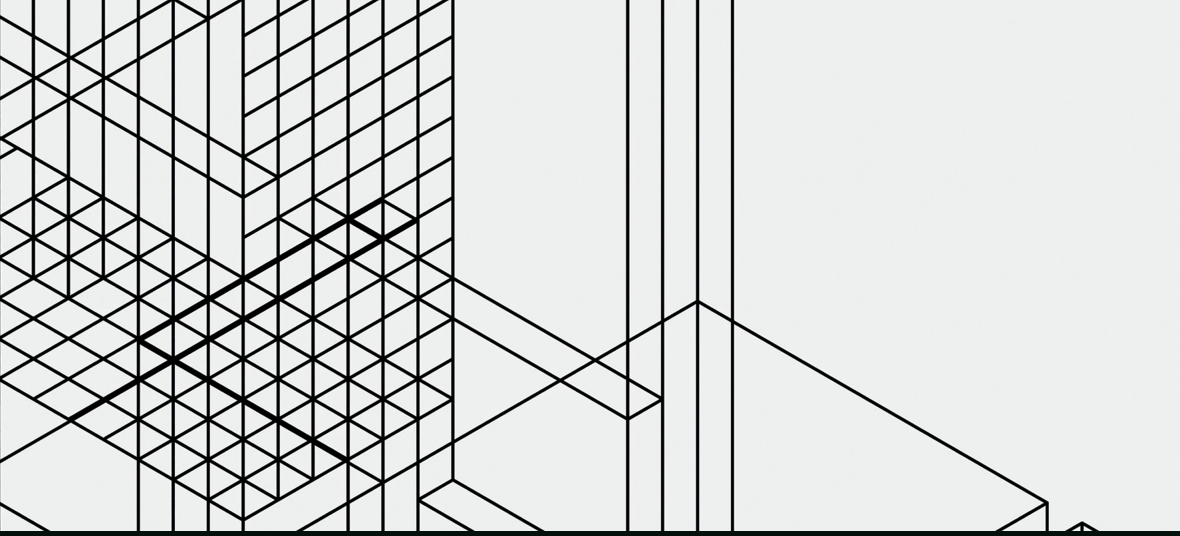
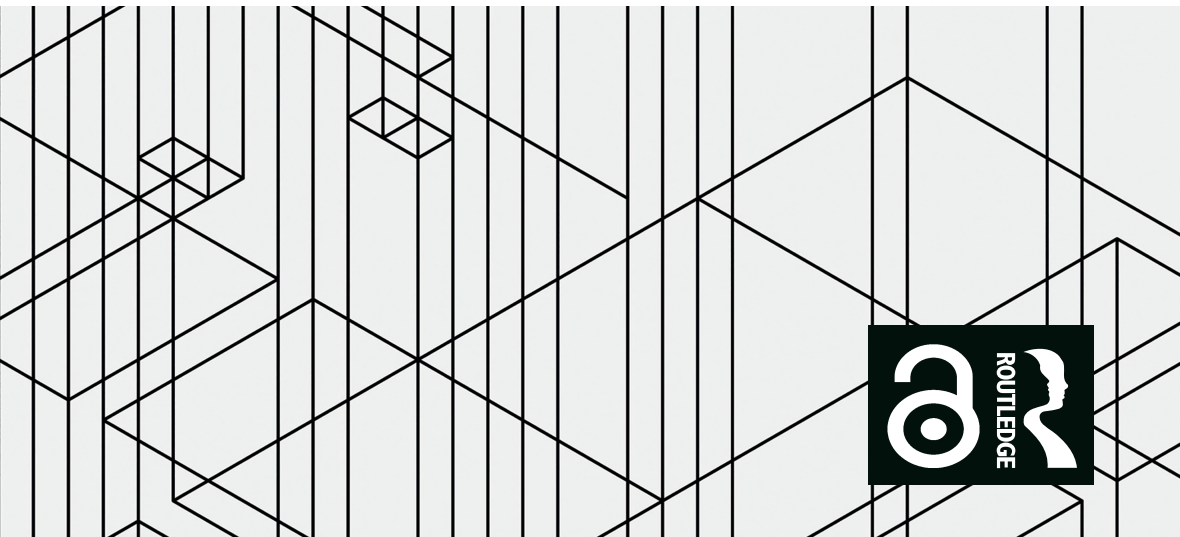


*Routledge Applied Corpus Linguistics*

# THE LINGUISTIC CHALLENGE OF THE TRANSITION TO SECONDARY SCHOOL

## A CORPUS STUDY OF ACADEMIC LANGUAGE

Alice Deignan, Duygu Candarli and Florence Oxley



# The Linguistic Challenge of the Transition to Secondary School

This book provides a unique analysis and description of the linguistic challenges faced by school students as they move from primary to secondary school, a major transition, which some students struggle with emotionally and academically. The study:

- draws on a bespoke corpus of 2.5 million words of written materials and transcribed classroom recordings, provided by the project's partner schools;
- combines quantitative and qualitative approaches to the corpus data to explore linguistic variation across school levels, registers and subjects;
- describes the procedures of corpus compilation and analysis of written and spoken academic language, showing how modern corpus tools can be applied to this far-reaching social and educational issue;
- uncovers differences and similarities between the academic language that school children are exposed to at primary and secondary school, contrasting this against the backdrop of the non-academic language that they encounter outside school.

This book is important reading for advanced students and researchers in corpus linguistics, applied linguistics and teacher education. It carries implications for policymakers and schools looking to support students at this critical point in their schooling.

**Alice Deignan** is Professor of Applied Linguistics in the School of Education, University of Leeds. She is the author of 'Metaphor and Corpus Linguistics' (2005, John Benjamins) and 'Figurative Language, Genre and Register' (2013, CUP, with Elena Semino and Jeannette Littlemore).

**Duygu Candarli** is currently Lecturer in Language Education at the University of Dundee. She specialises in academic discourse, corpus linguistics, second language writing and writing assessment. She has published research articles on these areas in international peer-reviewed journals.

**Florence Oxley** is a Research Assistant in the School of Education, University of Leeds and a PhD candidate in Linguistics, University of York. She is interested in corpus linguistics, literacy and first language acquisition, and has published on infant phonological development.

## Routledge Applied Corpus Linguistics
**Series Editor: Michael McCarthy**

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

**Series Editor: Anne O'Keeffe**

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

**Series Co-Founder: Ronald Carter**

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

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

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

### Other Titles in This Series

**Investigating a Corpus of Historical Oral Testimonies**
The Linguistic Construction of Certainty
*Chris Fitzgerald*

**Data-Driven Learning and Language Learning**
Corpus Use in Italian L2 Pedagogy
*Luciana Forti*

Inter-Varietal Applied Corpus Studies

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

# The Linguistic Challenge of the Transition to Secondary School

## A Corpus Study of Academic Language

**Alice Deignan, Duygu Candarli and Florence Oxley**

Routledge
Taylor & Francis Group
LONDON AND NEW YORK

First published 2023
by Routledge
4 Park Square, Milton Park, Abingdon, Oxon OX14 4RN

and by Routledge
605 Third Avenue, New York, NY 10158

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



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

*Library of Congress Cataloging-in-Publication Data*
Names: Deignan, Alice, author. | Candarli, Duygu, author. | Oxley, Florence, author.
Title: The linguistic challenge of the transition to secondary school: a corpus study of academic language/Alice Deignan, Duygu Candarli and Florence Oxley.
Description: Abingdon, Oxon; New York, NY: Routledge, 2023. |
Series: Routledge applied corpus linguistics | Includes bibliographical references and index. |
Summary: "This book provides a unique analysis and description of the linguistic challenges faced by school students as they move from primary to secondary school, a major transition, which some students struggle with emotionally and academically"–Provided by publisher.
Identifiers: LCCN 2022031250 (print) | LCCN 2022031251 (ebook) | ISBN 9780367534219 (hardback) | ISBN 9780367534226 (paperback) | ISBN 9781003081890 (ebook)
Subjects: LCSH: Academic language. | Corpora (Linguistics) | Students, Transfer of. | Student adjustment. | Education, Primary. | Education, Secondary. | Articulation (Education)
Classification: LCC P120.A24 D45 2023 (print) | LCC P120.A24 (ebook) | DDC 407.1–dc23/eng/20220831
LC record available at https://lccn.loc.gov/2022031250
LC ebook record available at https://lccn.loc.gov/2022031251

ISBN: 978-0-367-53421-9 (hbk)
ISBN: 978-0-367-53422-6 (pbk)
ISBN: 978-1-003-08189-0 (ebk)

DOI: 10.4324/9781003081890

Typeset in Sabon
by Deanta Global Publishing Services, Chennai, India

# Contents

# Extracts

# Figures

# Tables

# Acknowledgements

The research described in this book was part of a project called 'The linguistic challenges of the transition from primary to secondary school', funded by the Economic and Social Research Council, UK, whose support we gratefully acknowledge, initially from September 2018 to April 2021, grant number ES R006687/1. Alice was Principal Investigator, Duygu was Research Fellow and Florence was Research Assistant. The Co-Investigators were Gary Chambers and Michael Inglis from Leeds University and Vaclav Brezina and Elena Semino from Lancaster University. Robbie Love and Doğuş Öksüz also supported the research in the early stages of the project. We are very grateful to the whole project team for their ideas and insights. Marcus Jones, who is Literacy Lead at Huntington Research School, York, acted as a project consultant and gave us invaluable perspectives throughout. We could not have done this research without his guidance and the generous participation of our partner schools and their head teachers, teachers and students. We thank the series editors, Anne O'Keeffe and Michael McCarthy, for their feedback and encouragement. Alice would also like to thank her partner Tim for his support, and friends Kathryn Atkins and Jane Bradbury, whose perceptive practitioner experience informed and challenged this research over many discussions. Duygu would like to thank her mother Nilgün for her support and encouragement over the years and her colleagues at the University of Dundee for their patience and support. Florence would like to thank Tamar Keren-Portnoy and Eytan Zweig for their ongoing guidance and support and Jordan Hayward for listening while she thinks aloud.

# 1 Schools, the transition, students and teachers

*Alice Deignan*

## Introduction

This book is about research that has the ultimate goal of producing information and resources to support school students, especially in lower secondary school. School students report feeling intense academic pressure in today's competitive world. In an online article, Jasmine Savory reflects on her life as a 21st-century teenager in London, writing that 'students are finding it harder and harder to keep up with the growing amount of revision they face throughout their school life', in a world of 'target grades, league tables, and the persistent question that forever echoes around school corridors ("what did you get?" "what did you get?")' (2022). The pressure is felt by school students around the world, with negative consequences for mental and physical health (Pascoe et al., 2020). It is also found across achievement levels. While high achievers struggle with the strain of high-stakes examinations (Banks & Smyth, 2015), students who have been identified as 'disengaged' and under-achieving also feel under pressure. Duffy and Elwood (2013) interviewed a number of such students and reported that many admitted to worries about qualifications and their future lives. One student said they realised 'how hard it is to get a job and everything, so you just put your head down so you can have a good chance', while another is quoted as saying, 'There's no jobs so you're worrying about getting a job all the time' (2013, p. 121).

As a first step to succeeding academically, students need to be able to access the materials presented to them in class and through texts. Language and literacy thus have a central role in school success (Clark, 2019, p. 6). Difficulties with the language of school can present a significant barrier to academic achievement, resulting in young people performing below their intellectual potential. Language difficulties are not confined to students studying in a second or additional language. Many children and young people who are highly linguistically proficient in everyday situations find that their skill does not transfer to school (Gee, 2004). This is because the language of school is different to the language of home and the playground (Leung, 2014), and it becomes more different as children move up the school grades or years.

DOI: 10.4324/9781003081890-1

It is widely recognised that children from different backgrounds do not start school with the same levels of knowledge of academic language (Gee, 2004; Schleppegrell 2001). There is some degree of correlation between social class and confidence in academic language, with children from lower socio-economic status (SES) backgrounds tending to feeling less at ease in handling academic ways of writing and speaking (Gee, 2004; Patterson, 2020). This means that a group of young people are disadvantaged at school right from the start. As applied and corpus linguists, we want to make a contribution to supporting such students by using our knowledge and tools to develop a detailed description of the language of school, that is, to demystify it for outsiders.

## The transition and the context of this research

In many nations, children study for six or seven years in primary, or elementary school, then, at between ten and 14 years of age, move to secondary, or high school (Evans et al., 2018), a move that is often referred to simply as 'the transition'. For many children, this move to 'big school' is a large and psychologically profound change. Zeedyk et al. stated that 'this period is regarded as one of the most difficult in pupils' educational careers, and success in navigating it can affect not only children's academic performance but their general sense of well-being and mental health' (2003, p. 68). In England, where this study took place, the transition occurs at around the age of 11. Children usually (but not always) start formal schooling at the age of four, spending seven years in primary school, then move to secondary school, for five years in the first instance. Many students continue for a further two years at the same school, until the age of 18, while some leave at 16 to attend a different school or college, start vocational training or enter employment. Table 1.1 is a simplified presentation of the school structure for most students in England (adapted from www.gov.uk/national-curriculum).

England has a National Curriculum, which, as shown in Table 1.1, is divided into five Key Stages (KSs) plus the Early Years curriculum. The National Curriculum specifies in some detail the topics and schemes of work for each year of schooling. This is coordinated with national assessment. For mainstream school students, the most high-stakes assessments are GCSEs (General Certificate of Secondary Education), usually taken at the end of Year 11, when most children are 16, and A Levels or other qualifications taken at the end of Year 13. At the end of Year 6, that is, the end of primary schooling, and of KS2, students take assessments known as SATs (Standard Attainment Tests), which at the time of writing, are examinations in mathematics and English. The focus of the research described in this book is on the years around the transition at the end of Year 6, that is, the last two years of primary school, and the first two years of secondary school, the beginning of KS3, which are the shaded rows in the table.

*Table 1.1* Typical stages of schooling in England.

| *Age* | *Year* | *Key Stage* | *School* | *National assessment* |
|---|---|---|---|---|
| 3–4 | - | Early Years | Primary or nursery | |
| 4–5 | Reception | Early Years | Primary | |
| 5–6 | 1 | KS1 | Primary | |
| 6–7 | 2 | KS1 | Primary | KS1 teacher assessments |
| 7–8 | 3 | KS2 | Primary | |
| 8–9 | 4 | KS2 | Primary | |
| 9–10 | 5 | KS2 | Primary | |
| 10–11 | 6 | KS2 | Primary | KS2 tests and teacher assessments |
| 11–12 | 7 | KS3 | Secondary | |
| 12–13 | 8 | KS3 | Secondary | |
| 13–14 | 9 | KS3 | Secondary | |
| 14–15 | 10 | KS4 | Secondary | |
| 15–16 | 11 | KS4 | Secondary | GCSEs |
| 16–17 | 12 | KS5 | Secondary or college or employment | A levels or other |
| 17–18 | 13 | KS5 | | |

## Issues at transition

Around the world, there has been a good deal of interest in the transition, with research into transition issues in Scotland (Jindal-Snape et al., 2019; West et al., 2010), Australia (Hopwood et al., 2016), the United States (Felmlee et al., 2018), Finland (Virtanen et al., 2019; Eskelä-Haapanen et al., 2020), New Zealand (McGee et al., 2003), Canada (Serbin et al., 2013) and others. There have also been several meta-studies drawing together research on the transition in different contexts and from different disciplinary viewpoints (e.g., Jindal-Snape et al., 2019; Evans et al., 2018; van Rens et al., 2017).

For many students, moving to secondary school is a positive experience, which is anticipated with excitement (Jindal-Snape & Cantali, 2019; Coffey, 2013; Eskelä-Haapanen et al., 2020). Jindal-Snape and Cantali (2019) found that students look forward to developments such as meeting new people, having specialist teachers, new subjects such as cookery, a wider range of sports and clubs and more equipment. This sense of excitement was voiced by the primary school students who we spoke to for this project in interviews discussed later in this chapter. Longitudinal studies that followed students into secondary school found that many of them did indeed have the positive experiences that they had anticipated (Jindal-Snape et al., 2019). Symonds and Hargreaves (2016) also heard reports of positive feelings, including in academic work. They interviewed and compared two groups of Year 7 students in England (aged 11–12 years): a group who had transitioned into secondary school, and a group who had not changed schools. (The latter group were in a 'middle school', a system that has

become much less common in recent years.) The researchers found that only the students who had transitioned reported that they enjoyed classes in Year 7 more than in the previous year, their last year of primary school; Year 7s who had not transitioned did not report this. Further, they 'appreciated more advanced equipment and challenging work' (2016, p. 72), and some said they found the increased academic pressure stimulating. They reported feeling 'grown-up' and mature, having left the younger children behind at primary school.

However, the move can also cause minor or major problems for some students (West et al., 2010; Evans et al., 2018; Wilson, 2011), and a number of studies have focused on these problems. West et al.'s (2010) research shows that these problems can have long-lasting effects. They conducted a longitudinal study of 2000 Scottish school students from the age of 11 through to 18/19, when they left school. A poor school transition predicted lower attainment and well-being than their peers at age 15, and the effect was still detectable, albeit reduced, at age 18/19, after the participants had left school. Researchers tend to divide transition issues into two types: first, social, psychological or emotional, and second, academic; we now look at each of these.

### *Social, psychological and emotional issues*

Many studies have reported on social, psychological and emotional aspects of the transition, finding these to be the major concerns of many stakeholders. Jindal-Snape and Cantali (2019) found from their interviews that students and their parents are very focused on practical and social aspects of the transition. Zeedyk et al. (2003) found the major student concern was bullying, followed by getting lost, peer relationships and coping with the workload. Rice et al. (2015) found that the top five concerns for Year 6 students were the following: getting lost; being bullied; discipline and detentions; homework; losing old friends. All but 'homework' are about social, relationship and institutional aspects of the transition. Zeedyk et al. (2003) stated that 'academic performance' was mentioned only infrequently by parents and students, noting that 'children's most pressing concerns do not appear to be academic ones' (2003, p. 73). Topping (2011), Rice et al. (2011) and Jindal-Snape and Foggie (2008) also found that children were overwhelmingly concerned with socio-emotional and practical issues rather than academic ones. There is consensus then, that children are worried, to varying degrees, about what secondary school will be like, but that academic work is not at the front of their minds.

An important sub-section of this group of studies looks closely at relationships, with other students and with teachers, Coffey (2013) claiming that these are central to the process, noting that students 'are at a point in their lives when friendships and interactions with peers are of high importance' (2013, p. 264). There has been research citing (fears of) the loss of

friendship networks from primary school and/or difficulty in making new friends in secondary school (West et al., 2010; Rice et al., 2015). Worries about this were the most frequently cited concern in Jindal-Snape and Cantali's study (2019), closely followed by bullying. Rice et al. (2015) note the same concerns, alongside the generalised issue of 'older children'. This is to be expected given that the transition coincides with early adolescence, a time when children's focus is moving away from their immediate home and family life, and their peers begin to rival their caregivers as the most central relationships in their lives.

The other important school relationships for students are those with teachers. In primary schools, students usually study under just one, generalist teacher, who is likely to know them well and be a parent-like figure. This was reflected in a survey carried out by the *Times Educational Supplement*[1] in 2016, where 2500 primary school pupils were asked to name what every child should have done by age 11. Top of the list of 100 experiences was '[accidentally] call a teacher "mum" or "dad"', with a large number of pupils reporting that they had done this in primary school, even in Year 6. The relationship is very different in secondary school; Symonds and Hargreaves describe it as much more neutral (2016). Secondary school students are taught by a number of subject specialists and may see eight or more different teachers over a week.

In England and Wales, secondary school teachers usually have a bachelor's degree in the subject they teach, followed by a one-year teacher training qualification. It might be expected therefore for many secondary teachers to have a disciplinary orientation (Bru et al. 2010). Primary school teachers often, but not always, have spent longer studying education and correspondingly less time on a subject specialism. They are therefore likely to have a child-development orientation, and spending most of their day with the same class means they usually develop an overview of the overall progress and circumstances of each child. Year 6 students report that as the oldest in their primary school, they feel special, being allocated responsibilities and prestige. In contrast, in the context of a large secondary school, KS3 students are often not the main concern (Ofsted, 2015), perhaps because they are not taking national examinations. Some teachers have told us of their preference for teaching older students, ideally for the post-compulsory A Level qualifications, where the subject matter is enjoyably challenging even for the teacher. In contrast, Year 7 students, beginners in the subject, developmentally immature on many fronts, and still 'liable to fall off their chairs' as one teacher told us, may be seen as less rewarding to teach, and a Year 7 class as a less prestigious assignment.

Good teacher-student relationships are a factor in stress reduction for all students (Banks & Smyth, 2015). Coffey (2013) writes of teachers' pivotal role in supporting students in early secondary school, and vulnerable and disaffected students have told researchers that a good relationship with a teacher can make all the difference (Duffy & Elwood, 2013). Unfortunately

for these more vulnerable students, support from teachers tends to decrease as they progress up the school years (Evans et al. (2018). Symonds and Hargreaves (2016) found that sometimes students reported resentment at behaviour management and control from teachers, which often increases just at a point when adolescents are wanting to be independent and autonomous. Tobbell and O'Donnell (2013) found that the Year 7 students they interviewed found it difficult to manage relationships with multiple teachers, especially as their different teachers sometimes had different expectations about behaviour. Numerous studies have noted similar challenges, and as Goldstein et al. note, 'At a time when youth would benefit greatly from close and nurturing extra-familial relationships with adults, opportunities for developing these relationships decline' (2015, p. 21).

Pressure in school has been found to increase from elementary school to secondary school (Klinger et al., 2015; Strand, 2019). Rice (2001) found that a reduction in pressure from teachers just after the transition had a positive effect on achievement in science and mathematics 'it may be that students benefit from a short-term hiatus from overwhelming academic pressure while they adjust to the new school environment' (2001, p. 390). Rice et al. (2011), in a UK-based study, note that the transition is stressful to all students including those who adjust well. Goldstein et al. (2015) explored the relationship between transition stress and academic outcomes and found a strong association: 'greater stress was associated with increased test and performance anxiety, lower school bonding, and lower academic performance' (2015, p. 26). As they note, a causal relationship from stress to lower academic performance can't be assumed; poor academic performance could be the cause of stress, and there may be other factors that they did not study. Nonetheless, they recommend that measures to reduce stress could result in better academic outcomes, and a more positive start to their new schools can only be a good thing for students.

The schools in England that we have worked with for this project have also placed emphasis on social, psychological and emotional aspects of the transition, as well as on important practical points such as finding classrooms and bringing the right equipment for each lesson. Ofsted (2015) finds this contextual, non-academic focus widespread. The emphasis might stem from what research and primary schools report about Year 6 students' worries, but it may also end up reinforcing the idea that these will be the biggest hurdles. In the project described in this book, we did not specifically investigate the social and emotional aspects of the transition. Nonetheless, these aspects are relevant to us, because if students are preoccupied or stressed, there will inevitably be a knock-on effect on their academic studies.

### *Academic issues*

There is very widespread agreement that there is a decline in academic achievement in the early years of secondary school (Jindal-Snape & Cantali,

2019; McGee et al., 2003; Evans et al., 2018; Topping, 2011; Goldstein et al., 2015). Studies referencing such concerns date back at least to 1961 in the UK (Nisbet & Entwistle, 1969). McGee et al. (2003) and Virtanen et al. (2019) both reflect that the same pattern is found in the first year of secondary school regardless of age, which varies between countries. This suggests it is not primarily an age-specific phenomenon associated with, for example, the onset of puberty. Further, both McGee et al. (2003) and Evans et al. (2018) find evidence in studies from several different contexts showing that where students transition twice or more, for example, from elementary to middle school and then on to high school, they suffer multiple declines in academic achievement. Felmlee et al. (2018) directly compared school students who transitioned to high school in the United States, at around 12 years of age, with those who did not. They found that the ones who transitioned became more socially isolated than their peers and tended to get fewer high grades, an effect that remained throughout high school. This was not universal, however; a small proportion, just over 10%, improved academically following the transition.

McGee et al. (2003) put forward several possible reasons for this dip in attainment. They suggest that sometimes secondary school work increases in volume rather than difficulty, leading to rushed work. They also suggest a possible decline in intrinsic motivation as external pressures increase, with increased emphasis on test results and performance in relation to other students. Students who do not believe that they are likely to do well might be discouraged from investing effort, from a fear of looking foolish – thinking it better to seem cynical and disengaged than to look like someone who tries and fails. They also cite discontinuities in teaching styles. Evans et al. (2018) suggest that the dip may be partly due to new structures, such as moving from room to room over the day, switching teachers and new environments – a larger school with many classrooms and more students. They also consider that the performance-orientation of secondary schools, as opposed to the task orientation of primary schools, might be a factor in demotivating or intimidating some students. Bru et al. (2010) claim that it is around the age of 11 that children learn that different people have varying abilities, and this, as well as effort, is a factor in achievement; up to that age, children tend to focus on effort alone. This realisation might lead to a drop in confidence for many students, and an increase in criticality towards teachers, 'blaming the teachers for academic failure' (2010, p. 529). Wigfield et al. (1991) investigated students' self-concepts of ability before and after the transition to secondary school and found that after the transition, and continuing through the first year of the new school, there was a significant drop in students' beliefs in their ability in academic subjects, sport and socially.

A further factor, in England at least, could be the lack of value placed on KS3 by some secondary schools. Following around 2000 inspections, 10,000 online questionnaires and some interviews with students in KS3 and 100 interviews with senior leaders, Ofsted concluded in 2015 that secondary

schools prioritise pastoral aspects of the transition over academic and that this is academically detrimental to the most able students in particular (2015, p. 7). Ofsted found that 85% of secondary school leaders who they interviewed said that KS4 and KS5 were prioritised in allocating staff, and KS3 subject classes were more likely to be split across two teachers and/or taught by teachers with a different subject specialism (2015, pp. 6-7).

Evans et al. write of an 'interplay between academic achievement, social schema, learning schema, and academic self-concept' (2018, p. 5); that is, they write, each of these is not an isolated factor. They point out that students experiencing difficulties immediately post-transition may continue to struggle throughout secondary school, and 'these effects can snowball and lead to future decreases in student achievement or engagement' (2018, p. 5). Zeedyk et al. (2003) assumed that the 'stress and worry [of the transition] can lead to decreases or even reversals in academic performance, school attendance or self-image' (2003, p. 68). Rice puts it more starkly, 'Disruption causes distraction that can undermine academic progress' (2001, p. 389).

The drop in attainment at transition seems especially marked in students from lower SES backgrounds (Cook et al., 2020; Hopwood et al., 2016; Wilson, 2011; Serbin et al., 2013). McGee et al. (2003) write about 'readiness' for secondary school, a notion which covers academic qualities and social and psychological qualities such as self-esteem. They assert that some of these qualities are seen less frequently in schools with lower SES and significant EAL (English as an Academic Language) populations. Children from lower SES backgrounds are already at-risk academically and are likely to have developed fewer skills and have lower attainment at the point of transition, according to test results and teacher assessment (Higgins et al., 2016; DfE 2016; Nunes et al., 2017). They are also more likely to lack the resources and support to cope successfully with the stress (Serbin, 2013) in the way that more privileged young people can.

There is little specific description of the details of the academic issues in the studies that claim this, in contrast to the wealth of detail about social, psychological and emotional factors. The claimed 'dip' accords with many teachers' anecdotal experience. While we recognise that it is not universal, it is common and widespread enough to be a cause for concern, and its possible prevalence in lower SES students makes it a social justice issue. There is some awareness among teachers in England that language could be a part of the problem but to date relatively little research on this. The next section introduces the issue.

## Language and the transition

We noted earlier that there is evidence from many countries of an academic dip, and it is tempting to make the initial assumption that students new to secondary school are struggling to meet higher academic expectations. However, transition studies suggest that this is not the case; indeed, in the

UK, there are suggestions that KS3 work, in itself, is fairly undemanding for many students. The 2015 Ofsted report on KS3 cited above suggested that for some students and in some subjects, academic work tends to be repetitious and not challenging enough. Ofsted found that it was common for secondary school teachers to underestimate Year 7 students' potential because they did not appreciate the level of attainment reached in KS2. This could result in some Year 7 students becoming bored and a degree of stagnation. Ofsted's report echoes McGee et al.'s finding (2003) that secondary school work increases in volume but not difficulty. In a blog about literacy in Year 7, Durran, an experienced teacher and advisor, raises similar questions (2017). Chedzoy and Burden (2005) interviewed students before and after transition and report that while a majority (60%) had anticipated that they would have to work harder academically, only 40% reported in Year 7 that this had turned out to be the case, and over half of Year 7 students found work in their new school to be too easy. At the same time though, they 'felt over-burdened with homework, for which they could see little value, and which considerably restricted their out-of-school activities' (2005, p. 33). These studies would suggest that the cause of the widely seen KS3 dip is not the challenge of the new work, but rather its context, volume and perhaps its presentation, against the backdrop of known stressors such as social and structural issues. There have been changes to the National Curriculum since these studies were conducted, and growing awareness in both primary and secondary sectors of the desirability of continuity. It may well be that the issue of repetition of KS2 material and generally unchallenging Year 7 work has been largely resolved. Nonetheless, challenges at KS3 are stubbornly persistent, and we discuss another possible contributor: academic language.

There is growing awareness among teachers and researchers that part of the challenge for students moving into Year 7 is linguistic (e.g., Quigley, 2017a, 2017b). One of the secondary school teachers who we talked to before starting our project voiced this view: 'Children are able to think but they can't articulate their thoughts because of the lack of language. It is not the concepts they are finding difficult at KS3; it is the ability to access material given to them'.

In this section, we introduce the language issue, which is the topic of the rest of this book. KS2 data on English and language skills suggest that there is a specific language and literacy issue for some students. Higgins et al. (2016) reported that in 2013, 14% of students, or one in seven, made the transition to secondary school with reading below the nationally expected level, as measured by the KS2 SATs taken in the summer term of Year 6. In 2017, the proportion of students not meeting the nationally expected level of reading had increased to 28% (DfE, 2017). There is a persistent association between under-achievement and economic disadvantage. In 2017, 15% of school students in Year 6 were known to be eligible for Free School Meals (FSM) due to low family income (DfE, 2017). Of this FSM group, the DfE reported that only 43% achieved the expected standards on reading, writing

and mathematics, compared with 64% of other students (percentages were not broken down by subject). The association of academic under-achievement with economic disadvantage persists in secondary school. Cook et al. (2020) found that in KS3, economically disadvantaged students fell still further behind their peers.

Sadly, it is more usual than not for the gap between underachievers and the rest to widen rather than narrow as students continue through secondary school. Of those who are below the nationally expected levels in English overall at the end of KS2, typically only 11% will go on to gain what is regarded as the baseline level of achievement in the English and Welsh national examinations at age 16, that is, five good GCSE passes including English language and mathematics (Higgins et al., 2016). This underachievement carries the danger of being excluded from higher academic education, given that success on resits for English language and mathematics is only between 20% and 34% (Ofqual, 2019), and that for most higher education courses, passes in both are required, in addition to A Levels or equivalent qualifications. Ultimately, given the current polarisation in educational outcomes in the UK, these students run the risk of becoming 'part of a precariat class' (Roberts, 2019, p. 1).

The studies described above examined performance in KS2 English, which we have interpreted as a proxy for language skills. Spencer et al. (2016) used more diverse and specific measures of language ability, 'a battery of language assessments selected to investigate: receptive skills at word, sentence, and narrative level and expressive skills using a narrative task' (2016, p. 187), with students aged 13 and 14. They compared results from these with GCSE data from the same students two years later, in English language, English literature and mathematics. They found that language skills were associated with achieving grades A*– C, as was socio-economic background. (A*– C were the grades then regarded as a 'good' pass for employment and further and higher education purposes; in the current 1–9 GCSE grading system, 'good' is understood as 4 or above.) Vocabulary skills were particularly important, using a test on receptive vocabulary knowledge, for outcomes in GCSE mathematics as well as English. Spencer et al. concluded that language was 'strongly implicated in predicting educational outcomes' (2016, p. 194). Nunes et al. (2017) similarly found a strong link between literacy and achievement in science.

As they progress through the school system, the language that students encounter and need to be able to handle becomes more specialised, and increasingly less like non-academic language (Schleppegrell, 2001). Quigley, a UK-based teacher educator and practitioner-researcher, observes a change in genre, and how this results in changes in register or 'academic codes' over the journey through primary and secondary school. He writes:

> As children advance through primary school, they progressively move away from story-driven reading primarily based on action-filled lived

> experiences. What we read and how we write necessarily shifts to a more tricky, academic style […] by the time they reach secondary school, they are expected to move between multiple, discrete disciplines in a single day. For many young people, the complexity of the very different academic codes they need to crack in order to achieve and thrive is frankly bewildering.
>
> (2017a, p. 65)

While this change may be gradual for the most part, research in the UK and elsewhere has found that there is a noticeable difference in academic language between primary and secondary schools (Braund & Driver, 2005, p. 78). Martin writes that the transition sees a change from 'a concern with basic literacy and numeracy, often taught in general terms, to subject-based teaching and learning involving highly specialised discourse of various kinds' (2013, p. 23). In the history and biology lessons that Martin observed in Australian secondary schools, he saw no teaching of disciplinary reading or writing. It was apparently assumed that students would already have these skills, and the necessary language knowledge would be in place (2013, p. 34).

UK teachers have also begun to write about this challenge of the transition. Durran (2017) notes that students go from a single literacy teacher in primary school to subject specialist teachers who may have little awareness of the demands that their disciplinary genre places on children. Quigley has described a 'language leap' at transition (2017b). He notes that some KS2 primary English work is very sophisticated in its use of terminology, so this is not a simple picture. Part of the issue in his view is the breadth and detail of coverage of subjects at KS3, each subject bringing its own disciplinary language. In a series of articles and books for teachers, he draws attention particularly to vocabulary (2016, 2017a, 2017b, 2018, 2020), also arguing that the language needed in secondary school is becoming more challenging with new curricula and exam specifications in recent years. In Chapter 2, we discuss existing studies of the language of school, which frame the research findings that we describe in the later chapters of this book.

As well as changes to the genres, registers, grammar and vocabulary of the language of school, that is, the qualitative changes, we believe that there are changes to the quantity of language that students encounter and interaction patterns. Tobbell and O'Donnell (2013) followed some Year 7 students through their school day at various sites in England. They saw a common pattern of teachers talking at length to the students, meaning that 'students may have spent well over half their day sitting in silence and listening to teachers talk' (2013, p. 21). This is in contrast to a typical primary classroom, which is more likely to be task-focused, with students spending a considerable amount of time talking to themselves and each other. The data that we have gathered for this project, which we describe in Chapter 3, shows similar differences in quantity. Written texts that students need to read to access the curriculum

include worksheets, textbooks and PowerPoint presentations. We found that, on average, these contain many more words, much more densely crowded onto the page or screen in Year 7, in contrast to Year 6. Our spoken data consist of teacher talk (we did not transcribe student talk), and again, an average teacher presentation consists of many more words in Year 7 than in Year 6.

In the 13 schools that we worked with on this project, the working day in both primary and secondary schools consists of around five hours. In primary schools, this is largely taken by a single teacher, in the same classroom. By contrast, in the secondary schools, with very few exceptions, each of the five hours was in a different location, with a different teacher, at a very intense pace and with a smaller proportion of time spent working on solo or group tasks. The overall result is that secondary students have to cope with a very large increase in quantity of receptive language, compared with their experience in primary school. This increased volume of language also contains unfamiliar genres and vocabulary, and academic grammatical structures that they rarely encounter outside school.

Unfortunately, Year 7 students are especially poorly placed to cope with this quantitative and qualitative step change in language. We know that they are stressed by social and practical issues and that they are on the verge of adolescence which brings its own strains. Stress has been shown to negatively impact effective learning (Pascoe et al., 2020) and memory formation and retrieval (Vogel & Schwabe, 2016). Little can be done about some of the stressors, but a better understanding of the nature of the language challenge that students face would help educators to support them. We have shown that teachers and researchers recognise and are sympathetic to both transition issues and academic language issues. However, as yet there is no large-scale study bringing the two together to identify the nature of the challenge at the KS2 to KS3 transition. This is what our project aimed to do, through a corpus study that gathered and analysed around two and a half million words of written and spoken data from 13 schools in the north of England. Our data, methods and some of our findings are explained in later chapters of this book. First though, we turn to students from some of the 13 participating schools and present some of the ideas they shared with us about language in order to give readers a sense of their voices.

## The voices of students in our project schools

Earlier in this chapter, we mentioned some previous studies that elicited primary school students' thoughts about what the transition would be like. These found, on the whole, that if they were worried, it was mostly about social and structural issues. We found very little research that had investigated students' views on the language of school. Phillips Galloway et al.'s study (2015) is one of the few that do this, but it does not cover the transition. Meston et al. (2020) interviewed teachers and students about their views on academic talk, which yielded some student mentions of

language forms, again though not covering the transition specifically. We wanted to find out what students in our region, northern England, thought about the language of school, before and after the transition, and whether they are aware of it as a possible problem or not. We do not present the full results and analysis here; Chambers (2020) has conducted an initial overview of the interviews with Year 7 pupils, and we will present the Year 6 and 7 analysis in full elsewhere (Deignan & Oxley, in preparation). In this introductory chapter, we cite some short extracts from the interviews as part of setting the scene for the project, to give the reader a flavour of the students' voices.

Several writers have noted that transition, while a one-time event in terms of the physical relocation to a different school, is a longer process when considered as a period of adjustment (Jindal-Snape & Cantali 2019; Rice et al., 2011). Longitudinal studies suggest that when then they first start secondary school, students may experience a 'honeymoon' effect early on, feeling less positive a little later in Year 7 (Bagnall et al., 2019; Chedzoy & Burden, 2005). With this in mind, we planned a series of interviews with the same students across Years 6 and 7. The students who we interviewed were all in a relatively secure situation, in that each of their primary schools 'feeds' a particular secondary school, so by and large the class make the transition as a cohort. Nonetheless, they all moved to a much bigger secondary school, often had a longer journey to school and were put in classes with many children who they had not met before, with a wider socio-economic mix than they had previously experienced.

We interviewed 30 students in small focus groups when they were in Year 6 at five different primary schools.[2] Each group contained six students, and we conducted ten interviews, two with each group, the first in March and April 2019, and the second in June 2019, after they had taken KS2 SATs. In September 2019, the students started at three different secondary schools, where we interviewed them again in the autumn term. Some students were interviewed again in Year 7, in early March 2020, in total yielding seven Year 7 interviews. Not all the second Year 7 interviews were possible because schools in Britain closed to most children on 20 March 2020 due to the Covid-19 pandemic. (Only vulnerable children and children of key workers continued to attend school after that date, with other children learning at home.) Once schools reopened some months later, we felt that the online learning experience, and wider stresses from the pandemic, would most likely have eclipsed students' impressions of the transition. We therefore have a dataset of 17 group interviews of around 30–40 minutes each. The interviews were transcribed by different members of the project team, and each transcript was between 6500 and 9500 words in length. In addition to checking for accuracy by the original interviewer, they have been read by three of the project team and coded and themed using NVivo. Here we highlight some language-related points that they made.

In each of the Year 6 focus group interviews, the researcher asked students for their feelings about the upcoming transition, and about the transition activities that they had done so far. She asked about what they looked forward to and thought they might find difficult, then asked some focused questions on language. The students mentioned a number of subject-specific words that they found difficult to remember, including *rhombus*, *trapezium*, *isosceles* (mathematics), *aorta* (science) and *homophone* (English). The researcher asked them to read short passages from school texts, as a prompt for further discussion of language; students did not know *glucose*, *diffusion* and *respiration*. There was some developing awareness of polysemy, seen in the following extract. Here, the students and the researcher discuss meanings of *concentration*, which in the science text they are reading refers to the quantity of a substance in a solution.

| *line* | *speaker* | *utterance* |
|---|---|---|
| 721 | Researcher | it's quite normal not to understand by the way so this is not an exam we'll just talk about this what about 'concentration' in this text? |
| 722 | Elsie May | [whispers] where is it? |
| 723 | Eleanor | concentration |
| 724 | Researcher | so this is in the middle |
| 725 | Rosie | [whispers] concentration |
| 726 | Elsie May | I know what concentration is but like it's different in here I think |
| 727 | Researcher | what is concentration? |
| 728 | Elsie May | it's where you're like you're really focusing on something |
| 729 | Researcher | yeah you're focused on is this the same thing here in this text? |
| 730 | Elsie May | no? |
| 731 | Researcher | no? |
| 732 | Elsie May | wait is it? |
| 733 | Rosie | [whispering] it's <u=?> different |
| 734 | Elsie May | no it's different |

*Extract 1.1,* pupil interview, School F.

Students also referred to 'technical' words in a number of interviews, including *parallel*, and in the following brief discussion, *depth*.

| *line* | *speaker* | *utterance* |
|---|---|---|
| 513 | Researcher | why is it [depth] a technical word for you? |
| 514 | James | because there's height depth and width |
| 515 | Researcher | yeah |
| 516 | James | and it's confusing which is depth |

*Extract 1.2,* pupil interview, School F.

Students are also concerned to use what they see as more 'academic' vocabulary, of a more formal register, and are encouraged to do this. They have

various ways of describing this (Chambers, 2020), including 'higher level', and they talk about being told to 'up-level their words'. One example they gave was planning to write 'big, and then make it colossal', 'or gargantuan'. In another school, students talked about 'posh' words, in the following utterances:

| *line* | *speaker* | *utterance* |
|---|---|---|
| 408 | Maddie | climate because that's just like a posh word for weather |

*Extract 1.3,* pupil interview, School A.

| *line* | *speaker* | *utterance* |
|---|---|---|
| 435 | Chloë | yeah cos if you said I'm tired so I'm going to bed then that's what you would say normally but then if you were talking posh you'd be like I'm exhausted therefore I'm going to bed you'd speak it more formal |

*Extract 1.4,* pupil interview, School A.

Students in the same school mentioned 'ambitious words':

| *line* | *speaker* | *utterance* |
|---|---|---|
| 371 | Elsie May | yeah because we have our own personal targets and mine is to use more ambitious vocabulary |
| 372 | Researcher | what do you mean by ambitious vocabulary? |
| 373 | Students | harder words |
| 374 | Elsie May | words that you wouldn't use in year four or year five |
| 375 | Huxley | cos like we have the homework |
| 376 | Researcher | could you give us an example of better words |
| 377 | students | discrimination individuality inquisitively |

*Extract 1.5,* pupil interview, School F.

In many exchanges such as the one below students seem to regard 'higher-level' words as synonyms for more everyday words:

| *line* | *speaker* | *utterance* |
|---|---|---|
| 519 | Zair | it's got like more higher level words |
| 520 | Researcher | mhm what are the high level words here? |
| 521 | Zair | like |
| 522 | Brann | [whispering] unsub-tle sub-tle |
| 523 | Zair | like you could just say it was like horrible but they put traumatic |

*Extract 1.6,* pupil interview, School A.

In Meston et al.'s (2020) investigation of students' and educators' understanding of academic talk, a number of students mentioned 'fancy words', apparently roughly the same notion. The students who we interviewed talked about some words being 'better'.

These informal observations from our Year 6 interviews are consistent with the small number of studies in the literature on students' views of academic language. While not specifically about the transition, Phillips Galloway et al.'s study (2015) elicited students' metalinguistic reflections on academic language. They asked students in grades 4–8 in the United States to evaluate some sample texts. Their participants wrote about 'better words', 'longer words', 'detailed words' and words that 'explain more', which the authors interpret as referencing lexical precision. Phillips Galloway et al. (2015) write that in learning academic language, students learn on two levels: they learn the forms and meaning of new language, and simultaneously, they develop 'metalinguistic awareness of the academic register' (2015, p. 221); problems with academic language might stem from either of these. The academic metalanguage enables teachers and students to talk about academic language, how and when it facilitates communication and in which contexts. In our interviews with students, we were asking them to operate on both levels.

Our third round of interviews, when the students had transitioned to secondary school, tackled academic language in more detail. The students in each group were at the same primary school, but are no longer studying together, as they have been split into different tutorial forms and possibly different ability groups. They knew the researcher quite well by this point, as she interviewed them twice when they were in Year 6. As part of the interviews in the third round, the researcher again asked the students to read several short passages and identify words that they didn't know or found difficult. The data show many instances of groups working together to try to find the disciplinary meaning of a word from another meaning that they knew. We look at a few examples from an interview with six students that took place early in their Year 7. They begin by discussing a short passage about electricity. There are several examples of them working collaboratively to work out the disciplinary meaning of unknown words, through comparison with their more familiar meanings, showing that they are well aware of polysemy.

There is a short discussion about *conductor*, in which one student says, 'I think it's isn't that like that person who goes like like stands like their hand like', referring to the meaning denoting the conductor of an orchestra. The researcher says, 'oh in music' and another student says, 'this is science', apparently to orient her classmates to the discipline of the text and therefore a different meaning.

In the following extract later in the same interview, the researcher asks the students about *current*, and again, there is a discussion about the contextual, disciplinary meaning, and the more familiar meaning. All six students contribute, to a greater or lesser degree.

| *line* | *speaker* | *utterance* |
|---|---|---|
| 507 | Researcher | okay what about current? how much current will flow around the circuit? |
| 508 | Gabbie | oh I don't get that |
| 509 | Researcher | what is current? |
| 510 | Gabbie | I don't know |
| 511 | Brann | oh I do |
| 512 | Zair | like the climate |
| 513 | Chloë | is it like? No |
| 514 | Brann | the current in the sea like the current or the current |
| 515 | Students | oh |
| 516 | Brann | goes out... |
| 517 | Maddie | it's like a thing goes through something or something like that |
| 518 | Chloë | yeah |
| 519 | Brann | it's pushing you back out |
| 520 | Researcher | could you please repeat that again? |
| 521 | Brann | current like in the sea |
| 522 | Researcher | in the sea |
| 523 | Brann | sometimes like the current is really rough |
| 526 | Zair | uh? |
| 527 | Gabbie | and it pulls you out to sea |
| 528 | Brann | pulls you out to sea yeah |
| 529 | Researcher | Is this the same thing here though |
| 530 | Gabbie | the temperature |
| 531 | students | [laughing] |
| 532 | Maddie | no |
| 534 | Researcher | is this the same thing here though? |
| 535 | Asher | no |
| 536 | Gabbie | it could be |
| 536 | Asher | no |
| 538 | Brann | could be cos it's pushing it out |
| 539 | Gabbie | yeah like pushing the electrons |
| 540 | Asher | isn't it the way that like electricity moves through the wires here? |

*Extract 1.7,* pupil interview, School K.

We thought that these students are aware of the frequent tendency for specialist terms to be related by metaphorical comparison with meanings that might be more familiar to them. However, they struggle in identifying the 'grounds' for the metaphor, that is, the attribute of the literal meaning that is exploited. In the case of electrical *current*, this is the constant, unidirectional movement of a channel of water. Brann says that he knows the meaning (line 511), but in lines 514, 521 and 523, it becomes apparent that he is referencing the sea, as a body of water, rather than a river, or a current within the sea. He talks about the potential of a current to become rough and dangerous, in lines 523 and 528, and another student, Gabbie, joins in with the analogy. The confusion is resolved by Asher, in line 540, who may have already known the scientific meaning.

Students have often been told to work out meaning from context. In the following extract, from a different focus group and secondary

school, Jay, a Year 7 student, is confident that he understands the word *fertile* from context.

| *line* | *speaker* | *utterance* |
|---|---|---|
| 173 | Jay | I found a couple of words challenging shall I say which ones? |
| 174 | Researcher | yeah yeah yeah or you could show if you can't spell that's alright |
| 175 | Jay | there's one like romans like R O M A N |
| 176 | Researcher | Romanesque? |
| 177 | Jay | yeah |
| 178 | Researcher | yep what else? |
| 179 | Jay | edginess |
| 180 | Researcher | yeah |
| 181 | Jay | and that's it in the second one |
| 182 | Researcher | only that? |
| 183 | Jay | yeah |
| 184 | Researcher | there is a word Granada's fertile valley what is fertile? |
| 185 | Jay | fertile is like is it like a range of things it's like like there's a range of different things and it's kind of |
| 186 | Researcher | so you could infer the meaning of fertile here? |
| 187 | Jay | it's like so let's say there was like a restaurant like an Italian restaurant and different kinds it would kinda be like fertile cos like there's like different things there and it's not all like the same and it's like different like cultures and stuff |

*Extract 1.8,* pupil interview, School L.

Jay appears to be a confident and self-aware student, identifying some low-frequency vocabulary in lines 176 and 179. He does not see *fertile* as a difficult word, but his explanations in lines 185 and 187 show that he does not have an accurate understanding of it. It occurs in the sentence 'Granada's fertile valley and sweeping hills have attracted many different civilisations throughout the centuries'; his incorrect definition would be plausible, showing the limitations of inferring from context.

As well as subject-specific lexis, there is more general academic vocabulary that the students struggle with. Shortly after the exchange in Extract 1.7, the students move on to discuss *model*, which is used in the text in a metalinguistic sense, inviting the reader to compare a central heating system with an electrical circuit. The students attempt to interpret the academic meaning with reference to their everyday experience, one student suggesting that the meaning is 'like a model plane', but they do not come to a satisfactory understanding of its meaning in context. This is followed by an exchange over the meaning of *disprove*, which most of the group do not seem to understand. They discuss whether it means the same as *disapprove* or is related to *disappoint*. Other focus groups, discussing the same passages, also had difficulty with this word. General academic words such as these are often termed Tier 2 vocabulary, as opposed to

discipline-specific words such as *current* and *conductor*, termed Tier 3. Other examples of problematic Tier 3 vocabulary listed by Year 7 students in our interviews include: from science, *chloroplast*, *cell membrane*, from history, *motte and bailey castle*, from English, *oxymoron*, and from mathematics, *product*. The classification of vocabulary into tiers is discussed in more detail in Chapter 2.

These student interviews were consistent with our belief that students at the transition do not yet have the academic and specialist vocabulary that they will need to be successful academically as they progress through school. This view is shared by the teachers who we interviewed, from their primary and secondary schools, as well as by the wider teacher community, as discussed previously. They are also consistent with the views of writers such as Phillips Galloway et al. (2015) and Meston et al. (2020), who conducted similar research in the United States.

## Aims of this book

In this chapter, we have argued that academic language is a barrier to some students, making it more difficult for them to access the curriculum and achieve their intellectual potential. We have discussed the transition from primary to secondary school, suggesting that the language of secondary school may make KS3 more difficult and contribute to the academic dip often seen. The aim of the research described in the rest of this book was to describe how the language of secondary school differs from the language of primary school. Our main research method is corpus linguistics. We use a number of different techniques from the discipline to interrogate our data. We describe the corpora that we created and describe our methods in Chapter 3. In the following chapter, we review existing descriptions of the language of school.

## Notes

1 https://www.sundaypost.com/features/ever-called-teacher-mum-top-10-primary-school-rites-passage-revealed/
2 Ethical approval for this study, and for all the project activities described in this book, was given by the Ethics Committee of the Faculty of Social Sciences, University of Leeds. Students and their parents or guardians consented to the interviews twice, in Year 6 and again in Year 7.

## References

Bagnall, C. L., Skipper, Y., & Fox, C. L. (2019). "You're in this world now": Students', teachers', and parents' experiences of school transition and how they feel it can be improved. *British Journal of Educational Psychology*, *90*(1), 206–226. https://doi.org/10.1111/bjep.12273

Banks, J., & Smyth, E. (2015). "Your whole life depends on it": Academic stress and high-stakes testing in Ireland. *Journal of Youth Studies*, *18*(5), 598–616. https://doi.org/10.1080/13676261.2014.992317

Braund, M., & Driver, M. (2005). Pupils' perceptions of practical science in primary and secondary school: Implications for improving progression and continuity of learning. *Educational Research*, *47*(1), 77–91. https://doi.org/10.1080/0013188042000337578

Bru, E., Stornes, T., Munthe, E., & Thuen, E. (2010). Students' perceptions of teacher support across the transition from primary to secondary school. *Scandinavian Journal of Educational Research*, *54*(6), 519–533. https://doi.org/10.1080/00313831.2010.522842

Chambers, G. (2020, October 29). *Linguistic encounters: What Year 7 students say about* '*difficult words*'. https://linguistictransition.leeds.ac.uk/linguistic-encounters-what-year-7-students-say-about-difficult-words/.

Chedzoy, S. M., & Burden, R. L. (2005). Making the move: Assessing student attitudes to primary-secondary school transfer. *Research in Education*, *74*(1), 22–35. https://doi.org/10.7227/rie.74.3

Clark, U. (2019). *Developing language and literacy in English across the secondary school curriculum: An inclusive approach*. Palgrave Macmillan.

Coffey, A. (2013). Relationships: The key to successful transition from primary to secondary school? *Improving Schools*, *16*(3), 261–271. https://doi.org/10.1177/1365480213505181

Cook, W., Shaw, B., & Morris, S. (2020). *Disadvantage in early secondary school*. Nuffield Foundation.

Deignan, A., & Oxley, F. (in preparation). The language challenge of the primary-secondary transition: Students' voices.

Department for Education (DfE). (2016). *Understanding school and college performance measures*. https://www.gov.uk/government/publications/understanding-school-and-college-performance-measures

Department for Education (DfE). (2017). *National curriculum assessments at key stage 2 in England*. https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/667372/SFR69_2017_text.pdf

Duffy, G., & Elwood, J. (2013). The perspectives of "disengaged" students in the 14–19 phase on motivations and barriers to learning within the contexts of institutions and classrooms. *London Review of Education*, *11*(2), 112–126. https://doi.org/10.1080/14748460.2013.799808

Durran, J. (2017). *Avoiding a literacy dip in year 7*. https://jamesdurran.blog/2017/08/23/avoiding-a-literacy-dip-in-year-7/

Eskelä-Haapanen, S., Vasalampi, K., & Lerkkanen, M.-K. (2020). Students' positive expectations and concerns prior to the school transition to lower secondary school. *Scandinavian Journal of Educational Research*, *65*(7), 1252–1265. https://doi.org/10.1080/00313831.2020.1791248

Evans, D., Borriello, G. A., & Field, A. P. (2018). A review of the academic and psychological impact of the transition to secondary education. *Frontiers in Psychology*, *9*, 1482. https://doi.org/10.3389/fpsyg.2018.01482

Felmlee, D., McMillan, C., Inara Rodis, P., & Osgood, D. W. (2018). Falling behind: Lingering costs of the high school transition for youth friendships and grades. *Sociology of Education*, *91*(2), 159–182. https://doi.org/10.1177/0038040718762136

Gee, J. P. (2004). *Situated language and learning: A critique of traditional schooling*. Routledge.

