., & Candlin, C. (1993). *Exploring academic English: a workbook for student essay writing*. Sydney: National Centre for English Language Teaching and Research.

Thurston, J., & Candlin, C. (1998). Concordancing and the teaching of the vocabulary of academic English. *English for Specific Purposes, 17*, 267-280.

Tognini-Bonelli, E. (2001). *Corpus linguistics at work: Studies in corpus linguistics*. (Vol. 6). Amsterdam: J. Benjamins.

Tribble, C. (1997). Improvising corpora for ELT: Quick-and-dirty ways of developing corpora for language teaching. In B. Lewandowska-Tomaszczyk & P. J. Melia (Eds.), *Practical applications in language corpora*. Lodz: Lodz University Press.

Tribble, C. (1999). *Writing difficult texts*. Unpublished Ph.D. Thesis, Lancaster University, UK.

Tribble, C., & Jones, G. (1990). *Concordancing in the classroom*. Harlow: Longman.

Tribble, C., & Jones, G. (1997). *Concordancing in the classroom: a resource guide for teachers*. Houston: Athelstan.

Tribble, L. (1985). *English for science and technology: a discourse approach*. Cambridge: Cambridge University Press.

Tucker, R. G., & D'Anglejan, A. (1975). The acquisition of complex English structures by adult learners. *Language learning, 25* (2), 281-296.

van Leeuwen, T.,& Humphrey, S. (1996). On Learning to Look Through a Geographer's Eyes. In R. Hasan & G. Williams (Eds.), *Literacy in society*. London: Longman.

van Lier, L. (2002). An ecological-semiotic perspective on language and linguistics. In C. Kramsch (Ed.), *Language acquisition and language socialization: ecological perspectives.* New York: Continuum.

Vekic, K. (2003). *Unsettled hope: Unaccompanied minors in Ireland, from understanding to response*. Dublin: Marino Institute of Education.

Vollmer, H. (2007). *Language across the curriculum*. Strasbourg: Council of Europe.

Vollmer, H. (2009). *Language in other subjects*. Strasbourg: Council of Europe.

Vygotsky, L. S. (1986). *Thought and language*. Cambridge; MA: MIT Press.

Wallace, M. (1998). *Action research for language teachers.* Cambridge: Cambridge University Press.

Ward, T. (2004). *Education and language needs of separated children in Ireland*. Retrieved from http://www.cdvec.ie/uploads/publications/LanguageNeedsChildren.pdf

West, M. (1953). *A general service list of English words*. London: Longman.

Widdowson, H. G. (1990). *Aspects of language teaching*. Oxford: Oxford University Press.

Widdowson, H. G. (1998). Context, community and authentic language. *TESOL Quarterly, 32* (4), 705-716.

Widdowson, H. G. (2002). Corpora and language teaching tomorrow. Keynote lecture delivered at the Fifth Teaching and Language Corpora Conference, Bertinoro, Italy, 29 July.

Williams, G. (1998). *Collocations in science writing*. Tubingen: Günter Narr Verlag.

Willis, D. (1990). *The lexical syllabus: a new approach to language teaching*. London: HarperCollins.

Wittenburg, P. (2008). Preprocessing multimodal corpora. In A. Lüdeling & M. Kytö (Eds.), *Corpus linguistics: an international handbook.* (Vol. 1). Berlin: Mouton de Gruyter.

Wray, A. (2002). *Formulaic language and the lexicon*. Cambridge: Cambridge University Press.

Xiao, Z. and McEnery, T. (2005). Two approaches to genre analysis: three genres in modern American English. *Journal of English Linguistics, 33 (1)*, 62-82.

Xue, G., & Nation, I. S. P. (1984). A university word list. *Language Learning and Communication, 3*, 215-229.

Zipf, G. (1935). *The psychobiology of language: an introduction to dynamic philology*. Cambridge, MA: MIT Press.

Zipf, G. (1945). The meaning-frequency relationship of words. *Journal of General Psychology* 33, 251-256.

Yates, F. (1960). *Sampling methods for censuses and surveys*. (3rd Ed.). London: Charles Griffin & Company Limited.