Goldstein, S. E., Boxer, P., & Rudolph, E. (2015). Middle school transition stress: Links with academic performance, motivation, and school experiences. *Contemporary School Psychology*, *19*(1), 21–29. https://doi.org/10.1007/s40688-014-0044-4

Higgins, S., Katsipataki, M., & Coleman, R. (2016). *Reading at the transition: Interim evidence brief*. Education Endowment Foundation. https://educationendowmentfoundation.org.uk/education-evidence/evidence-reviews/reading-at-the-transition

Hopwood, B., Hay, I., & Dyment, J. (2016). The transition from primary to secondary school: Teacher perspectives. *Australian Journal of Educational Research*, *43*, 289–307.

Jindal-Snape, D., & Cantali, D. (2019). A four-stage longitudinal study exploring pupils' experiences, preparation and support systems during primary-secondary school transitions. *British Educational Research Journal*, *45*(6), 1255–1278.

Jindal-Snape, D., Cantali, D., MacGillivray, S., & Hannah, E. (2019). *Primary-secondary transitions: A systematic literature review*. University of Dundee, Scottish Government. https://www.gov.scot/publications/primary-secondary-transitions-systematic-literature-review/

Jindal-Snape, D., & Foggie, J. (2008). A holistic approach to primary-secondary transitions. *Improving Schools*, *11*(1), 5–18.

Klinger, D. A., Freeman, J. G., Bilz, L., Liiv, K., Ramelow, D., Sebok, S. S., Samdal, O., Dur, W., & Rasmussen, M. (2015). Cross-national trends in perceived school pressure by gender and age from 1994 to 2010. *The European Journal of Public Health*, *25*(suppl 2), 51–56. https://doi.org/10.1093/eurpub/ckv027

Leung, C. (2014). Researching language and communication in schooling. *Linguistics and Education*, *26*, 136–144.

Martin, J. R. (2013). Embedded literacy: Knowledge as meaning. *Linguistics and Education*, *24*, 23–37.

McGee, C., Ward, R., Gibbons, J., & Harlow, A. (2003). *Transition to secondary school: A literature review*. Ministry of Education, New Zealand. https://www.educationcounts.govt.nz/publications/schooling/5431

Meston, H. M., Phillips Galloway, E., & Brown McClain, J. (2020). 'They're the ones who hold the answers': Exploring educators' and students' conceptions of academic conversation. *Journal of Adolescent and Adult Literacy*, *64*(4), 409–491.

Nisbet, J. D., & Entwistle, N. J. (1969). *The transfer from primary to secondary education*. University of London Press.

Nunes, T., Bryant, P., Strand, S., Hillier, J., Barros, R., & Miller-Friedmann, J. (2017). *Review of SES and science learning in formal educational settings. A report prepared for the EEF and the royal society*. Education Endowment Foundation. https://royalsociety.org/news/2017/09/eef-royal-society-publish-evidence-review-science-attainment-gap/

Ofqual. (2019). *Guide to GCSE results for England, 2019*. https://www.gov.uk/government/news/guide-to-gcse-results-for-england-2019

Ofsted. (2015). *Key stage 3: The wasted years?* Crown Copyright. https://www.gov.uk/government/publications/key-stage-3-the-wasted-years

Pascoe, M. C., Hetrick, S. E., & Parker, A. G. (2020). The impact of stress on students in secondary school and higher education. *International Journal of*

*Adolescence and Health*, *25*(1), 104–112. https://doi.org/10.1080/02673843.2019.1596823

Patterson, A. (2020). Sustaining disciplinary literacy in science: A transformative just model for teaching the language of science. *Journal of Adolescent and Adult Literacy*, *64*(3), 333–336.

Phillips Galloway, E., Stude, J., & Uccelli, P. (2015). Adolescents' metalinguistic reflections on the academic register in speech and writing. *Linguistics and Education*, *31*, 221–237.

Quigley, A. (2016). *One word at a time - Teaching vocabulary*. https://www.theconfidentteacher.com/2016/11/8181/

Quigley, A. (2017a). Cracking the academic code. *Times Educational Supplement*, 7th Sept 2017, p. 65.

Quigley, A. (2017b). *The 'Language Leap' at transition*. https://www.theconfidentteacher.com/2021/04/the-language-leap-at-transition/

Quigley, A. (2018). *Closing the vocabulary gap*. Routledge, Taylor & Francis Group.

Quigley, A. (2020). *Teaching vocabulary –4 ways to boost secondary language skills*. https://www.teachwire.net/news/teaching-vocabulary-4-ways-to-boost-secondary-language-skills

Rice, J. K. (2001). Explaining the negative impact of the transition from middle to high school on student performance in mathematics and science. *Educational Administration Quarterly*, *37*(3), 372–400. https://doi.org/10.1177/00131610121969352

Rice, F., Frederiksson, N., & Seymour, J. (2011). Assessing pupil concerns about transition to secondary school. *British Journal of Education Psychology*, *81*, 244–263.

Rice, F., Frederickson, N., Shelton, K., McManus, C., Riglin, L., & Ng-Knight, T. (2015). *Identifying factors that predict successful and difficult transitions to secondary school*. Nuffield Foundation. https://nuffieldfoundation.org/wp-content/uploads/2019/11/STARS_report.pdf

Roberts, K. (2019). Routes through education into employment as Britain enters the 2020s. *European Journal of Educational Management*, *2*(1), 1–11. https://doi.org/10.12973/eujem.2.1.1

Savory, J. (2022). *Are your teenage years really the happiest?* https://www.thisislocallondon.co.uk/news/19866949.teenage-years-really-happiest--jasmine-savory-st-philomenas-high-school/

Schleppegrell, M. J. (2001). Linguistic features of the language of schooling. *Linguistics and Education*, *12*(4), 431–459. https://doi.org/10.1016/s0898-5898(01)00073-0

Serbin, L. A., Stack, D. M., & Kingdon, D. (2013). Academic success across the transition from primary to secondary schooling among lower-income adolescents: Understanding the effects of family resources and gender. *Journal of Youth and Adolescence*, *42*(9), 1331–1347. https://doi.org/10.1007/s10964-013-9987-4

Spencer, S., Clegg, J., Stackhouse, J., & Rush, R. (2016). Contribution of spoken language and socio-economic background to adolescents' educational achievement at age 16 years. *International Journal of Language & Communication Disorders*, *52*(2), 184–196. https://doi.org/10.1111/1460-6984.12264

Strand, G. M. (2019). Experiencing the transition to lower secondary school: Students' voices. *International Journal of Educational Research*, *97*, 13–21. https://doi.org/10.1016/j.ijer.2019.06.009

Symonds, J., & Hargreaves, L. (2016). Emotional and motivational engagement at school transition: A qualitative stage-environment fit study. *Journal of Early Adolescence*, *36*(1), 54–85. https://doi.org/10.1177/0272431614556348

Tobbell, J., & O'Donnell, V. L. (2013). The formation of interpersonal and learning relationships in the transition from primary to secondary school: Students, teachers and school context. *International Journal of Educational Research*, *59*, 11–23. https://doi.org/10.1016/j.ijer.2013.02.003

Topping, K. (2011). Primary–secondary transition: Differences between teachers' and children's perceptions. *Improving Schools*, *14*(3), 268–285. https://doi.org/10.1177/1365480211419587

van Rens, M., Haelermans, C., Groot, W., & Maassen van den Brink, H. (2017). Facilitating a successful transition to secondary school: (How) does it work? A systematic literature review. *Adolescent Research Review*, *3*(1), 43–56. https://doi.org/10.1007/s40894-017-0063-2

Virtanen, T. E., Vasalampi, K., Torppa, M., Lerkkanen, M.-K., & Nurmi, J.-E. (2019). Changes in students' psychological well-being during transition from primary school to lower secondary school: A person-centered approach. *Learning and Individual Differences*, *69*, 138–149. https://doi.org/10.1016/j.lindif.2018.12.001

Vogel, S., & Schwabe, L. (2016). Learning and memory under stress: Implications for the classroom. *Nature Partner Journals Science of Learning*, *1*, 16011. https://doi.org/10.1038/npjscilearn.2016.11

West, P., Sweeting, H., & Young, R. (2010). Transition matters: Pupils' experiences of the primary–secondary school transition in the West of Scotland and consequences for well-being and attainment. *Research Papers in Education*, *25*(1), 21–50. https://doi.org/10.1080/02671520802308677

Wigfield, A., Eccles, J. S., Mac Iver, D., Reuman, D. A., & Midgley, C. (1991). Transitions during early adolescence: Changes in children's domain-specific self-perceptions and general self-esteem across the transition to junior high school. *Developmental Psychology*, *27*(4), 552–565. https://doi.org/10.1037/0012-1649.27.4.552

Wilson, P. (2011). *A rapid evidence assessment investigating the drop in attainment during the transition phase with a particular focus on child poverty*. Welsh Assembly Government Social Research. https://gov.wales/rapid-evidence-assessment-investigating-drop-attainment-during-transition-phase-particular-focus-0

Zeedyk, S. M., Gallacher, J., Henderson, M., Hope, G., Husband, B., & Lindsay, K. (2003). Negotiating the transition from primary to secondary school: Perceptions of pupils, parents and teachers. *School Psychology International*, *24*(1), 67–79.

# 2 Academic language and the school transition

*Alice Deignan*

## Introduction

In this chapter, we summarise debates around language and school and descriptions of the characteristics of the language of school. We will describe the progress that has been made towards understanding the language challenges that school students face, and argue that there remain gaps, some of which our corpus project can fill. We begin by briefly describing the most influential bodies of work that have attempted to distinguish the language of school from everyday language.

## Perspectives on the language of school

### *Bernstein's language codes and the language of school*

Many researchers and teachers have believed for a long time that success in school needs a form of language different from everyday language. A well-known and much-contested view on the subject was put forward by the sociologist Basil Bernstein, in works written over a number of years from the late 1950s on. Bernstein claimed that there are contrasting ways of communicating, which he described as 'restricted code' and 'elaborated code'. Restricted code is predominantly routinised and bound to physical and social context. Utterances and their forms are predictable, with limited vocabulary and simple syntax (1966). Bernstein claimed that in restricted code 'the meanings are likely to be concrete, descriptive or narrative, rather than analytical or abstract' (ibid). Elaborated code, by contrast, is context-independent, allowing for talk about the abstract, ideas not in the immediate here and now, and for academic analysis. Bernstein argued that differences in the codes children have access to are a major factor in educational success; 'as a child progresses through school, it becomes critical for him to possess, or at least be oriented towards, an elaborated code, if he is to succeed' (1964, p. 67). He explicitly associated his ideas with social class (1964, 1966) but not with underlying intelligence (1964, p. 58). The argument runs that while all speakers have access to restricted code, only some have access to elaborated code: 'there is a relatively high probability

DOI: 10.4324/9781003081890-2

of finding children limited to this code among sections of the working class population' (Bernstein, 1964, p. 62).

If read as a dismissal of the intellectual potential of working-class children, of course, this is abhorrent to modern educationalists. Bernstein's many critics have identified his views as a deficit position, that is, one which claims that working-class children lack something that middle-class children have (Jones, 2013). However, some write that Bernstein was misunderstood, among them Jenks (2010). This is partly sometimes through Bernstein's own choices of expression. For instance, Jenks writes, his references to the speech patterns of areas of British cities known to be socially deprived – the Gorbals (Glasgow), Tiger Bay (Cardiff) and so on – make a modern reader feel uneasy, suggesting negative stereotyping, or even worse. Jenks argues that contrary to the impression this gives, 'Bernstein was actually outraged by the inequalities and indignities visited upon the educational experience of working class children' (2010, p. 73).

Jones (2013) summarises the history of debate over Bernstein's codes and draws out the argument made by a number of researchers about the key difference between the children who feel at ease in school and go on to succeed, and those who do not feel at ease, and often do not succeed academically. He writes that the difference is not a linguistic code, but, at root, literacy. Some children arrive at school already familiar with written materials and having some of the metalanguage for talking about reading and writing, words such as *sentence*, and are able to 'sound out' letters, as a precursor to decoding, while others do not. Once pointed out, the similarity between descriptions of restricted versus elaborated code and speech versus writing is clear. For example, in an early paper, Bernstein (1959, reprinted in 2010) described formal characteristics of 'public' and 'formal' language (later 'restricted' and 'elaborated' codes) and listed ten characteristics of 'public' language. The first four are:

1. Short, grammatically simple, often unfinished sentences, a poor syntactical construction with a verbal form stressing the active mood.
2. Simple and repetitive use of conjunctions (so, then, and, because).
3. Frequent use of short commands and questions.
4. Rigid and limited use of adjectives and adverbs.

(1959, p. 54)

These are almost identical to lists of features of spoken grammar (for example, Leech, 2000), as contrasted with written grammar. Unlike Leech's work on spoken grammar however, negative evaluation is found in Bernstein's lexical choices: *poor*, *repetitive*, *rigid* and *limited* in the above description. Just as 'elaborated code' was assumed to be superior in its expressive potential to 'restricted code', so a bias towards written language is still common in some circles. There is a widely held popular belief, albeit often unvoiced, that the spoken form is a degraded and inferior version of writing (Linell,

2019). This means that reframing the distinction as one of literacy does not necessarily remove the value judgements attached to the restricted/elaborated codes model.

The debate about language, academic achievement and class has continued in various forms. In the research described here, we aim simply to describe the features of secondary school academic language. We note that many teachers and researchers believe that these are less accessible to children from lower socioeconomic status (SES) backgrounds. While this belief has contributed to our motivation for this research, we have not studied the question ourselves.

### *The Systemic-Functional Linguistics approach*

Martin (e.g., 1985, 2009, 2013), Christie (e.g., 1992, Christie & Martin 1997) and their colleagues have done seminal work on the genres and registers of schooling. Originally based in Sydney, they work within the Hallidayan Systemic-Functional Linguistics (SFL) approach and were influenced by Bernstein. Christie (2009) describes in detail the dialogue between the two strands of research, which found common ground in concerns with the nature and construction of knowledge and with the underachievement of working-class children. Educational researchers within the SFL school take a genre and register approach, using a functionally driven understanding of genre. School genres that are identified include 'explanations, reports, procedures, and expositions, and various types of narratives' (Christie, 1992, p. 146).

The group worked on the implications for schools of their approach and developed interventions with educationalists to support the development of writing (Martin, 1999). In particular, they argued against the 'whole language' approach that was predominant in Australian classrooms at the time (Martin 1993), and which it was claimed promoted an 'invisible pedagogy' (ibid, 162), which advantaged, generally, middle class children who were able to understand what was required to succeed without explicit instruction. They argued for a visible, genre-based pedagogy (Martin, 1993, 2009), 'as an issue of social justice' (Martin, 2009, p. 11). Over a period of several decades, SFL researchers have undertaken a number of studies of the registers of specific disciplines in secondary school, such as Martin's descriptions of the registers of biology and history (2013) and Coffin's work on history (2006). Work such as this will be discussed later in this book when we turn to the language of different school subjects.

### *BICS and CALP*

The debate about academic language has also concerned clusters of researchers working with second language learners, centrally Jim Cummins, working in Canada. Cummins makes a distinction between two types of

language proficiency: Basic Interpersonal Communication Skills (BICS) and Cognitive Academic Language Proficiency (CALP) (Cummins, 1980, 2008). In children's first language, CALP 'becomes differentiated from BICS after the early stages of schooling to reflect primarily the language that children acquire in school and which they need to use effectively if they are to progress through the grades' (2008, p. 72). As Cummins notes, the distinction was originally developed through studies of second language learners of English in schools. His analysis of Canadian test data suggests that it takes at least five years for second language learners to develop CALP to the level of native English language speakers in their grade, as against just two years for BICS (1981). This is supported by data from other studies in different contexts (Cummins, 2008). If this difference is not understood, children may be assessed as competent in English on the basis of their BICS; in informal conversation, they will sound proficient. They may then be assessed as not needing ongoing language support, and struggle with academic language. Teachers may perceive them as weak academically when the problem is not a lack of ability, but their completely normal delay in developing CALP. Second language learners' academic potential is thus prone to being underestimated, leading ultimately to underachievement and disengagement from education.

Cummins notes that CALP is especially associated with literacy (1980). Leung (2014) also notes the association but points out that informal conversational language is often interleaved with CALP in lessons. This may indeed help to make lesson content more accessible to second language learners, but it could even compound their difficulties in understanding written academic texts and producing appropriate written language themselves when assessed. As empirical evidence for the existence of the distinction between BICS and CALP, Cummins (2008) notes work by the corpus linguists Biber (1986) and Coxhead (2000). He notes that their research demonstrates through analysis of naturally occurring language that there is a formal distinction between conversational and academic language.

### *CALS*

A group of researchers in the United States have developed the notion of CALS: Core Academic Language Skills: 'a set of high-utility cross-disciplinary skills that comprise school-relevant language proficiency, otherwise called academic language proficiency' (MacFarlane et al., 2020, p. 89). The group built on existing beliefs that academic language plays a key role in academic success, notably the BICS/CALP distinction, but they broaden the notion to include skills as well as knowledge (Uccelli, Barr et al. 2015). They argue that knowledge of academic vocabulary in successful students is most likely at least in part a proxy for the skills associated with the words, such as packing information densely through nominalisation, and connecting ideas logically. Part of their reasoning was previous research showing

disappointing results from vocabulary interventions, that is, simply teaching words did not prove sufficient to produce significant improvement in academic reading proficiency. At the same time though, the writers also reason that language skills and knowledge might support the development of conceptual understanding. For example, knowing connectives could help students to develop their understanding of these relationships in texts.

Uccelli, Barr et al. (2015) present a list of six Core Academic Language Skills, as follows: *unpacking complex words*; *comprehending complex sentences*; *connecting ideas*; *tracking themes*; *organising argumentative texts*; and *awareness of academic register*. More recently, two further skills have been added: *metalinguistic vocabulary* and *identifying epistemic stance* (MacFarlane et al., 2020). Each of the skills is linked to language knowledge; for example, connectives, abstract nouns and formal lexis, but emphasis is placed on the skill, that is, mastery of the academic function that the language is used for. In two studies each involving several hundred school students, Uccelli, Barr et al. (2015) and Uccelli, Philips Galloway et al. (2015) designed tests of the first six of the above CALS and compared the results with students' academic word knowledge, SES, word reading fluency and reading comprehension. They found CALS to be an independent predictor of reading comprehension test scores, even after controlling for the other variables including academic word knowledge. CALS are generic, or cross-disciplinary, and the researchers note the likely existence of discipline-specific academic language skills, such as understanding discourse structures of a story when found in a mathematics problem as opposed to in English (Uccelli, Barr et al., 2015, p. 1097). Like the researchers in the traditions described earlier, they note that some students have vastly greater opportunities to 'participate in school-like literacies at home and at school' (Uccelli & Phillips Galloway, 2017, p. 397).

In this section, we have traced an overview of thinking and research into the role of the language of school, focusing on four important approaches: Bernstein's elaborated and restricted codes; the SFL approach; Cummins' BICS and CALP and Uccelli et al.'s CALS. All agreed that school has a language variety of its own, which children need to master in order to succeed academically. Further, they all note that children's linguistic, social and economic backgrounds set them up to learn the language of school more or less easily. The next section picks up from the SFL concern with function, and the CALS emphasis on skill, to look at how academic language relates to educational and intellectual purpose.

## Academic language and function

### *Academic language and social prestige*

We begin by contesting the suggestion that academic language is simply a matter of different word and syntactic choices from everyday language. We

saw in the previous chapter that some of the primary school students that we spoke to talked about 'posh' synonyms for everyday words. A notion that we heard repeatedly is that writing academically is a kind of translation exercise, in which the writer converts their everyday, colloquial language into a 'better' form, word by word. For instance, in the early Year 6 interview round, the researcher asked students to read and comment on a science text. Several groups mentioned 'hard words' and 'advanced words' and one student gave the example of *environment*, saying:

| *line* | *speaker* | *utterance* |
|---|---|---|
| 443 | Chloë | it's a better word for land yeah so like environment is we have to take care of this environment it belongs to someone else |

*Extract 2.1,* pupil interview, School A.

The same student explains formal, or 'posh' registers as follows:

| *line* | *speaker* | *utterance* |
|---|---|---|
| 435 | Chloë | yeah cos if you said I'm tired so I'm going to bed then that's what you would say normally but then if you were talking posh you'd be like I'm exhausted therefore I'm going to bed you'd speak it more formal |

*Extract 2.2,* pupil interview, School A.

These students' views imply that academic value is tied purely to the linguistic form, rather than the content, of their speaking and writing. It would follow that they could be more successful academically by learning more academic-sounding synonyms for everyday words, and perhaps complexifying the grammatical structures they use. Of course, this is not so; there is general agreement among language instructors and researchers that an approach of simply substituting low-frequency synonyms for everyday, less prestigious vocabulary and deliberately complexifying syntax does not make for better academic writing and speaking (e.g., Bottomley, 2014), but, as university teachers, it is an approach we encounter right up to PhD level. We have worked with many students who have been trained to write in this way, and we have struggled to persuade them that strong academic thought should be expressed as clearly as possible, and that complexity for its own sake needs to be avoided.

An opposing view to the 'good language is complex' stance just described is held by many prominent researchers. They believe that academic and standard language forms are 'owned' by the middle classes, whose social status leads those language forms to be valued above others. A number of researchers have demonstrated that academic-sounding language is often

positively evaluated by listeners and readers but does not necessarily encode better content. In 1969, Labov analysed transcripts of informal interviews with a 15-year-old working-class black boy from Harlem, Larry, and contrasted this with an interview on a similar topic with a young middle-class black man, Charles. Charles uses standard English, and his discourse has many modifiers, hedges and abstract, educated-sounding words such as *culture* and *science*. Labov describes him as 'obviously a "good speaker" who strikes the listener as well-educated, intelligent and sincere' (1969, p. 40). However, Labov's detailed content analysis shows that Charles's line of argument is circular, and his reasoning mediocre:

> Our initial impression of him as a good speaker is simply our long-conditioned reaction to middle-class verbosity; we know that people who use these stylistic devices are educated people and we are inclined to credit them with saying something intelligent.
>
> (pp. 41–42)

By contrast, Larry, the working class boy from Harlem, uses non-standard language forms, which, when Labov was writing, were highly stigmatised, and he is not verbose. Labov's content analysis shows his argument to be tight and precise, as well as witty.

More recently, Schleppegrell (2001) noted that in the early stages of schooling, middle-class children produce language that teachers regard as acceptable, while their working-class peers tended to produce less well-regarded language. As for Labov's work though, deeper analysis showed that the content of the middle-class children's utterances was not superior. Bunch and Martin (2021) also argue for looking for the quality of ideas and thinking rather than over-focusing on formal academic vocabulary and structures.

It goes without saying that making negative evaluations on the basis of non-standard or informal language use, or regional accent is illogical and classist. The argument for looking beyond vocabulary and syntax and evaluating the quality of argument and thought in their own right is powerful. However, language that is complex for its own sake, 'academic gibberish, or unnecessarily dense and intricate structures that obscure communication' (Uccelli & Phillips Galloway, 2017, p. 396) should not be equated with genuinely good academic language, which is clear and precise. It is also, importantly, different from everyday language, not in order to impress or confuse others, but because it has to do different things.

### *Function: Academic language to facilitate and express academic thought*

Researchers are in general agreement that academic language has developed functionally to facilitate and express academic thought (Gee, 2004,

2008; Nagy & Townsend, 2012; Heller & Morek, 2015). It is difficult to learn, manipulate and reason about non-everyday material without the tool of academic language, and more so as children move up through school. School material becomes increasingly technical and nuanced, and concerns topics that are not normally the subject of everyday discourse (Nagy & Townsend, 2012, p. 92).

Gee (2008) demonstrates this by analysing utterances made by fourth-grade school students in a science lesson. The children conducted an experiment in which they submerged objects made from different materials in water to answer the question 'What makes things rust' (2008, p. 57). He quotes two children's responses: 'But if we didn't put the metal things on there, it wouldn't be all rusty', referring to a plastic plate which has become stained with rust from a metal bottle cap, and 'But if we didn't put the water on there, it wouldn't be all rusty', referring to the bottle cap itself. Gee claims that the phrase 'all rusty' fails to distinguish two different meanings: 'having rust on it' (a state, describing the plastic plate, which is unchanged) and 'having rusted' (a process, describing the bottle cap, which has changed)' (p. 57). Gee writes that one of the goals of the lesson was to help the children learn the difference between states and processes. This kind of thinking is more specific and precise than that needed in the children's everyday lives, and their everyday language is not yet sufficiently subtle to support it.

In addition to thinking with scientific precision, children need to develop language to handle abstract concepts and generalisations. Coffin (1997) writes that when children learn history, they need to move from 'commonsense' to abstract meaning, and to do so, they 'need to gain control of language which is highly abstract' (1997, p. 202).

As well as for thought, clear and appropriate academic language is needed for communication. Children need to understand academic speech and texts, which are composed following expectations and conventions often not made explicit, and eventually, they have to attempt to produce them. Schleppegrell (2001, 2012a) identifies purpose as the determiner of the features of the language of school, arguing, following the SFL school, that its lexical and grammatical choices realise the context of schooling. Communication with others requires additional language skills over and above those needed for academic thought – discussed above. These are concerned with managing physical and/or psychological distance or assumed distance.

Uccelli and Phillips Galloway (2017) write that academic language takes the shape that it does because it has to communicate complex and abstract ideas precisely to a distant audience. Snow (1983) writes that the process of learning academic language is one of increasing decontextualisation, and 'full blown adult literacy is the ultimate decontextualised skill' (1983, p. 175). Children start on this path by being exposed to narratives that have an impersonal voice and complex language, oriented towards a distant audience who do not share reference (ibid). Schleppegrell (2001) shows how

this is realised in the widely used primary-school activity 'sharing time', or 'show and tell' in which children bring in an object from home and describe it to the class. In this activity, children are expected to talk in a decontextualised way, not assuming shared knowledge, and being linguistically explicit. Success in this task is an early step towards the demands of literacy. She also writes of the need for primary-aged children to understand and produce narrative speech that has literate qualities, noting that they have to be able to understand such features in reading and listening (2012). Not all children are equipped for this by their home life, which may be proficient in different registers but not academic language. Addressing the question of why some children struggle to learn to read, Snow (1983) suggests 'Perhaps most children are not failing at reading and writing but at comprehending and producing decontextualised information' (1983, p. 186), a suggestion that echoes Bernstein's codes theory discussed above. Avenia-Tapper and Isacoff (2016) used text analysis to show that children from lower-income families used more deictic terms in their science writing than children from better-off families, who were linguistically more explicit, and whose texts tended to be scored more highly. The linguistically explicit texts were more successful at conveying expertise and authority.

The assumed distance in academic communication also requires speakers and writers to structure information conventionally, logically and overtly (Schleppegrell, 2001; Veel, 2005), in order to meet the expectations of their audience. A good deal of genre analysis has shown school texts to be structured along very prescriptive lines, which are, in general, not signalled explicitly to children (Schleppegrell, 2012a). Skilled academic communicators use discourse organisation, connectors, syntax and lexis to signal why a contribution is important, how it relates to the ongoing discourse and how ideas are connected. Not all students can interpret this; Duffy and Elwood (2013) found that 'disengaged' students reported not understanding the teacher's intentions or the lesson objectives. Schleppegrell (2001) notes that for success in 'sharing time', primary school children have to be able to signal linguistically why their information or story is important and how it relates to the task at hand. We have identified some macro-functions that school students need to be able to recognise and handle through the use of recognised language choices. These sets of functionally motivated choices result in genres and registers. In the next section, we explain the model of genre and register that we use to frame the studies described later in this book.

## Register and genre

Earlier in this chapter, we discussed the erroneous conflation of prestigious language forms with academic language. A genre and register approach provides clarity on the issue and moves it away from the evaluation that is implicit in much discourse about the language of school. Phillips Galloway

et al. (2015) write that Mainstream American English, or the standard, 'is a language *variety (or dialect)* and AL is a *register*' (emphasis in original) (2015, p. 223). They add that while societal values are often ascribed to language varieties, instruction in the academic language register assumes no value system beyond appropriacy to the task. Developing the ability to handle different registers is part of the ongoing language learning that takes place during adolescence (Uccelli, Barr et al., 2015). Swales and his co-researchers (e.g., 1990) have explored how related issues present challenges to students and early career academics in higher education using a genre framework.

Much of the research into the language of school that we have cited, such as the central work by Martin (e.g., 1989, 1993, 2013) and Schleppegrell (e.g., 2001, 2004, 2012a) has developed within and from the SFL tradition, in which the constructs of genre and register are central. In the SFL models, genres are defined as 'staged, goal-oriented social process[es]' (Martin, 1993, p. 142). Christie (1997) writes of 'curriculum genres', within which registers operate, and 'macrogenres', which are sequences of genres across time, working towards a macrofunction. Genres describe broad functions, such as narrative, exposition and recount (Schleppegrell, 2001), and are analysed into stages. In the SFL model, register is metaphorically nested within genre. Register describes the relationship between context and language choices, using the notions of field (topic), tenor (relationships between participants, attitude) and mode (how language is expressed, e.g., through speech or writing, 'the role language plays in the context') (Schleppegrell, 2012b, p. 22).

Hunston (2013) notes that SFL and corpus linguistics ask similar questions about register, and have much in common, but do not engage. She compares Biber and Conrad's (2009/2019) corpus approach with SFL and notes that in both approaches, the context of situation generates language choices. Both are probabilistic, using frequency to allow registers to emerge from data. There is no genuine contradiction between the approaches, and Hunston maps them onto each other coherently (2013), arguing that the differences between them are to do with emphasis and terminology. We have used the SFL framework in some of our earlier research, but in the analytical studies described later in this book, we have chosen to work with Biber and Conrad's construct of register. This is because it was specifically developed from and alongside Biber's leading corpus work on register (e.g., Biber, 1988, 1998). Further, it was developed not just for the description of registers, but for comparison between them, which is the central goal of our research.

Conrad (2019) summarises different approaches to register and genre used by different groups of researchers. While a few earlier approaches focused purely on the description of language features, approaches currently used have in common an understanding of register as 'a variety associated with a particular situation of use (including particular communicative purposes)' (Biber & Conrad 2019, p. 8). In their approach, a register analysis

has 'three components: the situation of use, including all aspects of the context of production or reception; the linguistic features; and the functional associations between the situational characteristics and the linguistic features' (Conrad, 2019, p. 140). The functional interpretation is generated through comparison of the linguistic and situational analyses. Biber and Conrad (ibid) write that situation of use, including function, is more basic than the linguistic features. That is, we find ourselves in a specific situation or context, with specific communicative needs, and certain linguistic features result from that, not the other way round. Conrad (2019) notes that with few exceptions, register analysis does not identify linguistic features that are unique to a register, but rather, ones that are more or less frequent. In mature academic prose for example, linguistic features such as nominalisations and dense noun phrases are observed. These also occur in other registers, but statistically less frequently.

Following Conrad (2019) and Conrad and Biber (2019), when we consider discourse community and rhetorical moves, we refer to genre analysis. Conrad and Biber (2019) note that some linguistic markers of genres are conventional rather than functional, such as the way that letters are set out, or the way we open and close conversations and service encounters. To analyse genres, they write, we need whole texts, as particular features are often confined to one part of the text, genres being staged. In contrast, register features are pervasive, so texts can be sampled and analysed using corpus linguistics, which does not support whole text analysis easily. Most of this book concerns register, though some of what we have done in our project more widely is genre analysis, such as parts of Candarli et al.'s analysis of PowerPoint presentations in KS2 and KS3 (2019).

Biber's approach to register, that is, the relationship between situation, function and linguistic features of text, is broadly set out in his 1988 book, which built on earlier studies of differences between speech and writing (e.g., 1986). (At that point, he used the term 'genre'; in more recent years, 'register' has been used by his school.) The approach has been tested and applied up to the time of writing. It starts by automatically identifying and counting features of texts and then interpreting them functionally (1988, p. 24), rather than using pre-determined, theoretically derived categories based on the assumed functions of a register. Biber (1988) identified a large number of linguistic features, such as past tense verbs, time and place adverbials, agentless passives and so on, in the main general purpose corpora available at that time, and in some additional written data. These linguistic features themselves are too numerous and detailed to be of use in determining registers. Therefore, factor analysis was used to find patterns of co-occurrence in the corpora. The clusters of linguistic features were then analysed qualitatively in context to establish functions that they are associated with. For example, past tense adverbials were found to co-occur with 'third person animate referents, reported speech, and depictive details', in narrative texts. This process led to the identification of

'Dimensions': 'bundles of linguistic features that co-occur in texts because they work together to mark some common underlying function' (1988, p. 55). In his 1988 work, Biber identified six Dimensions through this methodology, as follows (1988, p. 115):

Dimension 1: Informational versus Involved Production
Dimension 2: Narrative versus Non-Narrative Concerns
Dimension 3: Explicit versus Situation-Dependent Reference
Dimension 4: Overt Expression of Persuasion
Dimension 5: Abstract versus Non-Abstract Information
Dimension 6: On-Line Informational Elaboration

Each of these dimensions was derived from and is associated with clusters of linguistic features, which we list in Chapter 3. Dimensions can be used to describe how registers differ from each other (Biber & Conrad, 2019, p. 223). Other researchers have used the methodology to construct other dimensions in specialised corpora (e.g., Gardner et al., 2018). In Chapter 3, we describe how we have used Multidimensional analysis (MD analysis), using the first five of Biber's Dimensions, in some of our studies.

## Features of academic language

### *Overview*

In this section, we review existing descriptions of the language of school and discuss linguistic aspects that have proved particularly noteworthy or problematic. Schleppegrell (2001) gives a landmark and comprehensive description of the language of school. She begins with the constructs of register and genre, using the SFL framework and taking the starting point of function, as we discussed previously. She notes, like Biber (1988), that register description is frequency-based, identifying features more likely to occur in the language of school, rather than unique to it. She compares the language of school with spoken interaction, on the assumption that spoken interaction will be familiar to students, while the language of school may be less familiar to at least some of them. Her full description is given in Table 2.1 (2001, p. 438).

Schleppegrell's sources for the description of the language of school include corpus studies, largely from the SFL school, but also drawing on corpus work by Biber. She also used some of her own studies of school texts, which have used discourse analysis techniques. Schleppegrell (2004) challenges the characterisation of academic texts as decontextualised, explicit and complex, exploring and then hedging all of these terms. She argues that they are a property of the interaction between text and reader, and understood in comparison to other registers, which children may be more or less familiar with, rather than absolute.

*Table 2.1* Register features of spoken interaction and school-based texts (Schleppegrell, 2001, p. 438).

| | *Spoken interaction* | *School-based texts* |
|---|---|---|
| **Lexical features** | | |
| Lexical choices | generic | specific, technical |
| Lexical density | sparse | dense, elaboration of noun phrases through modifiers, relative clauses, and prepositional phrases |
| Subjects | pronominal, present or known participants | lexical, nominalisations, and expanded NPs |
| **Grammatical strategies** | | |
| Segmentation | prosodic segmentation: structure indicated prosodically | sentence structure: structure indicated syntactically |
| Mood | varied, attitude conveyed prosodically | mainly declarative, attitude conveyed lexically |
| Clause linkage and conjunction strategies | clause chaining with conjunctions, information added in finite segments, use of many conjunctions with generalised meanings | clause-combining strategies of embedding, use of verbs, prepositions, and nouns to make logical links, conjunctions have core (narrow) meanings |
| Organisational strategies | emergent structure, clause themes include conjunctive and discourse markers that segment and link part[s] of text | hierarchical structure, using nominalisation, logical links indication through nominal, verbal and adjectival expressions and thematic elements that structure discourse |

Snow and Uccelli (2009) also developed a description of the language of school, which they term 'an inventory of features' (p. 118), and which also contrasts the language of school with colloquial language. This draws on previous descriptions of academic writing, including Schleppegrell's, described above. They matched overarching characteristics of academic language, such as density, with linguistic features such as nominalisation. They emphasise in several places that language in itself is not sufficient for success: broader generic knowledge, argumentation skills and disciplinary knowledge are also essential. Students need a sense of self, audience and how to represent their material to their audience. Without this, the detailed linguistic choices cannot be understood. They present their inventory of features of academic language (Table 2.2, cited from Snow & Uccelli, 2009, pp. 119–120).

Snow and Uccelli comment on the length of the list. They also point out that while any of these traits might mark a text out as being a stretch of academic language, 'it is unclear that any of them actually defines the

*Table 2.2* Linguistic features and core domains of cognitive accomplishments involved in academic language performance (Snow & Uccelli, 2009).

| *More colloquial* | *More academic* |
|---|---|
| 1 **Interpersonal stance** | |
| Expressive/involved | ⇨ Detached/distant (Schleppegrell, 2001) |
| Situationally driven personal stances | Authoritative stance (Schleppegrell, 2001) |
| 2 **Information load** | |
| Redundancy (Ong, 1995)/ wordiness | Conciseness |
| Sparsity | Density (*proportion of content words per total words*) (Schleppegrell, 2001) |
| 3 **Organisation of information** | |
| Dependency (Halliday, 1993)/ addition (Ong, 1995) (*one element is bound or linked to another but is not part of it*) | Constituency (Halliday, 1994)/ Subordination (Ong, 1995) (*embedding, one element is a structural part of another*) |
| Minimal awareness of unfolding text as discourse (*marginal role of metadiscourse markers*) | Explicit awareness of organised discourse (*central role of textual metadiscourse markers*) (Hyland & Tse, 2004) |
| Situational support (exophoric reference) | Autonomous text (endophoric reference) |
| Loosely connected/ dialogic structure | Stepwise logical argumentation/unfolding, tightly constructed |
| 4 **Lexical choices** | |
| Low lexical density | High lexical density (Chafe & Danielewicz, 1987) |
| Colloquial expressions | Formal/prestigious expressions (e.g., *say/ like* vs. *for instance*) |
| Fuzziness (e.g., *sort of, something, like*) | Precision (*lexical choices and connectives*) |
| Concrete/common-sense concepts | Abstract/technical concepts |
| 5 **Representational congruence** | |
| Simple/congruent grammar (simple sentences, e.g., *You heat water and it evaporates faster*) | Complex/ congruent grammar (complex sentences, e.g., *If the water gets hotter, it evaporates faster*) / Compact/incongruent grammar (*clause embedding and nominalisation, e.g., The increasing evaporation of water due to rising temperatures*) (Halliday, 1993) |

(*Continued*)

*Table 2.2* (Continued)

| *More colloquial* | *More academic* | |
|---|---|---|
| Animated entities as agents (e.g., *Gutenberg invented printing with movable type*.) | Abstract concepts as agents (e.g., *Printing technology revolutionised European bookmaking*) (Halliday, 1993) | |
| **Genre mastery** | | |
| Generic values (Bhatia, 2002) (narration, description, explanation) | School-based genres (e.g., lab reports, persuasive essay) | Discipline-specific specialised genres |
| **Reasoning strategies** | | |
| Basic ways of argumentation and persuasion | Specific reasoning moves valued at school (Reznitskaya et al. (2001) | Discipline-specific reasoning moves |
| **Disciplinary knowledge** | | |
| • **Taxonomies** | | |
| Common-sense understanding | Abstract groupings and relations | Disciplinary taxonomies and salient relations |
| • **Epistemological assumptions** | | |
| Knowledge as fact | Knowledge as constructed | |

phenomenon' (2009, p. 121). Moving from these overviews, we now discuss more specific issues within academic language.

### *Disciplinary language*

A number of writers have argued that 'academic language' is not a single register but rather a cluster of related registers. The usual starting point for analysis is the school subject. Martin (2013) claims that the literacy that is required in secondary school is associated with the nature of knowledge in each subject or discipline, and there are differences between them. Schleppegrell (2001) argues that academic language is not a single register, but rather many registers, instantiating genres, which include narratives, descriptions and definitions. Gee (2005) also argues that academic language is associated with particular ways of thinking and acting, sometimes specific to a discipline, though sometimes crossing traditional disciplinary boundaries. Bower and Ellerton (2007) write that even within a discipline, there are numerous sub-genres. There have been many studies of the language of different school subjects and university disciplines, many, though not all, by researchers within the SFL tradition. There is a particularly strong tradition of studying the language of science (e.g., Norris & Phillips, 2003; Gee,

2008; Arya et al., 2011; Martin, 2013; Patterson Williams, 2020), claiming that doing science involves using the language of science. This is an area which science educators regularly research, as evidenced in key journals such as the *International Journal of Science Education*. History has been a focus in the SFL tradition (Coffin, 1997, 2006; Schleppegrell, 2012a), and there have been comparative studies such as Shanahan and Shanahan's (2008) study of mathematics, chemistry and history. Part of our research has looked at how the language of the core school subjects changes across the transition. We discuss the issue and describe the studies in Chapters 5, 6 and 7.

### *The vocabulary of school*

It has been known for some time that the language of school involves a significant amount of new and specialised vocabulary. Merzyn (1987) looked at vocabulary in a widely used secondary school physics textbook, estimating that it contained around 2000 words unknown to students, at least five on each page. He argued that in his context, Germany, students would therefore be expected to interact with more new words in physics per lesson than in a foreign language lesson. Coxhead et al. (2012) looked at the connection between secondary school science textbooks and vocabulary, attempting to determine what size of vocabulary is needed. They found that to read secondary school science textbooks, at least 3000 more words are needed than to read a novel, and that there is a steep increase in the vocabulary needed as students progress through secondary school. Nagy and Townsend (2012) reviewed the growing body of research demonstrating the importance of academic vocabulary to reading and academic progress more generally. They describe how academic vocabulary differs from everyday vocabulary, arguing that the two are interrelated. Like the work that we described earlier in this chapter, they trace academic vocabulary to its function, which they see as getting more abstraction and greater informational density.

A good deal of the current discussion about school language focuses on academic vocabulary and its importance to success in school. Quigley, a former teacher who writes about research for an education practitioner readership, asserts that rich vocabulary knowledge is essential for success in school (2016, 2017, 2018, 2020). He claims that there is a wide variation in levels of knowledge between different groups of children, which is correlated closely with social class and levels of parental education (2018). Research by Oxford University Press involved surveying secondary school teachers and found that the teachers who took part 'reported that 43% of Year 7 pupils have a limited vocabulary such that it affects their learning' (2018, p. 4).

Several studies have found that vocabulary knowledge predicts success or otherwise at school. Spencer et al. (2016) found that vocabulary knowledge at ages 13–14 had a strong association with good GCSE results two

years later in English language, English literature and mathematics. Schuth et al. (2017) found that academic vocabulary knowledge, as opposed to general vocabulary knowledge and other factors, correlated with academic performance in their study of 170 German students in grade 4 (aged 9–10). Townsend et al. (2012) tested 339 seventh and eighth grade students in the United States (aged approximately 12–14) on their knowledge of general vocabulary and academic vocabulary. Results showed correlation between performance on two state-wide tests and knowledge of academic vocabulary. As Schuth et al. (2017) found with a different age group and in a different context, academic vocabulary knowledge was more important for academic success than general vocabulary knowledge.

Interventions to support vocabulary development have had mixed results. Snow et al.'s (2009) vocabulary teaching experiment found that vocabulary instruction correlated with improved scores in a standardised test of achievement in English (Massachusetts Comprehensive Assessment System, English Language Arts) in the treatment group, sixth to eighth grade students (equivalent to Years 7–9 in England and Wales). However, Uccelli, Barr et al. (2015) review academic vocabulary interventions and find a more complicated picture, with improvements noted in some areas and studies but no change in others. It is very possible that some studies were carried out using a partial or flawed understanding of the nature of academic vocabulary. In the next sections, we look at some aspects of the vocabulary of school language.

### *Polysemy and homonymy*

The term 'polysemy' means multiplicity of meaning, and refers to many, even most words having more than one meaning. Polysemy is widely recognised as an issue in academic vocabulary (e.g., Nagy & Townsend 2012). A pattern that is frequently found is where a word has one meaning in colloquial language, which is likely to be familiar to students, but in academic language, it is used with another meaning. An example from our data is *volume*. In one of our interviews, a student told us that this was confusing as they tended to think of the 'sound' meaning, which they were familiar with from volume buttons on televisions and phones. In our secondary school corpus, *volume* is almost always used to refer to the physical space occupied by an object or entity, in biology, where it collocates with *lung* and *blood*, and in mathematics, where it collocates with words such as *cylinder*. Polysemy is documented by a number of writers; Fang's (2006, p. 494) examples are *school* (academic meaning: a group of fish), *fault* (academic meaning: break in rock formation) and *volume*. Patterson et al. (2018, p. 296) make the same point, citing the words *force*, *power* and *energy*, which have everyday meanings and specialised meanings in school science.

Polysemy is, potentially, a threat to a word list approach such as the Academic Word List (AWL) studies conducted by Coxhead and others (e.g.,

Coxhead 2000; Coxhead et al., 2012), because word lists are based on form and do not take account of items having multiple meanings. There has been relatively little exploration of this, one exception being Wang and Nation's study (2004) on a related topic. They investigated homography in the AWL (homographs are two or more words recognisably different in meaning and possibly from different roots etymologically but having the same written form). Wang and Nation aimed to establish whether homographs, rather than word forms, had been counted, some words would not have been frequent enough to be included in the AWL. They found that 60 words in the AWL have multiple meanings or around 10%. Of these, there were only three words – *intelligence*, *offset* and *panel* – where the colloquial meaning of the word had inflated its apparent frequency as an academic word. For example, *intelligence* has a general meaning describing a person's ability to learn, and a specialised one referring to information collected, perhaps covertly. Wang and Nation describe their findings as reassuring (2004, p. 309), in that homography did not seem to have had a statistical impact on the composition of the AWL. Nevertheless, we believe that the finding that 10% of words in the list are polysemous/homographs is of importance to educators.

Nation and Parent (2016) also discuss multiple meaning in word lists, writing that it may be useful to count homographs and similarly related pairs separately, but that senses of polysemous words, which have discernible semantic connections, should not be separated. They argue that 'Treating related sense of words as different words is not giving learners enough credit for what they are able to do with context while reading' (2016, p. 51). However, target users of the AWL are university-level students, with many years of language learning experience behind them, and therefore, one might assume, a good level of metalinguistic competence. They probably have knowledge and skills lacked by the 11–13 year-olds, mostly monolingual and many from socio-economically deprived backgrounds, whose language experience is the focus of our research.

There is broad consensus among education professionals that polysemy is a problem for learners. Todd (2017) writes that polysemous words present one of the greatest challenges for students in his field, engineering, and identifies 45 such words from his engineering corpus. Although the specialist meaning of terms such as *value* is related semantically to their everyday senses, Todd writes that students are often not able to guess this. Science educationalists have argued extensively that the polysemy between everyday words and their meanings as scientific terms poses a problem to students (Strömdahl, 2012). An experimental study by Logan and Kieffer (2017) found that the academic sense of polysemous words was a source of difficulty in reading for adolescents, regardless of whether they knew the everyday meaning of the word. Bower and Ellerton give the example of *function* used in school mathematics to 'special types of relationships between two sets' (2007, p. 336), having little relation to the everyday meaning of purpose. In the UK context, Deignan et al. (2019) found that some school

students were unable to use polysemous words such as *release* accurately in its scientific sense, referring to the generation of carbon dioxide through burning fossil fuels, even though the semantic relationship with the more colloquial sense is clear. Quigley (2022) writes about the issue in a widely read blog for teachers.

Another kind of polysemy is between the senses of a polysemous word in different disciplines. Hyland and Tse (2007) found that there are differences in use across academic subject areas, even for such apparently generic academic words as *attribute*, which can mean 'a characteristic of', or as a verb, meaning 'to accredit'. They also note that the same word used in different academic fields will tend to have different collocates. Collocations are associated with domains, meaning that words used in different domains will have different collocations (Taljard, 2016). Our research has found examples such as *concentration*, which students tend to be familiar with in its colloquial sense of 'think hard'. In school science, its meaning refers to the amount of a substance dissolved in water, while in English lessons, it occurs frequently in *concentration camp*, because many schools set 'The Boy in Striped Pyjamas' (a novel about a Nazi extermination camp, Boyne, 2006) as a core text for Year 7. While the meanings are etymologically related, we do not think that all children would be able to work them out without support, and our student interviews confirmed this.

### *Tiers*

Nation (2001) classified the vocabulary needed by second language learners into three types: general vocabulary, academic vocabulary and technical vocabulary. A distinction between the last two types had been made earlier by teachers in higher education, for example, Baker (1988), with Nation's 'academic vocabulary' termed 'sub-technical'. In a similar way, Beck et al. (2002) classified the language of school into 'Tiers', a notion which has struck a chord with many teachers and is now widely discussed in schools, though we have rarely seen it used in the research literature. They write:

> The first tier consists of the most basic words: *warm*, *dog*, *tired*, *run*, *talk*, *party*, *swim*, *look* and so on. These are the words that typically occur in oral conversations, and so children are exposed to them at high frequency from a very early age. […]
>
> Moving on to the third tier – this set of words has a frequency of use that is quite low and often limited to specific topics and domains. Some examples of Tier Three words might be *filibuster*, *pantheon* and *epidermis*. In general, a rich understanding of these words would not be of high utility for these learners. These words are probably best learned when a specific need arises, […]
>
> The second tier contains words that are of high utility for mature language users and are found across a variety of domains. Examples

> include *contradict*, *circumstances*, *precede*, *auspicious*, *fervent*, and *retrospect*.
>
> (2002, p. 9)

They go on to write that students are less likely to have encountered these Tier 2 words than Tier 1 words because they are infrequent in conversation, and that Tier 2 words play an important part in learning.

Tier 2 words seem to have an important role in accessing and explaining other knowledge. Tang and Rappa (2020) use the term 'scientific metalanguage' to refer to words that fall into Tier 2 in Beck et al.'s model. Tang and Rappa identified metalanguage associated with the four genres of science: 'experimental report, informational report, argument and explanation' (2020, p. 134). For example, vocabulary associated with the scientific report includes: *aim*, *method*, *procedure* and *observation* (p. 136). School students need to be able to handle these terms in order to understand and construct scientific criticality. Quigley (2018) writes that Tier 2 words 'make sense of' Tier 3 words (2018, p. 89), and also notes the important sub-group of Tier 2, academic discourse markers. Beck et al. (2002) write that they are of high utility, but do not provide a more detailed breakdown of their functions.

Tier 2 words are defined mainly in terms of their distribution rather than function, specifically that they tend to be found in written, academic texts rather than colloquial speech, and that they are found across subject areas rather than being specific to one discipline (Beck et al., 2002; Quigley, 2018). Both Beck et al. and Quigley cite Coxhead's (2000) AWL as an empirical resource for identifying Tier 2 words, though it was developed from a corpus of university-level texts rather than school texts. A definition built on frequency and distribution across texts lends itself very well to corpus methods. In 1988, Baker developed a method for identifying the 'subtechnical' vocabulary of medicine. Deignan and Love (2021) used the same methodology to develop a candidate list from educational texts. We have already mentioned, above, the challenge that polysemy presents for word lists; this was a major issue in the study. Collocation and qualitative analysis of the candidate Tier 2 words that were identified automatically showed that a large number of them were polysemous. Further, Deignan and Love found that many had senses that seemed to be in different tiers. For example, *found* has a Tier 1 meaning that students would be familiar with from early years, and a Tier 2 meaning reporting academic findings. *Energy* has a Tier 1 meaning, characterised by personal possessives ('my energy levels') and collocates referencing food and drink ('energy drinks'), but as used in physics is a highly specialised, abstract concept, perhaps indicating that it is in Tier 3. The more individual words are examined, the more subjective the classification seems to be, and we believe it is not possible to rigorously allocate the lexicon into these three tiers. Nonetheless, we view it as a very useful construct, which contains an important insight into educational

vocabulary. It is also widely known and talked about by teachers, as we have seen in teachers' tweets and blogs, and in our interviews.

### *Grammar and discourse*

At the level of syntax, university-level academic discourse has different characteristics from other registers (Biber 1988; Biber et al., 1998). For instance, Biber et al. (1998) analysed the rate of nominalisations in academic and fiction sub-corpora of the Longman-Lancaster corpus and speech from the London-Lund corpus. This showed that nominalisations occur at about the same rate in the speech and fiction sub-corpora, but almost four times as frequently in the academic sub-corpus. Another major difference between academic writing and conversation is the frequent use of post-modifying prepositional phrases in university-level academic writing (Biber & Gray, 2010). We noted earlier that Schleppegrell (2001) and Snow and Uccelli's (2009) inventories of the language of school claim that its grammatical structure is less complex than that of conversation. Biber and Gray's (2010) study suggests that the varieties are equally complex but in different ways.

Academic language has well-known differences at the level of word grammar. Fang (2006) writes that academic words are often different parts of speech from their everyday uses. For example, *young* is rarely used as a noun in colloquial discourse in the way that it is in academic language. Researchers in the SFL tradition call this kind of variance 'grammatical metaphor'. They argue that notions have a congruent part of speech, which is commonly used in everyday discourse. Nagy and Townsend write: 'Typically, nouns represent persons, places, or things; verbs represent actions, and identifiable agents (e.g., people) perform actions. However, in grammatical metaphor, nouns can represent complex processes, and abstract concepts can "perform" actions' (2012, p. 94). Christie (2002) is among many SFL scholars who see grammatical metaphor as a central feature of academic language.

At the discourse structure, academic language also differs from other registers that students may be more familiar with. MacFarlane et al. (2020) describe discourse-level features of academic language: at text organisation level, argumentative text (not narrative); epistemic stance; and being able to pick up clues to organisation and themes from linguistic signals such as connectives and anaphoric reference. Fang (2006) points out that even logical connectors such as *or* can be problematic, as, in scientific text, it often introduces a paraphrase.

### *Specific issues at the transition*

Work by researchers such as Snow (1983) and Schleppegrell (e.g., 2012a) has shown that there is an academic language that advantages some and disadvantages others from the earliest days of primary school. We noted above that Schleppegrell (2001) has described how the widely used primary school

activity of 'Show and Tell' brings expectations about language genre to be used which are not made explicit. Snow notes that by the age of 12 or 13, academically successful students have become proficient in handling decontextualised information. Fang (2006) argues that the narrative books that students read in lower years are closer to social language, and claims that expository text of specialist texts is not only unfamiliar but alienating. Fang et al. (2020) point out that factual writing does not draw on students' everyday linguistic resources in the same ways as fictional/imaginary writing does.

Martin (2013) is one of few linguists, as opposed to teachers, to discuss the issue of the transition in academic language. Christie (2002) writes that 'it is with the transition to secondary school that students must learn to handle the grammar of written English differently from the ways they handled it for primary schooling' (p. 45). She claims that a number of changes in the grammar of advanced literacy enable students to write about the abstract, to generalise, argument and reflect (ibid).

## Conclusion

Many of the studies described in this chapter have been based on relatively small-scale text studies. Some corpus work, especially word lists, has been quantitative, while the text studies are qualitative. The work of two of the project team on tiers (Deignan & Love, 2021) suggested that there is room for corpus studies that take both quantitative and qualitative approaches. Conrad (2019) notes two kinds of methodology that have been used. We use corpus methods, with a combination of quantitative (Brezina, 2018) and qualitative. These methods will also enable us to study in more depth the way that the language of school changes, something that a number of researchers have touched on but not specified in detail. Chapter 3 describes the research methods we have used.

## References

Arya, D. J., Hiebert, E. H., & Pearson, P. D. (2011). The effects of syntactic and lexical complexity on the comprehension of elementary science texts. *International Electronic Journal of Elementary Education*, *4*(1), 107–125.

Avenia-Tapper, B., & Isacoff, N. J. (2016). Explicitness in science discourse: A Gricean account of income differences. *Language in Education*, *30*(1), 58–71. https://doi.org/10.1080/09500782.2015.1088547

Baker, M. (1988). Sub-technical vocabulary and the ESP teacher: An analysis of some rhetorical items in medical journal articles. *Reading in a Foreign Language*, *4*(3), 91–105.

Beck, I., McKeown, M., & Kucan, L. (2002). *Bringing words to life: Robust vocabulary instruction*. Guilford Press.

Bernstein, B. (1959/2010). A public language: Some sociological implications of a linguistic form. *British Journal of Sociology*, *61*(Supplement 1), 153–69. https://doi.org/10.1111/j.1468-4446.2009.01240.x

Bernstein, B. (1964). Elaborated and restricted codes: Their social origins and some consequences. *American Anthropologist*, *66*(2), *55*–69. https://doi.org/10.1525/aa.1964.66.suppl_3.02a00030

Bernstein, B. (1966). Elaborated and restricted codes: An outline. *Sociological Inquiry*, *36*(2), 254–261. https://doi.org/10.1111/j.1475-682x.1966.tb00628.x

Bhatia, V. J. (2002). Applied genre analysis: Analytical advances and pedagogical procedures. In Johns, A. M. (Ed.), *Genre in the classroom: Multiple perspectives* (pp. 279–283). Lawrence Erlbaum Associates.

Biber, D. (1986). Spoken and written textual dimensions in English: Resolving the contradictory findings. *Language*, *62*, 384–414.

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

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

Biber, D., & Conrad, S. (2019). *Register, genre, and style* (2nd ed.). Cambridge University Press.

Biber, D., & Gray, B. (2010). Challenging stereotypes about academic writing: Complexity, elaboration, explicitness. *Journal of English for Academic Purposes*, *9*, 2–20. https://doi.org/10.1016/j.jeap.2010.01.001

Bottomley, J. (2014). *Academic writing for international students of science*. Routledge.

Bower, M. W., & Ellerton, N. F. (2007). Language genre transitions in a secondary school physics classroom. *International Journal of Mathematical Education in Science and Technology*, *38*(3), 338–351. https://doi.org/10.1080/00207390601116078

Boyne, J. (2006). *The boy in striped pyjamas*. David Fickling Books.

Brezina, V. (2018). *Statistics in corpus linguistics: A practical guide*. Cambridge University Press.

Bunch, G. C., & Martin, D. (2021). From 'academic language' to the 'language of ideas': A disciplinary perspective on using language in K-12 settings. *Language and Education*, *35*(6), 539–556. https://doi.org/10.1080/09500782.2020.1842443

Candarli, D., Love, R., & Deignan, A. (2019, August, 29–31). *Rhetorical variation in teachers' lesson preparations across key stage 2 and key stage 3* [Paper presentation] BAAL Corpus Linguistics Special Interest Group Annual Workshop 2019, Lancaster University.

Chafe, W., & Danielewicz, J. M. (1987). Properties of spoken and written language. In Horowitz, R., & Samuels, S. J. (Eds.), *Comprehending oral and written language* (pp. 83–113). Academic Press.

Christie, F. (1992). Literacy in Australia. *Annual Review of Applied Linguistics*, *12*, 142–155. https://doi.org/10.1017/S0267190500002191

Christie, F. (1997). Curriculum macrogenres as forms of initiation into a culture. In Christie, F., & Martin, J. R. (Eds.), *Genre and institutions: Social processes in the workplace and school* (pp. 134–160). Bloomsbury.

Christie, F. (2002). The development of abstraction in adolescence in subject English. In Schleppegrell, M., Colombi, M. C., & Baugh, J. (Eds.), *Developing advanced literacy in first and second languages: Meaning with power* (pp. 45–66). Taylor and Francis.

Christie, F. (2009). Ongoing dialogue: Functional linguistic and sociological perspectives on education. In Christie, F., & Martin, J. R. (Eds.), *Language,*

*knowledge and pedagogy: Functional linguistics and sociological perspectives* (pp. 3–13). Bloomsbury.

Christie, F., & Martin, J. R. (Eds.). (1997). *Genre and institutions: Social processes in the workplace and school*. Continuum.

Coffin, C. (1997). Constructing and giving value to the past: An investigation into secondary school history. In Christie, F., & Martin, J. R. (Eds.), *Genre and institutions: Social processes in the workplace and school* (pp. 196–230). Bloomsbury.

Coffin, C. (2006). Learning the language of school history: The role of linguistics in mapping the writing demands of the secondary school curriculum. *Journal of Curriculum Studies*, *38*(4), 413–429. https://doi.org/10.1080/00220270500508810

Conrad, S. (2019). Register in English for academic purposes and English for specific purposes. *Register Studies*, *1*(1), 168–198. https://doi.org/10.1075/rs.18008.con

Coxhead, A. (2000). A new academic wordlist. *TESOL Quarterly*, *34*(2), 213–238. https://doi.org/10.2307/3587951

Coxhead, A., & White, R. (2012). Building a corpus of secondary school texts: First you have to catch the rabbit. *New Zealand Studies in Applied Linguistics*, *8*(2), 67–73. https://search.informit.org/doi/abs/10.3316/informit.108520980622633

Cummins, J. (1980). The cross-lingual dimensions of language proficiency: Implications for bilingual education and the optimal age issue. *TESOL Quarterly*, *14*(2), 175. https://doi.org/10.2307/3586312

Cummins, J. (1981). Age on arrival and immigrant second language learning in Canada: A reassessment. *Applied Linguistics*, *2*(2), 132–149. https://doi.org/10.1093/applin/2.2.132

Cummins, J. (2008). BICS and CALP: Empirical and theoretical status of the distinction. In Street, B., & Hornberger, N. H. (Eds.), *Encyclopaedia of language and education*, *Volume 2: Literacy* (2nd ed., pp. 71–83). Springer Science + Business Media LLC. https://doi.org/10.1007/978-0-387-30424-3_36

Deignan, A., & Love, R. (2021). Using corpus methods to identify subject specific uses of polysemous words in English secondary school science materials. *Corpora*, *16*(2), 165–189. https://doi.org/10.3366/cor.2021.0216

Deignan, A., Semino, E., & Paul, S.-A. (2019). Metaphors of climate science in three genres: Research articles, educational texts, and secondary school student talk. *Applied Linguistics*, *40*(2), 379–403. https://doi.org/10.1093/applin/amx035

Duffy, G., & Elwood, J. (2013). The perspectives of 'disengaged' students in the 14–19 phase on motivations and barriers to learning within the contexts of institutions and classrooms. *London Review of Education*, *11*(2), 112–126. https://doi.org/10.1080/14748460.2013.799808

Fang, Z. (2006). The language demands of science reading in middle school. *International Journal of Science Education*, *28*(5), 491–520. https://doi.org/10.1080/09500690500339092

Fang, Z., Cao, P., & Murray, N. (2020). Language and meaning making: Register choices in seventh- and ninth-grade students' factual writing. *Linguistics and Education*, *56*, 100798. https://doi.org/10.1016/j.linged.2020.100798

Gardner, S., Nesi, H., & Biber, D. (2018). Discipline, level, genre: Integrating situational perspectives in a new MD analysis of university student writing. *Applied Linguistics*, *40*(4), 646–674. https://doi.org/10.1093/applin/amy005

Gee, J. P. (2004). *Situated language and learning: A critique of traditional schooling*. Routledge.

Gee, J. P. (2005). Language in the science classroom: Academic social languages as the heart of school-based literacy. In Yerrick, R., & Roth, W.-M. (Eds.), *Establishing scientific discourse communities: Multiple voices of teaching and learning research* (pp. 19—37). Lawrence Erlbaum Associates.

Gee, J. P. (2008). What is academic language? In Roseberry, A., & Warren, B. (Eds.), *Teaching science to English language learners: Building on students' strengths* (pp. 57–70) NSTA Press.

Halliday, M. A. K. (1993). Some grammatical problems in scientific English. In Halliday, M. A. K., & Martin, J. R. (Eds.), *Writing science: Literacy and discursive power* (pp. 101–118). Falmer.

Heller, V., & Morek, M. (2015). Academic discourse as situated practice: An introduction. *Linguistics and Education*, *31*, 174–186. https://doi.org/10.1016/j.linged.2014.01.008

Hunston, S. (2013). Systemic functional linguistics, corpus linguistics, and the ideology of science. *Text & Talk*, *33*(4–5). https://doi.org/10.1515/text-2013-0028

Hyland, K., & Tse, P. (2004). Metadiscourse in academic writing: A reappraisal. *Applied Linguistics*, *25*(2), 156–177. https://doi.org/10.1093/applin/25.2.156

Hyland, K., & Tse, P. (2007). Is there an "Academic Vocabulary"? *TESOL Quarterly*, *41*(2), 235–253. https://doi.org/10.1002/j.1545-7249.2007.tb00058.x

Jenks, C. (2010). Language speaks more than it says: Some reflections on the work of Basil Bernstein. *The British Journal of Sociology*, *61*(Supplement 1), 71–76. https://doi.org/10.1111/j.1468-4446.2009.01284.x

Jones, P. E. (2013). Bernstein's "codes" and the linguistics of "deficit." *Language and Education*, *27*(2), 161–179. https://doi.org/10.1080/09500782.2012.760587

Kenneth Logan, J., & Kieffer, M. J. (2017). Evaluating the role of polysemous word knowledge in reading comprehension among bilingual adolescents. *Reading and Writing*, *30*(8), 1687–1704. https://doi.org/10.1007/s11145-017-9745-1

Labov, W. (1969). The logic of non-standard English. *Georgetown Monographs on Language and Linguistics*, *22*, 1–31.

Leech, G. (2000). Grammars of spoken English: New outcomes of corpus-oriented research. *Language Learning*, *50*(4), 675–724. https://doi.org/10.1111/0023-8333.00143

Leung, C. (2014). Researching language and communication in schooling. *Linguistics and Education*, *26*, 136–144. https://doi.org/10.1016/j.linged.2014.01.005

Linell, P. (2019). The Written Language Bias (WLB) in linguistics 40 years after. *Language Sciences*, *76*, 101230. https://doi.org/10.1016/j.langsci.2019.05.003

MacFarlane, M., Barr, C., & Uccelli, P. (2020). Core academic language skills: Validating a construct in linguistically dissimilar settings. *The Language Learning Journal*, 1–14. https://doi.org/10.1080/09571736.2019.1705880

Martin, J. R. (1985). *Factual writing: Exploring and challenging social reality*. Deakin University Press.

Martin, J. R. (1993). Genre and literacy-modeling context in educational linguistics. *Annual Review of Applied Linguistics*, *13*, 141–172. https://doi.org/10.1017/s0267190500002440

Martin, J. R. (1999). Design and practice: Enacting functional linguistics. *Annual Review of Applied Linguistics*, *20*, 116–126. https://doi.org/10.1017/S026719050020007X

Martin, J. R. (2009). Genre and language learning: A social semiotic perspective. *Linguistics and Education*, *20*(1), 10–21. https://doi.org/10.1016/j.linged.2009.01.003

Martin, J. R. (2013). Embedded literacy: Knowledge as meaning. *Linguistics and Education*, *24*(1), 23–37. https://doi.org/10.1016/j.linged.2012.11.006

Merzyn, G. (1987). The language of school science. *International Journal of Science Education*, *9*(4), 483–489. https://doi.org/10.1080/0950069870090406

Nagy, W., & Townsend, D. (2012). Words as tools: Learning academic vocabulary as language acquisition. *Reading Research Quarterly*, *41*(1), 91–108. https://doi.org/10.1002/RRQ.011

Nation, I. S. P. (2001). *Learning vocabulary in another language*. Cambridge University Press.

Nation, I. S. P., & Parent, K. (2016). Homonyms and polysemes. In Nation, I. S. P. (Ed.), *Making and using word lists for language learning and testing* (41–53). John Benjamins.

Norris, S. P., & Phillips, L. M. (2003). How literacy in its fundamental sense is central to scientific literacy. *Science Education*, *87*(2), 224–240. https://doi.org/10.1002/sce.10066

Ong, W. J. (1995). *Orality and literacy: The technologizing of the world*. Routledge Press.

Oxford University Press. (2018). *Why closing the word gap matters: Oxford language report*. Oxford University Press. https://www.oup.com.cn/test/word-gap.pdf

Patterson, A., Roman, D., Friend, M., Osborne, J., & Donovan, B. (2018). Reading for meaning: The foundational knowledge every teacher of science should have. *International Journal of Science Education*, *40*(3), 291–307. https://doi.org/10.1080/09500693.2017.1416205

Patterson Williams, A. D. (2020). Sustaining disciplinary literacy in science: A transformative, just model for teaching the language of science. *Journal of Adolescent & Adult Literacy*, *64*(3), 333–336. https://doi.org/10.1002/jaal.1100

Quigley, A. (2016). *One word at a time - Teaching vocabulary*. https://www.theconfidentteacher.com/2016/11/8181/

Quigley, A. (2017). Cracking the academic code. *Teach Secondary Magazine*.

Quigley, A. (2018). *Closing the vocabulary gap*. Routledge, Taylor & Francis Group.

Quigley, A. (2020). *Teaching vocabulary –4 ways to boost secondary language skills*. https://www.teachwire.net/news/teaching-vocabulary-4-ways-to-boost-secondary-language-skills

Quigley, A. (2022). *Commonly confused academic vocabulary*. https://www.theconfidentteacher.com/2022/01/commonly-confused-academic-vocabulary/

Reznitskaya, A., Anderson, R., Nurlen, B., Nguyen-Jahiel, K., Archodidou, A., & Kim, S. (2001). Influence of oral discussion on written argument. *Discourse Processes*, *32*, 155–175. https://doi.org/10.1080/0163853X.2001.9651596

Schleppegrell, M. J. (2001). Linguistic features of the language of schooling. *Linguistics and Education*, *12*(4), 431–459. https://doi.org/10.1016/s0898-5898(01)00073-0

Schleppegrell, M. J. (2004). *The language of schooling: A functional linguistics perspective*. Lawrence Erlbaum Associates.

Schleppegrell, M. J. (2012a). Academic language in teaching and learning. *The Elementary School Journal*, *112*(3), 409–418. https://doi.org/10.1086/663297

Schleppegrell, M. J. (2012b). Systemic functional linguistics. In Gee, J. P., & Handford, M. (Eds.), *The Routledge handbook of discourse analysis* (pp. 21–34). Taylor and Francis.

Schuth, E., Köhne, J., & Weinert, S. (2017). The influence of academic vocabulary knowledge on school performance. *Learning and Instruction*, *49*, 157–165. https://doi.org/10.1016/j.learninstruc.2017.01.005

Shanahan, T., & Shanahan, C. (2008). Teaching disciplinary literacy to adolescents: Rethinking content - Area literacy. *Harvard Educational Review*, *78*(1), 40–59. https://doi.org/10.17763/haer.78.1.v62444321p602101

Snow, C. (1983). Literacy and language: Relationships during the preschool years. *Harvard Educational Review*, *53*(2), 165–189. https://doi.org/10.17763/haer.53.2.t6177w39817w2861

Snow, C., Lawrence, J. F., & White, C. (2009). Generating knowledge of academic language among urban middle school students. *Journal of Research on Educational Effectiveness*, *2*(4), 325–344. https://doi.org/10.1080/19345740903167042

Snow, C., & Uccelli, P. (2009). The challenge of academic language. In Olson, D. R., & Torrance, N. (Eds.), *The Cambridge handbook of literacy* (pp. 112–133). Cambridge University Press.

Spencer, S., Clegg, J., Stackhouse, J., & Rush, R. (2016). Contribution of spoken language and socio-economic background to adolescents' educational achievement at age 16 years. *International Journal of Language & Communication Disorders*, *52*(2), 184–196. https://doi.org/10.1111/1460-6984.12264

Strömdahl, H. R. (2011). On discerning critical elements, relationships and shifts in attaining scientific terms: The challenge of polysemy/homonymy and reference. *Science & Education*, *21*(1), 55–85. https://doi.org/10.1007/s11191-010-9335-y

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

Taljard, E. (2016). Collocations and grammatical patterns in a multilingual online term bank. *Lexikos*, *25*. https://doi.org/10.5788/25-1-1304

Tang, K. S., & Rappa, N. A. (2020). The role of metalanguage in an explicit literacy instruction on scientific explanation. *International Journal of Science and Mathematics Education*, *19*(7), 1311–1331. https://doi.org/10.1007/s10763-020-10121-6

Todd, R. W. (2017). An opaque engineering word list: Which words should a teacher focus on? *English for Specific Purposes*, *45*, 31–39. https://doi.org/10.1016/j.esp.2016.08.003

Townsend, D., Filippini, A., Collins, P., & Biancarosa, G. (2012). Evidence for the importance of academic word knowledge for the academic achievement of diverse middle school students. *The Elementary School Journal*, *112*(3), 497–518. https://doi.org/10.1086/663301

Uccelli, P., Barr, C. D. Dobbs, C. L., Phillips Galloway, E., Menesnes, A., & Sánchez, E. (2015). An expanded operational construct and a novel instrument to chart school-relevant language proficiency in pre-adolescent and adolescent learners. *Applied Psycholinguistics*, *36*, 1077–1109. https://doi.org/10.1017/S014271641400006X

Uccelli, P., & Phillips Galloway, E. (2017). Academic language across content areas: Lessons from an innovative assessment and students' reflections about language. *Journal of Adolescent and Adult Literacy*, *60*(4), 395–404. https://doi.org/10.1002/jaal.553

Uccelli, P., Galloway, E. P., Barr, C. D., Meneses, A., & Dobbs, C. L. (2015). Beyond vocabulary: Exploring cross-disciplinary academic-language proficiency and its association with reading comprehension. *Reading Research Quarterly*, *50*(3), 337–356. https://doi.org/10.1002/rrq.104

Veel, R. (2005). Learning how to mean – Scientifically speaking: Apprenticeship into scientific discourse in the secondary school. In Christie, F., & Martin, J. R. (Eds.), *Genre and institutions: Social processes in the workplace and school* (pp. 161–195). Bloomsbury.

Wang, M.-T. K., & Nation, I. S. P. (2004). Word meaning in academic English: Homography in the academic word list. *Applied Linguistics*, *25*(3), 291–314. https://doi.org/10.1093/applin/25.3.291

# 3 Corpus data and methods

*Duygu Candarli*

## Introduction

This chapter overviews the corpus linguistic data and methods used in the studies described in the remaining chapters of this book. Corpus linguistics is a methodology that analyses collections of language data known as corpora, which are normally too large to read in full and search by hand using traditional text analysis procedures. They are compiled in a principled way in an attempt to represent a language or variety of language (Baker, 2006). Corpus linguists use a range of computational techniques to examine recurrent linguistic patterns in these corpora. The advantage of corpus linguistics is that it allows the analyst to access reliable information on the frequency and nature of linguistic patterns in the corpora, which would be not possible through intuition or the use of a small number of text extracts (McEnery & Hardie, 2012). There are many corpora in existence, some very large ones attempting general coverage of a language and, increasingly, more specialised ones, enabling study of a variety or register.

In the field of education, a number of corpora of spoken and written university registers have been built (e.g., Biber et al., 2002; Biber, 2006; Nesi & Gardner, 2012; Römer & O'Donnell, 2011; Thompson & Nesi, 2001). These have been used to describe both the spoken and written language used in higher education, especially at English-medium universities. A growing body of the literature that used such corpora has made aspects of the language of university visible for instructors of English for academic purposes, content lecturers and materials writers. This can support first-year students to develop the skills required to understand a range of university registers and succeed in transitioning from school to university, and master's students from international backgrounds, among others.

At the school level, fewer corpora have been built, and none that we are aware of that cover the transition from primary to secondary school, but there is a small number representing other aspects of schooling. Durrant and Brenchley (2019) created a corpus of writing produced by students in Years 2, 6, 9 and 11 – that is, the ends of Key Stages 1–4 – in the subjects of English, science, history, geography and religious studies, which they used

DOI: 10.4324/9781003081890-3

to study the development of children's vocabulary use. Their corpus comprises a large number of texts, 2898, and contributors, 983 children, and it remains unique in representing pre-university students' writing at schools in England, but nonetheless, it is not large. The median token count of Year 2 texts is approximately 63, increasing across the years of data collection points (Durrant & Brenchley, 2019).

School textbook corpora are easier to compile in volume. Coxhead and White (2012) compiled corpora of textbooks used for English, science and social studies (a subject incorporating socially relevant themes from subjects such as history, geography and economics), at secondary schools in New Zealand, to create a relatively large corpus of 1,211,373 tokens. This is perhaps slightly unbalanced however, as more than half the tokens, 751,638, are from fiction registers in the sub-corpus of English textbooks. Green and Lambert (2018) built a corpus of 16,253,350 tokens from secondary school textbooks from the Singapore national syllabi. They used this to develop subject-specific word lists for eight secondary school subjects. Greene and Coxhead (2015) built a corpus of 18,202,382 tokens of middle school textbooks used by state school students in the United States. Middle school covers students aged approximately 10–14 years. They used their corpus as the basis for subject word lists, following Coxhead's methodology for the New Academic Word List (Coxhead, 2000). This focus on textbooks in school corpora is at odds with corpora of the language of university, which as well as student writing (Römer & O'Donnell, 2011; Nesi & Gardner, 2012), cover talk in lectures and other university speech registers (Thompson & Nesi, 2001; Simpson et al., 1999). The TOEFL 2000 Spoken and Written Language (T2K-SWAL) corpus (Biber, 2006) covers an extensive range of spoken and written university genres including course packs, course management and institutional writing texts as well as textbooks.

There are several possible reasons for the relatively small number of corpora and corpus studies of school language, and the limited number of registers that have been collected. Researchers tend to be university-based, often within language support centres, or with close links to them, so corpus research into the language of university study directly supports their teaching and students. They may also have ready access to texts from their own or co-researchers' institutions and may be able to discuss materials with discipline experts. By contrast, collecting school data requires making contacts across different educational cultures, often with more complex ethical considerations, as school students are not adults. Once identified, texts are less easy to prepare for corpus work. Although online copies of school textbooks are often available, they may not be straightforward to convert into data that can be accessed using corpus software, due to their presentation, with numerous boxed charts and figures, and embedded graphics. Collecting other kinds of data in schools can be even more challenging: it is time-consuming and resource-intensive to collect

and prepare teacher worksheets and PowerPoint presentations for corpus analysis. Spoken classroom data are very difficult to obtain, for practical and ethical reasons, and time-consuming to transcribe. A further issue, seen above in Durrant and Brenchley's (2019) study, is the small token counts of many school language texts and registers. For example, a mean token count (running words) of mathematics worksheets at Key Stage 2 in our corpus is only 240 tokens. This is symptomatic of one of the difficulties of compiling corpora of school-level texts, particularly at the younger end of the age range.

In the following sections, we describe how we constructed our corpus and the content and size of the final database. We then discuss our written and spoken corpora and the various techniques we used to analyse them.

## Constructing our corpus

### *Characteristics of our partner schools*

The data for our corpora were provided by our 13 partner schools, which we mentioned briefly in Chapter 1. We approached Huntington School in York, which at the time, in 2016, was one of five UK schools in the newly formed Research Schools network (https://researchschool.org.uk/). (The network has since developed considerably, at the time of writing comprising 28 Research Schools and ten Associate Research Schools (EEF, 2022)). Research Schools are state schools which have applied for and gained Research School status through a competitive process. The aim of Research Schools is 'to lead the way in the use of evidence-based practice and bring research closer to schools' (EEF). Their brief does not include being involved in primary research such as our project, but as we had hoped, the staff were enthusiastic to participate, and the Literacy Lead teacher agreed to be a consultant for the project (Jones & Deignan, 2021). His collaboration was invaluable at all stages, from recruiting additional schools to the project, through data collection and discussion, to disseminating findings through networks of education professionals. We approached a number of other schools known to us through our teacher training and university networks and visited schools that expressed an interest; 13 schools eventually participated and were paid an honorarium for their participation. Of the 13, eight were primary schools and five were secondary. Five of the participating primary schools directly 'feed' three of the secondary schools. That is, most or all of the students from the primary school move together to the same secondary school for Year 7. Secondary schools are considerably larger than primaries and may have around six or eight feeder primaries, with some other students coming from other primary schools in addition to the feeder schools. The relationships between our partner schools are represented in Figure 3.1.

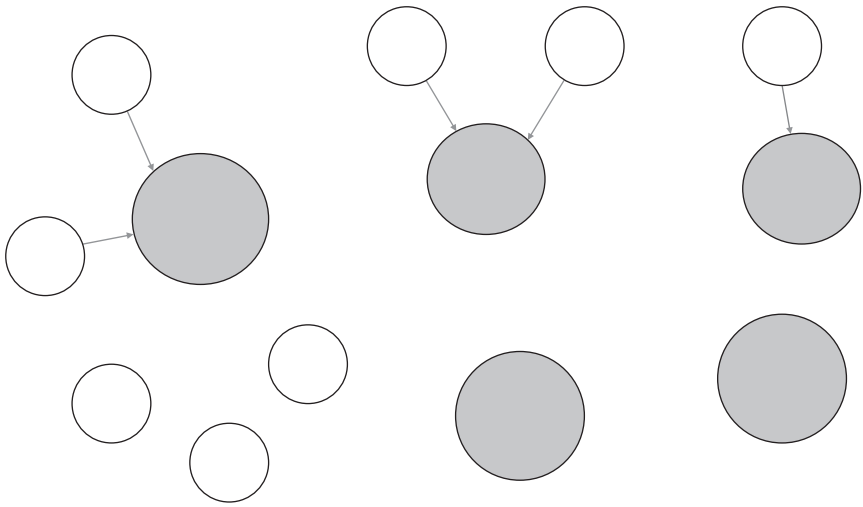


*Figure 3.1* Relationships between partner schools.

The large, grey circles represent secondary schools, and the smaller clear circles, primary schools. Arrows show where primary schools feed secondary schools. It can be seen that there are three small clusters of primary and secondary schools and five schools which have no connections to other partner schools. They are geographically dispersed across Yorkshire and the North East and include inner city, suburban and rural schools. All are state funded and non-selective.

All of the schools provided written and spoken data towards the corpus. As mentioned in Chapter 1, we also interviewed groups of students and teachers. These were from the five primary and three secondary schools that are part of clusters. We spoke to the students, six from each of the five primary schools, when they were in Year 6, and then after they had moved to Year 7, secondary school. The interview data is discussed very briefly in Chapter 1 and in more detail elsewhere (e.g., Chambers, 2020).

The characteristics of the schools are outlined here in terms of external measures as follows:

- The most recent ratings from the Office for Standards in Education, Children's Services and Skills (Ofsted) at the time of data collection;
- Eligibility for free school meals, a characteristic of the student population;
- Academic scores: for primary schools, we used the pupil progress score in reading, writing and mathematics, and for secondary schools, the Progress 8 score.

For state-funded schools in England and Wales, Ofsted inspectors make judgements on the following four areas:

(1) effectiveness of school management;
(2) quality of education provided;
(3) personal development of pupils;
(4) outcomes for pupils.

Inspectors use a four-tier rating scale: 'outstanding', 'good', 'requires improvement' and 'inadequate' (Ofsted, 2018).

Free school meals (FSM) eligibility is based on a low family income. FSM eligibility is widely correlated with potential disadvantage as well as attainment levels of pupils and schools (Gorard, 2012). We used FSM data from 2017–2018, the most recent data available at the time of the corpus compilation.

Pupil progress scores are concerned with the progress that pupils make between the end of Key Stage 1 and the end of Key Stage 2 and are used in assessing and comparing the performance of primary schools. They are calculated by comparing pupils' KS2 assessment and test results at one school with those of other schools' pupils at the national level. A score of 0 means that the students in the school perform at the same level at the end of KS2 as students with the same KS1 attainment nationally. Positive and negative scores indicate that students in the school make above average or below average progress respectively, relative to students nationally. In secondary schools, the Progress 8 score refers to the progress made between the end of Key Stage 2 and the end of Key Stage 4. It is based on GCSE results in up to eight qualifications, which include the core subjects of English, mathematics, sciences, history and geography. As for the KS2 pupil progress score, a score of 0 means that students have progressed in line with others with the same prior attainment nationally, and positive and negative scores indicate progress that is better or worse than comparable students nationally.

We gave the schools codes (school_a, school_b, school_c, etc.) to ensure their anonymity. Tables 3.1 and 3.2 show their characteristics.

As seen in Table 3.1, all the primary schools in our sample were rated 'good' at their most recent inspection at the time of data collection. Although there was no variation between the primary schools' ratings in our sample, this closely reflected the rating of the majority of the schools at the national level, since 69% of all primary schools were rated 'good' (Ofsted, 2018). At Key Stage 3, three different categories are represented in our sample, and the mean of the Ofsted ratings corresponded to 'good'; 53% of the secondary schools are rated 'good' nationally (Ofsted, 2018). It should be noted that Ofsted ratings remain highly controversial (see Perryman et al., 2018 for a discussion) and that they can only provide crude information about the overall effectiveness of schools.

*Table 3.1* Characteristics of our partner primary schools.

| *School code* | *Ofsted category* | *FSM* | *Pupil progress reading* | *Pupil progress writing* | *Pupil progress mathematics* |
|---|---|---|---|---|---|
| school_a | Good (2) | 8.9% | –0.9 (average) | –1.2 (average) | –0.8 (average) |
| school_b | Good (2) | 28.4% | –2.5 (below average) | –1.4 (average) | –2.6 below (average) |
| school_c | Good (2) | 9.4% | 0 (average) | 1 (average) | –0.4 (average) |
| school_d | Good (2) | 48.5% | –0.1 (average) | 2.2 (average) | –0.9 (average) |
| school_e | Good (2) | 8.3% | –1.8 (average) | 1.9 (above average) | –0.6 (average) |
| school_f | Good (2) | 9.5% | 3.4 (well above average) | –1.4 (average) | 1.1 (average) |
| school_g | Good (2) | 13.6% | –1.2 (average) | 0.5 (average) | 2.4 (average) |
| school_h | Good (2) | 17.3% | –3.7 (well below average) | –2 (average) | –2.8 (below average) |
| Mean | Good (2) | 17.99% | –0.85 (average) | –0.05 (average) | –0.58 (average) |
| Standard deviation | 0 | 14.05 | 2.11 | 1.65 | 1.73 |

*Table 3.2* Characteristics of our partner secondary schools.

| *School code* | *Ofsted category* | *FSM* | *Progress 8 score* |
|---|---|---|---|
| school_i | Outstanding (1) | 15.2% | –0.13 (average) |
| school_j | Outstanding (1) | 30.4% | 0.7 (well above average) |
| school_k | Outstanding (1) | 12% | 0.28 (above average) |
| school_l | Good (2) | 14.7% | 0.12 (average) |
| school_m | Inadequate (4) | 17.7% | –0.14 (average) |
| Mean | Good (1.8) | 18% | 0.17 (above average) |
| Standard deviation | 1.3 | 7.22 | 0.31 |

The mean percentage of the pupils who had been eligible for FSM at any time during the past six years in our primary school sample was 17.99%. This was below the national average, 24.3% at the time of data collection. The same figure in our secondary school sample was 18%, below the national average of 28.6%. The congruence between the primary and secondary schools' mean Ofsted ratings and mean percentage of pupils eligible for FSM made the profile of the schools similar at KS2 and KS3 levels.

As shown in Table 3.1, most of the scores in reading, writing and mathematics were average scores in our sample. The mean scores approximately correspond to the national average, equivalent to 64% of all schools in reading, 67% of all schools in writing and 57% of all schools in mathematics (DfE, 2016). Table 3.2 shows that the mean Progress 8 score in our secondary sample was 0.17, which corresponds to the above-average score that only 17% of all schools received nationally (DfE, 2016).

Taken together, the mean Ofsted ratings for both our primary and secondary school sample and pupil progress score for our primary school sample are very similar to the average school at the national level. The mean percentage of FSM was below the national average for both our primary and secondary school samples, meaning that the student population of our sample schools is probably slightly more advantaged than the national average. The mean Progress 8 score was slightly above the national average in our secondary school sample. It was not possible to recruit a more diverse sample of secondary schools even though we made multiple attempts to do so. Schools in the categories 'requires improvement' and 'inadequate' are subject to reinspection monitoring by inspectors (Ofsted, 2022), which arguably leaves little time for teachers and head teachers to collaborate with universities for research. We were told several times by leaders in such schools that while they were interested in our research, it could not be a priority for them.

### *Corpus design and representativeness*

As noted earlier, we built a corpus using data supplied by our 13 partner schools. The corpus can be split in two ways: into written and spoken texts, and into Key Stage 2 and Key Stage 3 texts. It consists of texts from the subjects of English, mathematics, science, history and geography, on the basis of the subjects used for Progress 8, which we took as a proxy for valued subjects. Each subject can be analysed separately. As we intended the corpus to represent the language that students encounter during the academic part of their schooling from teachers and other educationalists, it contains no student-produced texts. With our focus on the transition, we collected data from Years 5 and 6 for the Key Stage 2 corpus, and Years 7 and 8 for the Key Stage 3 corpus, although the complete Key Stages are comprised of additional years (see Chapter 1). We collected the data in the school year of 2018–2019.

Biber describes representativeness in corpus design as 'the extent to which a sample includes the full range of variability in population' (1993, p. 243). He notes that preconditions for achieving representativeness are that the population from which the corpus is sampled is clearly defined and that the range of text types that the population comprises is fully known. Taking the first of these, our population is the academic language encountered by students at English state schools in Years 5 to 8 in Progress 8 subjects, and the sample is texts sourced from the 13 schools that had agreed to be project partners. This leads to a restriction on the situational parameters, the geographical representation, as all our partner schools are located in northern England. We have reasonable confidence that our sample is representative of the population in terms of academic content because all state-funded schools in England are obliged to follow the detailed specifications of the National Curriculum, and

textbooks have national reach. The previous section outlines to what extent the schools are representative in other ways.

The second of Biber's preconditions, knowing the range of text types, is not straightforward, as students study multiple subjects, and within these, encounter many registers. We sought to ensure ecological validity (Stangor, 1998) as far as possible, that is, to ensure that the resources that were collected are similar to the everyday life experience of language users – school students. Some school subjects are taught almost daily and some less frequently. To ensure that the weighting of subject materials in the corpus approximately reflected the time students spent on each, we obtained sample timetables from the schools. Table 3.3 shows the timetable of a class of Year 7 students at one of our partner secondary schools.

We also discussed the composition of the corpus with teachers at our partner schools, and in particular with the project consultant from Huntington School. A sampling frame was designed to include both the written and spoken registers of the subjects of English, mathematics, science, history and geography that would reflect their class times to create a representative and balanced corpus as much as possible; however, no target was set for the number of texts or text length and resources were collected in 'an opportunistic mode' (McEnery & Hardie, 2012, p. 64). As McEnery and Brookes (2022, p. 37) note, 'balanced, representative corpora are best viewed as a theoretical ideal rather than being necessarily achievable in practice'.

In addition to the corpus design that involves representativeness and balance, ethics and copyright are the other important considerations in building a corpus (McEnery & Hardie, 2012; McEnery & Brookes, 2022). In our written corpus, textbooks and some of the commercial presentations and worksheets are subject to copyright restrictions, and they cannot be redistributed publicly. In teacher-created resources, such as assessments and worksheets, we anonymised the names of the schools when the school name was present. As we describe below, our spoken corpus only includes the anonymised transcriptions of teachers who provided written informed

*Table 3.3* A weekly timetable for Year 7 students.

| **Monday** | **Tuesday** | **Wednesday** | **Thursday** | **Friday** |
|---|---|---|---|---|
| Registration | Registration | Registration | Registration | Registration |
| Geography | ICT and Computing | Religion, Philosophy and Ethics | Mathematics | Science |
| Physical Education | Art | History | Geography | Physical Education |
| Tutor Report | French | Mathematics | Science | Drama |
| French | Science | Food & Textiles Technology | English | Music |
| Technology | English | English | History | Mathematics |

consent to participate in our project. We discuss the registers within the corpus in the following sections on the written and spoken corpora.

## The written corpus

### *Representativeness and data gathering*

We have discussed our use of the student timetables in our decisions about the overall balance of the corpus. In order to increase the degree of representativeness of the written corpus at the level of individual subject, we reviewed the Department for Education's (DfE) national curriculum documents for each subject for KS2 and KS3 in England (DfE, 2013, 2014). This gave us an understanding of attainment targets and topics as well as notes and guidance aimed at teachers and led to the inclusion of additional materials. For example, the programme of study of English has a word list for Years 5 and 6, and this was included in our corpus.

A particular issue for the written corpus was the wide range of written registers. We consulted the teachers in each school in order to determine registers that were used in each subject and identify the approximate extent of their use during lessons. For instance, we found that presentations and worksheets are central registers of academic language in lessons. Textbooks are used only occasionally, around 10% of the class time, or in some subjects, not used at all. These distributions differed from one subject to another. Naturally, teachers could only give us rough estimates, but they were nonetheless a useful guide to informing decisions about what proportions of each register to include for each subject. The practice of consulting informants is a crucial step in developing corpora for English for specific purposes and validating registers and their representation in accurate proportions in the corpus (Gray, 2015).

Where possible, a soft copy version of the written resources was collected. When no soft copy version of the resources was available, a hard copy of these resources was collected and scanned. Then, we used the software package *ABBYY PDF Transformer+* for optical character recognition (OCR). We manually checked all the scanned resources and corrected any OCR errors. All the written resources were converted to plain text files with UTF-8 encoding for corpus analysis, though it should be noted that some corpus software, including #LancsBox (Brezina et al., 2020) and AntConc v.4 (Anthony, 2022) can read PDF and Word files. We used #Lancsbox v.6.0 (Brezina et al., 2020) to calculate token counts.

### *Composition of the written corpus*

Tables 3.4 and 3.5 show the composition of the written corpora, divided into the five subject areas that we collected.

As we noted above, we sought to make the corpus ecologically valid through consulting with teachers about the balance of subjects and

*Table 3.4* KS2 written corpus.

| *Subject* | *Texts* | *Tokens* | *Mean length* | *SD text length* |
|---|---|---|---|---|
| English | 600 | 303,257 | 505 | 1381 |
| Mathematics | 614 | 174,337 | 284 | 904 |
| Science | 177 | 160,355 | 906 | 3069 |
| History | 140 | 83,998 | 600 | 683 |
| Geography | 152 | 62,300 | 410 | 541 |
| Total | 1683 | 784,247 | | |

*Table 3.5* KS3 written corpus.

| *Subject* | *Texts* | *Tokens* | *Mean length* | *SD text length* |
|---|---|---|---|---|
| English | 334 | 260,806 | 781 | 3552 |
| Mathematics | 872 | 257,459 | 295 | 353 |
| Science | 675 | 356,319 | 528 | 3046 |
| History | 156 | 233,600 | 1497 | 9141 |
| Geography | 170 | 70,503 | 415 | 346 |
| Total | 2207 | 1,178,687 | | |

consulting timetables. We did not therefore attempt to gather additional materials to increase the size of the small sub-corpora, as this might have distorted the importance of that subject in the corpus as a whole, threatening representativeness. In particular, the KS2 history and geography written sub-corpora are very small because these subjects are not taught explicitly in primary schools. Instead, students have a timetable slot for 'topics', which covers content related to science, geography and history. We classified these texts into the subjects of science, geography and history, consulting with the primary school teachers who used the materials, and who had sometimes designed them. The heavy weighting of English and mathematics in KS2 is almost certainly partly due to the amount of time that is spent in Year 6 on preparing for the national SATs (Standard Attainment Tests, see Chapter 1), which cover English and mathematics. It can be seen that there is a big increase in the relative size of the science sub-corpus at KS3, which may constitute one aspect of the linguistic challenge for students. In addition to the subject categorisation of the written school language registers, we also categorised them into sub-registers to explore lexico-grammatical variation in the written school language resources in terms of both subjects and sub-registers across the Key Stages.

### *Sub-registers*

Register studies use a number of situational characteristics to describe texts, including participants, relations among participants, channel/mode, setting,

communicative purposes and topic (Biber & Conrad, 2019). The participants of school registers were teachers and students in a classroom setting, and at the top level of analysis, the registers are the written and spoken resources of English, mathematics, science, history and geography. We conducted a systematic data-driven categorisation of the school sub-registers in our written corpus and focused on mode and communicative purposes of the texts in order to categorise them into sub-registers. Our findings are shown in Table 3.6. As can be seen, mode refers to the channel of the school sub-registers that were presented to students. Two resources were used for the identification of the school sub-registers and description of their situational characteristics, as recommended by Biber and Conrad (2019): (1) the insights that we gained from the teachers, expert informants in this context, into the school registers and the purposes of texts; (2) our examination of the texts within the registers that we conducted to identify their communicative purposes. With the exception of textbooks and fiction, all the texts in our written corpus were read and analysed inductively in order to describe their primary and other communicative purposes and identify their sub-registers.

In addition to their primary communicative purposes shown in Table 3.6, the school sub-registers served other purposes. For example, worksheets, which were presented both electronically and in written mode to students, contained exercises and questions that students were expected to complete in order to practise subject topics and strengthen their learning, and they also included short reading extracts, accompanied by questions related to them, to convey information. Presentations, which were electronic resources, primarily included informational subject-specific content on the topics but also contained warm-up questions and practice exercises to enable students to practise content. Like presentations, textbooks also served a multifunctional purpose at schools. The primary function of textbooks was to provide students with information on subject topics, and they included assessment tasks that assessed students' knowledge as well as exercises and

*Table 3.6* Written school language sub-registers and their situational characteristics.

| *Sub-registers* | *Mode* | *Primary communicative purpose* |
|---|---|---|
| Worksheets | Written/electronic written | Practising subject content and reinforcing learning |
| Presentations | Electronic written | Presenting subject content |
| Textbooks | Written | Presenting subject content |
| Assessment tasks | Written/electronic written | Assessing students' knowledge |
| Reading extract | Written | Presenting exposition |
| Glossary | Written | Presenting vocabulary and its definitions |

questions that were aimed to reinforce students' learning. Assessment tasks involved exams, quizzes, peer assessment tasks and self-assessment criteria that evaluated students' summative or formative progress. We describe reading extracts, unaccompanied by any exercises or questions, as non-fiction expository texts on subject-specific topics that introduced information to students. Similarly, the glossary included vocabulary and its definitions without any exercises or questions. It should be noted that there was also the register of fiction that students encountered in their English classes. This fiction register was in the form of novels and stories that students read as part of their English classes. Although we collected these written resources, we have excluded the register of fiction in this book, since it does not meet our definition of the academic language of school, introduced in Chapter 2.

## The spoken corpus

We also constructed a spoken corpus that comprised transcribed teacher talk in Years 5–8. As for the written corpus, this is divided into KS2 and KS3, and can be further sub-divided by year group and subject. We aimed to represent the teacher talk encountered by students in the subjects of English, mathematics, science, history and geography. In this book, we report our analysis of teacher talk in English, mathematics and science.

Audio recordings were collected from our partner schools. A number of teachers at our partner schools gave written informed consent to be recorded. We had anticipated some reluctance but did not find any. The teachers were provided with audio recorders and microphones worn on a lanyard, and they were asked to record their lessons themselves, without an observer. We did not set out to record student talk, in line with the project aims, so a lanyard microphone was ideal.

The teacher talk was transcribed by a professional agency. Any student contributions that happened to be audible were ignored and not transcribed. We did not have informed consent from students for their utterances to be recorded and transcribed and are not analysing these data. Occasionally, this makes interpreting the teacher utterances difficult, when they are responding to student questions, for example. Obtaining informed consent from all the students, usually around 30 per class, and from their parents or caregivers would have been unmanageable.

The transcribers used an orthographic transcription scheme adapted from the spoken British National Corpus (BNC) 2014 transcription scheme (see Love et al., 2017). The teachers were allocated codes to ensure anonymity. In order to ensure accuracy and consistency of the transcription, a research assistant manually checked all the transcribed texts and corrected any errors. Below is an example extract of an English lesson in Year 5.

| *line* | *speaker* | *utterance* |
|---|---|---|
| 1 | T061 | with your partner (.) there's a few tricky ones on here (.) today (.) okay and don't worry if you don't know it this is why we practise it isn't it so that we can go through it again and again (.) who's got is there anyone's that's got ten on one of our SPaG tests ye=yet? (.) okay look only a few very few people in the class (.) so well done if you have (.) but it's obviously very tricky (.) okay anybody now hasn't got their partner's test in front of them ready to mark it? (.) you might have two to mark somebody's gone (.) gone off somewhere (.) okay this one you should have been able to do I hope which sentence uses a relative clause the map that I brought with me is out of date or the map I bought yesterday is out of date (.) so which one is the relative clause? go on <name M>?# |

*Extract 3.1,* Year 5 English lesson recording, Teacher 061.

While transcribing the teacher talk, no punctuation marks were used, except for question marks. A short pause was marked by a tag (.). The sampling frame for the collection of audio recordings was designed to represent all five subjects in proportion to their distribution within the timetable for one week at school (see Figure 3.2). For instance, we collected three lesson recordings of English, mathematics and science separately in Years 7 and 8 (KS3) to represent teacher talk. A similar procedure was followed for the lesson recordings of Years 5 and 6 (KS2), taking into account the timetable of our partner primary schools. In total, we collected 218 audio recordings. Due to the time-consuming and resource-intensive nature of high-quality transcription procedures, to date, we have only 108 fully transcribed audio recordings of English, mathematics, science, geography and history subjects, as shown in Table 3.7. This means that the spoken corpus of teacher talk in this book is not a balanced corpus of teacher talk at the transition. To our knowledge, however, it is still the largest corpus of teacher talk at the transition from primary to secondary school.

The corpus size, at 506,517 tokens, was calculated using #Lancsbox v.6 (Brezina et al., 2020). The mean text length of teacher talk showed an increase in all the subjects from KS2 to KS3, suggesting that the volume of teacher talk that the students encountered in one lesson on average increased at KS3. The larger standard deviations in text lengths at KS2 than KS3 indicated that the length of the teacher talk varied to a greater extent at KS2 than KS3, except for the history subject.

Our written and spoken corpus for both KS2 and KS3 totals 2,469,451 tokens. We divide this up in a number of different ways for the various analyses presented in the following chapters. In Chapter 4, we analyse the written data only, and take out the texts that include fewer than 100 tokens; in Chapters 5, 6 and 7, we focus on specific subjects and treat written and spoken data together.

*Table 3.7* The spoken corpus of teacher talk.

| | *Number of texts* | | *Number of tokens* | | *Mean text length (tokens)* | | *Standard deviation text length (tokens)* | |
|---|---|---|---|---|---|---|---|---|
| | *KS2* | *KS3* | *KS2* | *KS3* | *KS2* | *KS3* | *KS2* | *KS3* |
| English | 15 | 8 | 72,475 | 47,595 | 4832 | 5949 | 1284 | 992 |
| Subtotal (English) | 23 | | 120,070 | | | | | |
| Mathematics | 18 | 11 | 82,031 | 51,171 | 4557 | 4652 | 2225 | 1467 |
| Subtotal (Mathematics) | 29 | | 133,202 | | | | | |
| Science | 18 | 10 | 62,375 | 47,772 | 3465 | 4777 | 2923 | 1410 |
| Subtotal (Science) | 28 | | 110,147 | | | | | |
| History | 6 | 7 | 22,114 | 39,319 | 3686 | 5617 | 1621 | 1354 |
| Subtotal (History) | 13 | | 61,433 | | | | | |
| Geography | 6 | 9 | 24,472 | 57,193 | 4079 | 6355 | 1345 | 2440 |
| Subtotal (Geography) | 15 | | 81,665 | | | | | |
| Grand total | 108 | | 506,517 | | | | | |

## Corpus analytical methods used

Thompson and Hunston give an elegant description of their use of corpus methods, as follows: 'We apply to the data that we have collected […] corpus investigation methods that rearrange and process that data. Our challenge then is to make sense of the rearranged data' (2019, p. 6). We have described above how we collected a relatively large quantity of school language data; we now describe how we rearranged and processed it, and in subsequent chapters, how we made sense of it. To explore our corpus, we have used a range of methods: quantitative, qualitative and mixed methods. Thompson and Hunston (2019, p. 6) place their corpus methods on a cline from qualitative to quantitative. Figure 3.2 is taken from their discussion.

We also use a range of corpus methods. As our study centrally concerns comparing two corpora, KS2 and KS3, in different ways, it is to be expected

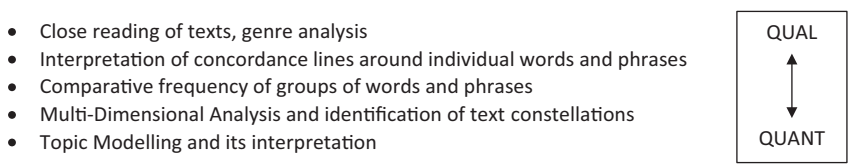



*Figure 3.2* Thompson and Hunston's representation of methods used in their studies of interdisciplinary genres (2019, p. 6).

that the central method, comparing the frequency of words and phrases, is at its heart. We also use multi-dimensional (MD) analysis and undertake detailed interpretation of concordance lines, thus covering the central three methods from Figure 3.2. We now overview these methods, starting from the quantitative end of the cline, with MD analysis, as this reflects the order of our chapters. Further methodological detail is given in individual chapters.

## Quantitative data analysis procedures

### *Multi-dimensional analysis*

MD analysis was originally developed by Biber (1988). It is a quantitatively driven analytical approach to corpus analysis, which aims to provide a comprehensive linguistic description of registers. MD analysis is 'derived from factor analysis […] which observes the sequential, partial, and observed correlations of a wide range of variables in order to produce groups of co-occurring factors' (Friginal, 2013, p. 138). It is based on the premise that texts in the same register will exhibit clusters of co-occurring linguistic features, which reflect the underlying communicative functions of the register (Biber, 1988; Friginal, 2013; Biber & Conrad, 2019). For example, 'private verbs', such as *assume*, *believe*, *doubt* and *know* are found through factor analysis to co-occur with a group of other linguistic characteristics that includes present tense, second person pronoun and use of *DO* as a pro-verb (Biber, 1988, p. 75). Taken together, qualitative analysis shows that these features are associated with 'involved' discourse. Their relative absence and the presence of other features such as agentless passives and attributive adjectives are associated with 'informational' discourse. This analysis leads to the construction of a 'dimension', 'involved vs informational'.

There are several steps to conducting MD analysis. First, frequencies of lexico-grammatical features are counted across the registers in the corpus. Linguistic co-occurrence patterns that constitute an underlying dimension of variation are identified quantitatively using factor analysis. Then, each dimension of variation, statistically determined, is analysed qualitatively to construct the underlying communicative functions associated with each dimension in different registers. In his seminal work, Biber (1988) found six main dimensions of variation in a general corpus of written and spoken registers, and a seventh, which was not matched to a functional interpretation. The first five dimensions that we focus on in this study are as follows:

#### *Dimension 1: Involved versus informational discourse*

A positive score on Dimension 1 indicates involved discourse (e.g., conversational registers), while a negative score indicates informational discourse

(written registers, such as academic prose). The positively loaded (involved) linguistic features are as follows (Biber, 1988, p. 102):

> private verbs, that-deletions, contractions, present tense verbs, second pronouns, *do* as pro-verb, analytic negations, demonstrative pronouns, first person pronouns, pronoun *it*, *be* as main verb, causative subordination, discourse particles, indefinite pronouns, general hedges, amplifiers, sentence relatives, wh-questions, possibility modals, non-phrasal coordination, wh-clauses, final prepositions.

The negatively loaded (informational) linguistic features include 'nouns, word length, prepositions, type/token ratio, attributive adjectives' (Biber, 1988, p. 102).

*Dimension 2: Narrative versus non-narrative discourse*

A positive Dimension 2 score represents narrative discourse marked by past events (e.g., fiction) whereas a negative Dimension 2 score represents non-narrative discourse (e.g., academic prose). The positively loaded (narrative) linguistic features are 'past tense verbs, third person pronouns, perfects aspect verbs, public verbs, synthetic negation, present participial clauses' (Biber, 1988, p. 102). The negatively loaded linguistic features are 'present tense verbs, attributive adjectives, past participial WHIZ deletions (past participial forms of verbs as post-nominal modifiers – the solution proposed by the team), and word length' (Biber, 1988, p. 102).

*Dimension 3: Situation-dependent versus elaborated reference*

A positive Dimension 3 score characterises discourse dependent on the situation (e.g., a sports broadcast) while a negative Dimension 3 score exhibits elaborated reference and independence of the context (e.g., academic prose). The positively loaded linguistic features are 'wh-relative clauses on object positions, pied piping constructions, wh-relative clauses on subject positions, phrasal coordination, nominalisations' (Biber, 1988, p. 102). The negatively loaded features are 'time adverbials, place adverbials, and adverbs' (Biber, 1988, p. 102).

*Dimension 4: Overt expression of persuasion*

A positive Dimension 4 score is characteristic of persuasive discourse (e.g., editorials). The positive linguistic features of this dimension are 'infinitives, prediction modals, suasive verbs, conditional subordination, necessity modals, split auxiliaries' (Biber, 1988, p. 103). There are no negatively loaded linguistic features of this dimension.

*Dimension 5: Abstract versus non-abstract information*

A positive Dimension 5 score denotes abstract discourse (e.g., scientific discourse) whereas a negative Dimension 5 score denotes non-abstract discourse. The positive linguistic features are 'conjuncts, agentless passives, past participial clauses, by-passives, past participial WHIZ deletions, other adverbial subordinators' (Biber, 1988, p. 103). A relatively low type/token ratio, that is, lack of lexical variation, is the only linguistic feature negatively loaded to this dimension. (Biber notes that although this may seem surprising, abstract discourse is often technical, and tends to repeat key terms rather than seeking stylistic variation.)

MD analysis is ideally suited to investigating register changes between KS2 and KS3, with its potential for finding subtle, measurable distinctions along a large number of linguistic features and dimensions. The method has been used with many different corpora to date (see Berber Sardinha & Veirano Pinto, 2019). In Chapter 4, we discuss MD analysis studies relevant to our own. We then report an MD analysis of our written corpus of English, mathematics and science subjects at KS2 and KS3.

## Mixed and qualitative data analysis

Towards the qualitative end of Thompson and Hunston's cline shown in Figure 3.2 is the method: 'Interpretation of concordance lines around specific words and phrases' (2019). This method was an integral part of our studies; in order to help teachers and students with the linguistic challenges of secondary school, we need to be able to provide details of usage and meaning. However, we needed a way into our corpus before examining concordances. Using concordance examination on its own is indicated when the central research questions entail the detailed analysis of pre-determined words and expressions. For example, Auge (2021) sought to identify the associations of the expression *greenhouse effect* across a range of registers. In other cases, studying one set of words and expressions can lead to further concordance analysis. For example, Islentyeva and Kafi (2021) studied attitudes towards the EU in the British press from 2016–2018. They began with the words *Britain*, *European* and the *EU*, and used corpus software to identify the most significant collocates immediately before and after each of the three words, finding words such as *voters*, *people*, *migrants* and *culture*. These were then classified semantically and analysed in detail using concordances. Another approach, if the corpus is fairly homogenous, has been to manually analyse a sample and identify candidates of interest. Charteris-Black (2004) has taken this approach in his study of the ideological use of metaphorical meanings of words in corpora.

With our corpora and for the research questions that we look at in Chapters 5, 6 and 7, none of these approaches would be sufficient as a starting point. We know from teachers' reports, and from examples that have

come to our notice, that KS2 and KS3 language use is likely to be different at the level of detail. However, we begin from the position of not knowing in advance which words would be significant, and our corpus is far from homogenous, so sampling would not be effective. We therefore began by using tools that showed us what was frequent in our corpus and sub-corpora, and what was more frequent in each sub-corpus relative to the others. The results of these analyses are valuable in themselves, and also give us the starting points for more detailed, qualitative studies. In discourse studies, corpus techniques increase the rigour of the analysis and minimise the researcher's subjective selection of texts and linguistic features for analysis since corpus techniques, such as keyness analysis that we use, point to frequent linguistic features that are important in the corpus and provide quantitative information on their frequency of occurrence that would underpin qualitative interpretations (e.g., Baker, 2006; Mautner, 2022). Hence, cherry-picking of texts and linguistic features is avoided in corpus-informed qualitative studies of the meaning and function of words and phrases.

A number of previous studies have taken a frequency-based approach, using wordlists and keyness analysis to compare different corpora. Deignan et al. (2019), comparing metaphorical uses in different corpora of texts on the topic of climate change, used word lists to identify the most frequent lexical words, and then studied concordances of these words in detail. Baker et al. (2013) conducted a detailed Critical Discourse Analysis of a corpus of British newspapers, following a number of steps. They began with word lists to get an overall sense of 'aboutness', and to look for expected and unexpected semantic domains. They then compared different sections of their corpus against each other, highlighting frequent words, followed by detailed concordance examination. We use variations of Baker et al.'s approach in Chapters 5, 6 and 7 when we study KS2 and KS3 English, science and mathematics sub-corpora.

Frequency was measured using two related tools. First, to produce a list of the most frequent words in each sub-corpus, we used the Words tool in #LancsBox 6.0 (Brezina et al., 2020). Second, to compare word frequencies across different sub-corpora, we used the keywords technique (Baker, 2006; Rayson, 2019), also available within #LancsBox 6.0. The keywords tool allows the researcher to compare the lexical make-up of two corpora, by showing us which words are significantly more frequent in one than the other. There are a number of different statistical options within the tool, and when we describe the studies in the following chapters, we explain the choices that we made. Using the keywords tool, we generated lists of words that were, for example, significantly more frequent in KS3 English than KS2 English, or than a general or reference corpus. We discuss the use of reference corpora, and how we developed a reference corpus for this project, in Chapter 5. The resulting list has to be carefully checked manually, as it will contain items that are not of interest, such as proper names from literature texts and text codes. Once irrelevant items have been deleted, the lists are of interest in themselves and also as the starting point for concordance analysis.

This follows the approach used by researchers, including Gabrielatos (2018) and Partington and Duiguid (2021), of using keywords as a way into a specialised corpus.

> [The researcher derives] a list of key items ranked according to the value of the keyness metric used in the study. At this point, the researcher may switch to a targeted approach and select particular types of items for concordance analysis according to explicit criteria, such as their normalised or raw frequency, part of speech, core sense, or relation to a particular topic.
>
> (Gabrielatos 2018, p. 3)

We studied concordance data for all words identified through frequency lists and the keywords tool, reflecting our aim to describe as fully as possible the language challenge of secondary school. Our concordance analyses followed well-established procedures such as those described by Sinclair (e.g., 1991, 2003). That is, we identified the different meanings of the words and phrases under study, and considered their function, using expanded context to support this. We examined the syntactic patterns and collocations that these words and phrases occur. As the following chapters show, this qualitative analysis often showed subtle but important differences in meaning and use between the different registers in our corpus, and sometimes between school registers and non-school language, as represented by a reference corpus.

## Conclusion

In this chapter, we have explained how we attempted to tackle our questions about the challenge of the language of secondary school faced by KS3 students. To our knowledge, this is the first corpus to represent both written and spoken school language at the transition from primary to secondary school. This is also the first study to explore to what extent, if any, school language changes from primary to secondary school in both spoken and written modes by using corpus methods and quantitative and qualitative corpus techniques in order to investigate lexico-grammatical variation across the subjects and Key Stages. In the following four chapters, we describe various studies that we have conducted using these data and methods.

## References

Anthony, L. (2022). *AntConc* (Version 4.0.3) [Computer Software]. Waseda University. https://www.laurenceanthony.net/software

Auge, A. (2021). From scientific arguments to scepticism: Humans' place in the greenhouse. *Public Understanding of Science*, *31*(2), 179–194. https://doi.org/10.1177/09636625211035624.

Baker, P. (2006). *Using corpora in discourse analysis*. Continuum.

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

Berber Sardinha, T., & Veirano Pinto, M. (Eds.). (2019). *Multi-dimensional analysis: Research methods and current issues*. Bloomsbury Academic. https://doi.org/10.5040/9781350023857

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

Biber, D. (1993). Representativeness in corpus design. *Literary and Linguistic Computing*, *8*(4), 243–257. https://doi.org/10.1093/llc/8.4.243

Biber, D. (2006). *University language: A corpus-based study of spoken and written registers* (Vol. 23). John Benjamins Publishing.

Biber, D., & Conrad, S. (2019). *Register, genre, and style* (2nd ed.). Cambridge University Press.

Biber, D., Conrad, S., Reppen, R., Byrd, P., & Helt, M. (2002). Speaking and writing in the university: A multidimensional comparison. *TESOL Quarterly*, *36*(1), 9–48. https://doi.org/10.2307/3588359

Brezina, V., Weill-Tessier, P., & McEnery, A. (2020). #LancsBox v. 6. [software]. http://corpora.lancs.ac.uk/lancsbox.

Chambers, G. (2020). *Linguistic encounters: What year 7 students say about 'difficult words'*. https://linguistictransition.leeds.ac.uk/linguistic-encounters-what-year-7-students-say-about-difficult-words/

Charteris-Black, J. (2004). *Corpus approaches to critical metaphor analysis*. Palgrave Macmillan.

Coxhead, A. (2000). A new academic word list. *TESOL Quarterly*, *34*(2), 213–238. https://doi.org/10.2307/3587951

Coxhead, A., & White, R. (2012). Building a corpus of secondary school texts: First you have to catch the rabbit. *New Zealand Studies in Applied Linguistics*, *18*(2), 67–73. https://search.informit.org/doi/abs/10.3316/informit.108520980622633

Deignan, A., Semino, E., & Paul, S.-A. (2019). Metaphors of climate science in three genres: Research articles, educational texts, and secondary school student talk. *Applied Linguistics*, *40*(2), 379–403. https://doi.org/10.1093/applin/amx035

Department for Education (DfE). (2013). *The national curriculum in England: Key stages 1 and 2 framework document*. https://www.gov.uk/national-curriculum

Department for Education (DfE). (2014). *The national curriculum in England: Key stages 3 and 4 framework document*. https://www.gov.uk/national-curriculum

Department for Education (DfE). (2016). *Understanding school and college performance measures*. https://www.gov.uk/government/publications/understanding-school-and-college-performance-measures

Durrant, P., & Brenchley, M. (2019). Development of vocabulary sophistication across genres in English children's writing. *Reading and Writing*, *32*(8), 1927–1953. https://doi.org/10.1007/s11145-018-9932-8

Education Endowment Foundation (EEF). (2022). https://educationendowmentfoundation.org.uk/support-for-schools/research-schools-network

Friginal, E. (2013). Twenty-five years of Biber's multi-dimensional analysis: Introduction to the special issue and an interview with Douglas Biber. *Corpora*, *8*(2), 137–152. https://doi.org/10.3366/cor.2013

Gabrielatos, C. (2018). Keyness analysis: Nature, metrics and techniques. In Taylor, C., & Marchi, A. (Eds.), *Corpus approaches to discourse: A critical review* (pp. 225–258). Routledge.

Gorard, S. (2012). Who is eligible for free school meals? Characterising free school meals as a measure of disadvantage in England. *British Educational Research Journal*, *38*(6), 1003–1017. https://doi.org/10.1080/01411926.2011.608118

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

Green, C., & Lambert, J. (2018). Advancing disciplinary literacy through English for academic purposes: Discipline-specific wordlists, collocations and word families for eight secondary subjects. *Journal of English for Academic Purposes*, *35*, 105–115. https://doi.org/10.1016/j.jeap.2018.07.004

Greene Wells, J., & Coxhead, A. (2015). *Academic vocabulary for middle school students: Research-based lists and strategies for key content areas*. Brookes Publishing.

Islentyeva, A., & Abdel Kafi, M. (2021). Constructing national identity in the British Press: The Britain vs. Europe dichotomy. *Journal of Corpora and Discourse Studies*, *4*(0), 68. https://doi.org/10.18573/jcads.64

Jones, M., & Deignan, A. (2021, September 4). *The linguistic challenges of the transition from primary to secondary school*. [Paper presentation] ResearchED National Conference 2021, London, United Kingdom.

Love, R., Dembry, C., Hardie, A., Brezina, V., & McEnery, T. (2017). The spoken BNC2014. *International Journal of Corpus Linguistics*, *22*(3), 319–344. https://doi.org/10.1075/ijcl.22.3.02lov

Mautner, G. (2022). What can a corpus tell us about discourse? In A. O'Keeffe & M. McCarthy (Eds.), *The Routledge handbook of corpus linguistics* (pp. 250–262). Routledge.

McEnery, T., & Brookes, G. (2022). Building a written corpus: What are the basics? In A. O'Keeffe & M. McCarthy (Eds.), *The Routledge handbook of corpus linguistics* (pp. 35–47). Routledge.

McEnery, T., & Hardie, A. (2012). *Corpus linguistics: Method, theory and practice*. Cambridge University Press.

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

Ofsted. (2018). *State-funded schools inspections and outcomes as at 31 August 2018*. https://www.gov.uk/government/statistics/state-funded-schools-inspections-and-outcomes-as-at-31-august-2018

Ofsted. (2022). *School inspections: A guide for parents*. https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/1044285/School_inspections_-_a_guide_for_parents_January_2022.pdf

Partington, A., & Duiguid, A. (2021). Political media discourses. In Friginal, E., & Hardy, J. (Eds.), *The Routledge handbook of corpus approaches to discourse analysis* (pp. 116–135). Routledge.

Perryman, J., Maguire, M., Braun, A., & Ball, S. (2018). Surveillance, governmentality and moving the goalposts: The influence of Ofsted on the work of schools in a post-panoptic era. *British Journal of Educational Studies*, *66*(2), 145–163.

Rayson, P. (2019). Corpus analysis of key words. In Chapelle, C. A. (Ed.), *The concise encyclopaedia of applied linguistics* (pp. 320–326). Wiley.

Römer, U., & O'Donnell, B. M. (2011). From student hard drive to web corpus (part 1): The design, compilation and genre classification of the Michigan corpus of upper-level student papers (MICUSP). *Corpora*, *6*(2), 159–177. https://doi.org/10.3366/cor.2011.0011

Simpson, R. C., Briggs, S. L. Ovens, J., & Swales, J. M. (1999). *The Michigan corpus of academic spoken English*. The Regents of the University of Michigan.

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

Sinclair, J. M. (2003). *Reading concordances: An introduction*. Longman.

Stangor, C. (1998). *Research methods for the behavioral sciences*. Houghton Mifflin Company.

Thompson, P., & Hunston, S. (2019). *Interdisciplinary research discourse: Corpus investigations into environment journals*. Routledge.

Thompson, P., & Nesi, H. (2001). The British academic spoken English (BASE) corpus project. *Language Teaching Research*, *5*(3), 263–264. https://doi.org/10.1177/136216880100500305

# 4 Written school language registers at the transition

*Duygu Candarli*

## Introduction

This chapter explores and describes lexico-grammatical variation in written school resources around the transition, using a corpus technique known as multi-dimensional (MD) analysis, which we described in Chapter 3. Our study drills down to the levels of subjects and sub-registers within the written corpora. There have been analyses of individual linguistic features of the language of school, such as Fang et al.'s study of nouns (2006) and García et al.'s study of rhetorical devices (2018). While these offer valuable insights into their detailed areas of focus, we sought a broader overview. MD analysis offers an advantage over an analysis of individual features (see Biber et al., 2016), because it allows the researcher to study co-occurrences, or clusters of linguistic features, and to relate these to functions within Biber's MD analysis framework (1988). This procedure enabled us to detect functional variation in school language, which can be traced back to register differences and situational characteristics of the written resources at primary and secondary schools.

MD analysis has been extensively used to investigate lexico-grammatical and functional variation in written and spoken university registers (e.g., Biber et al., 2002; Biber, 2006), including research articles (e.g., Gray, 2013; Thompson et al., 2017), university students' writing at different levels (e.g., Gardner et al., 2019; Hardy & Römer, 2013) and second language writing at different proficiency levels (e.g., Friginal & Weigle, 2014). Most of these previous studies have identified 'involved versus informational language production' (Dimension 1, see Chapter 3) or the equivalent of it as the first dimension that accounts for the largest variance in register studies (see Goulart et al., 2020, for an extensive overview). Written academic registers have been associated with informational discourse that is marked by the co-occurrence of nouns, nominalisations, attributive adjectives, prepositions and diverse vocabulary that function to package information densely (e.g., Goulart et al., 2020). Information density was found to increase in university students' writing as their year of study increased at UK universities, suggesting a greater complexity in written production (Gardner et al., 2019). Conversation registers, on the other hand, are characterised by involved

DOI: 10.4324/9781003081890-4

discourse that is signalled by the co-occurrence of personal pronouns, private verbs (*think*) and present tense constructions that convey an interactive style.

'Narrative versus non-narrative discourse' (Dimension 2) has also been consistently identified as one of the dimensions in previous register studies. The co-occurrence of linguistic features, including past tense constructions, third person pronouns and perfect aspect verbs is featured in fiction registers. According to Jeong (2017), narrative registers are one of the first registers that students encounter, suggesting that students may have less difficulty in understanding narrative registers than non-narrative registers at the school. Non-narrative features, such as present tense constructions, are found much more frequently in expository registers, including academic prose and official documents than narrative registers (e.g., Goulart et al., 2020).

More recently, Le Foll (2021) focused on English as a foreign language textbooks used in European countries at different proficiency levels and compared sub-registers of textbooks, including conversation, fiction, informative and instructional registers with reference to the target language corpora using Biber's (1988) Dimension 1 'involved versus informational production'. She concluded that additive MD analysis was a fruitful endeavour to examine to what extent the language encountered by language learners was naturally occurring for the purposes of the evaluation and revision of textbooks. Despite the use of MD analysis to examine the linguistic variation in university and learner language registers, very little attention has been paid to pre-university or school language registers within the MD analysis framework.

Until now, the only study we are aware of that has employed an MD analysis of elementary school language in different disciplines was Reppen's study (2001) in the US context. Reppen (2001) used a corpus of 62,000 words consisting of fifth-grade children's literature, social science and science textbooks, texts written and spoken by children and children's monologues, and employed a new MD analysis to compare elementary student and adult language. She identified five dimensions of variation in elementary school language:

(1) 'Edited informational versus online-informational discourse' (p. 192) that characterised carefully edited texts that convey edited information (social studies and science textbooks) and discourse that has both online and informational features (student monologues).
(2) 'Lexically elaborate narrative versus non-narrative' discourse (p. 192) that distinguished lexically diverse narrative discourse (e.g., children's literature) and non-narrative discourse (monologues and children's writing).
(3) 'Involved personal opinion versus non-personal uninvolved discourse' (p. 192) that was concerned with interactional discourse versus abstract discourse (social studies and science textbooks).
(4) 'Projected scenario' (p. 192) that referred to the hypothetical or imagined style in children's writing tasks.

(5) 'Other-directed idea justification/exploration' (p. 192) that described children's writing tasks, which involved reasoning, addressed to other people.

The first three dimensions identified in elementary school language in Reppen's study (2001) are similar to Biber's (1988) original dimensions in general registers of the English language. Reppen (2001) noted that multidimensional analysis is 'very productive for addressing developmental issues, in addition to creating a more complete picture of the school language' (p. 199). MD analysis, however, remains yet to be exploited in the description of the school language at the transition stage. The other novelty of our analysis is that we examined a range of registers, including worksheets and teacher presentations prepared through Microsoft PowerPoint or other software, which have not received attention in previous MD analysis studies. We aim to address the following research question:

To what extent, if any, is there functional variation among the sub-registers of English, mathematics and science subjects from primary (KS2) to secondary school (KS3)?

## The corpus

Our study focuses on the written registers of the three main subjects, namely English, mathematics and science at KS2 and KS3. For MD analysis, we only used texts comprising at least 100 tokens because quantitative frequencies derived from tagging grammatical features could only provide reliable results for texts of a minimum of 100 tokens (Biber et al., 2016). Therefore, the written corpus, shown in Table 4.1, is slightly smaller than the whole corpus presented in Chapter 3, as short texts have been taken out. The written corpus that we used for MD analysis consisted of 2607 texts of 1,468,657 tokens.

Table 4.2 shows the distribution of the sub-registers in the English, mathematics and science written sub-corpora. The sub-registers are the text types

*Table 4.1* The written corpus of academic school language registers.

| | *Number of texts* | | *Number of tokens* | | *Mean text length (tokens)* | | *Standard deviation text length (tokens)* | |
|---|---|---|---|---|---|---|---|---|
| | *KS2* | *KS3* | *KS2* | *KS3* | *KS2* | *KS3* | *KS2* | *KS3* |
| English | 488 | 298 | 295,714 | 258,869 | 606 | 869 | 1514 | 3752 |
| Subtotal (English) | 786 | | 554,583 | | | | | |
| Mathematics | 415 | 694 | 160,012 | 245,698 | 386 | 354 | 1086 | 373 |
| Subtotal (Mathematics) | 1109 | | 405,710 | | | | | |
| Science | 145 | 567 | 158,539 | 349,825 | 1093 | 617 | 3364 | 3316 |
| Subtotal (Science) | 712 | | 508,364 | | | | | |
| Grand total | 2607 | | 1,468,657 | | | | | |

*Table 4.2* Distribution of sub-registers in the written corpus.

| *Subject* | *Sub-register* | *Number of texts* | | *Number of tokens* | | *Mean text length (tokens)* | | *Standard deviation text length (tokens)* | |
|---|---|---|---|---|---|---|---|---|---|
| | | *KS2* | *KS3* | *KS2* | *KS3* | *KS2* | *KS3* | *KS2* | *KS3* |
| English | | | | | | | | | |
| | Assessment | 77 | 37 | 106,191 | 78,507 | 1379 | 2122 | 3298 | 10,421 |
| | Glossary | - | 3 | - | 1117 | - | 372 | - | 27 |
| | Presentation | 141 | 161 | 77,380 | 132,341 | 549 | 822 | 484 | 1027 |
| | Reading extract | 47 | 24 | 18,399 | 13,961 | 391 | 582 | 337 | 447 |
| | Textbook | 1 | - | 13,918 | - | NA | - | NA | - |
| | Worksheet | 222 | 73 | 79,826 | 32,943 | 360 | 451 | 233 | 563 |
| Mathematics | | | | | | | | | |
| | Assessment | 43 | 88 | 28,082 | 24,596 | 653 | 280 | 828 | 153 |
| | Presentation | 74 | 249 | 30,757 | 106,967 | 416 | 430 | 302 | 456 |
| | Reading extract | - | 3 | - | 631 | - | 210 | - | 100 |
| | Textbook | 2 | - | 29,699 | - | 14,850 | - | 2187 | - |
| | Worksheet | 296 | 354 | 71,474 | 113,504 | 241 | 321 | 260 | 339 |
| Science | | | | | | | | | |
| | Assessment | 12 | 59 | 7054 | 22,210 | 588 | 376 | 677 | 278 |
| | Presentation | 65 | 283 | 35,571 | 143,442 | 547 | 507 | 465 | 326 |
| | Reading extract | 24 | 29 | 16,835 | 15,724 | 701 | 542 | 679 | 436 |
| | Textbook | 4 | 2 | 74,192 | 112,072 | 18,548 | 56,036 | 9951 | 1990 |
| | Worksheet | 40 | 194 | 24,887 | 56,377 | 622 | 291 | 1122 | 241 |

that we collected on the advice of teachers at our partner schools and using timetables as an indication of the amount of class time spent on each subject, as discussed in Chapter 3.

In considering the number of tokens for each sub-register and subject, there are interesting patterns. In KS2 English, the most dominant sub-register is assessment, probably because Year 6 students were preparing for the Standard Assessment Tests (SATs) that they took in 2019. In KS2 science, students are mostly exposed to textbooks, and in KS2 mathematics, worksheets.

At KS3, presentations became the primary sub-register that the students encountered in both English and science classes. For mathematics, the most dominant sub-register is worksheets, as for KS2. Students tend to be grouped by prior attainment relatively early for mathematics, including within classes in KS2, which might account for the extensive use of worksheets. We note that textbooks were rarely or never used in English or mathematics classes, showing the importance of representing a wide range of sub-registers to approach a representative corpus, despite the challenges in collecting such data.

## Analytical steps

We used an additive MD analysis to investigate the lexico-grammatical variation in school language registers for the following reasons:

(1) Our written corpus only included one main written register: written resources of the academic language of school that pupils encountered at KS2 and KS3.
(2) A new multi-dimensional analysis that requires exploratory factor analysis would capture variation in internally well-stratified corpus that includes a number of different spoken and written registers (Gray, 2021; Nini, 2019). An additive MD analysis, which does not require the use of exploratory factor analysis, enabled us 'to apply existing dimensions of variation (Biber, 1988) to "new" registers' (Berber Sardinha & Veirano Pinto, 2019, p. 4), that is, in this study, school language registers. In this way, the additive MD analysis offered us a robust framework to explore both linguistic and functional differences and similarities of the school language sub-registers across the subjects and Key Stages within one main register.

The MD analysis Tagger (MAT) v 1.3.2 (Nini, 2019), which replicates Biber's tagger (1988), was utilised to tag 67 linguistic features in our corpus. The type-token ratio was calculated for the first 100 tokens for each text in our corpus, since the minimum token size for each text was 100, as explained above. The MAT utilises the Stanford part-of-speech (POS) tagger (Toutanova et al., 2003). It expands part-of-speech (POS) tagging by identifying features in Biber's (1988) study and calculates Biber's (1988) original dimension scores for each text. The Stanford tagger provides a '97.24%

accuracy' on the Penn Treebank Wall Street Journal data set (Toutanova et al., 2003, p. 1). Nini (2019) also tested MAT and found it reliable to replicate Biber's tagger and calculate equivalent dimension scores. These accuracy and reliability tests suggest that the use of the MAT provided a foundation for the rigorous multi-dimensional analysis of the written school language registers that do not include children's writing or second language data, which would have posed challenges for POS tagging.

### *Statistical analysis*

We used statistical analysis in order to find whether there were any statistically significant differences in the lexico-grammatical variation in the written corpus of school language registers across the subjects, sub-registers and, importantly, the Key Stages. We built linear mixed-effects models to explore this variation, manifested by dimension scores, across the Key Stages for two reasons. First, linear mixed-effects models were robust enough to handle unbalanced data sets (Linck & Cunnings, 2015) (unequal number of texts in different subjects and sub-registers in this study). Second, texts in our corpus belonged to only 13 schools, which means that the same school contributed to multiple texts, creating non-independence in the data set. Mixed-effects models allow researchers to make robust estimates and inferences by incorporating non-independent sets of data through random effects (Winter, 2019). The package *lme4* version 1.1-27.1 (Bates et al., 2015) in R, an open-source programming language (R Core Team, 2020), was used to build mixed-effects models to estimate dimension scores across the subjects, sub-registers and Key Stages. The dependent variable was dimension scores, and the predictor (fixed effects) variables were Key Stage (two levels – KS2 and KS3), subject (three levels – English, mathematics, science) and sub-register (three levels – assessment, presentation, worksheet). The school was added as a random effect (intercepts only). We started with this most complex model for each dimension score separately and reduced the model complexity, by comparing the model fits, utilising Akaike's information criterion (AIC) values. The smaller AIC value indicates a better model fit for the data (Maydeu-Olivares & Garcia-Forero, 2010). We also report effect sizes of the linear mixed-effects models, using the R package *MuMIn* (Bartoń, 2022) which provides marginal $R^2$ values that indicate variance explained by the fixed effects only and conditional $R^2$ values that indicate variance explained by both the fixed and random effects. We only used three levels of the sub-registers since the other sub-registers, including glossaries, reading extracts and textbooks, had data sparsity in terms of the number of texts, which would have caused unstable estimates. When we report the descriptive statistics of the dimension scores of three subjects overall across the Key Stages in the figures in the next section, we included all the sub-registers. The statistical analysis and figures used in this chapter follow those of Candarli's (2022) study.

MD analysis utilises normalised frequencies of lexico-grammatical features for the calculation of dimension scores. As seen in Table 4.1, there was a large standard deviation in terms of text lengths in our corpus. Therefore, in order to ensure that text length (tokens) had no effect on dimension scores, we correlated each dimension score of each text with their text lengths in our corpus, separately for five dimensions, using *Pearson's r* (see Clarke & Grieve, 2019). The correlations varied from –0.016 to 0.034. This suggests that the normalisation of occurrences of lexico-grammatical features was robust enough to mitigate against any possible text length effects on dimension scores in this study, despite the large standard deviation in text lengths in our corpus.

## Multi-dimensional analysis of school language registers

This section presents the findings of five dimensions of lexico-grammatical variation in the written corpus of school language registers. We only focus on the first five dimensions of Biber's (1988) original dimensions for the register variation in our analysis of the written school language registers because Dimension 6 'on-line informational elaboration' characterises spoken registers and computer-mediated conversations, such as online chats (Berber Sardinha et al., 2019).

### *Dimension 1: Involved versus informational discourse*

Dimension 1: Involved versus informational discourse distinguishes between oral (involved) and written (informational) discourse in the literature (e.g., Biber & Conrad, 2019). High Dimension 1 scores are marked by the combination of highly frequent oral features, such as private verbs (e.g., *think*, *feel*), present tense verbs, first and second person pronouns, which typically co-occur in conversations (Biber, 1988; Biber & Conrad, 2019). As Dimension 1 scores decrease, the frequent co-occurrence of nouns, prepositions, longer words, attributive adjectives and diverse vocabulary (high type/token ratio) increases. The more frequent use of such linguistic features is a manifestation of informational density. Fang (2006, p. 502) noted that informational density 'can result in cognitive overload and engender comprehension failure' for students. Written academic discourse is characterised by informational density, suggesting dense information-packaging in texts.

Figure 4.1 shows Dimension 1 scores across the Key Stages and subjects in the written corpus of school language registers. All the school language registers of this study were characterised as informational, as can be inferred from the negative mean Dimension 1 scores, illustrated in Figure 4.1. However, there was variation in Dimension 1 scores across the subjects and Key Stages. The biggest change in Dimension 1 scores occurred in science from KS2 to KS3, showing the greatest increase in informational density at KS3. Mathematics, on the other hand, underwent the smallest change in terms of informational discourse from KS2 to KS3. A small increase in

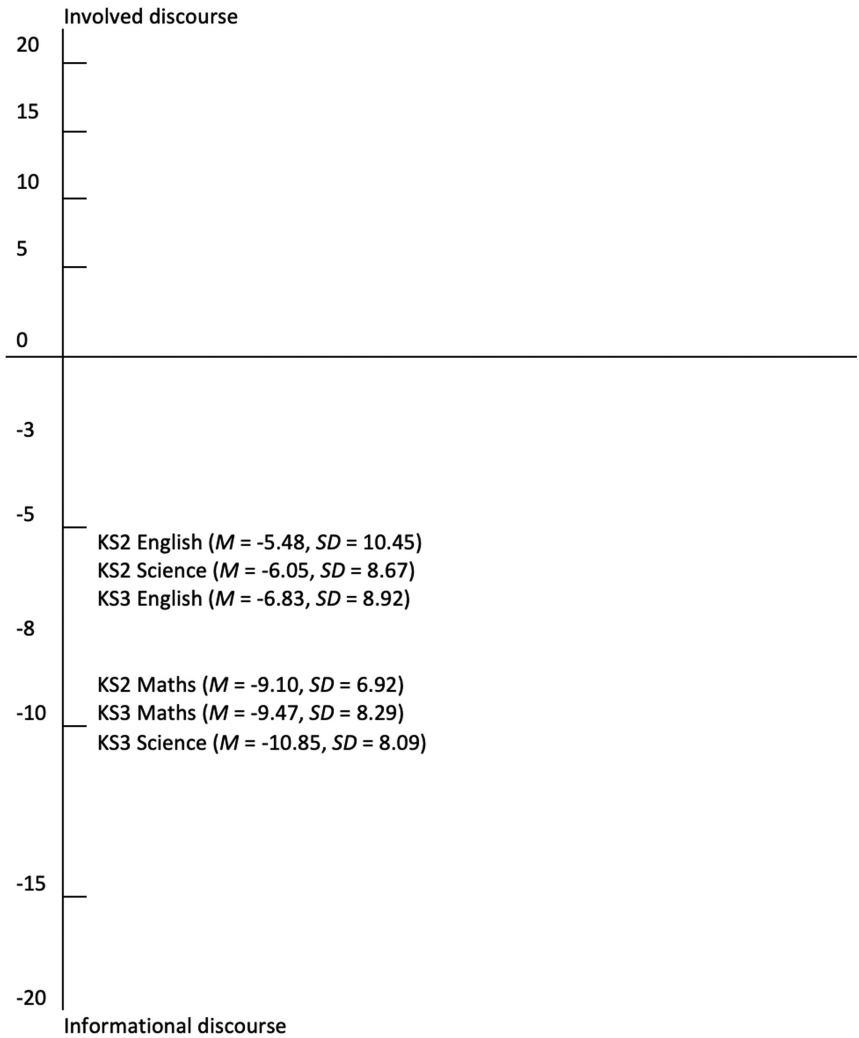



*Figure 4.1* Mean (*M*) and standard deviation (*SD*) of Dimension 1 scores across subjects and Key Stages.

informational density was observed in the written resources of English at KS3 in comparison with KS2, as Figure 4.1 illustrates. The large standard deviations in Dimension 1 scores suggest very large variation in informational density of the written resources in all three subjects.

Example 1, which shows a text with one of the highest Dimension 1 scores, illustrates involved discourse, marked by the co-occurrence of first person pronouns (*I*), present tense verbs (e.g., *recognize*), contractions

(*don't*) and a possibility modal (*can*) in a sub-register of assessment in the subject of English at KS2. The text in Example 1 included self-assessment criteria for vocabulary for Year 6 students, and it resembled oral registers in that the linguistic features written in bold functioned to involve students in a cognitive task. On the other hand, Example 2, which illustrates a text with one of the lowest Dimension 1 scores, represented great informational density. This high level of informational density was characterised by the co-occurrence of linguistic features (underlined in Examples 2 and 3), including longer words, nouns (e.g., *cartilage*), prepositions (e.g., *of*), diverse vocabulary (a greater number of different word types) and attributive adjectives (e.g., *inelastic*). In order for Year 7 students at the beginning of the secondary school to do the exercises featured in Examples 2 and 3, a number of noun phrases, such as 'a tough band of inelastic tissues' would require decoding and comprehension. Comprehension of such complex noun phrases may potentially create challenges for Year 7 students.

(1) **I don't know** this word.
**I recognize** this word.
**I know** this word and **can** use **it** in a sentence.

(English, Key Stage 2 – Year 6, assessment, Dimension 1 score: 36.59)

(2) Match the key term to the definition:
Ligament
Tendon
Cartilage
Synovial Fluid
A tough band of inelastic tissues attaching muscle to bone.
A smooth protective surface covers the bone ends, providing easy movement.
This tissue lines the joint capsule and secretes synovial fluid.
Bands of tough inelastic tissue holding bones to each other.

(Science, Key Stage 3 – Year 7, worksheet, Dimension 1 score: –33.83)

(3) Trapezium Kite
One pair of parallel sides
Isosceles quadrilaterals of this kind have one line of symmetry
Diagonals bisect each other at right angles

(Mathematics, Key Stage 3 – Year 7, worksheet, Dimension 1 score: –30.53)

The results of the mixed-effects model for Dimension 1 scores demonstrated statistically significant interactions between the predictors 'subject' and 'key

stage', indicating that the Dimension 1 scores of the school language registers at KS3 were shaped by the subjects differently, as shown in Table 4.1. A significant interaction between the science subject and KS3 indicated that Dimension 1 scores of the science registers at KS3 decreased to a larger extent than the English registers at KS3, significantly increasing informational density of the science school registers at KS3 ($t = -4.16$, $p < 0.001$). The pairwise comparisons, on the other hand, showed that there was no significant difference between the KS2 English and KS3 English registers ($t = 1.1$, $p = 0.87$) or the KS2 mathematics and KS3 mathematics ($t = 0.9$, $p = 0.93$) registers in terms of Dimension 1 scores. There was also a main effect of the sub-registers in that the assessment sub-registers had lower Dimension 1 scores than the presentation or worksheet sub-registers, irrespective of the subjects and key stages, as seen in Table 4.3. The pairwise comparisons showed that there was no difference between the worksheet and presentation sub-registers ($t = 1.2$, $p = 0.44$) with regard to Dimension 1 scores. This suggests that students encountered the most informationally dense texts within the assessment sub-registers. Informationally dense assessment tasks might further disadvantage students of low reading ability, potentially creating a risk for a drop in their attainment levels.

### *Dimension 2: Narrative versus non-narrative discourse*

Narrative discourse is characterised by people-oriented depiction of past events, manifested through the co-occurrence of past tense verbs, third person pronouns, perfect aspect verbs (e.g., *had been*), public verbs (e.g., *said*, *mentioned*), synthetic negation (e.g., *no response*) and present participial clauses (e.g., *Running in the park*, he lost his keys). Positive Dimension 2 scores indicate narrative discourse (e.g., fiction) while negative Dimension

*Table 4.3* Mixed-effects model results: Dimension 1 scores.

| *Predictors* | *Estimates* | ***SE*** | ***CI*** | ***t*** | ***p*** |
|---|---|---|---|---|---|
| (Intercept)* | −7.90 | 0.80 | −9.47 – −6.33 | −9.88 | **<0.001** |
| Mathematics | −3.70 | 0.60 | −4.88 – −2.52 | −6.14 | **<0.001** |
| Science | −0.13 | 0.88 | −1.86 – 1.60 | −0.15 | 0.883 |
| Key Stage 3 | −1.15 | 1.04 | −3.18 – 0.88 | −1.11 | 0.265 |
| Presentation | 3.26 | 0.55 | 2.19 – 4.33 | 5.96 | **<0.001** |
| Worksheet | 2.80 | 0.54 | 1.74 – 3.85 | 5.21 | **<0.001** |
| Mathematics * Key Stage 3 | 0.27 | 0.90 | −1.49 – 2.03 | 0.30 | 0.763 |
| Science * Key Stage 3 | −4.74 | 1.14 | −6.98 – −2.51 | −4.16 | **<0.001** |
| **Random effect** | Variance | SD | | | |
| School | 1.16 | 1.08 | | | |
| Marginal $R^2$/ Conditional $R^2$ | 0.08/0.09 | | | | |

* Reference level is English Key Stage 2 assessment.

2 scores represent non-narrative expository discourse realised through the co-occurrence of present tense verbs and attributive adjectives. Most school students are likely to be exposed to narrative discourse through storybooks outside of school. In fact, empirical research shows that students acquire narrative registers earlier than expository registers and that expository registers require higher cognitive load to understand than narrative registers (e.g., Berman & Nir-Sagiv, 2007; Joeng, 2017), probably because non-narrative expository registers are topic-oriented, relying on mostly abstract ideas. This may make non-narrative texts less accessible to students than narrative texts. Narrative registers rely on people-oriented and concrete past events that are familiar to schoolchildren; hence, it may be easier for students to comprehend narrative texts than non-narrative ones.

The written school language registers were non-narrative in all three subjects at both KS2 and KS3, as illustrated in Figure 4.2. Both English and mathematics registers showed a slight increase in non-narrativity from KS2 to KS3, though the registers of the English subject were overall less non-narrative than the mathematics registers. It should be noted that the literature that students read as part of their English curriculum did not form part of

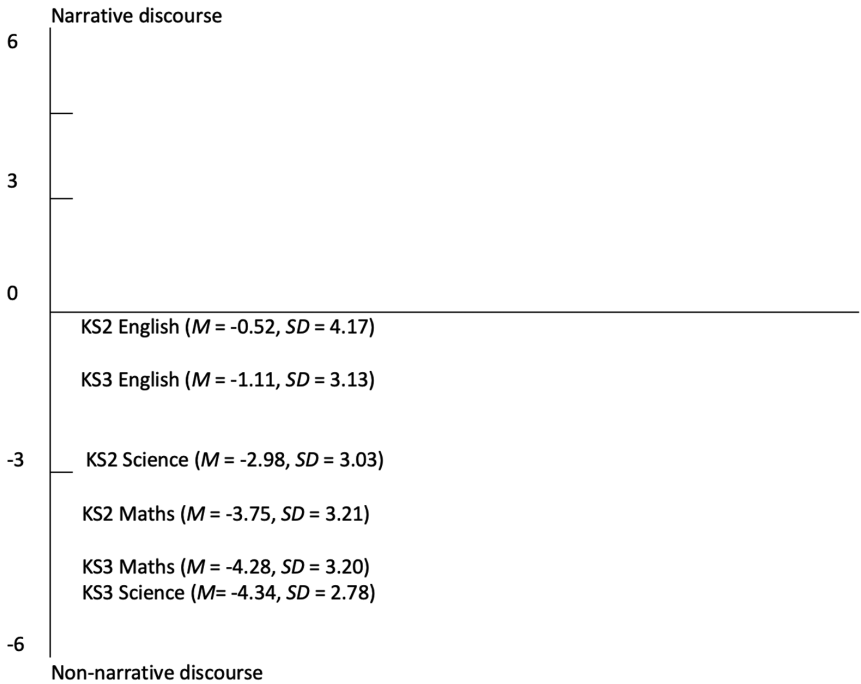



*Figure 4.2* Mean (M) and standard deviation (SD) of Dimension 2 scores across subjects and key stages.

this corpus. We have compiled a separate literature corpus for further study. Some literature extracts did feature in worksheets, however. The biggest increase in non-narrative discourse from KS2 to KS3 was seen in science, suggesting another layer of complexity for the science registers, in addition to informational density (the low Dimension 1 scores discussed above).

Example 4 shows an extract from a Year 6 worksheet that had a highly positive Dimension 2 score, indicating narrative discourse. Past tense verbs (*was*) and a third person pronoun (*her*) indicated an action-oriented past event that occurred around a specific person named Alundra. In this worksheet, Year 6 students were asked to respond to comprehension questions about a text of person-oriented chronological events. Such event-oriented discourse is likely to be more imageable than non-narrative discourse. On the other hand, Example 5, taken from an assessment task in science in Year 7, had the lowest Dimension 2 score in our written corpus, representing highly non-narrative discourse. The lack of narrative features and co-occurrence of present tense verbs (e.g., *contains*) and attributive adjectives (e.g., *various*) marked non-narrative and descriptive discourse, as shown in Examples 5 and 6. Such non-narrative discourse focused on description linked through logical relations within its discourse segments and relied on concepts (e.g., *organs* in Example 5) rather than proper nouns.

(4) To retrieve and infer
Why **was** Alundra confused about what **was going** on around **her**?

(English, Key Stage 2 – Year 6, worksheet, Dimension 2 score: 22.73)

(5) Organs and tissues test
Circle the correct answer
There are different colours in a kidney because it contains different
a) molecules b) tissues c) organs
Organs need more than one type of tissue because
a) variety is good b) each tissue does a different job c) each tissue is joined to the next one

(Science, Key Stage 3 – Year 7, assessment, Dimension 2 score: -7.38)

(6) How does the writer create tension in this extract from 'The Red Room'? In 'The Red Room' the author
wants to create a feeling of…… …… This is achieved through various techniques.

(English, Key Stage 3 – Year 8, worksheet, Dimension 2 score: -7.07)

The mixed-effects model results indicated a statistically significant three-way interaction between the predictors of science, Key Stage 3 and

*Table 4.4* Mixed-effects model results: Dimension 2 scores.

| *Predictors* | *Estimates* | *SE* | *CI* | *t* | *p* |
|---|---|---|---|---|---|
| (Intercept)* | –1.17 | 0.48 | –2.11 – –0.23 | –2.45 | **0.014** |
| Mathematics | –2.73 | 0.61 | –3.94 – –1.53 | –4.45 | **<0.001** |
| Science | –0.16 | 1.00 | –2.13 – 1.80 | –0.16 | 0.870 |
| Key Stage 3 | 0.54 | 0.78 | –1.00 – 2.07 | 0.68 | 0.495 |
| Presentation | –0.40 | 0.46 | –1.29 – 0.50 | –0.87 | 0.383 |
| Worksheet | 1.32 | 0.43 | 0.47 – 2.17 | 3.05 | **0.002** |
| Mathematics * Key Stage 3 | 0.51 | 0.89 | –1.24 – 2.26 | 0.57 | 0.567 |
| Science * Key Stage 3 | –3.54 | 1.22 | –5.93 – –1.15 | –2.90 | **0.004** |
| Mathematics * Presentation | 0.11 | 0.76 | –1.39 – 1.61 | 0.14 | 0.890 |
| Science * Presentation | –1.27 | 1.11 | –3.45 – 0.91 | –1.14 | 0.252 |
| Mathematics * Worksheet | –1.14 | 0.68 | –2.47 – 0.20 | –1.67 | 0.094 |
| Science * Worksheet | –3.84 | 1.14 | –6.08 – –1.60 | –3.36 | **0.001** |
| Key stage 3 * Presentation | 0.11 | 0.74 | –1.34 – 1.57 | 0.15 | 0.878 |
| Key stage 3 * Worksheet | –1.90 | 0.78 | –3.43 – –0.37 | –2.44 | **0.015** |
| Mathematics * Key Stage 3 * Presentation | –1.21 | 1.05 | –3.26 – 0.84 | –1.16 | 0.247 |
| Science * Key Stage 3 * Presentation | 1.19 | 1.34 | –1.44 – 3.81 | 0.89 | 0.376 |
| Mathematics * Key Stage 3 * Worksheet | –0.15 | 1.02 | –2.14 – 1.84 | –0.15 | 0.883 |
| Science * Key Stage 3 * Worksheet | 4.28 | 1.40 | 1.53 – 7.02 | 3.06 | **0.002** |
| **Random effect** | Variance | SD | | | |
| School | 0.39 | 0.62 | | | |
| Marginal $R^2$/Conditional $R^2$ | 0.21/0.24 | | | | |

* Reference level is English Key Stage 2 assessment.

sub-register (worksheet), as Table 4.4 shows. This three-way interaction showed that Dimension 2 scores of worksheets at KS3 for the science subject showed different patterns than the worksheets at KS3 for the English subject. Indeed, the post-hoc comparisons indicated that the only sub-register that showed significant differences between KS2 and KS3 for the subject of English was worksheets ($t = 2.21$, $p = 0.04$), increasing in non-narrativity at KS3. On the other hand, the worksheets of the science subject showed no significant differences between KS2 and KS3 in terms of Dimension 2 scores ($t = 0.88$, $p = 0.38$), whereas the sub-registers of assessment ($t = 2.7$, $p = 0.01$) and presentations ($t = 2.74$, $p = 0.01$) of the science subject became increasingly more non-narrative at KS3 in comparison to KS2. For the mathematics subject, none of the sub-registers showed any significant differences between KS2 and KS3 with regard to Dimension 2 scores.

### *Dimension 3: Explicit versus situation-dependent discourse*

The positive pole of Dimension 3, explicit reference, is concerned with elaborated text-internal reference that is manifested through the

co-occurrence of wh-relative clauses in subject and object positions, pied piping constructions (*the way in which this changed*), phrasal coordination (*and*) and nominalisations (*information*) (Nini, 2019). Explicit reference is one of the key characteristics of written registers, especially academic prose. Explicit reference features are associated with informational density and the presentation of information in written discourse (Biber, 1988). This suggests that written texts involving explicit discourse may create comprehension challenges for students, increasing their cognitive load (e.g., Fang, 2006). Situation-dependent discourse, on the other hand, relies on text-external references, including time and place adverbials and adverbs, which are typical of spoken registers (Biber & Conrad, 2019). In spoken registers, including sports broadcast, participants may share the same time or place and make reference to the temporal and physical aspects of the discourse during communication. Such situation-dependent discourse may be more accessible to students since its underlying linguistic features are concerned with the immediate environment or temporal space of the discourse.

The written school language registers were all characterised by explicit reference across the subjects and Key Stages, as illustrated in Figure 4.3. The written registers of both English and mathematics subjects became more explicit and elaborated at KS3 in comparison with KS2. It is interesting that the opposite pattern was observed for the written registers of the science subject in that KS2 science written registers included explicit and elaborated reference to a greater extent than KS3 science written registers. The registers of the science subject were overall marked by greater explicit and elaborated reference than the registers of English or mathematics subjects. This is not surprising since nominalisations and relative clauses characterise school science texts (Fang, 2006, 2012).

The extracts with one of the highest Dimension 3 scores exemplify explicit reference, involving wh-relative clauses in subject positions (*which makes*) and nominalisations (*fractions)*, as shown in Examples 7 and 8. In Example 7, the relative clause functioned as further elaboration for the concept of *friction* (nominalisation). The co-occurrence of nominalisations and wh-relative clauses in the subject position created highly explicit and elaborated discourse. Example 9, on the other hand, illustrates an extract taken from a text with one of the lowest Dimension 3 scores and depicts situation-dependent discourse. The adverb (*usually*) and time adverbial (*again*) referred to the text-external, temporal aspect of the discourse, which probably makes the text accessible to students.

(7) True or False
**Friction** is a contact force.
**Friction** is a force **which** makes objects move easier.

(Science, Key Stage 3 – Year 7, presentation, Dimension 3 score: 28)

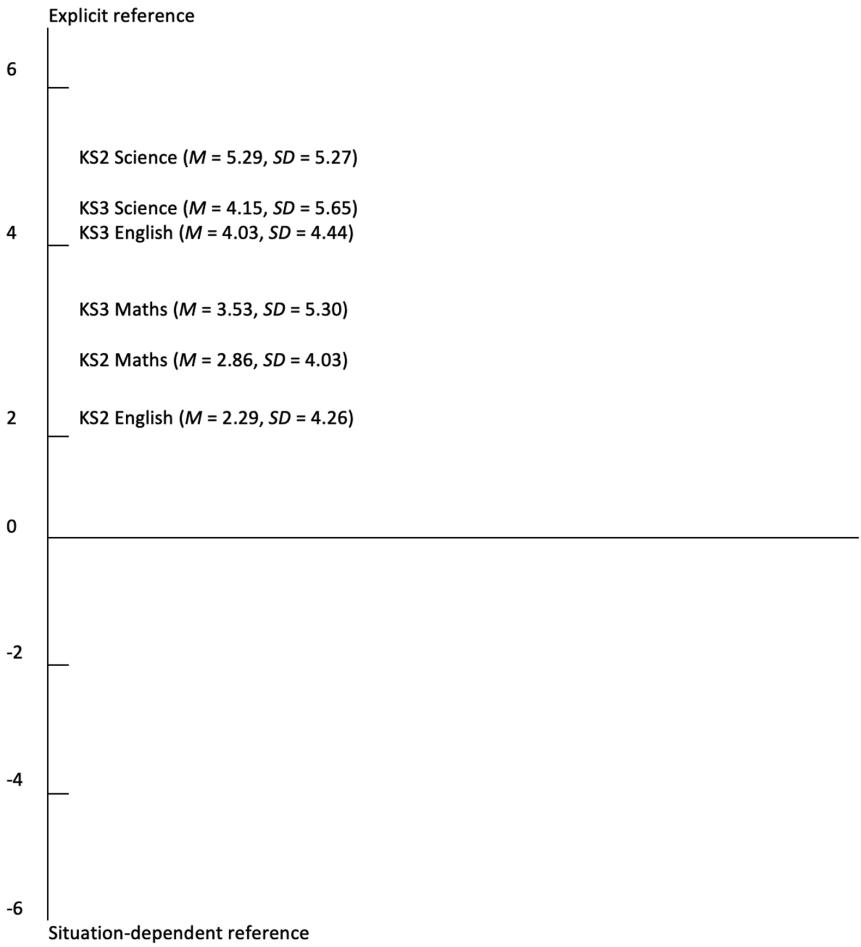



*Figure 4.3* Mean (M) and standard deviation (SD) of Dimension 3 scores across subjects and Key Stages.

(8) We can only add **and** subtract things if they have the same name. For **fractions**, this means if they have the same denominator.

(Mathematics, Key Stage 3 – Year 7, presentation, Dimension 3 score: 30)

(9) I <u>usually</u> walk to school in the morning.
OK!
Try <u>again</u>!

(English, Key Stage 2 – Year 5, assessment, Dimension 3 score: –8.5)

*Table 4.5* Mixed-effects model results: Dimension 3 scores.

| *Predictors* | *Estimates* | *SE* | *CI* | *t* | *p* |
|---|---|---|---|---|---|
| (Intercept)* | 1.25 | 0.69 | –0.10 – 2.60 | 1.82 | 0.069 |
| Mathematics | 0.41 | 0.93 | –1.42 – 2.24 | 0.44 | 0.662 |
| Science | 3.64 | 1.52 | 0.65 – 6.62 | 2.39 | **0.017** |
| Key Stage 3 | 5.52 | 1.14 | 3.29 – 7.75 | 4.85 | **<0.001** |
| Presentation | 0.98 | 0.69 | –0.38 – 2.34 | 1.42 | 0.156 |
| Worksheet | 0.53 | 0.66 | –0.76 – 1.82 | 0.80 | 0.421 |
| Mathematics * Key Stage 3 | –0.95 | 1.36 | –3.61 – 1.72 | –0.70 | 0.486 |
| Science * Key Stage 3 | –5.74 | 1.85 | –9.38 – –2.10 | –3.10 | **0.002** |
| Mathematics * Presentation | –0.15 | 1.16 | –2.43 – 2.13 | –0.13 | 0.896 |
| Science * Presentation | –1.24 | 1.69 | –4.55 – 2.07 | –0.74 | 0.462 |
| Mathematics * Worksheet | 0.55 | 1.03 | –1.48 – 2.57 | 0.53 | 0.596 |
| Science * Worksheet | –0.06 | 1.74 | –3.46 – 3.35 | –0.03 | 0.974 |
| Key Stage 3 * Presentation | –4.71 | 1.13 | –6.92 – –2.50 | –4.17 | **<0.001** |
| Key Stage 3 * Worksheet | –4.42 | 1.19 | –6.74 – –2.09 | –3.72 | **<0.001** |
| Mathematics * Key Stage 3 * Presentation | 1.12 | 1.59 | –2.00 – 4.25 | 0.71 | 0.480 |
| Science * Key Stage 3 * Presentation | 5.62 | 2.04 | 1.62 – 9.62 | 2.76 | **0.006** |
| Mathematics * Key Stage 3 * Worksheet | 0.83 | 1.54 | –2.20 – 3.86 | 0.54 | 0.592 |
| Science * Key Stage 3 * worksheet | 2.89 | 2.13 | –1.28 – 7.06 | 1.36 | 0.175 |
| **Random Effect** | Variance | SD | | | |
| School | 0.62 | 0.79 | | | |
| Marginal $R^2$/Conditional $R^2$ | 0.06/0.09 | | | | |

* Reference level is English Key Stage 2 assessment.

There was a statistically significant three-way interaction between the predictors of science subject, Key Stage 3 and presentation, as can be seen in Table 4.5. The science presentations exhibited different patterns in terms of Dimension 3 scores at KS3 than those of the English subject at KS3 since descriptive statistics showed that the presentation sub-registers increasingly showed explicit and elaborated reference at KS3 for both English and mathematics while this explicit and elaborated reference slightly decreased for the presentation sub-registers of the science subject. However, when we look at the inter-subject differences at KS3, the science presentation sub-registers showed elaborated and explicit discourse to a greater extent than those of English ($t = -4.2$, $p < 0.001$) or mathematics ($t = -4.1$, $p < 0.001$) subjects. At KS3, there were no between-subject differences for the assessment or worksheet sub-registers. Additionally, the post-hoc comparisons indicated that the only statistically significant difference between KS2 and KS3 occurred in the sub-registers of assessment for both English ($t = -4.8$, $p < 0.001$) and mathematics ($t = -4.2$, $p = 0.1$) subjects, while the assessment sub-registers showed no significant differences between KS2 and KS3 for the science subject ($t = 0.2$, $p = 1$). This means that the assessment sub-registers involved much more elaborated discourse at KS3 for these two subjects than at KS2.

### *Dimension 4: Overt expression of persuasion*

Dimension 4 is interpreted as overt persuasion and argumentation realised through the co-occurring features of infinitives (*happy to do*), prediction modals (*will*), suasive verbs (*agree*, *propose*), conditional subordination (*if you wish*) and necessity modals (*should*) (Nini, 2019). There are no negative features loaded to this dimension; therefore, the lack of these features signals non-persuasive or factual discourse. Overt expression of persuasion is salient in spoken registers and editorials that aim to change the addressee's opinion on a topic (Biber, 1988; Biber et al., 2002). Non-persuasive discourse, on the other hand, is associated with a factual or detached style that is expected in school contexts (e.g., Schleppegrell, 2001) although non-persuasive discourse may not suggest linguistic complexity for students per se. Biber et al. (2002) found that course packs and textbooks at university were non-persuasive, having negative Dimension 4 scores.

All the school registers in the subjects of English, mathematics and science were non-persuasive across the Key Stages, as Figure 4.4 shows. The

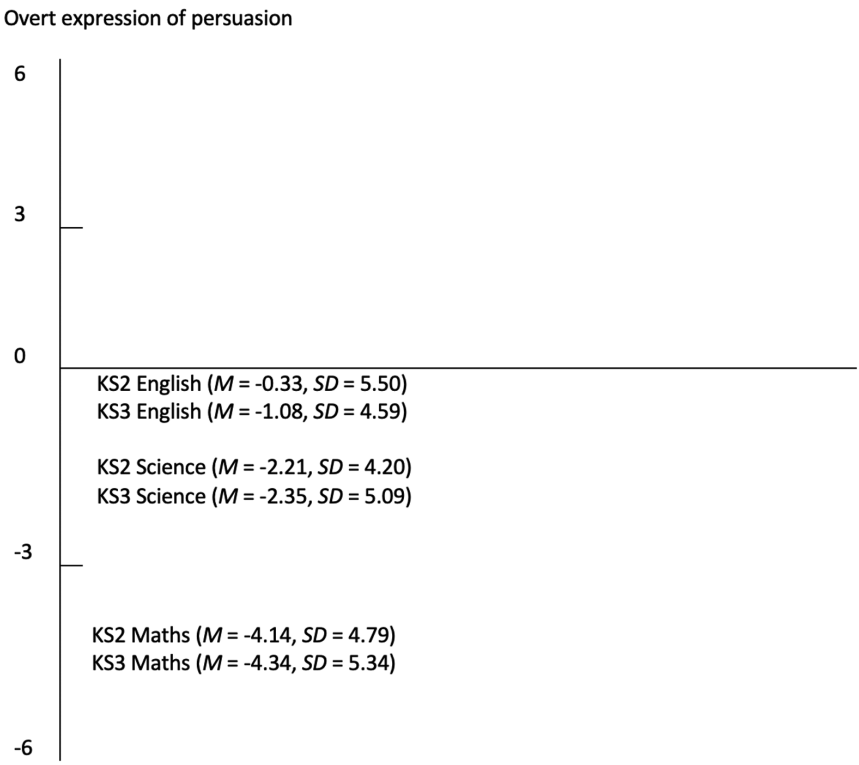



*Figure 4.4* Mean (M) and standard deviation (SD) of Dimension 4 scores across subjects and Key Stages.

large standard deviations suggest that there was a great deal of variation in terms of (non-)persuasion within these registers. Although all the subject registers became slightly more non-persuasive at KS3 in comparison with KS2, this difference at the transition stage remained negligible. It is unsurprising that the registers of mathematics subject were most non-persuasive among these three subjects since the registers of mathematics are characterised by 'technical vocabulary' with precise meanings and 'implicit logical relationships' (Schleppegrell, 2007, p. 141). The mathematics registers were followed by registers of science in terms of non-persuasive discourse. The registers of English subject depicted the least non-persuasive discourse.

Example 10 illustrates an extract from a text with one of the highest Dimension 4 scores, expressing persuasive discourse marked by the co-occurrence of a suasive verb (*demanded*), prediction modal (*would*) and infinitives (*to leave*). Our qualitative analysis of the written texts of the English registers at KS2 showed that such persuasive discourse occurred mostly due to teaching and practising the subjunctive form. As indicated in the national curriculum, one of the statutory requirements at primary school, KS2, is for pupils to be taught to '[recognise] vocabulary and structures that are appropriate for formal speech and writing, including subjunctive forms' (DfE, 2013, p. 48). In contrast, non-persuasive discourse illustrated in Examples 11 and 12, which had the lowest Dimension 4 scores, suggested a factual and detached style. As we can see in Examples 11 and 12, no persuasive linguistic features existed. Instead, factual information was presented in a detached tone through the use of the technical terms, including *mode*, *median* and *vacuole*.

(10) The subjunctive form
Example:
I **demanded** that she be quiet or else she **would** need **to** leave the room.

(English, Key Stage 2 – Year 6, worksheet, Dimension 4 score: 23.5)

(11) Types of average
The Mode: The Most Common is the most common piece of data
The Mean: The total divided by the number of pieces of data
The Median: The middle number (when in order)

(Mathematics, Key Stage 3 – Year 7, presentation, Dimension 4 score: –9.3)

(12) Which parts of a plant cell trap light energy?
Which parts of a plant cell trap light energy?
What is kept in the vacuole?

(Science, Key Stage 3 – Year 7, worksheet, Dimension 4 score: –9.27)

*Table 4.6* Mixed-effects model results: Dimension 4 scores.

| *Predictors* | *Estimates* | *SE* | *CI* | *t* | *p* |
|---|---|---|---|---|---|
| (Intercept)* | −0.31 | 0.53 | −1.35 – 0.74 | −0.57 | 0.568 |
| Mathematics | −5.16 | 0.67 | −6.47 – −3.86 | −7.76 | **<0.001** |
| Science | −3.35 | 0.79 | −4.90 – −1.79 | −4.23 | **<0.001** |
| Presentation | −0.06 | 0.56 | −1.16 – 1.03 | −0.11 | 0.909 |
| Worksheet | −1.52 | 0.56 | −2.63 – −0.42 | −2.70 | **0.007** |
| Mathematics * Presentation | 1.73 | 0.77 | 0.21 – 3.24 | 2.24 | **0.025** |
| Science * Presentation | 1.38 | 0.87 | −0.33 – 3.09 | 1.58 | 0.114 |
| Mathematics * Worksheet | 2.53 | 0.75 | 1.07 – 3.99 | 3.39 | **0.001** |
| Science * Worksheet | 3.45 | 0.89 | 1.70 – 5.20 | 3.86 | **<0.001** |
| **Random Effect** | Variance | SD | | | |
| School | 0.41 | 0.64 | | | |
| Marginal $R^2$ / Conditional $R^2$ | 0.09 / 0.10 | | | | |

* Reference level is English Key Stage 2.

The mixed-effects model indicated a statistically significant interaction between the sub-registers of presentation and mathematics subject as well as the sub-registers of the worksheet and the subjects of mathematics and science, as shown in Table 4.6. This suggests that the presentations of the mathematics registers, as well as the worksheets of both the mathematics and science registers, exhibited different patterns in terms of Dimension 4 scores than their counterparts in the subject of English. The post-hoc pairwise comparisons indicated that the sub-registers of worksheets were significantly more non-persuasive than presentations ($t$ = 3.4, $p$ = 0.2) in the subject of English. This may be traced back to the presentation and explicit teaching of the subjunctive forms that created relatively more persuasive discourse than worksheets. For the mathematics subject, presentations showed non-persuasive discourse to a lesser extent than the sub-registers of assessment ($t$ = −3.10, $p$ = 0.05). For all the other sub-registers and the sub-registers with the science subject, there were no significant differences, as the pairwise comparisons showed. The predictor 'key stage' was dropped from the model, indicating that no significant differences were observed between KS2 and KS3 in terms of Dimension 4.

### *Dimension 5: Impersonal versus non-impersonal style*

Dimension 5 'impersonal versus non-impersonal style' distinguishes between abstract, impersonal registers (academic prose) and non-impersonal, concrete registers (face-to-face conversations). This dimension is also described as 'abstract versus non-abstract information' (Biber, 1988; Nini, 2019). In this case, abstractness is not directly concerned with the content or the characteristics of vocabulary of the school registers. Instead, abstractness here lies in an impersonal style that is attributed to the co-occurrence of

lexico-grammatical features, including conjuncts (*however*), agentless passives, past participial clauses (*written* in two years, the book…), by passives, past participial WHIZ deletions (the curriculum *reviewed by* the majority of teachers) and other adverbial subordinators (*whereas*). There are no linguistic features loaded negatively to this dimension. The lack of the above-mentioned linguistic features marks a non-impersonal style in which there is active voice, or the addressees, students in this case, would be actively involved. In the English language, academic prose and official documents are characterised by an impersonal style, whereas spoken and fiction registers depict a non-impersonal style (Biber, 1988).

Figure 4.5 illustrates that the registers of English, mathematics and science across the Key Stages exhibited a non-impersonal style, as can be inferred from their mean scores. The large standard deviations indicate that substantial variation in Dimension 5 scores was observed, especially

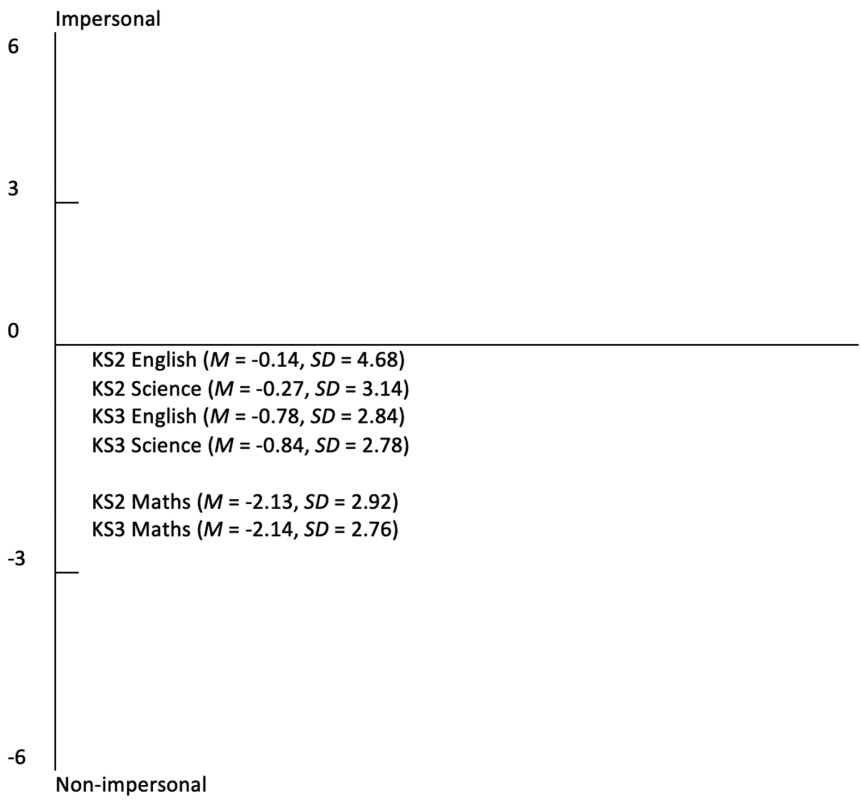



*Figure 4.5* Mean (M) and standard deviation (SD) of Dimension 5 scores across subjects and Key Stages.

for the registers of English. There was very little, if any, change in the degree of non-impersonal style of the registers across the Key Stages. Both English and science registers at KS2 and KS3 were very close to each other in terms of non-impersonal style. Interestingly, the registers of mathematics subjects were overall more non-impersonal than the two registers. It seems counterintuitive that the registers of school language in this study were found to be non-impersonal given that an impersonal style is claimed to be one of the characteristics of school language (e.g., Heller & Morek, 2015; Schleppegrell, 2012). This non-impersonal style can be attributed to the communicative purposes of the sub-registers that were represented in our corpus. Specifically, the majority of sub-registers were worksheets and presentations that served multiple pedagogic functions, including the presentation of the subject content and practice of the content that actively involved students. Even at the university level, Biber et al. (2002) found that passive constructions occurred less frequently in undergraduate-level textbooks than graduate-level ones. At the primary or secondary level of written resources, non-impersonal style seems to be aligned with the communicative purposes of the worksheets, presentations and assessments, most of which required students to actively engage with exercises, questions or other learning activities in this study.

An impersonal style is seen in Example 13. This presents information on impersonal writing to the students in a detached way within a presentation register in English at KS2. In this example, the impersonal style was a reflection of the explicit teaching of the conjunctions *however* and passive constructions, *covered*, rather than texts written in an impersonal tone for comprehension or any other communicative activities. Our qualitative analysis of the texts with high Dimension 5 scores also showed that the impersonal style of the KS2 English sub-registers was largely driven by the explicit presentation of passive constructions and practice activities related to conjunctions and passive constructions. Examples 14 and 15, on the other hand, illustrate a non-impersonal style that directed students to engage with reasoning and answer questions. As can be seen in Examples 14 and 15, there are no conjunctions, passive constructions or adverbial subordinators.

(13) Impersonal writing
- third person
- passive voice
- formal connectives, **e.g., however, therefore, furthermore, consequently**
- usually formal vocabulary **e.g., placed** in rather than **put**

**known** as rather than **called**.
This is **known** as...
The motor is **operated by**
The sides are **covered** in...

(English, Key Stage 2 – Year 6, presentation, Dimension 5 score: 10.9)

(14) How many different nets of a cube can you come up with?
Can you make a net of a Square based pyramid?
Can you make a net of a Triangular Prism?

(Mathematics, Key Stage 3 – Year 7, worksheet, Dimension 5 score: –3.9)

(15) Explain why mixtures have melting point ranges
Is dissolving a physical or a chemical reaction?

(Science, Key Stage 3 – Year 8, assessment, Dimension 5 score: –3.9)

Table 4.7 shows that there was a statistically significant three-way interaction between the predictors of Key Stage 3, the worksheet and the subjects of mathematics/science. This means that the worksheets became more impersonal at KS3 in comparison to KS2 for both the subjects of mathematics and science, unlike the worksheets of English at KS3 which showed a more non-impersonal style than at KS2 level. However, the pairwise

*Table 4.7* Mixed-effects model results: Dimension 5 scores.

| *Predictors* | *Estimates* | *SE* | *CI* | *t* | *p* |
|---|---|---|---|---|---|
| (Intercept)* | –0.78 | 0.41 | –1.58 – 0.03 | –1.89 | 0.058 |
| Mathematics | 0.89 | 0.57 | –0.23 – 2.00 | 1.56 | 0.119 |
| Science | 1.12 | 0.93 | –0.70 – 2.94 | 1.20 | 0.228 |
| Key Stage 3 | 0.80 | 0.68 | –0.54 – 2.14 | 1.17 | 0.241 |
| Presentation | 0.58 | 0.42 | –0.25 – 1.41 | 1.37 | 0.171 |
| Worksheet | 0.17 | 0.40 | –0.62 – 0.96 | 0.42 | 0.671 |
| Mathematics * Key Stage 3 | -3.40 | 0.83 | –5.02 – –1.77 | –4.11 | **<0.001** |
| Science * Key Stage 3 | –2.36 | 1.13 | –4.57 – –0.14 | –2.08 | **0.037** |
| Mathematics * Presentation | –2.98 | 0.71 | –4.37 – -1.58 | –4.19 | **<0.001** |
| Science * Presentation | –0.99 | 1.03 | –3.01 – 1.02 | –0.97 | 0.335 |
| Mathematics * Worksheet | –2.68 | 0.63 | –3.91 – –1.44 | –4.26 | **<0.001** |
| Science * Worksheet | –1.62 | 1.06 | –3.70 – 0.46 | –1.53 | 0.126 |
| Key Stage 3 * Presentation | –1.52 | 0.69 | –2.87 – –0.17 | –2.21 | **0.027** |
| Key Stage 3 * Worksheet | –1.09 | 0.72 | –2.50 – 0.33 | –1.50 | 0.132 |
| Mathematics * Key Stage 3 * presentation | 4.02 | 0.97 | 2.12 – 5.92 | 4.14 | **<0.001** |
| Science * Key Stage 3 * Presentation | 2.09 | 1.24 | –0.34 – 4.53 | 1.69 | 0.092 |
| Mathematics * Key Stage 3 * Worksheet | 3.66 | 0.94 | 1.82 – 5.51 | 3.89 | **<0.001** |
| Science * Key Stage 3 * Worksheet | 2.95 | 1.30 | 0.40 – 5.49 | 2.27 | **0.023** |
| **Random Effect** | Variance | SD | | | |
| School | 0.20 | 0.44 | | | |
| Marginal $R^2$/Conditional $R^2$ | 0.08/0.10 | | | | |

* Reference level is English Key Stage 2 assessment.

comparisons indicated that there was no statistical difference in Dimension 5 scores between the worksheets of KS2 and KS3 for each subject ($t = 0.10$, $p = 0.96$ for mathematics; $t = 0.60$, $p = 0.58$ for English; $t = -0.50$, $p = 0.62$ for science). As seen in Table 4.7, a statistically significant three-way interaction was also observed between the predictors of mathematics, Key Stage 3 and presentation. This suggests that the presentation registers of mathematics became more impersonal at KS3 in comparison with KS2 unlike the presentation registers of English that exhibited a more non-impersonal style at KS3 than at KS2. Nevertheless, the pairwise comparisons indicated no significant difference between KS2 and KS3 presentation registers for each subject ($t = 0.20$, $p = 0.86$ for mathematics; $t = 1.50$, $p = 0.15$ for English; $t = 1.90$, $p = 0.07$ for science). The only statistically significant difference for sub-registers of the same subject between the Key Stages was found in assessments for mathematics that showed a more non-impersonal style at KS3 than KS2 ($t = 4.00$, $p < 0.001$).

A relatively more impersonal style of mathematics assessments at KS2 may be traced back to the detached style of arithmetic and reasoning questions of SATs or practice tests that students did at KS2. As Example 16, which was taken from a practice test, shows, the passive constructions *made* and *needed* created an impersonal style that included no personal involvement. This impersonal style here remained at the syntactic level.

(16) This shape **is made** of wooden centimetre cubes.
How many more centimetre cubes **are needed** to make it into a solid cuboid 3 cm tall, 5 cm long and 5 cm wide?

(Mathematics, Key Stage 2 – Year 6, assessment, Dimension 5 score: 8.2)

## Discussion

This chapter has described the dimensions of linguistic variation in the written school language registers at the transition stage across subjects and sub-registers. The MD analysis indicated that both discipline-specific and sub-register-specific changes occurred in all dimensions except for Dimension 4 between KS2 and KS3. This suggests that these registers have unique ways of meaning-making, which necessitates not just discipline-specific literacy but also sub-register-specific literacy, for both students and teachers. It should be noted that the effect sizes of the mixed-effects models were mostly small, indicating that there are probably other variables that contribute to accounting for the variation in the school language registers of this study.

Overall, all the school language registers of this study were characterised as informational, non-narrative, explicit, non-persuasive and non-impersonal. Table 4.8 shows the changes in the functional variation of the sub-registers based on the statistically significant pairwise comparisons between KS2 and KS3 levels for all dimensions (written as abbreviations in

*Table 4.8* Changes in the functional variation of sub-registers across the Key Stages.

| | *English* | *Mathematics* | *Science* |
|---|---|---|---|
| Assessments | More explicit at KS3 (D3) | More explicit at KS3 (D3); more non-impersonal at KS3 (D5) | More informational at KS3 (D1); more non-narrative at KS3 (D2) |
| Presentations | | | More informational at KS3 (D1); more non-narrative at KS3 (D2) |
| Worksheets | More non-narrative at KS3 (D2) | | More informational at KS3 (D1) |

the table). The written registers of science underwent the most pronounced changes of all the subjects between primary and secondary school. The informational density of all science registers intensified at KS3 in comparison with KS2, increasing the reading demands on KS3 students substantially. This change may also be a reflection of the expectation that students should 'develop understanding of a range of scientific ideas… and use abstract ideas to develop explanations' (DfE, 2014, p. 58) at KS3. As the curriculum targets focus on 'ideas' and 'explanations', the co-occurrence of linguistic resources, including nouns, attributive adjectives, longer words and diverse vocabulary are necessary to make meaning, which increases informational production. Green (2019) found that secondary school science textbooks were more complex at the phrasal level than those for other subjects. This was attributed to science texts entailing 'procedure, report, explanation and exposition' (Fang, 2012, p. 24) and using 'nouns as key resources for compacting information' (Fang, 2012, p. 25). As we showed in Example 2, a cluster of nouns and noun phrases in the science registers may create challenges with decoding for students, especially when there is no wider co-text that could enable students to infer the meaning of these. We also found that the science sub-registers of assessments and presentations became more non-narrative at KS3 in relation to KS2. An increase in non-narrativity coupled with denser informational packaging for these sub-registers probably increased the reading demands for students, since narrative registers tend to be acquired earlier than non-narrative registers (e.g., Jeong, 2017). Chapter 6 reports a qualitatively focused comparison between KS2 and KS3 science texts.

There was an increasing trend towards informational discourse at KS3 for the English registers in comparison with KS2, though this trend was not statistically significant. The assessment sub-registers of English showed increasingly explicit discourse marked by the co-occurrence of relative clauses, nominalisations and phrasal coordination, suggesting an increase in complexity at the lexico-grammatical level. Previous research suggests that nominalisations are associated with comprehension difficulties at the secondary level

for students (Fang et al., 2006). Similarly, the English worksheets became more non-narrative at KS3 than KS2, which can potentially pose comprehension challenges for students. A decrease in narrativity of the worksheets of the English at KS3 may be a manifestation of the change in focus from 'predicting' and 'drawing inferences such as inferring characters' feelings' at KS2 (DfE, 2013, p. 44) to 'making critical comparisons' and 'making inferences and referring to evidence in the text' (DfE, 2014, p. 15) at KS3.

Interestingly, informational density (Dimension 1) showed almost no change in the mathematics registers across the Key Stages, but the assessment sub-registers became more explicit and context-independent at KS3. This increase in context-independent discourse may impose a higher cognitive load on students (see Sweller, 2011). This could potentially give rise to difficulties in understanding instructions and questions within the assessment sub-registers of mathematics at KS3. It seems counterintuitive that the assessment sub-registers of mathematics at KS2 were found to be more impersonal than those at KS3. This may be due to the impersonal nature of the SATs of mathematics at KS2 and practice tests that students were asked to complete in order to prepare for these tests, such as that shown in Example 16, a washback effect (Tennent, 2021). There may have been other changes in language demands of the registers of mathematics subjects between KS2 and KS3, which would not have been captured by the MD analysis. Wilkinson (2019) notes the multi-semiotic nature of mathematics, which involves mathematical symbols and visuals, especially at the secondary level. In Chapter 7, we further explore linguistic variation between the KS2 and KS3 mathematics registers, by combining quantitative corpus techniques with qualitative analysis, showcasing the usefulness of the mixed-method approach to the study of academic school language.

It is striking that the assessment sub-registers at KS3 were more informational, explicit and non-narrative than the presentations or worksheets at KS3, which suggested a greater linguistic complexity for assessment. When the assessment sub-registers become more complex in terms of the three dimensions than the other sub-registers, students would encounter more demanding written resources during assessment probably for the first time at secondary school level. We do not know whether teachers or students are aware of the higher reading demands of assessment sub-registers, but we argue that this finding potentially has important implications, especially for students with low reading abilities and/or for students from low SES backgrounds. 'International data revealing a dip in attainment levels' (West et al., 2010, p. 24) at the beginning of secondary school may partially be explained by the greater complexity and higher reading demands of assessment sub-registers in comparison with the other sub-registers.

Another striking finding of the MD analysis of the school language registers was that all the school language registers of this study were found to be non-impersonal or non-abstract. This is in contrast with previous studies that claimed that school language is impersonal (e.g., Heller & Morek,

2015; Schleppegrell, 2012). We offer two interpretations for this unexpected finding. First, most of the previous studies on academic language in the context of schooling were based on small sets of data or extracts of school texts, and they did not include a wide range of registers, such as worksheets and presentations that would make use of non-impersonal language resources to actively engage students in the content of the subject and activities. Second, the concept of 'impersonal language' was not delineated well in previous studies; hence, it is not possible to make direct comparisons with the findings of earlier studies. As discussed earlier, impersonal language was primarily associated with the detached language style at the syntactic level, attributed to the use of passive constructions and conjuncts in the present study.

## Conclusion

The findings of this chapter offer several theoretical and methodological implications for research on school language and register analysis more broadly. The significant effects of sub-registers or significant interactions between sub-registers and other predictors of subjects or Key Stages suggest that registers need to be categorised in a bottom-up and fine-grained manner, as we have attempted to do in this study. This bottom-up categorisation and situational analysis of sub-registers – assessment, presentations and worksheets – allowed us to develop more fine-tuned understandings of the linguistic and functional variation within the school language registers. For instance, it is notable that the sub-registers of assessment involved more informational density than the presentations or worksheets, irrespective of Key Stages and subjects. It would be impossible to capture this important finding without our novel approach to school sub-registers in this study. Hence, future studies of school language would benefit from further bottom-up situational analysis and categorisation to contribute to our understandings of school language. Moreover, our unexpected finding on the non-impersonal nature of school language registers in this study underlines the importance of corpus-based analysis of a wider range of registers in academic language in the context of schooling and operationalising the construct of 'impersonal' language at a fine-grained level in future register studies on school language.

## References

Bartoń, K. (2022). *MuMIn: Multi-model inference*. R package version 1.46.0. https://CRAN.R-project.org/package=MuMIn

Bates, D. M., Mächler, M., Bolker, B. M., & Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*, *67*(1), 1–48. https://doi.org/10.18637/jss.v067.i01

Berber Sardinha, T., Pinto, M. V., Mayer, C., Zuppardi, M. C., & Kauffmann, C. H. (2019). Adding registers to a previous multi-dimensional analysis. In T. Berber

Sardinha & M. Veirano Pinto (Eds.), *Multi-dimensional analysis: Research methods and current issues* (pp. 165–186). Bloomsbury Academic. https://doi.org/10.5040/9781350023857.0017

Berber Sardinha, T., & Veirano Pinto, M. (2019). Introduction. In T. Berber Sardinha & M. Veirano Pinto (Eds.), *Multi-dimensional analysis: Research methods and current issues* (pp. 1–8). Bloomsbury Academic. https://doi.org/10.5040/9781350023857.0007

Berman, R. A., & Nir-Sagiv, B. (2007). Comparing narrative and expository text construction across adolescence: A developmental paradox. *Discourse Processes*, *43*(2), 79–120. https://doi.org/10.1080/01638530709336894

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

Biber, D. (2006). *University language: A corpus-based study of spoken and written registers* (Vol. 23). John Benjamins Publishing.

Biber, D., & Conrad, S. (2019). *Register, genre, and style* (2nd ed.). Cambridge University Press.

Biber, D., Conrad, S., Reppen, R., Byrd, P., & Helt, M. (2002). Speaking and writing in the university: A multidimensional comparison. *TESOL Quarterly*, *36*(1), 9–48. https://doi.org/10.2307/3588359

Biber, D., Gray, B., & Staples, S. (2016). Predicting patterns of grammatical complexity across language exam task types and proficiency levels. *Applied Linguistics*, *37*(5), 639–668. https://doi.org/10.1093/applin/amu059

Candarli, D. (2022). Linguistic characteristics of online academic forum posts across subregisters, L1 backgrounds, and grades. *Lingua*, *267*, 103190. https://doi.org/10.1016/j.lingua.2021.103190

Clarke, I., & Grieve, J. (2019). Stylistic variation on the Donald Trump Twitter account: A linguistic analysis of tweets posted between 2009 and 2018. *PLoS ONE*, *14*(9), 1–27. https://doi.org/10.1371/journal.pone.0222062

Department for Education. (2013). *The national curriculum in England: Key stages 1 and 2 framework document*. https://www.gov.uk/national-curriculum

Department for Education. (2014). *The national curriculum in England: Key stages 3 and 4 framework document*. https://www.gov.uk/national-curriculum

Fang, Z. (2006). The language demands of science reading in middle school. *International Journal of Science Education*, *28*(5), 491–520. https://doi.org/10.1080/09500690500339092

Fang, Z. (2012). Language correlates of disciplinary literacy. *Topics in Language Disorders*, *32*(1), 19–34. https://doi.org/10.1097/TLD.0b013e31824501de

Fang, Z., Schleppegrell, M., & Cox, B. E. (2006). Understanding the language demands of schooling: Nouns in academic registers. *Journal of Literacy Research*, *38*(3), 247–273. https://doi.org/10.1207/s15548430jlr3803_1

Friginal, E., & Weigle, S. (2014). Exploring multiple profiles of L2 writing using multi-dimensional analysis. *Journal of Second Language Writing*, *26*, 80–95.

García, J. R., Montanero, M., Lucero, M., Cañedo, I., & Sánchez, E. (2018). Comparing rhetorical devices in history textbooks and teachers' lessons: Implications for the development of academic language skills. *Linguistics and Education*, *47*, 16–26. https://doi.org/10.1016/j.linged.2018.07.004

Gardner, S., Nesi, H., & Biber, D. (2019). Discipline, level, genre: Integrating situational perspectives in a new MD analysis of university student writing. *Applied Linguistics*, *40*(4), 646–674. https://doi.org/10.1093/applin/amy005

Goulart, L., Gray, B., Staples, S., Black, A., Shelton, A., Biber, D., Egbert, J., & Wizner, S. (2020). Linguistic perspectives on register. *Annual Review of Linguistics*, *6*(1), 435–455. https://doi.org/10.1146/annurev-linguistics-011718-012644

Gray, B. (2013). More than discipline: Uncovering multi-dimensional patterns of variation in academic research articles. *Corpora*, *8*(2), 153–181. https://doi.org/10.3366/cor.2013.0039

Gray, B. (2021, July 13–16). *Using multi-dimensional analysis to describe variation in specialised domains*. [Paper presentation]. Corpus Linguistics International Conference 2021. University of Limerick, Ireland.

Green, C. (2019). A multilevel description of textbook linguistic complexity across disciplines: Leveraging NLP to support disciplinary literacy. *Linguistics and Education*, *53*, 100748. https://doi.org/10.1016/j.linged.2019.100748

Hardy, J. A., & Römer, U. (2013). Revealing disciplinary variation in student writing: A multi-dimensional analysis of the Michigan corpus of upper-level student papers (MICUSP). *Corpora*, *8*(2), 183–207. https://doi.org/10.3366/cor.2013.0040

Heller, V., & Morek, M. (2015). Academic discourse as situated practice: An introduction. *Linguistics and Education*, *31*, 174–186. https://doi.org/10.1016/j.linged.2014.01.008

Jeong, H. (2017). Narrative and expository genre effects on students, raters, and performance criteria. *Assessing Writing*, *31*, 113–125. https://doi.org/10.1016/j.asw.2016.08.006

Le Foll, E. (2021). Register variation in school EFL textbooks. *Register Studies*, *3*(2), 207–246. https://doi.org/10.1075/rs.20009.lef

Linck, J. A., & Cunnings, I. (2015). The utility and application of mixed-effects models in second language research. *Language Learning*, *65*(S1), 185–207. https://doi.org/10.1111/lang.12117

Maydeu-Olivares, A., & Garcia-Forero, C. (2010). Goodness-of-fit testing. *International Encyclopedia of Education*, *7*(1), 190–196.

Nini, A. (2019). The multi-dimensional analysis tagger. In T. Berber Sardinha & M. Veirano Pinto (Eds.), *Multi-dimensional analysis: Research methods and current issues* (pp. 67–96). Bloomsbury Academic. https://doi.org/10.5040/9781350023857.0012

R Core Team. (2020). *A language and environment for statistical computing*. R Foundation for Statistical Computing, Vienna, Austria. https://www.R-project.org/

Reppen, R. (2001). Register variation in student and adult speech and writing. In S. Conrad, & D. Biber (Eds.), *Variation in English: Multi-dimensional studies* (pp. 187–199). Pearson. https://doi.org/10.4324/9781315840888

Schleppegrell, M. (2001). Linguistic features of the language of schooling. *Linguistics and Education*, *12*(4), 431–459. https://doi.org/10.1016/S0898-5898(01)00073-0

Schleppegrell, M. (2007). The linguistic challenges of mathematics teaching and learning: A research review. *Reading and Writing Quarterly*, *23*(2), 139–159. https://doi.org/10.1080/10573560601158461

Schleppegrell, M. (2012). Academic language in teaching and learning. *The Elementary School Journal*, *112*(3), 409–418. https://doi.org/10.1086/521238

Sweller, J. (2011). Cognitive load theory. *Psychology of Learning and Motivation*, *55*, 37–76. https://doi.org/10.1016/B978-0-12-387691-1.00002-8

Tennent, W. (2021). The assessment of reading comprehension in English primary schools: investigating the validity of the key stage 2 reading standard assessment test (SAT). *Education 3-13*, *49*(4), 481–494. https://doi.org/10.1080/03004279.2020.1735472

Thompson, P., Hunston, S., Murakami, A., & Vajn, D. (2017). Multi-dimensional analysis, text constellations, and interdisciplinary discourse. *International Journal of Corpus Linguistics*, *22*(2), 153–186. https://doi.org/10.1075/ijcl.22.2.01tho

Toutanova, K., Klein, D., Manning, C., & Singer, Y. (2003). Feature-rich part-of-speech tagging with a cyclic dependency network. In *Proceedings of the 2003 conference of the North American chapter of the association for computational linguistics on human language technology* (pp. 252–259). Association for Computational Linguistics.

West, P., Sweeting, H., & Young, R. (2010). Transition matters: pupils' experiences of the primary–secondary school transition in the West of Scotland and consequences for well-being and attainment. *Research Papers in Education*, *25*(1), 21–50. https://doi.org/10.1080/02671520802308677

Wilkinson, L. C. (2019). Learning language and mathematics: A perspective from linguistics and education. *Linguistics and Education*, *49*, 86–95. https://doi.org/10.1016/j.linged.2018.03.005

Winter, B. (2019). *Statistics for linguists: An introduction using R*. Routledge.

# 5 The language of English at the transition

*Alice Deignan and Florence Oxley*

## Introduction

This chapter investigates the change in the language of school English at the transition. We begin by overviewing the curriculum goals at each Key Stage, and central issues in each of the main areas of English teaching. This leads to our corpus investigation into the academic school language of the discipline, and what this reflects back on the nature of the subject and how it changes.

### *The KS2 and KS3 curricula*

As for other subjects, the National Curriculum for English in England and Wales was reviewed in 2013, and the version currently in force has been taught since September 2014. The curriculum proceeds from Key Stage 1 through to Key Stage 4, that is, spanning primary and secondary school and ostensibly aiming at an integrated and coherent curriculum. The following description of the goals of Years 5 and 6 is taken from the DfE document 'English programmes of study: key stages 1 and 2' (2013a, p. 31) and is slightly edited for reasons of space.

> By the beginning of year 5, pupils should be able to read aloud a wider range of poetry and books written at an age-appropriate interest level with accuracy and at a reasonable speaking pace. They should be able to read most words effortlessly and to work out how to pronounce unfamiliar written words with increasing automaticity. […]
>
> They should be able to prepare readings, with appropriate intonation to show their understanding, and should be able to summarise and present a familiar story in their own words. They should be reading widely and frequently, outside as well as in school, for pleasure and information. They should be able to read silently, with good understanding, inferring the meanings of unfamiliar words, and then discuss what they have read.

DOI: 10.4324/9781003081890-5

> Pupils should be able to write down their ideas quickly. Their grammar and punctuation should be broadly accurate. Pupils' spelling of most words taught so far should be accurate and they should be able to spell words that they have not yet been taught by using what they have learnt about how spelling works in English.
>
> During years 5 and 6, teachers should continue to emphasise pupils' enjoyment and understanding of language, especially vocabulary, to support their reading and writing. Pupils' knowledge of language, gained from stories, plays, poetry, non-fiction and textbooks, will support their increasing fluency as readers, their facility as writers, and their comprehension. As in years 3 and 4, pupils should be taught to enhance the effectiveness of their writing as well as their competence.
>
> It is essential that pupils whose decoding skills are poor are taught through a rigorous and systematic phonics programme so that they catch up rapidly with their peers in terms of their decoding and spelling. […]
>
> By the end of year 6, pupils' reading and writing should be sufficiently fluent and effortless for them to manage the general demands of the curriculum in year 7, across all subjects and not just in English, but there will continue to be a need for pupils to learn subject-specific vocabulary. They should be able to reflect their understanding of the audience for and purpose of their writing by selecting appropriate vocabulary and grammar. Teachers should prepare pupils for secondary education by ensuring that they can consciously control sentence structure in their writing and understand why sentences are constructed as they are. Pupils should understand nuances in vocabulary choice and age-appropriate, academic vocabulary. This involves consolidation, practice and discussion of language.

This is followed by a more detailed description of objectives in reading and writing, including an inventory of grammar forms, words to spell and punctuation. There is focus on reading fluency, word decoding, spelling, morphological awareness, reading for pleasure and comprehension, and some focus on inference, purpose and audience. The mechanical and functional skills associated with reading and writing are expected to be secure by the end of KS2.

A corresponding overview of the KS3 curriculum is given in the DfE publication 'English programmes of study: key stage 3' (2013b, pp. 2–3), as follows (slightly edited).

> **Spoken language**
>
> The national curriculum for English reflects the importance of spoken language in pupils' development across the whole curriculum

– cognitively, socially and linguistically. Spoken language continues to underpin the development of pupils' reading and writing during key stage 3 and teachers should therefore ensure pupils' confidence and competence in this area continue to develop. Pupils should be taught to understand and use the conventions for discussion and debate, as well as continuing to develop their skills in working collaboratively with their peers to discuss reading, writing and speech across the curriculum.

**Reading and writing**

Reading at key stage 3 should be wide, varied and challenging. Pupils should be expected to read whole books, to read in depth and to read for pleasure and information.

Pupils should continue to develop their knowledge of and skills in writing, refining their drafting skills and developing resilience to write at length. They should be taught to write formal and academic essays as well as writing imaginatively. They should be taught to write for a variety of purposes and audiences across a range of contexts. This requires an increasingly wide knowledge of vocabulary and grammar.

Opportunities for teachers to enhance pupils' vocabulary will arise naturally from their reading and writing. Teachers should show pupils how to understand the relationships between words, how to understand nuances in meaning, and how to develop their understanding of, and ability to use, figurative language.

Pupils should be taught to control their speaking and writing consciously, understand why sentences are constructed as they are and to use Standard English. They should understand and use age-appropriate vocabulary, including linguistic and literary terminology, for discussing their reading, writing and spoken language. This involves consolidation, practice and discussion of language. It is important that pupils learn the correct grammatical terms in English and that these terms are integrated within teaching.

As was the case for the KS2 curriculum, this is followed by more detail, including a glossary of grammatical and other linguistic terminology, but unlike KS2, without a list of language features to be learned; Verhoeven (2021) writes that this represents discontinuity. The KS3 reading and writing specifications include a strong focus on register, purpose and audience. Comprehension, inference and awareness of purpose and audience are expected to be developed further. Additionally, there is an emphasis on conscious metalinguistic knowledge, awareness of and

ability to engage with stylistic choices and the ability to critically analyse and consciously produce effect.

### *Assessment*

In England and Wales, students take National Curriculum Tests (commonly called SATs – Standard Attainment Tests) in English and mathematics (see Chapter 1) at the end of KS2, in the May when students are in Year 6. Currently, there are three separate English tests: two on grammar, punctuation and spelling, abbreviated to GPS (also known as SPaG) and one on reading. Writing is assessed over the year by the class teacher. The tests are marked externally and are used to evaluate and compare the effectiveness of schools. For students, SATs results are in principle not important, but in reality, they can matter, because secondary schools may use them as the basis for setting incoming students into ability groups (Tennent, 2021), potentially leading to self-fulfilling expectations of later attainment. Students in our interviews and in our wider experience reported caring about their results for their own sake. SATs are thus important for both teachers and students, and have a significant washback effect on teaching, especially in Year 6 (Cushing & Helks, 2021; Tennent, 2021), whereby 'Effectively, teachers "teach to the test"' (Tennent, 2021, p. 482). The impact of this on teaching and learning will come up later in this chapter.

In most schools in England and Wales, KS3 covers the first three years of secondary school. Formal SATs at the end of KS3 were abolished in October 2008, replaced by a requirement for schools to monitor progress and inform parents. There is therefore no external measure of school and student performance against the KS3 objectives. In contrast, KS4, when students are aged 14–16 years, leads up to national examinations, GCSEs. Performance in GCSEs is of great importance for students and their schools, as discussed in Chapter 1. Teachers have told us that KS3 goals may therefore be seen as less important, and there is a danger of the time that should be spent on KS3 goals being seen as an opportunity for early groundwork for KS4. NATE (National Association for Teaching English) refers to 'the colonisation of KS3 by KS4' (2022, p. 14).

Smith and her co-writers, all experienced teachers and teacher trainers, interviewed a number of teachers about KS3 English teaching (Smith et al., 2021). They argue that KS4 assessment impacts KS3 English negatively. For example, Smith et al. describe some schools introducing GCSE texts such as 'A Christmas Carol' (Charles Dickens) as early as Year 7, when, they claim, many students are not ready emotionally and intellectually. Teachers who we spoke to have made the same point about the use of Shakespeare and Brontë's Jane Eyre in KS3, which are taught, as one of them told us, 'either using a few difficult extracts or a (horrible) abridged version'. One teacher we spoke to told us that pressure to set her class difficult poetry from the KS4 set texts so that they could 'get ahead' was

leading to KS3 being 'starved of creativity and joy'. Lawrence (2020) found that some English teacher trainees they worked with did not see value in trying to engage students with poetry unless the activity overtly developed skills that would be tested in formal assessment. KS3 English is thus taught under conflicting objectives and pressures. We now explore issues in learning English further, grouped around the themes of reading, writing and spoken language, and then present our analyses of the academic language used across this period of schooling.

## Reading in Years 5–8

Key themes through the KS2 and KS3 curricula in reading are: reading for pleasure; making inferences; understanding genre, purpose and audience, and criticality.

### *Reading for pleasure*

The Year 5 and 6 requirements state that students should 'maintain positive attitudes to reading', while the KS3 curriculum requires that they should develop 'an appreciation and love of reading'. Cremin (2015) discusses the benefits of reading fiction and poetry for pleasure for children's personal, emotional and imaginative development as well as the development of their literacy skills. However, Hempel-Jorgensen et al. (2018) identified poor practice, reminiscent of points made by the teachers that we quoted in the previous section. They studied Year 5 reading in four low-SES primary schools in England which had invested resources in reading for pleasure; 12 children took part in focus groups and class teachers were interviewed and observed. The teachers' pedagogical practices were observed to be rooted in their notions of proficiency, and they sometimes restricted children's choices about what to read. They seemed to have low expectations of the children's potential to engage with literature, all suggesting a limited appreciation of the possibilities of reading for pleasure, and pedagogical outlooks and habits formed by the demands of external assessment.

At the secondary level, Cremin and Swann (2016) found that reading for pleasure was not perceived by students as part of their English subject learning, but something distinctly extra-curricular. Reading within lessons was focused on set books, and demonstrating the ability to read. Students often worried that their peers would label them 'geeky' for reading more widely (Cremin & Swann, 2016; Warsop, 2015).

Being able to talk about their reading with adults, especially in their family, is important to many students' enjoyment (Maynard, 2011). Parental involvement naturally declines in secondary school as students grow past the stage of being read aloud to (Maynard, 2011). Nottingham Education Partners (n.d.) also note that parental involvement in reading declines around the time when children transition to secondary school and claim

that this loss of support is detrimental to children's motivation and resilience with reading for pleasure, particularly among boys. Warsop's (2015) longitudinal study of students in Years 6 and 7 found that having an 'enabling adult' was important to continuing reading for pleasure in secondary school. In her data, this was often the school librarian, sometimes an English teacher. However, teachers do not always have the time or background knowledge to provide this support. Cremin et al. (2008) surveyed 1200 primary teachers about reading and literature. Many reported enjoying reading for pleasure themselves, but few were able to name six 'good' children's authors and poets, and tended to name the same, fairly narrow selection of well-known writers, such as Roald Dahl and Jacqueline Wilson, authors that were also named in Hempel-Jorgensen et al.'s study (2018). Cremin et al. (2008) suggest that this gap in teachers' knowledge may limit their ability to adequately teach reading for pleasure, particularly poetry. Hanratty and McPolin (2018) found that of 32 primary school teachers they surveyed, only four reported reading poetry or other literature for pleasure themselves; 30 of the 32 had not studied English Literature during their undergraduate and postgraduate degrees. Overall, research suggests a mixed picture around children's reading for pleasure: its importance and potential for educational and emotional benefits are generally agreed and there have been well-intentioned initiatives, but there is only patchy success, particularly in KS3, when the high-stakes KS4 assessments already cast a long shadow.

### *Making inferences*

Drawing inferences when reading is emphasised in both curricula. Kispal (2008) reviewed the research literature on teaching inferencing with reference to KS2 and KS3. She found that a child's ability to infer from their reading predicts their general reading comprehension. She also notes that students find it easier to generate inferences from narrative texts. Expository texts were more difficult to generate inferences from, probably because they have a generic structure that is less understood by non-experts, and because children might not have enough background knowledge. She also found that adult modelling of inferential skills in class and peer discussion is effective in demonstrating to children not only what to infer, but how to infer information from literature.

Phillips (2013) studied children at the beginning of KS2, finding that inferential reasoning skills are most effectively supported by teacher-led and peer discussion of literature. More generally, research has investigated the types of questioning most often used by teachers and the types of questioning that are most effective for children when analysing texts. Parker and Hurry (2007) investigated Key Stage 2 teachers' use of questioning during literacy lessons. Interviews and lesson observations were conducted with a sample of 51 teachers across 13 primary schools. They found that teachers

modelled sophisticated comprehension skills and strategies such as summarising, inferencing using contextual cues or identifying unknown language, but they often did not explicitly explain how, when or why these strategies should be employed. As a result, pupils' ability to assimilate and use these strategies was limited. Durran (2017) also suggested that implicit rather than explicit teaching of literacy may result in incomplete assimilation of knowledge and skills.

### *Understanding genre, purpose and audience; criticality*

The KS3 curriculum expects students to be able to consider the genre, purpose and audience of a text that they read, as part of the comprehension process. What this means in reality is rarely explored for reading, most discussion concerning writing. Several writers note a shift in the KS3 English curriculum away from a language focus in KS2 to a literature focus (e.g., Verhoeven, 2021), which brings an increasingly specialised approach to critiquing texts.

## Writing in Years 5–8

Key themes in writing in KS2 are presentation, including spelling, accuracy in grammar and vocabulary choice and use of Standard English; mastery of the writing process, from first draft to evaluating one's own work; and attention to the audience, purpose and genre of writing. Key themes in KS3 are accuracy in organisation, grammar and vocabulary, and in use of Standard English; attention to audience, purpose and genre; and argumentation. The detailed descriptive focus on grammar from KS2 does not appear in the KS3 curriculum document. Language work is discussed in the next section.

### *Understanding genre, purpose and audience*

A key aim of the Key Stage 2 and 3 National Curricula is to develop an awareness of a growing range of purposes and audiences, and to adapt one's writing accordingly. Jones (2021) observed that, among primary and secondary school children in Years 4–6 and 7–9, no 'audience' beyond the class teacher was imagined for pieces of written work. In one sense, children in this study did show evidence of making linguistic choices that tailored their written work to this 'audience' – children consciously chose to vary sentence structure and use what they perceived to be 'good' vocabulary in order to demonstrate their writing proficiency and please their teacher. However, many children found it difficult to explain how and why their language choices were appropriate to the genre or register of their written work, or what effect they might have on the 'intended' audience of such writing. This difficulty was mirrored in teaching staff involved in this study and Jones (2021, p.17) suggests that statutory grammar testing may encourage

memorising and using formulaic structures at the expense of classroom dialogue about writing for effect. Verhoeven (2021) suggests that this encourages writing habits that later have to be unlearned.

## Language and metalanguage in Years 5–8

### *Vocabulary*

Vocabulary is important for all school subjects, and as students progress through secondary school, their writing evidences increased register-appropriate use of academic vocabulary (Durrant & Brenchley, 2019). Unfamiliarity with the words used to carry meaning and subject content can slow or, sometimes, impede children's access to subjects across the curriculum (Coleman, 2017). Researchers and teachers have noted significant discrepancies among Year 7 pupils' vocabularies (e.g., Quigley 2016). We have discussed vocabulary issues across the curriculum in Chapter 2, and the vocabulary of science and mathematics will be discussed in later chapters; here we discuss vocabulary within the English curriculum.

Both KS2 and KS3 English curricula emphasise the importance of developing a wide vocabulary year on year. Both mention nuance of meaning, figurative language and age-appropriate metalanguage for discussing linguistics and literature. Learning vocabulary is exponential and closely linked to reading, in that a wider vocabulary enables students to read more widely, and to enjoy that reading, which then exposes them to still more vocabulary. This means that the gap between those who have an extensive vocabulary and those who do not tends to widen over time (Quigley, 2018).

There is research evidence that KS2 teachers pay a good deal of attention to vocabulary, but this tends to have a rather specific focus on enriching descriptions through the use of low-frequency words. The KS2 statutory word list for Years 5 and 6 contains a high volume of adjectives and adverbs, and these seem to occupy the most attention of both teachers and students. Jones (2021) describes a Year 6 student talking about his word choice in a piece of his writing and singling out adverbs he was pleased with. She writes that he seemed unaware that his choice of the nouns 'chasm' and 'canopy' were 'perhaps the most evocative in terms of setting the scene as being in the rain forest' (2021, p. 12) because they are nouns. Both Jones (2021) and Barrs (2019) refer to primary school students' use of 'wow' words, that is, colourful, low-frequency words, often adjectives, to replace more high-frequency words. There are numerous online resources available for teaching 'wow' words, defined on one such website as 'advanced adjectives, verbs and adverbs, which are good vocab to use in creative writing and make a piece of written work more vivid and interesting' (www.twinkl.co.uk/teaching-wiki/wow-words).

Barrs (2019) argues that primary and secondary school students have been over-encouraged to vary their vocabulary, often through use of online

thesauri, to the extent that they produce texts that are difficult to understand. She describes students choosing infrequent near-synonyms of a more familiar word, not understanding restrictions on use or connotations. For instance, she quotes a Year 9 student misusing 'briskly' in 'Briskly, the amount of alcohol intake for young people is rising' (2019 p. 13); we assume that the student had searched for a synonym for 'rapidly', and had also misplaced the adverb, possibly as a result of classroom focus on fronted adverbials. Both Barrs (2019) and Jones (2021) found that the students in their studies were not able to give details about why they had chosen particular words, beyond saying that they were 'better' or 'more advanced'. This echoes what we were told by the Year 6 students whom we interviewed, as reported in Chapter 1.

### *Grammar teaching*

Several writers note that there is a strong focus on grammar teaching in KS2, which is not followed through in KS3 (e.g., NATE, 2022). Verhoeven writes 'Year 6 pupils are expected to identify word classes and comment on a writer's use of fronted adverbials, but there is then no follow-up in Years 7 to 9 English at all' (2021, n.p.). Cushing also notes discontinuity in approaches to grammar: while primary school grammar teaching prioritises a view of grammar as a set of features and rules which are often taught in isolation and must be followed, secondary school grammar teaching takes a more descriptive approach to analysing grammar in use (Cushing, 2018a). He writes that primary school grammar teaching focuses on rapid recall of 'correct' ideas', while secondary school grammar teaching is intended to promote creative and analytical thought and engagement with authorial choice. Interview and survey data from 299 secondary school English teachers revealed that teachers believed that their incoming students viewed grammar as a list of terms and had a limited understanding concerning their effect or application (Cushing, 2018a).

Cushing (2018a, 2019), Jones (2021) and Safford (2016) have suggested that preparing children for the SPaG/GPS testing component of the KS2 SATs can promote prescriptivism, authority and performativity in grammar teaching at the expense of contextualised approaches. In his interview study of 22 primary school teachers, Cushing (2019) found that teachers felt that preparing Year 6 students to undertake GPS tests required them to characterise Standard English as 'correct' language, with non-Standard features representing 'incorrect' usage. The design of the SPaG test was also found to significantly influence the content of and pedagogical approaches taken to primary school grammar teaching in a study by Safford (2016). In interviews and surveys, 186 primary school teachers and teaching staff reported that they now spent much more time teaching both contextualised and decontextualised grammatical terminology explicitly, often using quizzes, drills and short writing activities similar

to those they used in teaching mathematics and phonics. Jones (2021) warns that focusing too heavily on explicitly teaching language features and linguistic terminology can distract children from communicating or interpreting the intended meaning of a text. In terms of their writing, she found that children made certain grammatical choices based on 'rehearsed classroom mantras' (Jones, 2021, p. 18). For instance, children reported using rhetorical questions and short sentences to provoke thought and create tension but were not able to explain how or why these choices had these effects.

Cushing and Helks' (2021) focus group study investigated Year 6 and 7 children's experiences of grammar teaching. They found that children's representations of grammar centred around ideas of correctness and terminology for concepts like 'word classes, phrases, and clauses' (p. 242), showing a strong influence of the National Curriculum. Absent from children's reports were ideas about how grammar relates to meaning, effect or authorial choice (p. 243). Other researchers have suggested that teaching grammar explicitly in terms of choices can benefit children's engagement with English literature. Myhill and colleagues (Myhill, 2021; Myhill & Watson, 2017; Myhill & Newman, 2019) have argued that using authentic texts and 'high quality [class] discussion' (Myhill & Watson, 2017, n.p.) about grammar, authorial choice and effect has long-lasting effects in terms of students' academic self-esteem and autonomy. They claim that explicitly teaching linguistic analysis as well as literary analysis can help children to access literature and learn more deeply about how grammar can be used to create meaning.

Teaching this kind of stylistic analysis requires confidence and some level of specialist knowledge. Cushing (2019) notes that a relatively small number of primary teachers have a linguistic background and suggests that teachers' own prescriptive understanding of grammar may influence their teaching. Further, some secondary school English teachers in Cushing's study did not feel well-equipped to implement the contextualised, explicit and prescriptive grammar teaching that they felt was required of them by educational policy (2018a). Cushing writes that most English teachers in the UK come from a literature background, and he has found some opposition to linguistics, teachers perceiving it as rule-bound and at odds with their identity as specialists in literature (2018b). In contrast, Bell (2016) found that while primary school teachers needed to work on their knowledge base of grammar, they nonetheless had positive attitudes towards it. Durran (2017) suggests that the move from explicit teaching of literacy at primary school to the largely implicit literacy teaching at secondary school may pose an additional difficulty for some pupils.

We have briefly overviewed studies of English teaching in KS2 and KS3, focusing on the core curriculum areas, and have identified some central curriculum themes and areas raised as possible concerns. In the second half of this chapter, we move on to our corpus studies.

## Corpus studies of the language of English in Years 5–8

Our corpus studies compared the KS2 English corpus with the KS3 English corpus, informed by a reference corpus. We consider what these can tell us about how the academic language of studying English changes, and what this might tell us about the changes in focus and emphasis as students move from primary to secondary school through the following questions:

Which words are significantly more frequent in KS3 English than in KS2 English?
Which words are significantly more frequent in KS3 English than in everyday English that students might have encountered outside the classroom?
What are their main meanings and functions of the words in these contexts?
What can this tell us about the nature of studying English in primary and secondary school and how this seems to change with the transition?

## Method

### *The corpora used*

In Chapter 3, we described how our school corpus data were collected. Table 5.1 contains figures from Tables 3.3 and 3.4 to show the composition of the written English corpus. Table 5.2 is extracted from Table 3.6 and shows the composition of the spoken (teacher talk) English corpora.

The study described here did not separate written and spoken data, but it did separate KS2 and KS3; in other words, the data were sliced differently from in the previous tables. Table 5.3 gives the same information about

*Table 5.1* Written English corpus from KS2 and KS3.

| *Subject* | *Texts* | *Tokens* | *Mean length* | *SD text length* |
|---|---|---|---|---|
| KS2 English | 600 | 303,257 | 505 | 1381 |
| KS3 English | 334 | 260,806 | 781 | 3552 |
| Total | 934 | 564,063 | | |

*Table 5.2* Spoken English corpus from KS2 and KS3.

| | *Number of texts* | | *Number of tokens* | | *Mean text length (tokens)* | | *Standard deviation text length (tokens)* | |
|---|---|---|---|---|---|---|---|---|
| | *KS2* | *KS3* | *KS2* | *KS3* | *KS2* | *KS3* | *KS2* | *KS3* |
| English | 15 | 8 | 72,475 | 47,595 | 4832 | 5949 | 1284 | 992 |
| Subtotal (spoken) | 23 | | 120,070 | | | | | |

*Table 5.3* Division of English corpus by Key Stage.

| | *Key Stage 2* | | *Key Stage 3* | |
|---|---|---|---|---|
| | *Number of texts* | *Tokens* | *Number of texts* | *tokens* |
| Written | 600 | 303,257 | 334 | 258,869 |
| Spoken | 15 | 72,475 | 8 | 47,595 |
| Total | 615 | 375,732 | 306 | 306,464 |

texts and tokens as Tables 5.1 and 5.2, but has been reorganised to show how we conducted these studies.

For one of the studies, we used reference corpora. A reference corpus is a general corpus that is used as a 'baseline' to compare a specialised corpus with (Hunston, 2002, p. 15), and is generally used to find out what is special about a specialised corpus. Scott writes: 'by comparing the frequency of each item in turn with a known reference, one may identify those items which occur with unusual frequency' (2009, p. 80). We used a reference corpus to find out what is special about texts that students encounter in English in Years 7 and 8. McEnery et al. (2006) give evidence from a small-scale study that the size of a reference corpus is not very important, so we decided not to prioritise size, but rather to use a reference corpus that was as current and balanced as possible from the various general corpora that are freely available. We chose the BNC2014 Baby+ corpus as our starting point (CASS, n.d.). We also consulted concordance data from the Oxford English Corpus, where we were interested in details of usage of less frequent words through the Sketch Engine software (Kilgarriff et al., 2014; www.sketchengine.eu).

The BNC2014 Baby+ was released in 2019 and was extracted from the data gathered for the British National Corpus 2014, which was released in two stages, the spoken sub-corpus in 2017 and the written sub-corpus in 2021. The full BNC2014 Baby+ contains 13 files, with 5,024,072 tokens (Table 5.4).

Scott (2009) argues that for detecting 'aboutness' of a specialised corpus, very similar results are found almost regardless of the composition of the reference corpus. For us though, the concern is not what is special relative to the language as a whole, but what is special relative to language that

*Table 5.4* Composition of the BNC2014 Baby+.

| | |
|---|---|
| Academic books | e-language social media |
| Academic journals | Fiction |
| e-language blogs | News: mass market |
| e-language email | News: regional |
| e-language forums | News: serious newspapers |
| e-language reviews | Speech |
| e-language SMS | |

students encounter outside the academic sections of the classroom. There may be everyday features of English as encountered by adults that are still unfamiliar to children aged 11–13. In starting with the language user, we are taking a slightly different approach from the usual goal of representing a language, text type or register (Baker, 2006). In this approach, 'corpora are created for the purpose of better understanding a particular type of discourse', and need 'specific texts that together can serve as a characteristic example of the target variety or target domain' (Friginal & Hardy, 2020: 2). Clancy (2010) makes a distinction between Variety and variety; the former is 'defined geographically' and related to users. His corpus of Irish English thus contains samples of different Irish dialects. Our approach is closer to Clancy's 'Variety', that is, tied to a specific group of users.

We therefore modified the BNC2014 Baby+ in an attempt to bring it closer to a reference corpus approximating to the adult language they encounter in everyday life. Following our interviews with teachers and students, we decided that the users we are concerned with, students in Years 5-8, would be less likely to encounter four of these text types – academic books and journals, emails and serious news – than the others, and we therefore removed these files. The serious newspaper sub-corpus consists of texts from the *Financial Times*, *Guardian*, *Observer*, *Times* and *Sunday Times*. The academic section consists of university-level texts, from disciplinary areas such as medicine and social sciences. The email section is small, at 24,333 tokens, so the decision about whether or not to include it is not significant in terms of volume, but we decided to exclude it, from conversations with students and teachers stating that students of this age rarely engage with email. We hypothesised that these four files could share features of the KS3 school corpus that would be unfamiliar to students, such as some specialised academic and technical terminology. Including the files in the reference corpus would have meant that the comparison would not have highlighted these features.

This left us with a reference corpus that we named 'BNC2014 Baby+ (Modified)', abbreviated to 'BNCBM', composed as shown in Table 5.5.[1] Numbers of types, tokens and lemmas are from the data downloaded with the files through #LancsBox v.6. Information and quotations in column 2 are taken from CASS (n.d.).

The total number of types in BNCBM is 94,335, as calculated by the #LancsBox v.6 software (Brezina et al., 2020). (The total number of types is considerably less than the total of types in each sub-corpus because many are duplicated across the nine sub-corpora.)

### *Frequent word analysis*

We began by examining the frequent words in the KS3 English corpus to get a sense of 'aboutness', following Baker et al. (2013). Baker et al.'s corpus consisted of approximately 143,000,000 tokens and was compiled with the

*Table 5.5* Composition of the BNCBM.

| *File/sub-corpus* | *Description* | *Number of tokens* | *Number of types* |
|---|---|---|---|
| e-language blogs | 295 blogs | 209,360 | 17,523 |
| e-language forums | 39 discussion forums | 195,363 | 14,896 |
| e-language reviews | 39 product reviews | 195,622 | 14,012 |
| e-language SMS | 22 files | 182,996 | 13,752 |
| e-language social media | Facebook and Twitter | 196,200 | 24,496 |
| fiction | Approx. 15,000 word samples from 69 books published 2010–2017 | 1,007,907 | 36,418 |
| news mass media | *Daily Star*, *Daily Star Sunday*, *Sunday Express*, *Sunday Mirror*, *The Express*, *The Mirror*, *The Sun* | 365,418 | 29,759 |
| news regional | 13 regional newspapers from England, Wales, Scotland, Northern Ireland | 361,076 | 25,941 |
| speech | A subset of the Spoken BNC2014 (Love et al. 2017), spoken data 'broadly representative of the UK population in terms of age, gender, region and class' | 932,820 | 21,407 |
| **Total** | | **3,646,762** | **94, 335** |

goal of understanding how Moslems are written about in different British newspapers. Baker et al. gathered articles published between 1993 and 2009 from a database of British newspapers through the use of around 40 query terms closely associated with Islam, such as *Koran*, *Mecca*, *Moslem* and *Muslim*, yielding just over 200,000 articles containing one or more of them. These query terms were therefore frequent in the resultant corpus. Baker et al.'s initial searches identified some other non-query content types, such as *terror* (found especially in texts after the 9/11 events). *Terror* and related words (*terrorist* etc) occurred around 40,000 times each. Baker et al. decided to investigate all non-query content words occurring at this level of frequency or higher. They found 147 such content types, termed '40K types', which covered 15.1% of tokens. Of these 147, 85 types indicate specific content, the others being more general words such as *come*, *take*, *good*, *little* and *new* (2013, p. 52). The 85 types include *war*, *government*, *police*, *military* and *attacks*. Baker et al. argue that these '40K' types 'reflect the most frequent topics in the corpus' (2013, p. 52).

We used LancsBox v.6 to create a word list for the KS3 English and KS2 corpora. Following Baker et al. (2013), we did not lemmatise but looked at types. In Baker et al.'s corpus, a raw frequency of 40,000 is equivalent to a normalised frequency of 27.9 times per 100,000 words. The same normalised frequency cut-off for our KS3 English corpus would have taken us to words that occurred 86 times, and at this point, manual inspection showed little interest in the aboutness of the corpus. Further, it would have generated 760 content types, which would have been unmanageable, in contrast

with the 147 types in Baker et al.'s (2013) study. It is well-established that type-token ratio decreases as corpus size grows and that more narrowly focused corpora have a lower type-token ratio (Baker, 2006). This means that we could expect to see a greater variety of types at the same normalised frequency level in our small KS3 English corpus, which covered a range of topics, than Baker et al. found in their corpus, which was both much larger overall and narrower in terms of topic. We therefore set our cut-off point higher, at types that occurred 100 times or more in the KS3 corpus, or 32.6 times per 100,000 words. This gave us 207 types for the KS3 corpus. The KS2 corpus is slightly larger, and we set the same normalised frequency cut-off point of 32.6 occurrences per 100,000 words, which took us to types that occurred 123 times or more, of which there are 363. Both of these word lists included a small number of transcription codes, numbers from pages and other non-words. We also considered dispersion, that is, the extent to which types are evenly, or unevenly, distributed across the texts in the corpus. We used $DP_{norm}$ as a measure of dispersion (Lijffijt & Gries 2012). This produces a value from 0, or perfectly distributed, to 1, that is, very unevenly distributed. We eliminated types with a value of over 0.95. We used #LancsBox v.6 for this calculation.

We manually checked the resulting lists, following Baker et al. in removing those whose meanings and functions were general. The examples of general words that they give are

> general lexical verbs (e.g., *come*, *say*, *take*), lexical markers of modality (e.g., *think*, *want*, *need*), and general adjectives (e.g., *good*, *little*, *new*). (2013, p. 52)

Baker et al. do not provide a complete list of the general words that they eliminated. To establish a list for our studies, we used the 'New General Service List' (*New GSL*) (Brezina & Gablasova, 2015) as a guide. Brezina and Gablasova list 2500 general words, in rank order, based on detailed corpus analysis. The full list of 2500 types would have covered a large proportion of our word lists, including some types that we could see, through manual inspection, were topic-specific. We therefore needed to decide on a cut-off point to define 'general words' for our studies. We also took account of Stubbs (2001, p. 42), who notes that the most frequent words in almost any corpus will be function words, followed by 'a few content words such as *think*, *know*, *time*, *people*, *two*, *see*, *way*, *first*, *new*, *say*, *man*, *little*, *good*'. We then consulted Brezina and Gablasova's ranking of the words listed by Stubbs and by Baker et al., finding that the majority are in the top 100, with four exceptions: *man* (105), *want* (106), *need* (117) and *little* (174). We therefore decided to set our cut-off at the top 200 words in the *New GSL* to capture the notion of general, non-topic-specific words in an as accurate and replicable way as possible. This results in a few anomalies; for instance, the exclusion of *people* (ranked 79) but not *person* (ranked 329), and the

inclusion of *down* (ranked 201) and *both* (ranked 202), but has the important advantage of replicability.

We also reduced our word lists by removing proper names and titles (*George*, *Lennie*, *King*, *Macbeth*, *Mr*) and *'name'*, which was used by transcribers to anonymise students when teachers had nominated one by name. We removed references to the physical surroundings of the classroom: *class*, *room*, *door*, *school*, to the social and temporal context: *shush*, *okay*, *please*, *er*, *today*, *minutes*, *lesson*, and to words in the titles or sub-headings of commercial publications: *crack*, *code*, *step*, and numbers. In the KS2 corpus, this led to us removing days of the week, numbers and months, which were frequent due to teachers dating materials. For the KS3 corpus, the procedure left 129 types, a few of which, such as *else*, *ever*, *yet*, *down* and *dog*, are not topic-specific, but do not meet our *New GSL* criterion for 'general'. For the KS2 corpus, the procedure left 126 topic-specific types. This seemed initially surprising, as the number of frequent types at the cut-off point of 32.6 per 100,000 in KS2 is around 75% more than that of KS3 (KS2, p. 363; KS2, p. 207). The KS2 corpus was found to have a much larger number of words in the top 200 of the *New GSL*, as well as words referring to the physical, social and temporal context. However, on reflection, we realised that this may be an indicator that the KS3 corpus is more specialised than KS2, which is not unexpected.

Following Baker et al.'s (2013) analysis, we then classified the remaining, topic-specific types into themes. We adapted a procedure used by metaphor researchers working with medium to large corpora (Cameron & Maslen, 2010; Deignan & Semino, 2010), using Microsoft Excel, which facilitates this kind of work (Cameron & Maslen, 2010). Each word under study was analysed in context, using concordance data, and notes were made. An extract of the analysis partway through is shown in Table 5.6.

We then assigned thematic labels to semantic and/or functional groupings, in a bottom-up, reiterative process. We did not use pre-assigned themes, but rather allowed these to emerge from our studies of the words in context.

### *Keyword analysis*

For our keyword studies, we used the Words tool in #LancsBox v.6 (Brezina et al., 2020), with its keywords facility. We set Words to calculate lemmas (that is, it grouped inflections of words together under the head word, so *sentence* and *sentences* were grouped together rather than appearing as separate entries). For the identification of keywords, Cohen's *d* was used because it takes dispersion into account (Brezina, 2014, 2018). We did this because dispersion statistics, as well as manual inspection, showed that the school corpora include some unevenly distributed words, associated with topics covered in a small number of lessons. If this was not taken into account, it could lead to artificially high counts for some words. For example, the words *iambic* and *pentameter* both occurred 43 times, as an

*Table 5.6* Extract of semantic and functional analysis-in-progress of frequent words in the KS3 English corpus.

| *Type* | *Normalised frequency per 100,000 words* | *Notes* | *Function* |
|---|---|---|---|
| use | 290.4 | Use of language, e.g., genre features, punctuation, tenses | Text analysis |
| think | 246.6 | Eliciting students' views. Reporting characters' feelings | S-elicit |
| write | 215.3 | Directions to students. Discussing how authors write. | S-direct |
| words | 180.1 | Analysing text with students | Text analysis |
| writing | 178.1 | Instructing and teaching students to do pieces of writing | S-direct |
| sentence | 150.7 | Instructions for students to do things with sentences; analysing texts | S-direct |
| word | 143.9 | Analysis, reasons for choice | Text analysis |
| know | 143.9 | Eliciting from students how they know/what characters know | S-elicit |
| gothic | 139.0 | Used to describe genre of text | Text, genre |
| key | 135.4 | Central terms = 'key vocabulary'. Also, occasionally, key stage | Text analysis |
| story | 133.7 | Analysing literature. Talking about genre (detective story) | Text, genre |
| work | 123.0 | Mostly used to talk about what students are writing | S-direct |
| read | 113.8 | Instruction to students | S-direct |
| poem | 113.5 | Analysing, commenting on poems. Students creating poems | Text analysis; S-instruct |
| create | 100.1 | Talk about text creating tension, fear etc., authors creating. Students creating | Text analysis; S-instruct |
| language | 94.6 | Classifying language into different genres, or by its effect – e.g., emotive. Language choices and features. | Text, genre |
| text | 93.6 | Analysing and evaluating texts | Text, analysis |
| explain | 93.3 | Close synonym with justify, in directions to students | S-instruct |

adjacent collocation. Because they are so infrequent in general English, they are ranked very highly on a raw keyword list. However, the 43 examples all occurred in multiple headers in two files: a presentation and a worksheet – associated with the same lesson. This means that they are unlikely to be characteristic of the language of school English lessons in Years 7 and 8 more broadly. Taking dispersion into account meant that words like this did not distort our overall picture of significant words in our corpora. As an additional check on dispersion, we again used $DP_{norm}$ and eliminated words that had a value of greater than 0.95.

*Table 5.7* Keyword studies.

| | *Focus corpus* | *Reference corpus* |
|---|---|---|
| Study 1 | KS3 English | KS2 English |
| Study 2 | KS2 English | KS3 English |
| Study 3 | KS3 English | BNCBM |

We conducted three keyword studies, as follows:

The raw output from the keywords procedure contained some words that we decided not to investigate further in the following groups.

- Grammatical words: *how*, *my*, *every* and other grammatical words emerged as key in comparison to KS2E due to minor genre differences. *My* for instance appears in target descriptors such as 'I develop both character and setting in my narrative writing'. *How*, *of what*, *why* and *your* are key in comparison to BNCBM, due to the number of direct questions in KS3 data, and the tendency to address students directly.
- We excluded the lemmas *be*, *have* and *do*, as concordance inspection showed that the majority of their uses were grammatical.
- Proper names: in KS3E, fictional names such as *Heathcliffe* and *Lennie* are frequent, although accounting for dispersion largely eliminated these in any case.
- Words associated with publications, worksheets and class management: The corpus includes multiple worksheets entitled 'Crack the Code', leading to *crack* and *code* being key in both comparisons. A number of worksheets include many examples of *step* when describing procedures.

Gabrielatos writes that keyness is a blunt tool because 'it does not cater for a host of linguistic features, most notably homography, polysemy, part of speech, multi-word units, and syntactic relations' (2018, p. 2). We noted in Chapter 2 and elsewhere that polysemy is an issue in academic language. Further, a close understanding of function and meaning is necessary for word lists to be of use to practitioners. We therefore followed the keyword analysis with qualitative analysis of concordance data.

## Results

### *Word frequency: Aboutness*

Table 5.8 shows the most frequent content words in the KS3 English corpus, with general words and words specific to context removed, as explained above. The frequency figure is normalised per 100,000 words.

Table 5.9 shows the corresponding words in the KS2 English corpus.

*Table 5.8* Most frequent topic-specific content words in the KS3 English corpus.

| *Rank* | *Type* | *Freq* | $DP_{norm}$ | *Rank* | *Type* | *Freq* | $DP_{norm}$ |
|---|---|---|---|---|---|---|---|
| 1 | write | 215.3 | 0.51 | 66 | form | 43.0 | 0.63 |
| 2 | words | 180.1 | 0.48 | 67 | british | 43.0 | 0.90 |
| 3 | writing | 178.1 | 0.63 | 68 | techniques | 42.7 | 0.84 |
| 4 | down | 157.9 | 0.38 | 69 | night | 42.4 | 0.60 |
| 5 | sentence | 150.7 | 0.67 | 70 | narrator | 42.0 | 0.83 |
| 6 | word | 143.9 | 0.54 | 71 | punctuation | 41.7 | 0.77 |
| 7 | gothic | 139.0 | 0.85 | 72 | description | 41.4 | 0.63 |
| 8 | key | 135.4 | 0.54 | 73 | mind | 41.4 | 0.51 |
| 9 | story | 133.7 | 0.68 | 74 | talk | 41.4 | 0.74 |
| 10 | read | 113.8 | 0.51 | 75 | identify | 41.1 | 0.61 |
| 11 | poem | 113.5 | 0.89 | 76 | sure | 40.5 | 0.50 |
| 12 | create | 1001 | 0.66 | 77 | wife | 40.1 | 0.69 |
| 13 | language | 94.6 | 0.69 | 78 | else | 39.4 | 0.55 |
| 14 | text | 93.6 | 0.68 | 79 | texts | 38.8 | 0.72 |
| 15 | explain | 93.3 | 0.60 | 80 | effective | 38.5 | 0.83 |
| 16 | effect | 92.0 | 0.70 | 81 | light | 38.2 | 0.61 |
| 17 | features | 90.0 | 0.73 | 82 | fear | 38.1 | 0.71 |
| 18 | reader | 89.0 | 0.73 | 83 | dream | 37.8 | 0.80 |
| 19 | understand | 83.8 | 0.64 | 84 | reading | 37.8 | 0.55 |
| 20 | ideas | 83.5 | 0.64 | 85 | quotation | 37.8 | 0.78 |
| 21 | range | 82.2 | 0.82 | 86 | white | 37.5 | 0.65 |
| 22 | example | 76.0 | 0.66 | 87 | yet | 37.5 | 0.58 |
| 23 | important | 72.7 | 0.66 | 88 | tension | 37.5 | 0.82 |
| 24 | question | 66.8 | 0.61 | 89 | dog | 37.2 | 0.66 |
| 25 | paragraph | 66.2 | 0.65 | 90 | structure | 37.2 | 0.75 |
| 26 | evidence | 66.4 | 0.73 | 91 | young | 37.2 | 0.51 |
| 27 | person | 64.6 | 0.57 | 92 | round | 36.9 | 0.60 |
| 28 | extract | 62.3 | 0.65 | 93 | head | 36.5 | 0.60 |
| 29 | setting | 62.0 | 0.79 | 94 | nothing | 36.2 | 0.58 |
| 30 | character | 61.6 | 0.65 | 95 | ever | 36.2 | 0.61 |
| 31 | questions | 60.6 | 0.70 | 96 | death | 35.8 | 0.62 |
| 32 | characters | 59.7 | 0.57 | 97 | quite | 35.8 | 0.57 |
| 33 | describe | 59.3 | 0.60 | 98 | full | 35.8 | 0.57 |
| 34 | vocabulary | 59.3 | 0.81 | 99 | lady | 32.5 | 0.67 |
| 35 | identity | 58.0 | 0.94 | 100 | consider | 35.2 | 0.62 |
| 36 | explore | 57.4 | 0.71 | 101 | hair | 35.2 | 0.66 |
| 37 | remember | 56.7 | 0.57 | 102 | poems | 35.2 | 0.91 |
| 38 | quote | 56.7 | 0.80 | 103 | hear | 35.2 | 0.50 |
| 39 | love | 56.4 | 0.72 | 104 | verb | 34.9 | 0.81 |
| 40 | chapter | 55.8 | 0.67 | 105 | learn | 34.9 | 0.65 |
| 41 | meaning | 55.1 | 0.68 | 106 | present | 34.6 | 0.60 |
| 42 | sentences | 54.8 | 0.70 | 107 | let | 34.6 | 0.59 |
| 43 | noun | 54.4 | 0.83 | 108 | types | 34.6 | 0.85 |
| 44 | audience | 54.4 | 0.83 | 109 | dark | 34.6 | 0.60 |
| 45 | writer | 54.4 | 0.75 | 110 | level | 34.2 | 0.82 |
| 46 | job | 53.2 | 0.86 | 111 | anything | 34.2 | 0.53 |
| 47 | short | 51.2 | 0.65 | 112 | purpose | 33.9 | 0.78 |
| 48 | piece | 51.2 | 0.70 | 113 | war | 33.9 | 0.89 |

(*Continued*)

*Table 5.8* (Continued)

| *Rank* | *Type* | *Freq* | $DP_{norm}$ | *Rank* | *Type* | *Freq* | $DP_{norm}$ |
|---|---|---|---|---|---|---|---|
| 49 | book | 50.9 | 0.61 | 114 | theme | 33.9 | 0.86 |
| 50 | written | 50.2 | 0.68 | 115 | kind | 33.6 | 0.54 |
| 51 | letter | 49.9 | 0.76 | 116 | speech | 33.6 | 0.81 |
| 52 | learning | 49.6 | 0.73 | 117 | skills | 33.6 | 0.79 |
| 53 | black | 48.3 | 0.61 | 118 | success | 33.6 | 0.79 |
| 54 | bit | 48.0 | 0.73 | 119 | interesting | 33.6 | 0.80 |
| 55 | able | 47.6 | 0.63 | 120 | heart | 33.6 | 0.60 |
| 56 | adjectives | 46.3 | 0.77 | 121 | simple | 33.6 | 0.85 |
| 57 | answer | 46.3 | 0.59 | 122 | monster | 33.6 | 0.86 |
| 58 | discuss | 46.0 | 0.67 | 123 | technique | 32.9 | 0.88 |
| 59 | eyes | 46.0 | 0.57 | 124 | someone | 32.9 | 0.64 |
| 60 | face | 45.4 | 0.58 | 125 | challenge | 32.9 | 0.71 |
| 61 | horror | 45.0 | 0.75 | 126 | towards | 32.9 | 0.54 |
| 62 | information | 44.7 | 0.75 | 127 | marks | 32.9 | 0.76 |
| 63 | books | 44.0 | 0.56 | 128 | least | 32.6 | 0.55 |
| 64 | goal | 44.0 | 0.92 | 129 | poetry | 32.6 | 0.92 |
| 65 | list | 39.1 | 0.65 | | | | |

For each of these lists, we grouped the words into themes, as described in the previous section, using Microsoft Excel. Our corpus contains a number of instructions to students, issued by teachers verbally, and in written form on worksheets and presentations and in textbooks. This means that speech acts are frequent, as can be seen in Table 5.6, and meant that we had to be very aware of function as well as denotation in classifying the words. We studied the concordances for each of the types listed, and identified a number of meanings and functions, as follows.

In the KS3 corpus, a number of words are used to organise and direct teaching and learning. These include verbs such as *write*, *read*, *describe*, *explain* and *list*, which direct the students to do something, sometimes as part of assessment, and nouns such as *sentence* and *paragraph*, which occur in instructions to 'write a sentence' or 'a paragraph'. A few words are associated with learning and the curriculum more generally, such as *level*, *teacher* and *learn*. *British* is in the list largely because of the 2014 direction from the government to 'promote British values', which include 'appreciation that living under the rule of law protects individual citizens and is essential for their wellbeing and safety' and religious tolerance (DfE, 2014, p. 5). A further, small group of words are linguistic terms and used in language analysis and error correction. These are *punctuation*, *noun*, *adjective(s)*, *verb*, [quotation] *marks*, *vocabulary* and *form*.

The majority of words on the list are associated with the analysis of text in some way. Some are the central vocabulary of text analysis: *text*, *character(s)*, *set*, *setting*. Some refer to major themes or strong emotions in literature, such as *war*, *death*, *fear*, *identity*, *love* and *horror*; and *theme*

*Table 5.9* Most frequent topic-specific content words in the KS2 English corpus.

| *Rank* | *Type* | *Freq* | $DP_{norm}$ | *Rank* | *Type* | *Freq* | $DP_{norm}$ |
|---|---|---|---|---|---|---|---|
| 1 | word | 372.6 | 0.52 | 64 | reader | 50.5 | 0.75 |
| 2 | sentence | 357.5 | 0.55 | 65 | letters | 50.0 | 0.67 |
| 3 | words | 339.0 | 0.53 | 66 | reading | 49.7 | 0.70 |
| 4 | mark | 284.0 | 0.75 | 67 | adverb | 49.0 | 0.78 |
| 5 | write | 245.1 | 0.54 | 68 | sense | 48.7 | 0.67 |
| 6 | text | 203.3 | 0.65 | 69 | phrases | 48.7 | 0.75 |
| 7 | sentences | 201.5 | 0.63 | 70 | nouns | 46.8 | 0.74 |
| 8 | read | 167.1 | 0.54 | 71 | correctly | 46.8 | 0.73 |
| 9 | verb | 135.1 | 0.69 | 72 | stop | 46.3 | 0.58 |
| 10 | correct | 124.5 | 0.66 | 73 | important | 46.0 | 0.65 |
| 11 | clause | 124.0 | 0.72 | 74 | table | 45.5 | 0.65 |
| 12 | correct | 124.5 | 0.66 | 75 | least | 44.7 | 0.80 |
| 13 | clause | 124.0 | 0.72 | 76 | present | 44.4 | 0.76 |
| 14 | right | 121.1 | 0.59 | 77 | grammar | 44.4 | 0.82 |
| 15 | paragraph | 111.9 | 0.67 | 78 | form | 44.2 | 0.73 |
| 16 | answer | 111.9 | 0.69 | 79 | adjectives | 43.9 | 0.76 |
| 17 | writing | 105.2 | 0.58 | 80 | line | 42.8 | 0.65 |
| 18 | noun | 103.4 | 0.66 | 81 | copy | 41.5 | 0.78 |
| 19 | question | 103.1 | 0.61 | 82 | english | 41.5 | 0.73 |
| 20 | information | 101.3 | 0.68 | 83 | missing | 41.5 | 0.72 |
| 21 | spelling | 100.7 | 0.71 | 84 | someone | 41.3 | 0.65 |
| 22 | relative | 98.1 | 0.78 | 85 | together | 41.3 | 0.68 |
| 23 | down | 97.8 | 0.51 | 86 | quite | 40.2 | 0.64 |
| 24 | evidence | 90.2 | 0.79 | 87 | play | 39.9 | 0.66 |
| 25 | punctuation | 86.7 | 0.73 | 88 | create | 39.7 | 0.81 |
| 26 | marks | 86.2 | 0.71 | 89 | pronoun | 39.7 | 0.77 |
| 27 | add | 82.5 | 0.65 | 90 | head | 39.4 | 0.65 |
| 28 | bit | 82.2 | 0.62 | 91 | quick | 39.1 | 0.77 |
| 29 | clauses | 80.1 | 0.79 | 92 | both | 39.1 | 0.62 |
| 30 | box | 78.8 | 0.67 | 93 | check | 39.1 | 0.65 |
| 31 | sure | 75.9 | 0.51 | 94 | comma | 38.9 | 0.71 |
| 32 | explain | 74.6 | 0.69 | 95 | yourself | 38.9 | 0.67 |
| 33 | describe | 74.0 | 0.75 | 96 | let's | 38.3 | 0.73 |
| 34 | main | 72.5 | 0.69 | 97 | dog | 38.1 | 0.74 |
| 35 | commas | 71.7 | 0.72 | 98 | under | 38.1 | 0.65 |
| 36 | able | 71.4 | 0.70 | 99 | piece | 37.3 | 0.68 |
| 37 | test | 71.4 | 0.82 | 100 | extra | 37.3 | 0.73 |
| 38 | section | 67.2 | 0.75 | 101 | conjunctions | 37.3 | 0.75 |
| 39 | example | 66.1 | 0.65 | 102 | identify | 36.0 | 0.87 |
| 40 | book | 64.8 | 0.64 | 103 | subject | 36.0 | 0.76 |
| 41 | speech | 62.9 | 0.76 | 104 | else | 35.4 | 0.69 |
| 42 | past | 62.9 | 0.68 | 105 | water | 35.4 | 0.68 |
| 43 | adjective | 60.8 | 0.75 | 106 | description | 35.2 | 0.83 |
| 44 | remember | 60.0 | 0.58 | 107 | written | 34.9 | 0.65 |
| 45 | circle | 59.0 | 0.74 | 108 | family | 34.6 | 0.72 |
| 46 | vocabulary | 59.0 | 0.84 | 109 | eat | 34.4 | 0.66 |
| 47 | letter | 59.0 | 0.68 | 110 | similar | 34.4 | 0.76 |
| 48 | complete | 59.0 | 0.74 | 111 | eyes | 34.4 | 0.72 |
| 49 | verbs | 58.7 | 0.69 | 112 | learn | 34.4 | 0.73 |

(*Continued*)

*Table 5.9* (Continued)

| *Rank* | *Type* | *Freq* | $DP_{norm}$ | *Rank* | *Type* | *Freq* | $DP_{norm}$ |
|---|---|---|---|---|---|---|---|
| 50 | story | 58.7 | 0.72 | 113 | often | 34.4 | 0.65 |
| 51 | answers | 58.2 | 0.70 | 114 | examples | 34.4 | 0.75 |
| 52 | sound | 56.8 | 0.67 | 115 | formal | 33.8 | 0.73 |
| 53 | full | 56.1 | 0.59 | 116 | night | 33.6 | 0.75 |
| 54 | person | 56.1 | 0.61 | 117 | big | 33.6 | 0,65 |
| 55 | character | 56.1 | 0.80 | 118 | pronouns | 33.3 | 0.79 |
| 56 | tense | 55.3 | 0.74 | 119 | match | 33.3 | 0.76 |
| 57 | phrase | 55.3 | 0.68 | 120 | draw | 32.8 | 0.76 |
| 58 | list | 53.4 | 0.71 | 121 | retrieve | 32.8 | 0.91 |
| 59 | language | 52.6 | 0.80 | 122 | rules | 32.8 | 0.84 |
| 60 | change | 52.1 | 0.63 | 123 | root | 32.5 | 0.81 |
| 61 | version | 51.8 | 0.90 | 124 | subordinate | 32.5 | 0.79 |
| 62 | meaning | 51.6 | 0.68 | 125 | conjunction | 32.5 | 0.73 |
| 63 | author | 51.6 | 0.88 | 126 | ending | 32.5 | 0.84 |

itself; *heart* is a peripheral member of this group as it tends to stand for feelings. There is a small group associated with the head: *head*, *eyes*, *face*. Concordance analysis shows that these are used in the extracts of literary texts that students read, often as indicators of characters' emotions. Some frequent words refer to genre or are used in genre analysis discussions; some examples of *horror* overlap with this group. The 'genre' group includes *features*, *poems*, *poetry*, *purpose*, *language*, *story*, *gothic*. A large, related group is used in discussion of the effect that writing is intended to have on the reader and how this is achieved. This includes *create*, *tension*, *reader*, *effect*, *audience*, *effective*, *technique(s)*, *language*. A few words are associated with prompts to support students inferencing about characters and events in texts. These include some uses of *learn*:

(1)

> What do we *learn* about this narrator? What clues are given about the content of the story that follows?

*Mean* and *know*, which are not in our main word list because they are in the 200 *New GSL*, are also used in this way:

(2)

> he's poisoned with guilt what a fantastic way of putting it so (.) it could *mean* that he's full of evil thoughts that he wants to do more and more evil things like kill Banquo or it could *mean* that he can't help thinking about guilty terrible things [...] it could *mean* both.
>
> How is Bruno feeling at the end of the chapter? How do you *know*?

What has happened to the man? How do you *know*? How could the poem's title be ambiguous?

Learning to draw inferences is a very prominent objective in the KS3 curriculum, but we found relatively little lexical evidence for activities to support it in our data.

This analysis is obviously subjective, but it was supported with detailed concordance analysis. For instance, we classified *create* in the 'effect' group on the basis of concordance lines such as:

(3)

Are there any words or phrases that help to *create* tension and suspense?
Write some notes to explain how they help to *create* a scary mood.

Many of the words could, potentially, have multiple uses, and a few do. For instance, we classified *language* as belonging to both the 'genre' and 'effect' groups after analysing concordance lines such as:

(4)

… the *language* and structural features of a formal letter. (genre)
… using *language* to create an effect on a target audience. (effect)

*Speech* is used both in analysing the effectiveness of speeches by public figures and in phrases such as *reported speech* discussing grammar.

In the KS2 corpus, as was the case for the KS3 corpus, there is a large group of words associated with assessment. These include *correct*, *answer*, *explain*. The group also includes *circle* and *match*, in concordance lines such as:

(5)

*Circle* the four prepositions in the sentence below.
*Draw* a line to match each word to the correct suffix.

More of the KS2 corpus is concerned with assessment preparation than the KS3 corpus, which is not unexpected given the high-stakes KS2 SATs. We found it striking nonetheless to see what a large proportion of our corpus texts are concerned with preparing for assessment and practising assessment tasks.

We also found a large group of types concerned with grammatical description. Among these are *verb*, *noun*, *adjective*, *relative*, *pronoun*, *clause*, *past* [tense], *used* and *word* [class]. A closely related group are concerned with punctuation and spelling, including *punctuation*, *comma*, *speech* [marks], *spelling*, *ending* and *sound*. The last two words appear in spelling rules such as:

(6)

This week we are looking at words *ending* in -ible. It is often easy to confuse words that *end* in -ible and -able.

> The 'ie' spelling pattern is used most of the time if those letters make the *sound* 'igh' or 'ee' in a word.

A smaller number of types are concerned with reading literature and writing creatively, including *meaning*, *reader* and *author*.

If these frequent words do give an indication of 'aboutness', as argued by Baker et al. (2013), then we have found that both corpora are about assessment, while KS2 is also about grammar, punctuation and spelling. KS3 is also partly about language analysis, but much more about reading and writing for effect and across a range of genres.

### *Keywords*

The following tables give the most key lemmas, excluding the groups listed in the methodology section, for each of the three key word analyses. The second column gives the frequency, normalised to per thousand words. The third column gives the Cohen's *d* statistic. Keywords showing Cohen's *d* below 0.2 were not included in our study, as the effect size is considered too low. Cohen (1988, p. 40) offers definitions of effect size as: small: 0.2; medium: 0.5; large, 0.8. Table 5.10 shows the 30 most key lemmas in KS3 English, using the KS2 English corpus as a reference corpus, that is, the lemmas encountered significantly more frequently in Years 7 and 8 than in Years 5 and 6.

Table 5.11 shows the 30 most key lemmas in KS3 English, using the BNCBM as the reference corpus. In other words, these are words that are encountered much more frequently in KS3 than in everyday life outside school. There is some overlap between these and the keywords listed in Table 5.10, lemmas that are key in relation to KS2. However, it can be seen from the Cohen's *d* column that the effect sizes are much greater than in the comparisons between KS3 and KS2. That is, word choice and frequency in KS3 are much more like that of KS2 than that of everyday English. Both comparisons are relevant for students starting KS3.

The keywords procedure is usually considered a first step, giving indications as to where to look in more detail (Culpeper & Demmen, 2015; Baker, 2004). In the next step, we analysed concordance data for each of lemmas listed in Tables 5.10 and 5.11. That is, we studied concordance data in all three corpora for all lemmas that are key in the KS3 English corpus in relation to the KS2 English corpus, and to general English as represented by the BNCBM. The KS3 English and KS2 English corpora were small enough for us to analyse every concordance line. In the larger BNCBM, where there were more than 200 examples of a lemma, we analysed a random sample of 200. None of the keywords that we studied are completely unique to any of the corpora. However, we found that all of them differ in use and function across the corpora, often markedly. We do

*Table 5.10* Focus corpus KS3 English: Reference corpus KS2 English (ranked by Cohen's *d*).

| *Keyness rank* | *Lemma* | *Freq/1000* | *Cohen's* ***d*** | $DP_{norm}$ |
|---|---|---|---|---|
| 1 | create | 1.57 | 0.49 | 0.62 |
| 2 | explore | 0.75 | 0.48 | 0.67 |
| 3 | gothic | 1.24 | 0.45 | 0.85 |
| 4 | effect | 1.09 | 0.43 | 0.68 |
| 5 | feature | 0.92 | 0.38 | 0.73 |
| 6 | meet | 0.42 | 0.37 | 0.65 |
| 7 | writing (n) | 0.72 | 0.34 | 0.74 |
| 8 | technique | 0.76 | 0.34 | 0.80 |
| 9 | quotation | 0.57 | 0.32 | 0.73 |
| 10 | important | 0.73 | 0.32 | 0.66 |
| 11 | poem | 1.49 | 0.32 | 0.89 |
| 12 | work | 0.96 | 0.32 | 0.56 |
| 13 | audience | 0.57 | 0.32 | 0.83 |
| 14 | fear | 0.37 | 0.31 | 0.70 |
| 15 | castle* | 0.22 | 0.31 | 0.91 |
| 16 | develop | 0.39 | 0.31 | 0.72 |
| 17 | tension | 0.36 | 0.31 | 0.89 |
| 18 | key | 1.21 | 0.31 | 0.85 |
| 19 | supernatural | 0.29 | 0.3 | 0.85 |
| 20 | structure (n) | 0.36 | 0.3 | 0.76 |
| 21 | comment (v) | 0.21 | 0.29 | 0.85 |
| 22 | structure (v) | 0.11 | 0.29 | 0.88 |
| 23 | creative | 0.26 | 0.29 | 0.87 |
| 24 | focus (v) | 0.17 | 0.29 | 0.75 |
| 25 | writer | 0.68 | 0.28 | 0.73 |
| 26 | idea | 1.08 | 0.28 | 0.76 |
| 27 | historical | 0.10 | 0.27 | 0.83 |
| 28 | poetry | 0.33 | 0.27 | 0.92 |
| 29 | analysis | 0.32 | 0.26 | 0.83 |
| 30 | tone | 0.16 | 0.26 | 0.65 |

**castle* occurs frequently in both 'Macbeth' and 'Dracula', and is in the title of the Gothic novel 'The Castle of Otranto', which is discussed. Other words that occur in specific novels tend to be filtered out by the dispersion measure, but because *castle* occurs in a number of different sources, this did not happen.

not have space for a full account of our findings for each word analysed, so we describe and illustrate the patterns of variation. We describe four patterns, and then a general, widespread shift in meaning. There is overlap between the patterns, that is, they are tendencies rather than clearcut, exclusive categories. In the following, we give brief definitions of the uses we found in each corpus, with illustrative examples. The corpus examples are unedited, except where indicated.

(1) In the first pattern we illustrate, the main meanings or contextual uses of a lemma are different in one or more corpora. Typically, the KS3 and KS2 meanings are more clearly related, and the everyday meaning less

*Table 5.11* Focus corpus KS3 English: Reference corpus BNCBM (ranked by Cohen's *d*).

| *Keyness rank* | *Lemma* | *Freq/1000* | *Cohen's* ***d*** | $DP_{norm}$ |
|---|---|---|---|---|
| 1 | write | 3.6 | 0.96 | 0.48 |
| 2 | use | 4.9 | 0.93 | 0.91 |
| 3 | word | 3.2 | 0.78 | 0.46 |
| 4 | create | 1.57 | 0.7 | 0.62 |
| 5 | reader | 0.99 | 0.62 | 0.70 |
| 6 | text | 1.32 | 0.58 | 0.65 |
| 7 | language | 0.95 | 0.56 | 0.69 |
| 8 | describe | 1.05 | 0.53 | 0.52 |
| 9 | explore | 0.75 | 0.53 | 0.69 |
| 10 | effect | 1.09 | 0.53 | 0.82 |
| 11 | sentence | 2.04 | 0.5 | 0.64 |
| 12 | explain | 1.11 | 0.5 | 0.55 |
| 13 | different | 1 | 0.5 | 0.79 |
| 14 | key | 1.21 | 0.48 | 0.71 |
| 15 | gothic | 1.24 | 0.43 | 0.85 |
| 16 | setting | 1.47 | 0.47 | 0.76 |
| 17 | writing (n) | 0.72 | 0.45 | 0.74 |
| 18 | feature | 0.92 | 0.45 | 0.73 |
| 19 | story | 1.34 | 0.45 | 0.75 |
| 20 | character | 1.2 | 0.45 | 0.56 |
| 21 | show | 1.28 | 0.45 | 0.44 |
| 22 | follow | 0.73 | 0.43 | 0.52 |
| 23 | structure (n) | 0.36 | 0.43 | 0.76 |
| 24 | evidence | 0.63 | 0.43 | 0.74 |
| 25 | description | 0.5 | 0.43 | 0.62 |
| 26 | choose | 0.6 | 0.42 | 0.66 |
| 27 | analyse | 0.47 | 0.42 | 0.80 |
| 28 | annotate | 0.21 | 0.41 | 0.83 |
| 29 | technique | 0.76 | 0.41 | 0.80 |
| 30 | writer | 0.68 | 0.41 | 0.73 |

so. Lemmas in this group include the following: *explain*, *evidence*, *setting*, *feature*, *device*, *audience*, *show*.

| | ***Explain*** |
|---|---|
| KS3 | *Explain* is used in instructions and assessment rubrics, directing students to justify their ideas.<br>'Write a comparative paragraph *explaining* the different uses of language in each of the […]'<br>'be ready to *explain* your choices' |
| KS2 | *Explain* is used in a similar way to KS3.<br>'*Explain* two features of her character using evidence'. |
| BNCBM | *Explain* is most frequently used to report direct speech or paraphrase speech.<br>'It's made from the seeds', he *explained*, 'the most poisonous part of the plant'. |

| | ***Evidence*** |
|---|---|
| KS3 | *Evidence* refers to quotations and paraphrases from literature to support a point a student is making in their analysis.<br>'You can't do the analysis without the *evidence*'<br>'… ensure we use relevant textual *evidence* to support our points' |
| KS2 | *Evidence* is used in a similar way as in KS3, but almost entirely in written assessment rubrics.<br>'Discuss two aspects of her character using *evidence* to support your answer'. |
| BNCBM | The most frequent, and probably the most salient use is the criminal sense, collocating with words such as *forensic* and *against*.<br>'He was giving *evidence* at the trial of his uncle'. |

| | ***Setting*** |
|---|---|
| KS3 | *Setting* is used as a specialised literary and theatre term, as follows:<br>'… the use of *setting* within a Gothic horror story'<br>'… describe the *setting* and atmosphere' |
| KS2 | The KS3 use occurs rarely but has a similar meaning.<br>'you should have described the *setting* of the story so far'. |
| BNCBM | In everyday language, the main use is to describe a level on a mechanical device. There is a less frequent meaning resembling the academic use, but this tends to belong to the estate agent or travel writing genres.<br>'It uses less battery if it's on its lowest *setting*'.<br>'… in a beautiful studio *setting*'. |

| | ***Feature*** |
|---|---|
| KS3 | *Feature* refers to a characteristic of a genre<br>'Detective fiction shares some similar *features* with adventure stories'. |
| KS2 | *Feature* is less frequent but has a similar meaning to the KS3 one.<br>'… check you have included all the *features* of suspense'.<br>'… identify some *features* of a newspaper text'. |
| BNCBM | *Features* are characteristics or components of a product, or parts of the human face. Occasionally, it refers to part of a newspaper or magazine.<br>'... *features* like aperture, exposure, focus, special photographic effects'.<br>'He had neat *features* and he dressed well'. '… a *features* editor'. |

| | ***Device*** |
|---|---|
| KS3 | A *device* is a type of phrase or trope that is used to achieve a particular effect.<br>'… key poetic *devices*'.<br>'… rhetorical questions we know are persuasive *devices*'.<br>'… language *devices*'. |
| KS2 | *Device* is much less frequent in KS2 data, and examples mention cohesive, and descriptive devices, but not language or poetic ones.<br>'… a range of cohesive *devices*' |
| BNCBM | In the everyday corpus, a *device* is a small machine, especially a mobile phone.<br>'taking a small recording *device* from her pocket'<br>'… if you've logged into the iCloud on your *device*'. |

(2) In the next pattern, the KS3 meaning is also found in KS2, but is much less common, and is very infrequent in BNCBM, at least with its academic meaning. In KS2, our examples may represent just a few schools or teachers, or simply be rare. This means that many students might not encounter this meaning, or the word at all, until KS3. Lemmas with this pattern include: *effect*, *technique*, *gothic*, *annotate*, *analyse*, *theme*.

| | ***Effect*** |
|---|---|
| KS3 | *Effect* occurs in around a quarter of texts in the KS3 corpus. It denotes the reactions or emotions created in the reader by a writer's literary choices.<br>'What is the *effect* of the change to third person'.<br>'comment on the *effect* on the audience/ reader' |
| KS2 | *Effect* is used in 5% of texts in the KS2 corpus. There are four examples resembling the KS3 meaning.<br>'what is the *effect* of the word swarms in this sentence?' |
| BNCBM | *Effect* most typically occurs in collocations such as *sound effect*; *side effect*; *special effects*; *dramatic effect*; *have the opposite effect*. The KS3 use is very infrequent. |

| | ***Technique*** |
|---|---|
| KS3 | *Technique* occurs in nearly 20% of texts. It collocates with words such as *language*, *persuasive* and *writing*, referring to a practised and expert way of writing.<br>'You may want to refer to the writing *techniques* he has used'.<br>'These are known as imagery *techniques*'.<br>'… typical of the genre advert because it uses a persuasive *technique*'. |
| KS2 | Technique occurs in only 2.5% of texts.<br>'remember to use persuasive *technique*'. |
| BNCBM | *Technique* collocates with a wide range of words, including words referring to art, sport, parenting, spying and many other activities. No concordance lines refer to writing. |

*Gothic* is unusual here, in that it has a technical meaning in literature, and would be considered Tier 3 in Beck et al.'s (2002) classification (as discussed in Chapter 2).

| | ***Gothic*** |
|---|---|
| KS3 | All examples refer to a literary genre.<br>'Dracula is a *gothic* story'.<br>'The term *Gothic* fiction refers to a style of writing that is characterised by elements of fear, horror, death and gloom'. |
| KS2 | There is one example, referring to art.<br>'… *Gothic* art'. |
| BNCBM | *Gothic* is infrequent and most concordance lines refer to architecture. |

It is clear from the KS3 examples of *gothic* that teachers and materials writers recognised that the term will be new, as they define it carefully.

(3) In the third pattern we found, the KS3 meaning is apparently the same, but more nuanced, subtle, or specific than the KS2 and BNCBM meaning. This group includes: *language*, *discuss*, *explore*, *create*.

| | ***Language*** |
|---|---|
| KS3 | *Language* is a tool that writers use to create effects and interact with readers.<br>'Consider the effect that this *language* would have'.<br>'How does the writer use *language* to …?'<br>'By the end of the extract, Wells' use of *language* becomes more intense'. |
| KS2 | In KS2, the main use is to refer to different types of language, such as formal, informal and figurative.<br>'Formal *language* often uses longer words' |
| BNCBM | The most frequent use is to refer to different *languages*: English, French, Chinese etc. |

| | ***Discuss*** |
|---|---|
| KS3 | *Discuss* is used frequently in assessment rubrics or tasks, where it means to describe, compare, contrast and evaluate in a formal way, in writing. It is very occasionally used to direct students to speak, again in a formal academic way.<br>'*Discuss* how Charles Dickens creates tension and suspense'<br>'*Discuss* the roles and relative importance of the woman, the girl and the boy in this passage'<br>'*Discuss* with the person next to you what you think is meant by authorial intent' |
| KS2 | *Discuss* usually refers to oral, possibly less structured discourse.<br>'Which words were you *discussing* there?'<br>'We're just *discussing* being a young person'. |
| BNCBM | In everyday language, *discuss* does not usually refer to writing. It sometimes implies formality, and interpersonal tension.<br>'to *discuss* terms and conditions',<br>'… meet to *discuss* their concerns about CCTV'<br>'… a matter of some urgency that I must *discuss* with you'. |

| | ***Explore*** |
|---|---|
| KS3 | The object of *explore* tends to be abstract. The subject can be animate, usually a writer, or a text, such as a poem.<br>'Today we are going to *explore* the context of this novel'<br>'How did the poems *explore* the theme of identity?' |
| KS2 | There are few examples of *explore* in KS2 data. The subject is usually animate, and the meaning is concrete.<br>'[He] *explored* the empty room in the farmhouse' |
| BNCBM | In general language, *explore* is much less frequent and tends to be concrete.<br>'We hoped to hire a boat to *explore* all those tiny beaches'. |

(4) In some cases, the collocates are different, leading to a different aspect of meaning being emphasised in each corpus. Keywords in this group include: *develop*, *writer*, *vocabulary*, *supernatural*, *character*, *writing*.

| | ***Develop*** |
|---|---|
| KS3 | *Develop* is used to talk about how a piece of writing is crafted, and about students' becoming proficient in new skills.<br>'Take one of our ideas and *develop* a fabulous paragraph describing that particular feature…' |
| KS2 | In KS2 the use of *develop* is more general, shown by a wider range of collocates.<br>'Camels have had to *develop* special characteristics to survive in these challenging conditions'<br>'This could be achieved by *developing* cleaner fuels and electrically powered cars'. |
| BNCBM | In the general corpus, *develop* is used much more widely, about a range of entities. Collocates include *cancer*, *resistance*, *product*, *talent*, *digital technology*. |

| | ***Writer*** |
|---|---|
| KS3 | Collocates include *create*, *tension*, *suspense* and *choose*, and concordance lines show discussion of how writers create/ craft/ use language/ choose/ or want the reader to feel.<br>'How do *writers* create tension and suspense in a story?' |
| KS2 | *Writer* is much less frequent, and is often used to refer to their lives rather than their choices. The rare mentions of authorial intentions are much less sophisticated.<br>'a jealous *writer* named Roger Green wrote…',<br>'a group of British *writers* started the detection club'.<br>'What is this *writer* trying to say?' |
| BNCBM | *Writer* is mostly used to name someone's job. This is sometimes specified through collocations such as *script writer*, *cookery writer*, *crime writer*. |

Finally, there is also a general shift in linguistic metalanguage, from describing sentence grammar in KS2, towards describing text-level phenomena in KS3. This is seen in words such as: *structure*, *use*, *word*.

| | ***Structure*** |
|---|---|
| KS3 | *Structure* is found in analysis of sentences and texts.<br>'look at sentence *structure*, repetition and sound techniques'<br>'… questions to ask about a text when analysing language or *structure*' |
| KS2 | In KS2, *structure* is much less frequent, and tends to refer to grammar.<br>'… grammar *structures*'. '… the subjunctive verb *structure*'. |
| BNCBM | In the everyday corpus, *structure* is much less frequent than in KS3 and meanings are mostly literal. A few refer to organisations.<br>'bone *structure*', 'a massive wooden *structure*'<br>'departmental *structures*' |

| | ***Use*** |
|---|---|
| KS3 | *Use* describes language choices in terms of tools for particular effects.<br>'Each stanza *uses* a different image to explain feelings'.<br>'Can you think of a way to *use* emotional language?' |
| KS2 | *Use* tends to collocate with words for punctuation or sentence grammar features.<br>'*Use* a comma to separate the clauses'. |
| BNCBM | *Use* tends to have a concrete sense, and collocates with a wide range of words denoting objects.<br>'I'm happy with this bag, it's easy to *use*'.<br>'By *using* public transport you'll save money'. |

These findings are consistent with the frequency-based discussion of 'aboutness' above. They suggest that in formal terms, English shifts from a focus on sentence grammar to a focus on text analysis in KS3, as would be expected from the National Curriculum goals. Perhaps confusingly for students, some of the same lexis is used, but in different ways. There is another shift, from the KS2 perspective of comprehending a text as if it had a single meaning, to seeing it as the result of a writer-audience interaction. The concordance analyses show the increased precision and subtlety with which many semi-technical words are used in KS3. They also bring out the KS3 themes of effect and genre, and as for the shift in form focus from sentence grammar to text structure, this is sometimes handled using the same lexis as in KS2, used with different emphasis and more precise meaning.

## Conclusion

In this chapter, we have discussed the English curriculum in Key Stages 2 and 3. We looked at the central themes across the curricula, and how these

shift. We then used the corpus tools of word frequency lists, keywords and concordance analysis to establish the aboutness of each corpus, and what is different about KS3 as opposed to KS2 and everyday language. Qualitative analysis explored what lies behind the quantitative findings. By and large, the findings described here suggest that major issues for students are likely to be increased nuance, subtlety and precision of words that both they and their teachers may think they already know. This issue is within the broader backdrop of a change in the goals of the discipline, which, judging from our corpus data, is not always articulated explicitly to students.

## Note

1 We are indebted to Doğuş Öksüz for his input on modifying the BNC2014 Baby+ for this study.

## References

Baker, P. (2004). Querying keywords: Questions of difference, frequency and sense in keyword analysis. *Journal of English Linguistics*, *32*(4), 346–359. https://doi.org/10.1177/0075424204269894

Baker, P. (2006). *Using corpora in discourse analysis*. Continuum.

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

Barrs, M. (2019). Teaching bad writing. *English in Education*, *53*(1), 18–31. https://doi.org/10.1080/04250494.2018.1557858

Beck, I. L., Mckeown, M. G., & Kucan, L. (2002). *Bringing words to life: Robust vocabulary instruction* (2nd ed.). Guilford Press.

Bell, H. (2016). Teacher knowledge and beliefs about grammar: A case study of an English primary school. *English in Education*, *50*(2), 148–163. https://doi.org/10.1111/eie.12100

Brezina, V. (2014, May 4). *Effect sizes in corpus linguistics: Keywords, collocations and diachronic comparisons* [Paper presentation] ICAME 2014, University of Nottingham, United Kingdom.

Brezina, V. (2018). *Statistics in corpus linguistics: A practical guide*. Cambridge University Press.

Brezina, V., & Gablasova, D. (2015). Is there a core general vocabulary? Introducing the new general service list. *Applied Linguistics*, *36*(1), 1–22. https:doi.org/10.1093/applin/amt018

Brezina, V., Weill-Tessier, P., & McEnery, A. (2020). *#LancsBox v. 6.x* [software package].

Cameron, L., & Maslen, R. (2010). Identifying metaphors in discourse data. In Cameron, L., & Maslen, R. (Eds.), *Metaphor analysis: Research practice in applied linguistics, social sciences and the humanities* (pp. 97–115). Equinox.

CASS (Corpus Approaches to the Social Sciences Centre, Lancaster University). (n.d.). *The BNC2014 Baby+*. http://corpora.lancs.ac.uk/lancsbox/docs/pdf/BNC2014Baby.pdf

Clancy, B. (2010). Building a corpus to represent a variety of language. In O'Keeffe, A., & McCarthy, M. (Eds.), *Routledge handbook of corpus linguistics* (pp. 80–92). Routledge.

Cohen, J. (1988). *Statistical power analysis for the behavioural sciences*. Lawrence Erlbaum Associates.

Coleman, R. (2017). *EEF blog: Literacy at the transition – A research summary for teachers*. https://educationendowmentfoundation.org.uk/news/eef-blog-literacy-at-the-transition-a-research-summary-for-teachers

Cremin, T. (2015). Imaginatively exploring fiction. In Cremin, T., with Reedy, D., Bearne, E. & Dombey, H. (Eds.), *Teaching English creatively* (pp. 117–133). Routledge. https://doi.org/10.4324/9781315766904

Cremin, T., Mottram, M., Bearne, E., & Goodwin, P. (2008). Exploring teachers' knowledge of children's literature. *Cambridge Journal of Education*, *38*(4), 449–464. https://doi.org/10.1080/03057640802482363

Cremin, T., & Swann, J. (2016). Literature in common: Reading for pleasure in school reading groups. In Rothbauer, P., Skjerdingstad, K., McKechnie, L., & Oterholm, K. (Eds.), *Plotting the reading experience: Theory/practice/politics* (pp. 279–300). Wilfrid Laurier University Press.

Culpeper, J., & Demmen, J. (2015). Keywords. In Biber, D., & Reppen, R. (Eds.), *The Cambridge handbook of corpus linguistics* (pp. 90–105). Cambridge University Press.

Cushing, I. (2018a). Grammar policy and pedagogy from primary to secondary school. *Literacy*, *53*(3), 170–179. https://doi.org/10.1111/lit.12170

Cushing, I. (2018b). Stylistics goes to school. *Language and Literature: International Journal of Stylistics*, *27*(4), 271–285. https://doi.org/10.1177/0963947018794093

Cushing, I. (2019). Prescriptivism, linguicism and pedagogical coercion in primary school UK curriculum policy. *English Teaching: Practice & Critique*, *19*(1), 35–47. https://doi.org/10.1108/etpc-05-2019-0063

Cushing, I., & Helks, M. (2021). Exploring primary and secondary students' experiences of grammar teaching and testing in England. *English in Education*, 1–12. https://doi.org/10.1080/04250494.2021.1898282

Deignan, A., & Semino, E. (2010). Corpus techniques for metaphor analysis. In Cameron, L., & Maslen, R. (Eds.), *Metaphor analysis: Research practice in applied linguistics, social sciences and the humanities* (pp. 161–179). Equinox.

Department for Education (DfE). (2013a). *English programmes of study: Key stages 1 and 2. National curriculum in England*. https://www.gov.uk/government/publications/national-curriculum-in-england-english-programmes-of-study

Department for Education (DfE). (2013b). *English programmes of study: Key stage 3*. National curriculum in England. https://www.gov.uk/government/publications/national-curriculum-in-england-english-programmes-of-study

Department for Education (DfE). (2014). *Promoting fundamental British values as part of SMSC in schools: Departmental advice for maintained schools*. https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/380595/SMSC_Guidance_Maintained_Schools.pdf

Durran, J. (2017). *Avoiding a "literacy dip" in year 7*. https://jamesdurran.blog/2017/08/23/avoiding-a-literacy-dip-in-year-7/

Durrant, P., & Brenchley, M. (2019). Development of vocabulary sophistication across genres in English children's writing. *Reading and Writing*, *32*, 1927–1953. https://doi.org/10.1007/s11145-018-9932-8

Friginal, E., & Hardy, J. A. (2020). Corpus approaches to discourse analysis. In Friginal, E., & Hardy, J. A. (Eds.), *The Routledge handbook of corpus approaches to discourse* (pp. 1–4). Routledge.

Gabrielatos, C. (2018). Keyness analysis: Nature, metrics and techniques. In Taylor, C., & Marchi, A. (Eds.), *Corpus approaches to discourse: A critical review* (pp. 225–258). Routledge.

Hanratty, B., & McPolin, P. (2018). Re-focusing the teaching of poetry at key stage two in Northern Ireland: Some literary-critical and pedagogical explorations. *Changing English*, *26*(1), 63–76. https://doi.org/10.1080/1358684x.2018.1500443

Hempel-Jorgensen, A., Cremin, T., Harris, D., & Chamberlain, L. (2018). Pedagogy for reading for pleasure in low socio-economic primary schools: Beyond "pedagogy of poverty"? *Literacy*, *52*(2), 86–94. https://doi.org/10.1111/lit.12157

Hunston, S. (2002). *Corpora in applied linguistics*. Cambridge University Press.

Jones, S. (2021). Young writers 'learning to mean': From classroom discourse to personal intentions. *L1 Educational Studies in Language and Literature*, *21*(Running issue), 1–23. https://doi.org/10.17239/l1esll-2021.21.01.15

Kilgarriff, A., Baisa, V., Bušta, J., Jakubíček, M., Kovář, V., Michelfeit, J., Rychlý, P., & Suchomel, V. (2014). The Sketch Engine: Ten years on. *Lexicography*, *1*(1), 7–36. https://doi.org/10.1007/s40607-014-0009-9

Kispal, A. (2008). *Effective teaching of inference skills for reading: Literature review*. Research Report DCSF-RR031. ERIC. National Foundation for Educational Research. https://eric.ed.gov/?id=ED501868

Lawrence, C. (2020). "What's the point if it isn't marked?" Trainee teachers' responses to concepts of authentic engagement with poetry text. *English in Education*, *55*(2), 120–130. https://doi.org/10.1080/04250494.2019.1588071

Lijffijt, J., & Gries, S. T. (2012). Correction to Stefan Th. Gries' "Dispersions and adjusted frequencies in corpora", *International Journal of Corpus Linguistics*. *International Journal of Corpus Linguistics*, *17*(1), 147–149. https://doi.org/10.1075/ijcl.17.1.08lij

Love, R., Dembry, C., Hardie, A., Brezina, V., & McEnery, T. (2017). The Spoken BNC2014. *International Journal of Corpus Linguistics*, *22*(3), 319–344. https://doi.org/10.1075/ijcl.22.3.02lov

Maynard, S. (2011). Children's reading habits and attitudes. In Baker, D., & Evans, W. (Eds.), *Libraries and society: Role, responsibility and future in an age of change* (pp. 219–234). Chandos Publishing.

McEnery, T., Xiao, R., & Tony, Y. (2006). *Corpus-based language studies: An advanced resource book*. Routledge.

Myhill, D. (2021). Grammar re-imagined: Foregrounding understanding of language choice in writing. *English in Education*, *55*(3), 1–14. https://doi.org/10.1080/04250494.2021.1885975

Myhill, D., & Newman, R. (2019). Writing talk: Developing metalinguistic understanding through dialogic teaching. In N. Mercer, R. Wegerif, and L. Major (Eds.), *The Routledge international handbook of research on dialogic education* (pp. 306–372). Routledge.

Myhill, D. A., & Watson, A. M. (2017). 'The Dress of Thought': Analysing literature through a linguistic lens. https://ore.exeter.ac.uk/repository/handle/10871/29970

NATE (National Association for the Teaching of English). (2022). KS3 English in England: A NATE position paper. *Teaching English*, *22*, 14–16.

Nottingham Education Partners. (n.d.). *Strategic school improvement project 1: Reading transition Toolkit KS2 - KS3*. http://www.nottinghamschools.org.uk/media/1536189/transition-toolkit-v6.pdf

Parker, M., & Hurry, J. (2007). Teachers' use of questioning and modelling comprehension skills in primary classrooms. *Educational Review*, *59*(3), 299–314. https://doi.org/10.1080/00131910701427298

Phillips, E. (2013). A case study of questioning for reading comprehension during guided reading. *Education 3-13*, *41*(1), 110–120. https://doi.org/10.1080/03004279.2012.710106

Quigley, A. (2016). *One word at a time - Teaching vocabulary*. https://www.theconfidentteacher.com/2016/11/8181/

Quigley, A. (2018). *Closing the vocabulary gap*. Routledge.

Safford, K. (2016). Teaching grammar and testing grammar in the English primary school: The impact on teachers and their teaching of the grammar element of the statutory test in spelling, punctuation and grammar (SPaG). *Changing English*, *23*(1), 3–21. https://doi.org/10.1080/1358684x.2015.1133766

Scott, M. (2009). In search of a bad reference corpus. In Archer, D. (Ed.), *What's in a word list: Investigating word frequency and keyword extraction* (pp. 79–92). Ashgate.

Smith, L., Thomas, H., Chapman, S., Foley, J., Kelly, L., Kneen, J., & Watson, A. (2021). The dance and the tune: A storied exploration of the teaching of stories. *Changing English*, *29*(1), 40–52. https://doi.org/10.1080/1358684x.2021.1957669

Stubbs, M. (2001). *Words and phrases: Corpus studies of lexical semantics*. Blackwell.

Tennent, W. (2021). The assessment of reading comprehension in English primary schools: Investigating the validity of the key stage 2 reading standard assessment test (SAT). *Education 3-13*, *49*, 1–14. https://doi.org/10.1080/03004279.2020.1735472

Verhoeven, B. (2021). The politics of GCSE English language: Popular language ideology's influence on English's National Curriculum English language qualification. *English Today*, 1–10. https://doi.org/10.1017/S0266078421000110

Warsop, A. (2015). *'Why has she stopped reading': The case for supporting reading for pleasure in secondary schools*. University of East Anglia [unpublished doctoral thesis]. https://ueaeprints.uea.ac.uk/id/eprint/53427

# 6 The language of science at the transition

*Alice Deignan and Florence Oxley*

## Introduction

In this chapter, we present findings from a study of science in KS2 and KS3. We begin by overviewing the curricula for KS2 and KS3, then consider some of the central literature. We then present our corpus data and case study.

## The KS2 and KS3 curricula

Along with English and mathematics, science is 'core' until the end of compulsory schooling at the age of 15–16, after KS4 (DfE, n.d.). Standard Attainment Tests (SATs) were introduced in the 1990s, and for a number of years, English, mathematics and science were all examined at the end of KS2. There were ongoing concerns about the amount of compulsory testing for younger children, with the associated stress and learning time taken up by practice papers. As a result, the science SATs were removed and were last taken by all Year 6 students in 2009, with just a small number of schools administering the science test as a sampling exercise after that. The science SATs were replaced by teacher assessment of Year 6 students' progress, to be reported to parents and their destination secondary school. It became apparent fairly quickly that one outcome of this change was that by and large, a lot less time was spent teaching science in upper KS2 (Wellcome Trust, 2011) than before 2009. In KS3, science tends to have a similar amount of timetabled class time as English and mathematics. This means that many students experience a very significant increase in time spent on science in Year 7 compared with their recent primary school years. This was reflected in our corpus data, as mentioned in Chapter 3.

The upper KS2 science curriculum (Years 5 and 6) aims to develop students' abstract thinking and scientific reasoning. Extracts are as follows:

> The principal focus of science teaching in upper key stage 2 is to enable pupils to develop a deeper understanding of a wide range of scientific

DOI: 10.4324/9781003081890-6

> ideas. They should do this through exploring and talking about their ideas; asking their own questions about scientific phenomena; and analysing functions, relationships and interactions more systematically. […] they should encounter more abstract ideas and begin to recognise how these ideas help them to understand and predict how the world operates. They should also begin to recognise that scientific ideas change and develop over time. They should select the most appropriate ways to answer science questions using different types of scientific enquiry, including observing changes over different periods of time, noticing patterns, grouping and classifying things, carrying out comparative and fair tests and finding things out using a wide range of secondary sources of information. Pupils should draw conclusions based on their data and observations, use evidence to justify their ideas, and use their scientific knowledge and understanding to explain their findings. […]
>
> Pupils should read, spell and pronounce scientific vocabulary correctly.
>
> DfE 2013a, p. 24

The KS3 curriculum builds on this and branches into physics, chemistry and biology:

> The principal focus of science teaching in key stage 3 is to develop a deeper understanding of a range of scientific ideas in the subject disciplines of biology, chemistry and physics. Pupils should begin to see the connections between these subject areas and become aware of some of the big ideas underpinning scientific knowledge and understanding. Examples of these big ideas are the links between structure and function in living organisms, the particulate model as the key to understanding the properties and interactions of matter in all its forms, and the resources and means of transfer of energy as key determinants of all of these interactions. They should be encouraged to relate scientific explanations to phenomena in the world around them and start to use modelling and abstract ideas to develop and evaluate explanations.
>
> Pupils should understand that science is about working objectively, modifying explanations to take account of new evidence and ideas and subjecting results to peer review. Pupils should decide on the appropriate type of scientific enquiry to undertake to answer their own questions and develop a deeper understanding of factors to be taken into account when collecting, recording and processing data. They should evaluate their results and identify further questions arising from them.
>
> DfE 2013b, p. 2-3

Both the KS2 and KS3 curricula also mention the development of scientific language. Both include the following sentences in their introductions: 'Pupils should be able to describe associated processes and key characteristics in common language, but they should also be familiar with, and use, technical terminology accurately and precisely. They should build up an extended specialist vocabulary' (2013a, p. 3; 2013b, p. 2).

## Language and learning science at school

Many of the key issues facing humanity are scientific, for example, climate change, the custody and care of the planet's resources and public health. School students need a solid foundation in science in order to participate in decisions about such issues in the future, as citizens and voters (Xiao & Sandoval, 2017). Millar (2014, p. 16), in a discussion of why science is taught, terms this 'the democratic argument'. Science education, and more broadly, the public understanding of science, have thus been of concern to scientists themselves, as well as to educationalists. This has led to a relatively large number of studies in the language of science, compared with other school disciplines.

### *Scientific thinking and the language of science*

A number of researchers argue that learning the language of science is inseparable from learning science. In Chapter 2, we discussed Gee's (2008) analysis of children's talk about how metal rusts when in contact with water. Gee showed that children's everyday language was not sufficiently subtle and precise to distinguish important differences, and argued that without learning scientific language, the development of their scientific thinking would be limited.

Also, with a focus on younger learners and talk, Dawes (2004) writes that in order to access scientific ideas, children need to learn scientific language and scientific concepts, and this is a part of learning how to think scientifically at the same time. Fang (2005) describes writing in school science from a systemic-functional linguistics perspective and writes that developing scientific literacy is inseparable from learning science. He claims that language is a necessary part of making meaning in science – scientific language is used to express crucial ideas such as hypotheses, to reason scientifically and to justify interpretations. Similarly, Patterson Williams (2020) writes that scientific literacy is vital for engaging with and learning science since science texts are central to doing and learning science. The Education Endowment Foundation (2018) claims that there is a strong correlation between scientific literacy and attainment in school science and suggests that teachers should work to develop their students' fluency in scientific language. Bower and Ellerton (2007) found that access to the language of physics and mathematics is essential for access to concepts in physics.

Gerde and Wasik (2021) turn the relationship around, arguing that learning about the world through science introduces opportunities to learn language that children do not encounter in other aspects of their life. This includes words such as *experiment*, *cause* and *effect* (p. 535); learning science thus gives an opening into abstract and academic vocabulary more broadly. As Tang and Rappa (2020, p. 1312) write, 'the importance of embedding literacy instruction within science classroom teaching and learning is well-acknowledged among education researchers and practitioners'. Tang and Rappa (2020) also note that the language of science is not transparent from the genres used in classrooms, and often remains implicit in classroom genres. They argue for the use of scientific metalanguage to help students to deconstruct and interpret scientific language and ideas and to make conscious decisions in constructing their own scientific talk and writing. In this book, we are also starting from the premise that the language of science needs to be brought to the surface and examined in its own right in order to support students in learning science.

### *School science, language and socio-economic status*

Tang and Rappa (2020, p. 1311) argue that the language of science makes accessing and learning science difficult for most children. Some writers go further, claiming that the language of science can be alienating for some children (e.g., Halliday & Martin, 1993; Merzyn, 1987; Fang, 2005). Patterson Williams writes that the language of science is off-putting because language is closely bound up with identity, and for many children, learning scientific language involves putting another part of their identity 'on hold' (2020, p. 334). She believes this to be especially true for children from lower SES backgrounds, writing that the language of science is associated with middle-class values. This means that to compromise and discard 'lifeworld language' are necessary for non-middle-class children to engage with science (p. 333).

Early experience outside school can make science seem less alien and easier to engage with. Some research suggests that the kind of talk, specifically causal talk, that children are exposed to through their families can influence their scientific literacy and causal reasoning.

Booth et al. (2020) observed 153 American dyads of caregivers with three-year-old children talking about science in museums and in a laboratory setting. They found that the more caregivers talked about causal relationships between objects and phenomena, the more interested their child was in further information about causal relationships. They also found that the more caregivers asked their child to construct causal explanations, the stronger their child's scientific literacy skills were. These effects were observed regardless of children's cognitive ability or access to science-related resources at home such as books, toys, games and trips. Junge et al. (2021) also found that caregiver-child interaction and the home learning environment shape

children's interest in science. They found that as a group, children from higher SES backgrounds have more chances to learn about science, such as through books and trips. However, there was a lot of variability between individual families in both high and low SES groups. They suggested that caregivers' own interest in science, regardless of SES, and their inclination to do scientific activities with their children may strongly influence their children's interest and attainment in science.

Nunes et al. (2017) note a consistent link between SES and science attainment, with children from low SES backgrounds performing less well than their peers in science throughout school. Possible reasons, according to Nunes et al. (2017) are the lack of resources outside school available to children from less well-off backgrounds, and issues with teaching staff. There is a national shortage of specialist science teachers (Sims, 2019), and research in England has found that schools with the highest proportion of deprived pupils, measured by entitlement for free school meals, also have the highest proportion of unqualified teachers (Allen & Sims, 2018). Less experienced teachers are more likely to work in schools in deprived areas, and the gap in teachers' qualifications and experience between deprived and affluent areas is most extreme for mathematics and science (ibid). Nunes et al. (2017) write that lower attainment in science has little to do with individual children's interest in science since students from lower SES backgrounds show weaker attainment even when they have chosen science as an option.

## Features of the language of school science

Research into the language of professional and academic science language, with an aim of supporting science writers at university level, has a long tradition. Halliday and Martin (1993) extended this study to school science, working within the systemic-functional linguistic approach, discussed in Chapter 2. There are a number of overviews of the language of school science, including by Fang (2005), Gee (2008) and Patterson et al. (2018). We describe the central points that have been found under the headings of discourse, grammar, vocabulary and polysemy.

### *Discourse*

Patterson Williams (2020) and Fang (2005) write that scientific discourse attempts an authoritative tone to convey accuracy and objectivity of information. This results in the downplaying of personal, vague and subjective language. Such expressions in other genres would be used as politeness markers (Brown & Levinson, 1987), and their absence may contribute to the experience of some students that scientific language is 'impersonal and alienating' (Fang, 2005, p. 343). Snow writes: 'Maintaining the impersonal authoritative stance creates a distanced tone that is often puzzling to adolescent readers and is extremely difficult for adolescents to emulate in writing'

(2010, p. 451). We note though that our MD analysis in Chapter 4 suggested that not all sub-registers of school science have these qualities.

Tang and Rappa (2020) write that there are four overarching genres in the language of science: experimental report, informational report, argument and explanation. Tolmie et al. (2016) propose three core skills for learning science – children must be able to: make accurate observations explicitly; make accurate inferences about causal relationships between objects and phenomena, rejecting inaccurate inferences and irrelevant information; and use existing knowledge to explain the nature of these causal relationships. These are genres and clausal relationships that may be unfamiliar to KS3 students, as KS2 texts tend to have a narrative structure, even in science (Quigley, 2022). In Chapter 4, we showed that the sub-registers of presentations and assessments in science became significantly more non-narrative at KS3 than at KS2. Patterson et al. (2018, p. 297) claim that in school science texts the logical relationships between ideas are often not made explicit; connectives, that is, indicators of the relationships between ideas, are often absent. Although their presence would make texts longer, it would reduce the burden of inferencing on students.

Several writers note the multimodal and multisemiotic nature of many science texts (e.g., Fang, 2005; Norris & Phillips, 2003). Norris and Phillips (2003) write that children need scientific literacy skills to interpret written text in interaction with tables, graphs, diagrams and drawings. Visual material of this kind may support an experienced reader, but our conversations with students suggested that such materials may present additional difficulties in interpretation unless teachers overtly train students to interpret them.

### *Grammar*

Probably the earliest observation made about science texts is the tendency for processes to become nouns, that is, nominalisation (Halliday & Martin, 1993). Fang et al. (2006, p. 254) write: '*to attend* is a verb, but it can be turned into a noun as *attendance*, and that enables it to be modified and expanded (e.g., *perfect attendance*, *attendance at every session*)'. Fang (2005, p. 340) gives a number of examples from textbooks in which a clause is rephrased into a nominal group, for example: 'As winter begins, the first frost kills many of the insects. This sudden rise in the death rate causes the insect population to decrease'. Abstract nouns in scientific contexts can be particularly challenging as they are often nominalised forms of concrete verbs or of adjectives (Fang et al., 2006). They are useful in classifying the world and explaining hierarchical relationships between entities and ideas, for example, 'the process of cell division' (2006, p. 500). Fang et al. (2006) analysed nouns and nominalisation in a school science text in detail and found that this led to very dense text. The abstract nouns *change* and *pattern* were used to refer back to complex processes that had been described in a previous sentence.

Snow (2010) notes that scientific language tends to be concise, leading to a high density of information. Similarly, Fang (2006) writes that scientific language is economical, but this can make information more difficult to interpret. He notes that the grammar of scientific language tends to include complex sentences, subordinate clauses and use of the passive voice, all of which may make processing more difficult for young secondary school students, as we argued in Chapter 4.

### *Vocabulary*

Fang (2006) writes that science has its own lexicon and semantic systems. This includes discipline-specific specialist vocabulary that cannot be replaced by synonyms, such as *deciduous*. Specialist vocabulary is used to 'construct classes and categories and to establish taxonomic relationships' (p. 494) between entities and phenomena. He points out that many science vocabulary items are multi-morphemic and have classical roots that would be unknown to school students, and are thus very difficult for students to unpack for meaning. Another key strategy for dealing with unknown vocabulary is to use context, but Arya et al. (2011) point out that if there are too many unknown words, it becomes impossible for readers to infer the meaning of any of them, as they simply do not have a base from which to work. Given how lexically dense science writing is (Snow, 2010; Schleppegrell, 2001), it seems possible that this will happen for school students.

In Chapter 2, we discussed the notion of tiers of vocabulary. Words such as *deciduous* would be classified as Tier 3, that is, subject-specific and technical. Teachers and students are aware of Tier 3 vocabulary, and effort is spent in drawing attention to it and explaining it. There are many examples of definitions of Tier 3 words in the concordance for *is called* in our KS3 science corpus, such as:

(1)

> What you look at under a microscope is called a specimen. (Year 7 presentation)
> The random mixing and moving of particles is called diffusion. (Year 7 textbook)
> Together, the breaking of rocks into sediments and their moving away is called erosion. (Year 8 textbook)

As we wrote in Chapter 2, many teachers believe that a significant challenge is presented by what is termed Tier 2 vocabulary (Quigley, 2018). This is defined as academic vocabulary that is not specific to a single discipline, and has multiple discourse and reference functions (Beck et al., 2002). Researchers such as Norris and Phillips (2003) and Snow (2010) take the same view. While they recognise that specialist scientific vocabulary is of

great importance, they critique a view of scientific literacy that focuses solely on teaching scientific vocabulary, arguing that other general and academic vocabulary is also vital to learning science. Tier 2 vocabulary has been identified through intuitive study of individual words in use. In the absence of a systematic and generally established procedure for identifying Tier 2 words, we prefer to term the general set of vocabulary that we identified as 'general science words' to indicate the partial and preliminary status of the grouping.

### *Polysemy*

Many words have uses in everyday language as well as specific scientific meanings, and the everyday sense can interfere with understandings of science. Dawes (2004) discusses how the experimental group children in her study aged nine and ten (Year 5 in the UK) formed a misconception around the word *vibrate/ion* with reference to creating sound. They held the idea, based on their everyday knowledge of the word, that vibration is created by visually observable movement. This affected their understanding of what materials can and cannot vibrate and, in turn, their idea of what materials will make good conductors or insulators of sound. The children believed that because metal is 'strong', it would not vibrate. The children's concept of 'vibrate/ion' was underextended, and this impeded them applying it to scientific talk. Conversely, an everyday meaning may have additional nuances, which interfere with their understanding of the scientific sense. Dawes (2004) writes that for a child, *force* 'may be synonymous with aggression' (p. 678).

In many cases, words have separate meanings in the language of schooling, as discussed in Chapter 2. Polysemy has received particular attention in literature on school science. The Education Endowment Foundation (2018) write: 'it is familiar words used in unfamiliar contexts that cause most difficulty' (p. 32). Various writers cite a number of polysemous science words. For example, Strömdahl (2012) notes polysemy between everyday words such as *heat* and *work* and their scientific uses, while (Fang 2006) notes *school*, *fault* and *volume*. Chan (2015) writes that a number of verbs used in science and maths texts, including *find*, *give*, *convert* and *simplify*, are polysemous with everyday meanings. Bower and Ellerton (2007) note that some polysemous words have different meanings in different subjects, like *vector* in mathematics and physics. While a number of such individual examples are cited in the literature on the language of schooling, there has been no systematic study of polysemy across the most frequent words in any discipline. There is one recent and systematic study of polysemy in the academic language of higher education: Skoufaki and Petrić (2021) analysed dictionary entries for each of the words in the Academic Vocabulary List (AVL) (Gardner & Davies, 2014) to establish how many of these have multiple meanings. They found that of the most frequent 1000 AVL lemmas, 66.05% are polysemous. The authors acknowledge that because they

analysed dictionary entries, rather than concordance lines from relevant corpora, it is not possible to establish which meanings are more likely to occur in academic texts, nor whether they are restricted to specific disciplines.

Meyerson et al. (1991) designed a study to investigate the science vocabulary knowledge of students in third and fifth grades (equivalent to Years 4 and 6 in the UK system). Children were asked to classify words into conceptual groups. Fifth graders were more likely to recognise that some words would fit into more than one group, while third graders were more likely to put polysemous words into more general, non-science conceptual groups, suggesting a developing awareness of polysemy.

Various researchers have explored how polysemy can be defined (e.g., Deignan, 2005, with reference to metaphor in corpora; Gries, 2019, as a notion in cognitive linguistics). Moon (1987) writes that there is no external, fixed measure of how many senses a word may have, and decisions on splitting senses should be made according to purpose. For her work, the purpose was writing dictionaries to help language learners; for us, the purpose is trying to identify potential difficulties with the language of school. Gries (2019) notes that approaches range from 'extreme splitters' to 'extreme lumpers' (p. 474). Deignan and Love (2021) used a narrow 'splitter' understanding to separate senses on this basis, for example, finding a distinction between the everyday sense of *ice* 'piece of frozen water used to cool drinks' and the sense found in their climate science texts, 'large stretch of frozen water, part of the sea, a lake or river'. Using this understanding, Deignan and Love (2021) found polysemy to be very widespread in their climate data.

For this case study, our research questions were as follows:

1. What are the most significant keywords in KS3 science, using KS2 and BNCBM as reference corpora?
2. What are the most frequent content-specific words in the KS2 and KS3 science corpora, and what do these indicate about the 'aboutness' of the corpora? Which of these are the general science words, as opposed to words associated with specific science topics?
3. To what extent are the general science words in each corpus polysemous?

## Method

### *The corpora*

In Chapter 3, we described how our corpus was collected and built. Table 6.1 contains figures from Tables 3.3 and 3.4, to show the composition of the written science corpora. Table 6.2 is extracted from Table 3.5, to show the spoken science corpora.

As for the study of the English corpora, described in Chapter 5, the study described here did not separate written and spoken data, but it did separate

*Table 6.1* Written science corpus from KS2 and KS3.

| *Subject* | *Texts* | *Tokens* | *Mean length* | *SD text length* |
|---|---|---|---|---|
| KS2 science | 177 | 160,355 | 906 | 3069 |
| KS3 science | 675 | 356,319 | 528 | 3046 |
| Total | 852 | 516,674 | | |

*Table 6.2* Spoken science corpus from KS2 and KS3.

| | *Number of texts* | | *Number of tokens* | | *Mean text length (tokens)* | | *Standard deviation text length (tokens)* | |
|---|---|---|---|---|---|---|---|---|
| | *KS2* | *KS3* | *KS2* | *KS3* | *KS2* | *KS3* | *KS2* | *KS3* |
| Science | 18 | 10 | 62,375 | 47,772 | 3465 | 4777 | 2923 | 1410 |
| Subtotal (spoken) | 28 | | 110,147 | | | | | |

*Table 6.3* Division of science corpus by Key Stage.

| | *Key Stage 2* | | *Key Stage 3* | |
|---|---|---|---|---|
| | *Number of texts* | *Tokens* | *Number of texts* | *tokens* |
| Written | 177 | 160,355 | 675 | 356,319 |
| Spoken | 18 | 62,375 | 10 | 47,772 |
| Total | 195 | 222,730 | 685 | 404,091 |

KS2 and KS3. Table 6.3 gives the same information about texts and tokens as Tables 6.1 and 6.2, but has been reorganised.

The compositions of the written science corpora are given in Chapter 4. As for the other subjects, the spoken corpora consist of transcripts of audio recordings of lessons, with student contributions omitted.

While polysemy has often been observed in previous work on the language of school science, as we noted previously, with the exception of Deignan and Love's (2021) study, corpus data are rarely used. Deignan and Love (2021) used a corpus of science materials around the theme of climate science, consisting of 214,858 tokens, to study the materials that young people access on the topic of climate change both in and out of school. Of these, 22,416 were from school science textbooks, and 192,442 were from websites that students told researchers that they would consult for further information about climate change. Our school science corpora are thus nearly three times larger and are more diverse in terms of scientific topics covered. They are also more representative of students' school

experience in terms of sub-registers, comprising assessments, presentations, worksheets, textbooks and reading extracts, as discussed in Chapter 3, and by not using website material. Website material comprised nearly 90% of Deignan and Love's corpus but is rarely accessed by students in class.

### *Focus and tools*

In order to get a sense of the challenges of KS3 science, we began by using the keywords procedure in #LancsBox 6.0 (Brezina et al., 2020) to identify lemmas that are key to KS3 science with relation to first, KS2 science, and second, our general corpus, BNCBM, described in Chapter 5. As in Chapter 5, we used the keyness statistic Cohen's *d* because it takes dispersion into account.

We then used #LancsBox 6.0 to identify the most frequent types in the KS2 and KS3 science corpora, with a normalised frequency of 32.6 words per 100,000 words as our cut-off point, as for the studies reported in Chapter 5. We again used the $DP_{norm}$ dispersion measure (Lijffijt & Gries 2012), eliminating any words with a value of over 0.95, to ensure that we did not capture words that are very unevenly distributed, that is, occurring in a very small number of texts in our corpora. We eliminated words that are in the top 200 New GSL (Brezina & Gablasova, 2015), and words for class and lesson infrastructure, such as *class*, *door* and *name*, which are common across all lessons and subjects. We grouped the types semantically, as described in Chapter 5, to establish the 'aboutness' of each corpus. As for the English corpora that we discussed in Chapter 5, we found that we needed to read concordance lines for this stage. As part of this process, for each of the corpora, we identified words that occurred across a range of texts, which did not appear to have a highly specialised meaning, and which were not a clear fit with any of the other semantic groups we had identified due to their general meaning. We describe these as 'general scientific words', as described earlier.

To answer our third research question, we analysed these for polysemy. We followed the standard procedure undertaken by lexicographers, as described by Skoufaki and Petrić (2021). That is, we studied the available concordance lines and grouped occurrences of each word according to meaning and use. We tended towards splitting meanings finely because we wanted to identify cases in which a word might have been encountered in an everyday context but have a subtly different use in science, which could potentially be problematic for some students.

As part of the polysemy analysis, we considered collocation, as it can give strong clues to meaning (e.g., Sinclair, 2004). Moon (2010, pp. 203–204) illustrates this using randomly sampled concordance lines for *race*, *saw* and *colourful*. In her data, examples where *race* means 'competitive activity' show collocates such as '*champ*, *won*, *title*, *victory*', while those with *race* meaning 'ethic grouping' show the collocates '*relations*, *religion*, *gender*, *human*'. For *colourful*, the concrete meaning collocates with physical objects, and places, while the abstract meaning collocates with *history*,

*culture* and words denoting people (ibid). Similarly, Baker (2006) finds that collocates of *bachelor* include *eligible* and *degree*, the former associated with the meaning 'unmarried man', and the latter with the meaning 'first level of university education'. We compared the collocational profiles of our target words across KS3, KS2 and the BNCBM corpora, using the log Dice statistic (Brezina, 2018a, p. 274; 2018b), in #LancsBox 6.0. This statistic was chosen because it offers a standardised measure, 'which makes Log Dice directly comparable across different corpora' (Gablasova et al., 2017, p. 10), and because it 'highlights exclusive but not necessarily rare combinations' (ibid). We then examined concordance lines in detail to identify any cases of polysemy between KS3 general science uses of words and uses that students might be more familiar with from KS2 or from language outside the classroom.

## Results

### *Keywords*

The following tables give the most key lemmas for each keyword analysis. The third column gives the frequency, normalised to frequency per 100,000 words in the KS3 science corpus. The fourth column gives the Cohen's *d* statistic, and the fifth, dispersion, using the $DP_{norm}$ statistic. Keywords showing Cohen's *d* below 0.2 were not included in our study, as the effect size is considered too low (see Chapter 5). Table 6.4 shows the 30 most key lemmas in KS3 science, using the KS2 science corpus as a reference corpus, and Table 6.5 shows the 30 most key lemmas in KS3 science with the BNCBM as reference corpus.

As noted in Chapter 5, Cohen (1988, p. 40) gives definitions of effect size as: small: 0.2; medium: 0.5; large, 0.8. The fourth column in each table, showing Cohen's *d* for each keyword, indicates that effect sizes are greater in the second table. This suggests that KS3 science is more like KS2 science than it is like everyday language. This was also the case for English, as reported in Chapter 5, but the effect sizes for both keyword studies are slightly smaller for science than for English.

### *Frequent words*

Using the words tool in #LancsBox 6.0 to identify frequent types and removing words that are in the top 200 of the *New GSL*, and words associated with the lesson context, resulted in a list of 131 types for KS2 science and 231 types for KS3 science. The KS3 science corpus is nearly twice as big as the KS2 one, but the same normalised frequency cut-off point was used. This suggests a much wider range of topics in KS3, but this has to be a tentative finding because of the disparity in corpus size. It would not be unexpected, given that science starts to diverge into the three separate disciplines of physics, chemistry and biology in KS3. The lists are also longer than for the other subjects looked at, but again, with different corpus sizes, this is

*Table 6.4* Keywords in KS3 science, reference corpus KS2 science.

| *Keyness rank* | *Lemma* | *Freq* | *Cohen's d* | $DP_{norm}$ |
|---|---|---|---|---|
| 1 | force (n) | 374 | 0.4 | 0.68 |
| 2 | equation | 82 | 0.4 | 0.67 |
| 3 | cell | 470 | 0.38 | 0.69 |
| 4 | chemical (n) | 178 | 0.37 | 0.53 |
| 5 | calculate | 71 | 0.37 | 0.72 |
| 6 | oxygen | 170 | 0.37 | 0.57 |
| 7 | mass (n) | 155 | 0.34 | 0.68 |
| 8 | reaction | 237 | 0.32 | 0.58 |
| 9 | speed | 148 | 0.31 | 0.67 |
| 10 | carbon | 152 | 0.31 | 0.58 |
| 11 | weight | 104 | 0.3 | 0.73 |
| 12 | measure (v) | 116 | 0.3 | 0.58 |
| 13 | iron (n) | 695 | 0.28 | 0.73 |
| 14 | dependent | 36 | 0.28 | 0.81 |
| 15 | chemical (adj) | 51 | 0.28 | 0.6 |
| 16 | metal | 153 | 0.28 | 0.28 |
| 17 | copper | 105 | 0.28 | 0.72 |
| 18 | nerve | 428 | 0.28 | 0.8 |
| 19 | control (n) | 22 | 0.27 | 0.81 |
| 20 | magnesium | 71 | 0.26 | 0.67 |
| 21 | acid | 145 | 0.26 | 0.66 |
| 22 | energy | 447 | 0.26 | 0.65 |
| 23 | car | 78 | 0.26 | 0.67 |
| 24 | substance | 152 | 0.26 | 0.61 |
| 25 | nucleus | 45 | 0.25 | 0.77 |
| 26 | vacuole | 20 | 0.25 | 0.83 |
| 27 | distance (n) | 113 | 0.25 | 0.77 |
| 28 | dioxide | 99 | 0.24 | 0.61 |
| 29 | experiment | 62 | 0.24 | 0.79 |
| 30 | tube | 64 | 0.24 | 0.74 |

suggestive, rather than conclusive, that the vocabulary challenge for science may be greater than for English and mathematics. This is consistent with the findings from the MD analysis reported in Chapter 4.

Here, just the top 120 frequent types are shown for each key stage, in Tables 6.6 and 6.7, though we read concordance lines for all words that met our cut-off level. The normalised frequency per 100,000 words and dispersion statistics are given.

### *Aboutness and general science words*

Our analysis of the KS2 science corpus suggested clusters of words including the following:

Living things: *animals, plants, living, seeds, leaves, soil, insects*
Human or animal bodies: *animals, egg, heart, cells, body, fish, teeth, wings, blood*

*Table 6.5* Keywords in KS3 science, reference corpus BNCBM.

| *Keyness rank* | *Lemma* | *Freq* | *Cohen's* ***d*** | $DP_{norm}$ |
|---|---|---|---|---|
| 1 | water | 347 | 0.55 | 0.5 |
| 2 | explain | 217 | 0.53 | 0.49 |
| 3 | diagram | 98 | 0.47 | 0.57 |
| 4 | object (n) | 216 | 0.47 | 0.61 |
| 5 | force (n) | 374 | 0.46 | 0.68 |
| 6 | oxygen | 170 | 0.45 | 0.59 |
| 7 | measure (v) | 116 | 0.44 | 0.59 |
| 8 | word (n) | 178 | 0.44 | 0.51 |
| 9 | draw (v) | 123 | 0.44 | 0.66 |
| 10 | cell | 470 | 0.43 | 0.69 |
| 11 | table (n) | 127 | 0.43 | 0.51 |
| 12 | chemical (n) | 178 | 0.42 | 0.54 |
| 13 | describe | 193 | 0.4 | 0.5 |
| 14 | type (n) | 117 | 0.4 | 0.51 |
| 15 | equation | 82 | 0.4 | 0.67 |
| 16 | carbon | 152 | 0.39 | 0.58 |
| 17 | result (n) | 106 | 0.39 | 0.67 |
| 18 | contain | 107 | 0.38 | 0.57 |
| 19 | gas | 130 | 0.38 | 0.58 |
| 20 | produce (v) | 105 | 0.37 | 0.57 |
| 21 | calculate | 71 | 0.36 | 0.72 |
| 22 | body | 154 | 0.36 | 0.56 |
| 23 | mass | 155 | 0.36 | 0.68 |
| 24 | plant (n) | 138 | 0.35 | 0.6 |
| 25 | energy | 447 | 0.35 | 0.6 |
| 26 | move (v) | 167 | 0.34 | 0.51 |
| 27 | substance | 152 | 0.33 | 0.61 |
| 28 | dioxide | 99 | 0.33 | 0.6 |
| 29 | metal | 153 | 0.33 | 0.73 |
| 30 | example | 114 | 0.33 | 0.49 |

Evolution: *species, darwin, evolution*
The solar system: *earth, space, sun, moon, gas, shadow, gravity*
Physics: *energy, friction, force*
Electricity and light: *light, shadow, circuit, edison*
Elements: *carbon, oxygen*
Assessment and objectives: *write, describe, explain, identify*

In itself, this is not of major interest for two reasons. The first is that despite close reading of concordance lines, it is often not possible to reliably disentangle distinct groups, as can be seen above. Second, the words and groups closely reflect the National Curriculum topics and are thus almost entirely expected. The topics are, for Year 5, 'Properties and changes of materials', 'Earth and space' and 'Forces'; for Year 6, 'Evolution and inheritance, 'Light' and 'Electricity'; and for both years: 'Living things and their habitats' and 'Animals including humans' (DfE, 2013a).

*Table 6.6* Most frequent topic-specific content words in the KS2 science corpus.

| *Rank* | *Type* | *Freq* | $DP_{norm}$ | *Rank* | *Type* | *Freq* | $DP_{norm}$ |
|---|---|---|---|---|---|---|---|
| 1 | water | 472.4 | 0.46 | 61 | together | 47.1 | 0.43 |
| 2 | light | 266.7 | 0.56 | 62 | able | 46.7 | 0.48 |
| 3 | plants | 210.2 | 0.52 | 63 | objects | 46.7 | 0.59 |
| 4 | food | 207.5 | 0.47 | 64 | lots | 46.2 | 0.42 |
| 5 | animals | 206.6 | 0.48 | 65 | word | 45.8 | 0.58 |
| 6 | plastic | 186.8 | 0.83 | 66 | key | 45.3 | 0.61 |
| 7 | living | 172 | 0.59 | 67 | ideas | 45.3 | 0.75 |
| 8 | down | 149.5 | 0.35 | 68 | fish | 45.3 | 0.56 |
| 9 | earth | 145.9 | 0.5 | 69 | side | 44.5 | 0.56 |
| 10 | materials | 118.1 | 0.58 | 70 | friction | 44.5 | 0.66 |
| 11 | body | 118.1 | 0.54 | 71 | pollution | 43.6 | 0.93 |
| 12 | sun | 113.6 | 0.52 | 72 | sugar | 43.6 | 0.58 |
| 13 | plant | 105 | 0.59 | 73 | results | 43.6 | 0.61 |
| 14 | air | 100.6 | 0.58 | 74 | amount | 43.1 | 0.54 |
| 15 | change | 99.7 | 0.51 | 75 | else | 42.2 | 0.57 |
| 16 | explain | 97.9 | 0.62 | 76 | types | 42.2 | 0.58 |
| 17 | blood | 96.6 | 0.64 | 77 | cold | 42.2 | 0.58 |
| 18 | grow | 93.4 | 0.5 | 78 | remember | 41.8 | 0.57 |
| 19 | electricity | 89.8 | 0.69 | 79 | micro-organism | 41.8 | 0.67 |
| 20 | eat | 87.1 | 0.46 | 80 | evidence | 41.8 | 0.74 |
| 21 | evolution | 86.2 | 0.85 | 81 | once | 41.8 | 0.42 |
| 22 | liquid | 80.4 | 0.63 | 82 | pollen | 41.8 | 0.67 |
| 23 | animal | 79.9 | 0.5 | 83 | natural | 41.8 | 0.69 |
| 24 | moon | 75.4 | 0.64 | 84 | teeth | 41.3 | 0.64 |
| 25 | humans | 74.1 | 0.69 | 85 | forces | 40.4 | 0.65 |
| 26 | gas | 73.6 | 0.59 | 86 | habitat | 40 | 0.61 |
| 27 | shadow | 72.3 | 0.62 | 87 | left | 40 | 0.55 |
| 28 | species | 70.9 | 0.8 | 88 | lot | 40 | 0.48 |
| 29 | egg | 69.1 | 0.69 | 89 | moving | 40 | 0.54 |
| 30 | move | 69.1 | 0.37 | 90 | let | 39.5 | 0.52 |
| 31 | seeds | 67.8 | 0.61 | 91 | sound | 39.1 | 0.66 |
| 32 | force | 67.8 | 0.66 | 92 | type | 38.6 | 0.52 |
| 33 | object | 67.8 | 0.6 | 93 | mirror | 38.6 | 0.64 |
| 34 | shape | 66.5 | 0.51 | 94 | insects | 38.6 | 0.6 |
| 35 | bit | 65.6 | 0.57 | 95 | temperature | 38.2 | 0.68 |
| 36 | science | 64.7 | 0.54 | 96 | read | 38.2 | 0.59 |
| 37 | human | 62.9 | 0.69 | 97 | ago | 38.2 | 0.76 |
| 38 | circuit | 61.1 | 0.67 | 98 | glass | 37.7 | 0.55 |
| 39 | heat | 59.3 | 0.62 | 99 | stop | 37.7 | 0.42 |
| 40 | write | 58.4 | 0.51 | 100 | size | 37.7 | 0.53 |
| 41 | environment | 57.9 | 0.68 | 101 | line | 37.7 | 0.49 |
| 42 | important | 57.5 | 0.5 | 102 | top | 37.3 | 0.47 |
| 43 | changes | 57 | 0.54 | 103 | less | 37.3 | 0.49 |
| 44 | sea | 55.7 | 0.57 | 104 | surface | 37.3 | 0.58 |
| 45 | energy | 55.2 | 0.73 | 105 | diagram | 36.8 | 0.57 |
| 46 | solid | 55.2 | 0.59 | 106 | identify | 36.8 | 0.76 |
| 47 | source | 55.2 | 0.62 | 107 | dissolve | 36.8 | 0.65 |
| 48 | material | 55.2 | 0.55 | 108 | fossils | 36.8 | 0.9 |

(*Continued*)

*Table 6.6* (Continued)

| *Rank* | *Type* | *Freq* | $DP_{norm}$ | *Rank* | *Type* | *Freq* | $DP_{norm}$ |
|---|---|---|---|---|---|---|---|
| 49 | heart | 54.3 | 0.67 | 109 | eggs | 36.8 | 0.74 |
| 50 | gravity | 53.4 | 0.65 | 110 | add | 36 | 0.54 |
| 51 | eyes | 53.4 | 0.61 | 111 | kind | 36 | 0.53 |
| 52 | darwin | 53 | 0.9 | 112 | oxygen | 35.5 | 0.67 |
| 53 | leaves | 53 | 0.58 | 113 | scientific | 35.5 | 0.76 |
| 54 | soil | 52.1 | 0.64 | 114 | white | 35.5 | 0.7 |
| 55 | cells | 51.6 | 0.78 | 115 | sand | 35 | 0.6 |
| 56 | hard | 50.7 | 0.41 | 116 | enough | 35 | 0.45 |
| 57 | information | 50.3 | 0.68 | 117 | salt | 34.6 | 0.65 |
| 58 | legs | 49.4 | 0.66 | 118 | ground | 34.6 | 0.55 |
| 59 | process | 48.9 | 0.61 | 119 | big | 34.6 | 0.44 |
| 60 | words | 47.1 | 0.53 | 120 | shadows | 34.1 | 0.63 |

We also identified 'general science words'. We did this by close reading of concordances and noting the range of texts that the types appeared in. As described above, they are used in a variety of topic areas across the corpus. We considered the dispersion statistic, but without a much larger corpus, with an even number of text types and sub-registers, this is of limited value. The process therefore is more subjective than would be ideal. The types identified appear to constitute a core KS2 vocabulary for talking and writing about fundamental entities in science and science processes. This group is shown in Table 6.8.

Our analysis of the KS3 corpus also suggested groups of words reflecting the National Curriculum (DfE, 2013b), including:

Introductory physics: *energy, light, force, mass, thermal*
Introductory chemistry: *chemical, [periodic] table*
Introductory biology: *cell, body*
Elements: *carbon, copper, oxygen, atoms, magnesium, hydrogen*
Assessment and objectives: *explain, describe, write, words, diagram*

The 'general science words' in this group are shown in Table 6.9.

It is to be expected that as students progress through the school years, there will be an increase in the amount of core, general vocabulary needed to structure content. Nonetheless, this seems a very large expansion. Most or all of these words may be familiar, but if their everyday use is very different from their science use, they may still present problems. We therefore studied concordance data for each of these KS3 general scientific types.

### *Polysemy*

As explained in the methodology section, we examined concordances for each of the listed 64 types in the KS3 science corpus, the KS2 science corpus

*Table 6.7* Most frequent topic-specific content words in the KS3 science corpus.

| *Rank* | *Type* | *Freq* | $DP_{norm}$ | *Rank* | *Type* | *Freq* | $DP_{norm}$ |
|---|---|---|---|---|---|---|---|
| 1 | energy | 447.0 | 0.65 | 61 | type | 75.2 | 0.58 |
| 2 | water | 353.9 | 0.50 | 62 | around | 75 | 0.5 |
| 3 | cell | 259.1 | 0.74 | 63 | iron | 75 | 0.72 |
| 4 | light | 257.4 | 0.69 | 64 | word | 74.5 | 0.61 |
| 5 | force | 250.9 | 0.68 | 65 | element | 73.7 | 0.65 |
| 6 | cells | 210.8 | 0.67 | 66 | metals | 71.5 | 0.78 |
| 7 | explain | 195 | 0.50 | 67 | liquid | 71 | 0.65 |
| 8 | chemical | 191.8 | 0.54 | 68 | complete | 71 | 0.57 |
| 9 | describe | 177.2 | 0.52 | 69 | magnesium | 70.5 | 0.67 |
| 10 | down | 175.7 | 0.42 | 70 | current | 70.5 | 0.83 |
| 11 | oxygen | 179.5 | 0.57 | 71 | car | 70.5 | 0.73 |
| 12 | state | 167.5 | 0.57 | 72 | heat | 70 | 0.67 |
| 13 | mass | 166 | 0.68 | 73 | solution | 69.3 | 0.67 |
| 14 | speed | 160.6 | 0.68 | 74 | ph | 68 | 0.83 |
| 15 | store | 151.9 | 0.78 | 75 | changes | 67.8 | 0.63 |
| 16 | carbon | 151.4 | 0.58 | 76 | able | 67.3 | 0.58 |
| 17 | acid | 149.4 | 0.70 | 77 | colours | 67 | 0.78 |
| 18 | body | 145 | 0.56 | 78 | friction | 66.8 | 0.78 |
| 19 | reaction | 144.2 | 0.62 | 79 | measure | 66.6 | 0.6 |
| 20 | blood | 141.5 | 0.73 | 80 | data | 66.5 | 0.66 |
| 21 | forces | 140.5 | 0.73 | 81 | thermal | 65 | 0.74 |
| 22 | food | 139.3 | 0.69 | 82 | hydrogen | 64.8 | 0.64 |
| 23 | object | 136.3 | 0.64 | 83 | white | 63.8 | 0.76 |
| 24 | write | 130.6 | 0.48 | 84 | heart | 63.1 | 0.82 |
| 25 | table | 122.2 | 0.51 | 85 | equation | 62.3 | 0.66 |
| 26 | red | 121 | 0.74 | 86 | substances | 61.6 | 0.66 |
| 27 | air | 120.2 | 0.57 | 87 | plants | 60.6 | 0.64 |
| 28 | graph | 115 | 0.69 | 88 | increases | 60.4 | 0.68 |
| 29 | draw | 110.6 | 0.65 | 89 | surface | 59.1 | 0.61 |
| 30 | temperature | 108.6 | 0.61 | 90 | difference | 58.6 | 0.65 |
| 31 | particles | 108.6 | 0.65 | 91 | oxide | 58.6 | 0.67 |
| 32 | distance | 107.9 | 0.77 | 92 | ball | 58.1 | 0.69 |
| 33 | copper | 106.6 | 0.72 | 93 | muscle | 56.9 | 0.77 |
| 34 | weight | 104.9 | 0.76 | 94 | summary | 55.9 | 0.69 |
| 35 | elements | 104.2 | 0.63 | 95 | experiment | 54.9 | 0.79 |
| 36 | words | 103.9 | 0.57 | 96 | moving | 54.9 | 0.65 |
| 37 | earth | 101.2 | 0.64 | 97 | line | 54.7 | 0.63 |
| 38 | system | 100.2 | 0.68 | 98 | circuit | 54.7 | 0.81 |
| 39 | move | 99.4 | 0.57 | 99 | contains | 54.2 | 0.62 |
| 40 | change | 99.4 | 0.56 | 100 | calculate | 53.9 | 0.72 |
| 41 | gas | 98.5 | 0.58 | 101 | tube | 53.9 | 0.76 |
| 42 | dioxide | 98.5 | 0.61 | 102 | small | 53.7 | 0.64 |
| 43 | variable | 97 | 0.79 | 103 | moon | 53.7 | 0.79 |
| 44 | key | 93.8 | 0.54 | 104 | sun | 53.7 | 0.69 |
| 45 | reactions | 92.3 | 0.6 | 105 | mixture | 53.7 | 0.66 |
| 46 | sound | 92 | 0.75 | 106 | tissue | 52 | 0.81 |
| 47 | substance | 90.3 | 0.64 | 107 | bone | 52 | 0.82 |
| 48 | plant | 89.3 | 0.66 | 108 | sodium | 52 | 0.66 |

(*Continued*)

*Table 6.7* (Continued)

| *Rank* | *Type* | *Freq* | $DP_{norm}$ | *Rank* | *Type* | *Freq* | $DP_{norm}$ |
|---|---|---|---|---|---|---|---|
| 49 | pressure | 86.3 | 0.7 | 109 | blue | 51.5 | 0.79 |
| 50 | happens | 86.3 | 0.54 | 110 | transferred | 51.2 | 0.72 |
| 51 | diagram | 86.1 | 0.6 | 111 | solid | 51 | 0.65 |
| 52 | metal | 86.1 | 0.72 | 112 | green | 51 | 0.76 |
| 53 | results | 86.1 | 0.74 | 113 | muscles | 50.5 | 0.74 |
| 54 | objects | 85.1 | 0.65 | 114 | add | 50.5 | 0.6 |
| 55 | colour | 84.4 | 0.69 | 115 | rock | 49.2 | 0.7 |
| 56 | atoms | 82.9 | 0.71 | 116 | inside | 49.2 | 0.58 |
| 57 | properties | 82.4 | 0.62 | 117 | less | 48.7 | 0.64 |
| 58 | gravity | 81.4 | 0.81 | 118 | material | 48.7 | 0.63 |
| 59 | test | 77.2 | 0.69 | 119 | atom | 48.5 | 0.75 |
| 60 | together | 75.9 | 0.54 | 120 | organs | 48 | 0.81 |

*Table 6.8* General science types in KS2 science corpus ranked by frequency.

| *Rank* | *Type* | *Freq* | $DP_{norm}$ | *Rank* | *Type* | *Freq* | $DP_{norm}$ |
|---|---|---|---|---|---|---|---|
| 1 | materials | 118 | 0.58 | 13 | objects | 46.7 | 0.59 |
| 2 | change | 99.7 | 0.51 | 14 | results | 43.6 | 0.61 |
| 3 | shape | 66.4 | 0.51 | 15 | amount | 43.1 | 0.54 |
| 4 | heat | 59.3 | 0.62 | 16 | evidence | 41.8 | 0.74 |
| 5 | important | 57.5 | 0.50 | 17 | natural | 41.8 | 0.69 |
| 6 | changes | 57 | 0.54 | 18 | moving | 40 | 0.54 |
| 7 | energy | 55.2 | 0.73 | 19 | sound | 39.1 | 0.66 |
| 8 | solid | 55.2 | 0.59 | 20 | temperature | 38.2 | 0.69 |
| 9 | material | 55,2 | 0.55 | 21 | size | 37.7 | 0.53 |
| 10 | hard | 50.7 | 0.41 | 22 | surface | 37.3 | 0.58 |
| 11 | information | 50.3 | 0.68 | 23 | weight | 33.7 | 0.68 |
| 12 | process | 49 | 0.61 | | | | |

and our reference corpus, BNCBM. In several cases, this included more than one part of speech. For every type, there were differences in collocational profiles in each of the three corpora. This is to be expected given that the topics of each were different, but where the collocates seem to reflect a difference in meaning, they are noted below. The existing literature on school language does not offer a classification of the different kinds of polysemy, and we therefore developed our own through our examination of concordance lines. From this bottom-up analysis, we found five kinds of polysemy, that is, ways in which meanings of words differed from each other. These are: (1) contextual differences; (2) fine-grained differences in use; (3) meaning differences; (4) lexico-grammatical differences; and (5) frequency differences. The first three groups represent our attempt to divide the cline from near-identical meanings through to distinguishable polysemy. The analysis of a large number of concordance lines was essential to

*Table 6.9* General science types in KS3 science corpus ranked by frequency.

| *Rank* | *Type* | *freq* | $DP_{norm}$ | *Rank* | *Type* | *Freq* | $DP_{norm}$ |
|---|---|---|---|---|---|---|---|
| 1 | mass | 166 | 0.68 | 33 | increases | 64 | 0.68 |
| 2 | speed | 161 | 0.68 | 34 | moving | 55 | 0.65 |
| 3 | store | 152 | 0.8 | 35 | mixture | 54 | 0.66 |
| 4 | acid | 149 | 0.8 | 36 | transferred | 51 | 0.72 |
| 5 | reaction | 144 | 0.62 | 37 | solid | 51 | 0.65 |
| 6 | forces | 141 | 0.73 | 38 | material | 49 | 0.63 |
| 7 | object | 136 | 0.64 | 39 | stores | 48 | 0.76 |
| 8 | particles | 109 | 0.65 | 40 | boiling | 47 | 0.71 |
| 9 | temperature | 109 | 0.61 | 41 | materials | 47 | 0.68 |
| 10 | distance | 108 | 0.77 | 42 | independent | 46 | 0.81 |
| 11 | weight | 105 | 0.76 | 43 | surroundings | 45 | 0.81 |
| 12 | system | 100 | 0.68 | 44 | increase | 44 | 0.73 |
| 13 | move | 99 | 0.57 | 45 | measured | 43 | 0.66 |
| 14 | change | 99 | 0.56 | 46 | decreases | 43 | 0.82 |
| 15 | gas | 99 | 0.58 | 47 | melting | 43 | 0.68 |
| 16 | reactions | 92 | 0.61 | 48 | form | 42 | 0.61 |
| 17 | sound | 92 | 0.75 | 49 | resistance | 41 | 0.63 |
| 18 | substance | 90 | 0.64 | 50 | effect | 40 | 0.76 |
| 19 | pressure | 86 | 0.70 | 51 | volume | 40 | 0.75 |
| 20 | metal | 86 | 0.72 | 52 | produce | 39 | 0.65 |
| 21 | objects | 85 | 0.65 | 53 | produced | 39 | 0.69 |
| 22 | properties | 82 | 0.62 | 54 | organisms | 39 | 0.69 |
| 23 | test | 77 | 0.69 | 55 | variables | 38 | 0.79 |
| 24 | metals | 72 | 0.78 | 56 | balanced | 38 | 0.70 |
| 25 | liquid | 71 | 0.65 | 57 | model | 38 | 0.69 |
| 26 | heat | 70 | 0.67 | 58 | affect | 37 | 0.75 |
| 27 | solution | 69 | 0.67 | 59 | size | 37 | 0.64 |
| 28 | changes | 68 | 0.63 | 60 | dependent | 36 | 0.80 |
| 29 | data | 67 | 0.66 | 61 | image | 36 | 0.74 |
| 30 | measure | 67 | 0.60 | 62 | physical | 36 | 0.66 |
| 31 | equation | 62 | 0.67 | 63 | gases | 34 | 0.71 |
| 32 | substances | 62 | 0.66 | 64 | function | 33 | 0.71 |

try to classify words in one of these three, but there are inevitably borderline cases. Some words in these three groups also show lexico-grammatical differences, group (4), as is to be expected, given the regular association of content, meaning and grammar that has been observed in numerous corpus studies (e.g., Sinclair, 2004; Hunston & Su, 2019). Examples of each are now given.

### *Group 1: Contextual differences*

In the first group, there is some difference in meaning but this seems largely due to the topics of the texts in each corpus and context. This seems unlikely

to be problematic for KS3 students. Word types in this group include: *gas, speed, temperature, weight, system, move, change(s), liquid, moving, boiling, materials, surroundings, measured, melting, effect, affect, physical, function.*

| | ***Gas*** | | |
|---|---|---|---|
| *Corpus* | *Context or meaning* | *Example* | *Sub-register* |
| KS3 | specialised scientific | When methane is burned in oxygen, it produces a gas, carbon dioxide, and water.<br>… the organs involved in gas exchange. | Y8 presentation<br>Y7 textbook |
| KS2 | scientific, including everyday examples | Natural gas is found deep underground and is pumped into our homes.<br>Yeast… feeds on the sugar in the dough mixture and makes bubbles of a gas called carbon dioxide. | Y5 presentation<br>Y6 textbook |
| BNCBM | source of energy for personal or business use | … electricity and gas seem to be more expensive in the regions. | e-language forums |

| | ***Speed*** | | |
|---|---|---|---|
| *Corpus* | *Context or meaning* | *Example* | *Sub-register* |
| KS3 | specialised scientific; a quality to be calculated precisely | Here is some data for the speed of the Apollo spacecraft as it moved away from the Earth.<br>… a force can change the speed of an object. | Y7 assessment<br>Y7 presentation |
| KS2 | specialised scientific | Everything would fall at the same speed if there was no air resistance. | Y6 presentation |
| BNCBM | vehicles, also used non-literally | … 149 mph top speed.<br>… the speed was around 50 mbps. | news, mass market<br>e-language reviews |

*Group 2: Fine-grained differences in use*

The words in the second group show different uses, associated with different contexts, and these are on the border of having different meanings.

'Splitters' – analysts whose purpose means that they split meaning and use very finely (Gries, 2019; Skoufaki & Petrić, 2021) – would be likely to classify these as distinct senses, while 'lumpers' (ibid) would not. The core reference appears to be the same across meanings, but in our view, there is enough difference in context of use for the science meanings to seem unfamiliar for some students. Types in this group are: *distance, sound, substance, pressure, test, heat, measure, equation, mixture, transferred, solid, produce(d), balanced, size.*

| | ***Substance*** | | |
|---|---|---|---|
| *Corpus* | *Context or meaning* | *Example* | *Sub-register* |
| KS3 | specialist scientific term for material | A base is a substance that neutralises an acid. | Y7 textbook |
| | | An element is a substance that cannot be broken down into other substances. | Y7 textbook |
| KS2 | specialist scientific term for material | Red blood cells contain a substance called haemoglobin. | Y6 reading |
| BNCBM | the core of something | … a triumph of style over substance. | e-language reviews |
| | a generic word for addictive drugs | … substance abuse. | news, regional |

| | ***Pressure*** | | |
|---|---|---|---|
| *Corpus* | *Context or meaning* | *Example* | *Sub-register* |
| KS3 | specialist scientific meaning, calculated precisely | The pressure at a particular depth in a liquid depends on the weight of the water above it. | Y8 textbook |
| | | …using a pressure of 0.5 N/cm$^2$ on her book. | Y8 worksheet |
| KS2 | scientific meaning in the context of liquid: blood, water. Infrequent. | … it causes your blood pressure to rise. | Y6 textbook |
| BNCBM | the most frequent meaning refers to psychological stress. The collocation *blood pressure* occurs less frequently. | … a competitive environment adds a lot more pressure. | news, regional |

| | ***Solid*** | | |
|---|---|---|---|
| *Corpus* | *Context or meaning* | *Example* | *Sub-register* |
| KS3 | specialist scientific meaning, contrasted with liquid and gas | light is stopped when it meets a solid object.<br>Is it solid, liquid or gas at room temperature? | Y8 assessment<br>Y7 presentation |
| KS2 | specialist scientific meaning. Some occurrences mean '3 dimensional' | Lots of small solid pieces together behave like a liquid.<br>A sphere is a round, solid shape. | Y6 textbook<br>Y5 presentation |
| BNCBM | the most frequent meaning is metaphorical; some literal occurrences meaning 'dense, strong, heavy' | You have a solid proposal, it's interesting and relevant.<br>… a pretty solid foundation for any business relationship.<br>The box is solid and very well packed. | e-language, SMS<br>e-language, blogs<br>e-language, reviews |

| | ***Balanced*** | | |
|---|---|---|---|
| *Corpus* | *Context or meaning* | *Example* | *Sub-register* |
| KS3 | specialist scientific meaning referring to forces. Some occurrences of everyday use. | Balanced forces do not change the direction an object is moving in or its speed.<br>… some problems with not having a balanced diet. | Y7 presentation<br>Y8 presentation |
| KS2 | specialist scientific meaning, with forces and diet. | … stationary objects have balanced forces on them. | Y6 textbook |
| BNCBM | a wide range of contexts, collocating with view, diet, lifestyle, with metaphorical meanings. No scientific uses. | … read all the information and take a balanced view.<br>… help you to live a more balanced lifestyle.<br>She's very balanced and a very happy child. | e-language, reviews<br>e-language, SMS news, mass |

*Group 3: Meaning differences*

In our third group, the dominant meanings differ between two or more corpora, and would probably be described as different even by 'lumpers' (Gries, 2019; Skoufaki & Petrić, 2021). In most cases, the BNCBM meaning is different from that found in the KS3 and KS2 science corpora. Word types in this group include: *mass, store(s), reaction (s), forces, properties, solution, independent, resistance, volume, variables, model, dependent.*

| | ***Mass*** | | |
|---|---|---|---|
| *Corpus* | *Main meaning/use* | *Examples* | *Sub-register* |
| KS3 | measurable quality | As the mass increases, so does the height. | Y7 presentation |
| KS2 | a large unorganised quantity | … the soil was packed with a twisted mass of white roots. | Y6 textbook |
| BNCBM | on a large scale; a large unorganised quantity | … the lie that mass immigration is happening now.<br>… a floppy mass of unruly hair. | speech<br>fiction |

| | ***Store*** | | |
|---|---|---|---|
| *Corpus* | *Meaning/use* | *Examples* | *Sub-register* |
| KS3 | reserve, usually of energy | Your body's chemical store of energy decreases.<br>… depleting the battery's energy store. | Y8 teacher talk<br>Y7 presentation |
| KS2 | reserve more generally | … most oceans that store most of our planet's water. | Y6 presentation |
| BNCBM | shop | Does anyone know of an Apple store in London? | e-language, social media |

| | ***Reaction*** | | |
|---|---|---|---|
| *Corpus* | *Meaning/use* | *Example* | *Sub-register* |
| KS3 | Specialist scientific meaning, frequently collocating with *chemical* and *combustion*. | … no atoms are lost or made during a chemical reaction so the mass of the result equals the mass of the reactants. | Y7 presentation |
| KS2 | Specialist scientific meaning as for KS3, but very infrequent. | … refresh the vinegar to start the reaction again. | Y7 teacher talk. |
| BNCBM | The most frequent meaning denotes an immediate response, often emotional, to an event. | … they kind of looked at me for a reaction. | speech |

| | ***Forces*** | | |
|---|---|---|---|
| *Corpus* | *Meaning/use* | *Example* | *Sub-register* |
| KS3 | specialist scientific meaning. | The regular pattern of particles and strong forces explain why solids keep their shape and cannot flow. | Y7 presentation |
| KS2 | specialist scientific meaning. | The forces at work are the same. | Y5 presentation |
| BNCBM | Nearly all occurrences are found in fixed collocations, often referring to the army or police. | armed forces<br>special forces<br>security forces<br>join forces<br>market forces | wide range of sub-corpora |

| | ***Independent*** | | |
|---|---|---|---|
| *Corpus* | *Meaning/use* | *Example* | *Sub-register* |
| KS3 | specialist scientific and research meaning, always with *variable*. | The independent variable is the variable we change in the experiment. | Y7 presentation |
| KS2 | Very infrequent. Split between scientific and everyday meanings. | Adolescents are increasingly independent.<br>The independent variable is the one which you decide to change. | Y5 presentation<br>Y6 presentation |
| BNCBM | Wide range of collocates. The most frequent meaning is 'not attached or affiliated'. | Dozens of independent businesses will be taking part.<br>The government has now set up an independent review. | news, regional news, mass |

*Group 4: Lexico-grammatical differences*

Some of the frequent general KS3 science types take a grammatical form that is very infrequent in one or more of the corpora. Examples include are *metals* and *gases*, reflecting a specialised discourse in which the distinctions between these kinds of substance are central.

| | ***Metals*** | | |
|---|---|---|---|
| *Corpus* | *Frequency per 100k* | *Example* | *Sub-register* |
| KS3 | 72 | The colour of fireworks comes from the reaction of two different metals with oxygen. | Y8 worksheet |
| KS2 | 13 | Most metals are good thermal conductors. | Y6 textbook |
| BNCBM | 0.08 | … precious metals. | fiction |

| | ***Gases*** | | |
|---|---|---|---|
| *Corpus* | *Frequency per 100k* | *Example* | *Sub-register* |
| KS3 | 34 | Light can travel through gases like the air. | Y7 textbook |
| KS2 | 28 | Most water contains gases from the air that have dissolved in it. | Y6 textbook |
| BNCBM | 0.1 | … greenhouse gases. | fiction |
| | | … exhaust gases. | fiction |

*Measure* is more usually a noun in BNCBM, whereas it is more usually a verb in the school corpora.

| | ***Measure*** | | |
|---|---|---|---|
| *Corpus* | *Frequency per 100k* | *Example* | *Sub-register* |
| KS3 | 86 | You can measure force with a newton meter. | Y7 textbook |
| KS2 | 29.2 | We measure our heart rate by measuring our pulse. | Y6 reading |
| BNCBM | 2.5 | … ridiculed and feared in similar measure. | news, regional |

*Group 5: Frequency differences*

For the words in the fourth group, there appears to be little difference in meaning, but the BNCBM occurrences, and sometimes also KS2 ones, are very infrequent. This may mean that KS3 students could find words difficult not because of difference in meaning, but because the uses are unfamiliar. Word types in this group include: *acid, object(s), particles, increases, decreases, organisms*.

| | ***Object*** | | |
|---|---|---|---|
| *Corpus* | *Frequency per 100k* | *Example* | *Sub-register* |
| KS3 | 136.4 | Friction can make an object get hotter. | Y7 assessment |
| KS2 | 67.8 | The light can't get through the object and that's when we end up with a shadow. | Y5 presentation |
| BNCBM | 1.7 | He padded forward and sniffed at the curious object. | fiction |

| | ***Increases*** | | |
|---|---|---|---|
| *Corpus* | *Frequency per 100k* | *Example* | *Sub-register* |
| KS3 | 60.4 | As the human population increases, more space is needed to meet our needs. | Y7 presentation. |
| KS2 | 2.2 | Our pulse increases during exercise. | Y6 reading |
| BNCBM | 0.4 | Never use a barbecue indoors – this increases the risk of both fire and carbon monoxide... | news, regional |

| | ***Particles*** | | |
|---|---|---|---|
| *Corpus* | *Frequency per 100k* | *Example* | *Sub-register* |
| KS3 | 108.7 | What are particles themselves made of? | Y7 presentation |
| KS2 | 13.5 | The rest of the soil particles float around in the water. | Y6 textbook |
| BNCBM | 0.2 | … so you're not breathing in dust particles. | speech |

We also noted that while the KS2 uses of these word types show the same meaning as in the KS3 data, there is a tendency for examples to come from everyday, rather than scientific contexts. This is seen to varying extents in the above examples, and also in the following concordance examples for *produce*.

| | ***Produce*** | | |
|---|---|---|---|
| *Corpus* | *Frequency* | *Example* | *Sub-register* |
| KS3 | 39.4 | Some micro-organisms have to produce their own food. | Y8 textbook |
| KS2 | 32.3 | Many plants and trees produce their seeds in the autumn. | Y6 textbook |
| BNCBM | 2.9 | … if someone's taken the trouble to produce something. | speech |

### *Metaphorical uses*

As well as considering the extent of polysemy, we also analysed its nature, and specifically metaphorical polysemy. Metaphor is a major way in which new meanings of words develop (Moon, 1987; Nerlich, 2003), for some cognitive linguists, the central mechanism behind polysemy (e.g., Gibbs, 1999). Metaphor has also been discussed with regard to science. For example, Brown (2003), has discussed the centrality of using metaphor to think and communicate scientifically, using examples drawn from physics, chemistry and biology, and including models of the atom, proteins and climate change. Knudsen (2003) has written about the changes in meaning and use when scientists' technical metaphors are used in non-scientific communication. There is a smaller literature on the use of metaphor in school science. Cameron (2002) took a discourse approach to analyse the use of metaphor in science lessons by Y5 and Y6 teachers. Deignan and Semino (2019) used corpora and interview data to identify problems for students aged 11–16 in interpreting metaphors of climate change, and Lancor (2012) examined metaphors for energy in pedagogical discourse, identifying how they highlight and hide aspects of the topic. There have also been studies on the use of conceptual metaphors to teach scientific concepts (e.g., chapters in Amin et al., 2018), and on students' uses of metaphors (e.g., Lancor, 2015).

The above studies all seem to tacitly assume that any issues that school students may have arise from the metaphorical meanings in science of words that they have only encountered with literal meanings. For example, Lancor (2012) writes of students' failing to understand fully some metaphors used to describe energy, while Deignan and Semino (2019) discuss limitations of the *greenhouse* metaphor. We found some evidence of metaphorical meanings of scientific words. Lancor's example *store* was frequent in our data. Size in the BNCBM tends to be used in a concrete sense, while it is more often abstract in the KS3 corpus in examples such as: 'The *size* of the force depends on the mass of the objects' (Y8 presentation). However, our study found more instances of the opposite phenomenon, hitherto unremarked in the literature: a number of the words frequent in our KS3 science corpus are

used with a literal meaning that is rare in everyday discourse. These words have an everyday meaning that is metaphorical, and which is therefore likely to be better known to students. Words that this applies to include: *pressure*, *substance*, *measure*, *equation*, *mixture*, *solid*, *image*. The following table gives examples of the dominant meanings of each in the KS3 science corpus and the BNCBM.

| *Word type* | *KS3 science corpus* | *BNCBM* |
|---|---|---|
| *pressure* | The pressure inside the container increases.<br>… this liquid pressure acts in all directions. | His team struggled in the face of pressure.<br>[She] claimed she was under pressure to help students … |
| *substance* | They contain a sticky substance dissolved in a solvent.<br>The new substance is iron sulfide. | He stands for ideas and substance over sound bites.<br>[It] produced nothing of substance.<br>(also a small number of examples of *substance* as a synonym for drugs) |
| *measure* | The PH scale is a measure of how acid or alkaline a solution is. (N)<br>… to measure the current flowing through a component in a circuit. (V)<br>Watch what happens and measure the temperature of the solution. | Everyone would laugh and cry in equal measure. (N)<br>… a true measure of his value.<br>You can measure the time pretty accurately. (V)<br>You want to measure yourself against the best. (V) |
| *equation* | … state the general equation for combustion reactions.<br>You can calculate weight by using an equation. | … take [him] out of the equation because he's the new kid on the block.<br>If you have a family, it's an even tougher equation to manage. |
| *mixture* | Granite is a mixture of compounds.<br>The reaction mixture gets so hot that the iron melts. | I just stood there, a mixture of horror and relief.<br>His face is a mixture of compassion and fear. |
| *solid* | Iodine is a brittle solid at room temperature.<br>When a substance is in the solid state, its particles touch each other. | The acting is very solid throughout.<br>It wants good schools, good hospitals, solid economy… |
| *image* | This magnifies the image using lenses.<br>A camera produces an image, just like your eye. | … his clean image as a player.<br>She has been accused of carefully stage-managing her image. |

The fact of scientific meanings being uncommon in everyday discourse does not necessarily entail that they will cause problems. Nonetheless, it is possible that the literal, scientific meanings of such words may be poorly understood, and students may not be aware of their specific scientific denotations.

### *Discussion and conclusion*

The series of related studies described in this chapter has shown how much new scientific vocabulary students encounter as they move up the school years, particularly at the beginning of secondary school. The corpora that we collected for different school subjects are not directly comparable with each other, being of different sizes and composition, attempting to replicate students' experience of each register, as described in Chapter 3. This means that we cannot draw direct comparisons, but it is suggestive that the number of different content word types meeting our threshold of 32.6 occurrences per 100,000 is much greater for our science corpora than for the English corpora analysed in Chapter 5 (Tables 6.6 and 6.7). Further studies with directly comparable corpora would be needed to confirm this possibility.

Our detailed studies of concordance lines have shown that the meanings and uses of words in KS3 science are often different from those in the BNCBM, notably tending towards being highly specific. This is consistent with discussion in the literature about the widespread nature of polysemy in science discourse; in fact, it suggests that some degree of polysemy is the norm for virtually every word that is used in both science and everyday discourse. Although scientific discourse is characterised by being abstracted from everyday life, this does not seem to result in more metaphorical uses of words, as would have been expected from previous research. Our analysis of concordance lines for the 64 most frequent general scientific words in KS3 has suggested that the reverse is the case: a number of words that are mainly used with a metaphorical sense in the BNCBM tend to be used literally in the KS3 science corpus. Again, a more detailed study, covering more of the most frequent words in school science, would be needed to confirm if this is a widespread pattern.

The keyword analysis comparing the KS3 and KS2 science corpora showed a smaller effect size than the one described in Chapter 5, which compared KS3 and KS2 English corpora. This suggests that word use across KS2 and KS3 is more similar in science than in English. This implies that there is a fairly consistent building of concepts from KS2 through to KS3 in science. Nonetheless, the dramatic increase in word types at KS3 in science shows a significant challenge for students. We have noted a tendency for examples of words to be drawn from everyday, observable experience in KS2, and to be abstracted and specialised at KS3. This is a normal aspect of academic progression but also an additional aspect of the challenge of transition to KS3.

The studies described here suggest that the vocabulary of school science is different in multiple ways from everyday vocabulary, even where words appear to have related uses.

## References

Allen, R., & Sims, S. (2018). Do pupils from low-income families get low-quality teachers? Indirect evidence from English schools. *Oxford Review of Education*, *44*(4), 441–458. https://doi.org/10.1080/03054985.2017.1421152

Amin, T. G., Jeppsson, F., & Haglund, J. (Eds.). (2018). *Conceptual metaphor and embodied cognition in science learning*. Routledge.

Arya, D. J., Hiebert, E. H., & Pearson, P. D. (2011). The effects of syntactic complexity on the comprehension of elementary science texts. *International Electronic Journal of Elementary Education*, *4*(11), 107–125. https://www.iejee.com/index.php/IEJEE/article/view/216

Baker, P. (2006). *Using corpora in discourse analysis*. Continuum.

Beck, I., McKeown, M. G., & Kucan, L. (2002). *Bringing words to life: Robust vocabulary instruction*. Guilford Press.

Booth, A. E., Shavlik, M., & Haden, C. A. (2020). Parents' causal talk: Links to children's causal stance and emerging scientific literacy. *Developmental Psychology*, *56*(11), 2055–2064. https://doi.org/10.1037/dev0001108

Bower, M. W., & Ellerton, N. F. (2007). Language genre transitions in a secondary school physics classroom. *International Journal of Mathematical Education in Science and Technology*, *38*(3), 338–351. https://doi.org/10.1080/00207390601116078

Brezina, V. (2018a). Statistical choices in corpus-based analysis. In Taylor, C., & Marchi, A. (Eds.), *Corpus approaches to discourse: A critical review* (pp. 271–280). Routledge.

Brezina, V. (2018b). *Statistics in corpus linguistics: A practical guide*. Cambridge University Press.

Brezina, V., & Gablasova, D. (2015). Is there a core general vocabulary? Introducing the new general service list. *Applied Linguistics*, *36*(1), 1–22. https://doi.10.1093/applin/amt018

Brezina, V., Weill-Tessier, P., & McEnery, A. (2020). *#LancsBox v. 6.x* [software package].

Brown, P., & Levinson, S. (1987). *Politeness: Some universals in language usage* (2nd ed.). Cambridge University Press.

Brown, T. (2003). *Making truth: Metaphor in science*. University of Illinois Press.

Cameron, L. (2002). Metaphors in the learning of science: A discourse focus. *British Educational Research Journal*, *28*(5), 673–688. https://doi.10.1080/0141192022000015534

Chan, S. (2015). Linguistic challenges in the mathematical register for EFL learners: Linguistic and multimodal strategies to help learners tackle mathematics word problems. *International Journal of Bilingual Education and Bilingualism*, *18*(3), 306–318. https://doi.org/10.1080/13670050.2014.988114

Dawes, L. (2004). Talk and learning in classroom science. *International Journal of Science Education*, *26*(6), 677–695. https://doi.org/10.1080/0950069032000097424

Deignan, A. (2005). *Metaphor and corpus linguistics*. John Benjamins.

Deignan, A., & Love, R. (2021). Using corpus methods to identify subject specific uses of polysemous words in English secondary school science materials. *Corpora*, *16*(2), 165–189. https://doi.org/10.3366/cor.2021.0216

Deignan, A., & Semino, E. (2019). Translating science for young people using metaphor. *The Translator*, *25*(4), 369–384. https://doi.org/10.1080/13556509.2020.1735759

Department for Education (DfE). (2013a). *Science programs of study: Key stages 1 and 2. The National Curriculum in England*. Crown copyright. https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/425618/PRIMARY_national_curriculum_-_Science.pdf

Department for Education (DfE). (2013b). *Science programs of study: Key stage 3. The National Curriculum in England*. Crown copyright. https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/335174/SECONDARY_national_curriculum_-_Science_220714.pdf

Department for Education (DfE). (n.d.). *The national curriculum*. Crown Copyright. https://www.gov.uk/national-curriculum/key-stage-3-and-4

Education Endowment Fun (EEF). (2018). *Improving secondary science*. https://educationendowmentfoundation.org.uk/education-evidence/guidance-reports/science-ks3-ks4

Fang, Z. (2005). Scientific literacy: A systemic functional linguistics perspective. *Science Education*, *89*, 335–347. https://doi.org/10.1002/sce.20050

Fang, Z. (2006). The language demands of science reading in middle school. *International Journal of Science Education*, *28*(5), 491–520. https://doi.org/10.1080/09500690500339092

Fang, Z., Schleppegrell, M. J., & Cox, B. E. (2006). Understanding the language demands of schooling: Nouns in academic registers. *Journal of Literacy Research*, *38*(3), 247–273. https://doi.org/10.1207/s15548430jlr3803_1

Gablasova, D., Brezina, V., & McEnery, T. (2017). Collocations in corpus-based language learning research: Identifying, comparing and interpreting the evidence. *Language Learning*, *67*(SI), 130–154. https://doi.org/10.1111/lang.12226

Gardner, D., & Davies, M. (2014). A new academic vocabulary list. *Applied Linguistics*, *35*, 305–327. https://doi.org/10.1093/applin/amt015

Gee, J. P. (2008). What is academic language? In Roseberry, A., & Warren, B. (Eds.), *Teaching science to English language learners: Building on students' strengths* (pp. 57–70). NSTA Press.

Gerde, H. K., & Wasik, B. A. (2021). Developing language through science. *The Reading Teacher*, *75*(5), 535–544. https://doi.org/10.1002/trtr.2075

Gibbs, R. (1999). Researching metaphor. In Cameron, L., & Low, G. D. (Eds.), *Researching and applying metaphor* (pp. 29–47). Cambridge University Press.

Gries, S. (2019). Polysemy. In Dabrowska, E., & Divjak, D. (Eds.), *Cognitive linguistics: Key topics* (pp. 473–490). Mouton de Gruyter. https://doi.10.1515/9783110626438-002

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

Hunston, S., & Su, H. (2019). Patterns, constructions and local grammar: A case study of 'evaluation'. *Applied Linguistics*, *40*(4), 567–593. https://doi.org/10.1093/applin/amx046

Junge, K., Schmerse, D., Lankes, E.-M., Carstensen, C. H., & Steffensky, M. (2021). How the home learning environment contributes to children's early science knowledge—Associations with parental characteristics and science-related activities. *Early Childhood Research Quarterly*, *56*, 294–305. https://doi.org/10.1016/j.ecresq.2021.04.004

Knudsen, S. (2003). Scientific metaphors going public. *Journal of Pragmatics*, *35*(8), 1247–1263. https://doi.org/10.1016/S0378-2166(02)00187-X

Lancor, R. (2014). Using metaphor theory to examine conceptions of energy in biology, chemistry, and physics. *Science and Education*, *23*, 1245–1267. https://doi.org/10.1007/s11191-012-9535-8

Lancor, R. (2015). An analysis of metaphors used by students to describe energy in an interdisciplinary general science course. *International Journal of Science Education*, *37*(5–6), 876–902. https://doi.org/10.1080/09500693.2015.1025309

Lijffijt, J., & Gries, S. T. (2012). Correction to Stefan Th. Gries' "Dispersions and adjusted frequencies in corpora", *International Journal of Corpus Linguistics*. *International Journal of Corpus Linguistics*, *17*(1), 147–149. https://doi.org/10.1075/ijcl.17.1.08lij

Merzyn, G. (1987). The language of school science. *International Journal of Science Education*, *9*(4), 483–489. https://doi.10.1080/0950069870090406

Millar, R. (2014). Designing a science curriculum fit for purpose. *School Science Review*, *95*(352), 15–20.

Moon, R. (1987). The analysis of meaning. In Sinclair, J. (Ed.), *Looking up: An account of the COBUILD project in lexical computing* (pp. 86–103). Collins ELT.

Moon, R. (2010). What can a corpus tell us about lexis? In O'Keeffe, A., & McCarthy, M. (Eds.), *The Routledge handbook of corpus linguistics* (pp. 197–211). Routledge.

Meyerson, M., Ford, M. S., Jones, W. P., & Ward, M. A. (1991). Science vocabulary knowledge of third and fifth graders. *Science Education*, *75*(4), 419–428. https://doi.org/10.1002/sce.3730750404

Nerlich, B. (2003). Polysemy, past and present. In Nerlich, B., Todd, Z., Herman, V., & Clarke, D. D. (Eds.), *Polysemy: Flexible patterns of meaning in mind and language* (pp. 49–76). Mouton de Gruyter.

Norris, S. P., & Phillips, L. M. (2003). How literacy in its fundamental sense is central to scientific literacy. *Science Education*, *87*, 224–240. https://doi.org/10.1002/sce.10066

Nunes, T., Bryant, P., Strand, S., Hillier, J., Barros, R., & Miller Friedman, J. (2017). *Review of SES and science learning in formal educational settings: A report prepared for the EEF and the Royal Society*. https://educationendowmentfoundation.org.uk/public/files/Review_of_SES_and_Science_Learning_in_Formal_Educational_Settings.pdf

Patterson Williams, A. (2020). Sustaining disciplinary literacy in science: A transformative just model for teaching the language of science. *Journal of Adolescent and Adult Literacy*, *64*(3), 334–338. https://doi.org/10.1002/jaal.1100

Patterson, A., Roman, D., Friend, M., Osborne, J., & Donovan, B. (2018). Reading for meaning: The foundational knowledge every teacher of science should have. *International Journal of Science Education*, *40*(3), 291–307. https://doi.org/10.1080/09500693.2017.1416205

Quigley, A. (2018). *Closing the vocabulary gap*. Routledge.

Quigley, A. (2022). *The problem with reading informational texts*. https://www.theconfidentteacher.com/2022/01/the-problem-with-reading-informational-texts/

Schleppegrell, M. (2001). Linguistic features of the language of schooling. *Linguistics and Education*, *12*(4), 431–459. https://doi.org/10.1016/s0898-5898(01)00073-0

Sims, S. (2019). *Increasing the quality and quantity of science teachers in schools: Eight evidence-based principles*. Gatsby Foundation.

Sinclair, J. (2004). *Trust the text*. Routledge.

Skoufaki, S., & Petrić, B. (2021). Exploring polysemy in the academic vocabulary list: A lexicographic approach. *Journal of English for Academic Purposes*, *54*, 101038. https://doi.org/10.1016/j.jeap.2021.101038

Snow, C. E. (2010). Academic language and the challenge of reading for learning about science. *Science*, *328*, 450–453. https://doi.10.1126/science.1182597

Strömdahl, H. R. (2012). On discerning critical elements, relationships and shifts in attaining scientific terms: The challenge of polysemy/ homonymy and reference. *Science and Education*, *21*, 55–85. https://doi.org/10.1007/s11191-010-9335-y

Tang, K. S., & Rappa, N. A. (2020). The role of metalanguage in an explicit literacy instruction on scientific explanation. *International Journal of Science and Mathematics Education*, *19*(7), 1311–1331. https://doi.org/10.1007/s10763-020-10121-6

Tolmie, A. K., Ghazali, Z., & Morris, S. (2016). Children's science learning: A core skills approach. *The British Journal of Educational Psychology*, *86*(3), 481–497. https://doi.org/10.1111/bjep.12119

Wellcome Trust. (2011). *Primary science survey report*. https://wellcomecollection.org/works/jdtkftd2

Xiao, S., & Sandoval, W. A. (2017). Associations between attitudes towards science and children's evaluation of information about socio-scientific issues. *Science and Education*, *26*, 247–269. https://doi.10.1007/s11191-017-9888-0

# 7 The language of mathematics at the transition

*Duygu Candarli and Florence Oxley*

## Introduction

This chapter presents a case study focusing on mathematics at the transition from primary to secondary school. The topic is especially urgent because it has been argued that the negative effects of the transition to secondary school are 'more pronounced for mathematics than for any other subject' (O'Meara et al., 2020, p. 497). Learning mathematics is inseparable from learning the language of mathematics (Cruz Neri & Retelsdorf, 2022), and there is widespread empirical evidence that children's understanding of mathematical language is associated with their performance in mathematics in pre-school years (Turan & Smedt, 2022) and in later schooling (Riccomini et al., 2015). In this chapter, we review the KS2 and KS3 curricula in mathematics and previous findings on the language of mathematics. Although earlier studies have identified the characteristics of school language in the discipline of mathematics overall (e.g., Schleppegrell, 2007; Wilkinson, 2019), there has been no systematic study that explores the differences between primary and secondary school mathematics. In this study, we used a relatively new corpus technique, key feature analysis (Biber & Egbert, 2018), to explore lexico-grammatical differences between KS2 and KS3 mathematics. Then, we analysed keywords as a window into differences at levels of lexico-grammar, semantics and discourse between KS2 and KS3 mathematics.

## The KS2 and KS3 curricula

Along with English, mathematics is tested formally in the Year 6 SATs. As reported in earlier chapters, this has resulted in a good deal of class time in Year 6 being spent on these two subjects. This can be even more the case for lower attaining students, who are sometimes taken out of other subjects for additional literacy and mathematics tuition in an attempt to boost their KS2 SATs results (Hutchings, 2015). In KS3, as for KS2, there is no statutory requirement to spend a fixed amount of time on each subject. A survey of 619 secondary teachers found that between 3.5 and 4 hours per week are timetabled for mathematics in KS3 (Stone, n.d.), of a typical 25-hour week.

DOI: 10.4324/9781003081890-7

The primary curriculum (DfE, 2013a) focuses on foundational arithmetic operations and introductory geometry and algebra. The programme of study for Years 5 and 6 is described as follows:

> The principal focus of mathematics teaching in upper key stage 2 is to ensure that pupils extend their understanding of the number system and place value to include larger integers. This should develop the connections that pupils make between multiplication and division with fractions, decimals, percentages and ratio.
>
> At this stage, pupils should develop their ability to solve a wider range of problems, including increasingly complex properties of numbers and arithmetic, and problems demanding efficient written and mental methods of calculation. With this foundation in arithmetic, pupils are introduced to the language of algebra as a means for solving a variety of problems. Teaching in geometry and measures should consolidate and extend knowledge developed in number. Teaching should also ensure that pupils classify shapes with increasingly complex geometric properties and that they learn the vocabulary they need to describe them.
>
> By the end of year 6, pupils should be fluent in written methods for all four operations, including long multiplication and division, and in working with fractions, decimals and percentages.
>
> Pupils should read, spell and pronounce mathematical vocabulary correctly.
>
> DfE 2013a, p. 30

For Year 6, the following topics are specified: number; ratio and proportion; algebra, measurement, geometry and statistics. Under each topic heading, there is a list of specifications. For example, geometry is divided into 'properties of shapes' and 'position and direction', the first having a list of five requirements, the second, two.

The KS3 programme of study follows the following topics: number; algebra; ratio, proportion and rates of change; geometry and measures; probability; statistics (DfE 2013b, pp. 4–9). Under each topic heading, there is a detailed specification of the skills and knowledge students are required to learn. For example, under 'geometry and measures', there is a list of 16 bullet points, the first four of which read:

> - derive and apply formulae to calculate and solve problems involving: perimeter and area of triangles, parallelograms, trapezia,

volume of cuboids (including cubes) and other prisms (including cylinders)
- calculate and solve problems involving: perimeters of 2-D shapes (including circles), areas of circles and composite shapes
- draw and measure line segments and angles in geometric figures, including interpreting scale drawings
- derive and use the standard ruler and compass constructions (perpendicular bisector of a line segment, constructing a perpendicular to a given line from/at a given point, bisecting a given angle); recognise and use the perpendicular distance from a point to a line as the shortest distance to the line.(DfE 2013b, p. 8)

The KS2 SATs are important for schools, and also indirectly important for students as the results frequently inform ability grouping in KS3. Currently, there are three mathematics papers: Paper 1 covers arithmetic, and Papers 2 and 3 'reasoning'. In Papers 2 and 3, problems are often expressed in relation to real-world objects and processes, as in the following questions from the 2022 papers:

(1) 'Adam has a bag of fruit that weighs 1.25 kilograms. He takes out a banana. Now the bag of fruit weighs 1.1kg. Next, he takes out an orange. Now the bag weighs 920g. How much more does the orange weigh than the banana?' (Paper 2)
(2) 'The full price of a T-shirt is £15. The price is reduced by 30%. What is the reduced price?' (Paper 3).

Following KS2 SATs, which are normally held in the May of Year 6, students have a further two months of primary school, which is often non-academic in focus, and includes activities preparing for the transition. Their next encounter with mathematics as a formal school subject is likely to be in the following September, at the start of secondary school. As for other subjects, students do not have national assessments at the end of KS3, so KS3 tends to be treated as preparation for KS4 and the high-stakes GCSE examinations taken at the end of Year 11. This may mark a change in the tenor of mathematics lessons for many students.

## Learning mathematics and language

### *Mathematics, anxiety and the transition*

Underachievement in mathematics is a widespread problem in the UK and one which impacts negatively on many people's adult lives (Evans & Field,

2020a). Litster (2013) cites research showing that adult numeracy skills in the UK are weaker than literacy skills, with 8.1 million adults having low numeracy skills in 2011. She found that poor numeracy is associated with a number of negative life outcomes and with a lack of confidence, which often prevents people from seeking opportunities to improve their mathematics. There is also a significant cost to the national economy (Evans & Field, 2020a).

Mathematics as a subject is known to induce anxiety among large numbers of people, with records of mathematics anxiety going back centuries (Dowker et al., 2016). This can cause people to avoid mathematical tasks, and when they do undertake them, can lead to short-term memory overload (ibid). In school children, anxiety about mathematics has been found to increase at the time of the primary-secondary transition (Madjar et al., 2018), with a significant increase towards the end of primary education, which remains high for some time. Evans and Field (2020b) identify a number of adverse effects of the transition on mathematics achievement and attitudes towards the subject. Evans et al. (2018) found that students tend to have more negative attitudes towards mathematics and science after the transition to secondary school. In the Australian context, Deieso and Fraser (2019) also found less positive attitudes towards and less enjoyment of mathematics following the primary-secondary transition.

Rice (2001) found a decline in achievement in mathematics and science following the transition, which was exacerbated by academic push from teachers. This finding might be explained in terms of the psychological stress caused by academic pressure. In relation to pressure, Evans et al. (2018) point out that students in UK secondary schools are often assigned to numbered teaching groups (usually known as 'sets') on the basis of prior attainment or perceived ability, and they are more likely to be 'setted' in this way in mathematics than in other subjects, often from early in Year 7. In a detailed case study of a large UK secondary school, Neumann (2021) reports interviews with Year 8 students about their experience of setting by perceived ability, finding that students feel labelled, stereotyped and embarrassed, especially if assigned to lower ability sets, which were associated with low motivation. This widespread use of setting, albeit a practice that many teachers believe to be pedagogically essential, is likely to make the transition in mathematics additionally stressful.

KS3 students are likely to be taught mathematics by a non-specialist because there is a longstanding shortage of specialist teachers of mathematics in the UK (Allen & Sims, 2018). Only 44% of mathematics teachers overall have a degree in the subject, many having degrees in adjacent subjects such as physics and economics, and those with mathematics degrees are much more likely to be allocated to teach KS5, then KS4, than KS3 (ibid). KS3 mathematics teachers are also more likely to be inexperienced, that is, either without Qualified Teacher Status or within two years of gaining it, than teachers of KS4 and KS5 (ibid). Noyes (2012) argues that there is a wider range of quality and approach within the teaching of mathematics

in secondary schools than for other subjects, with low-ability sets often negatively impacted.

Noyes (2012) reports a survey of over 3000 Year 7 students from diverse schools; 19.9% of respondents said that mathematics was their least favourite subject – the highest score across all subjects. This was the case for both boys and girls but especially marked for girls, with 23% telling researchers that mathematics was their least favourite subject. Noyes (2012) found a correlation between the perceived quality of the teacher, which included subject knowledge, and enjoyment of mathematics. Evans et al. (2018) found that effective teaching makes a difference in motivation, and Noyes (2012) found the same for more student-centred practices, with a wide variation across and within schools in both practices and enjoyment. Studies such as the one we describe in this chapter, aiming to describe aspects of the communication systems of mathematics, hope to tackle these issues.

### *Talking about mathematics*

Several studies note the technical nature of the mathematics register but also argue that for school students this needs to be scaffolded by everyday explanatory language. Leung (2005) writes that technical mathematical language and informal language may be used complementarily to give children more access to the register of mathematics. He discusses a transcript from a Year 5 mathematics lesson and notes that pupils often used informal language to discuss mathematical ideas insightfully (Leung, 2005, p.128). Byrne and Prendeville (2020) also find that semi-formal talk plays a very important role in mathematics learning and supports the learning of technical language. They conducted research into the role of discussion in primary school pupils' learning about mass and weight, finding that peer discussion was associated with more accurate use of more specifically relevant mathematical vocabulary. Schleppegrell (2007) argues that mathematics, more than any other discipline in school, is dependent on teachers explaining and translating concepts alongside more technical language. Leung (2005) writes that over-prioritising technical mathematical vocabulary use too early could lead to the loss of some of this learning process.

## Features of the language of mathematics

Simpson and Cole (2015) note that the language of mathematics can be thought of as akin to a foreign language, involving knowledge of 'vocabulary, syntax, word order and abbreviations unique to mathematics' (2015, p. 370), as well as an understanding of audience and appropriacy.

### *Discourse*

Mathematics communication is multisemiotic (Wilkinson, 2019; Riccomini et al., 2015). That is, meaning is communicated in multiple ways in

mathematics: through written and oral language; through mathematical symbols; and through graphs and mathematical models (Wilkinson, 2019, p. 88; Schleppegrell, 2007). Students who are used to seeing non-verbal material in texts as having an illustrative function, and therefore not essential to understanding core concepts, will need to develop an understanding of how materials such as graphs and diagrams in fact work with text to create meaning in mathematics.

Mathematical assessment requires students not just to be able to provide answers to problems but to be able to communicate a series of logical steps to show how they reached their answers (Simpson & Cole, 2015). Within our spoken data, transcriptions of teacher talk in our partner schools, we have numerous examples of mathematics classes in which the teacher stresses the importance of this, such as the following:

| *line* | *speaker* | *utterance* |
|---|---|---|
| 227 | T039 | no it's the right it's all the right answers but you didn't show how you got that thirty (…) |

*Extract 7.1,* Year 7 mathematics lesson recording, Teacher 039.

| *line* | *speaker* | *utterance* |
|---|---|---|
| 214 | T007 | … in your book I wanna see working out (.) what do I want to see <name>? |
| 215 | student | [not transcribed] |
| 216 | | no I want to see the working out and then the answer yeah (.) Mr notorious for doing everything in his head and then just writing down the answer (.) okay? |

*Extract 7.2,* Year 8 mathematics lesson recording, Teacher 007.

| *line* | *speaker* | *utterance* |
|---|---|---|
| 227 | T068 | you have to take me through the step you can't just give me an answer that's a big no-no |

*Extract 7.3,* Year 7 mathematics lesson recording, Teacher 068.

This requires from students, first, contextual knowledge of the norms of mathematics discourse, and second, the ability to produce logical sequences of meaning, akin to those produced in well-organised writing.

The nature of the meanings communicated in mathematics materials seems to change with the transition. Candarli et al. (2019) conducted preliminary analyses on parts of the corpora compiled for this project, including a keyword comparison of KS2 and KS3 written mathematics texts. This analysis led them to conclude that the language of mathematics in KS2

realised the personal, imageable and concrete nature of the mathematical problems that are set, as can be seen above in examples from the 2022 SATs in mathematics. The analysis suggested that in the KS3 corpus, in contrast, the language realises technical and abstract problems, sometimes expressed through algebraic symbols, although the underlying mathematics might not be significantly more challenging. We explore this in more detail below.

### *Grammar*

Shanahan et al. (2011) report that mathematicians tell them that close reading, not reading for gist, is essential because every word can matter. This detailed and slow reading may be at odds with reading styles for other subjects; Shanahan et al.'s research compared expert reading styles for chemistry and history, neither of which demanded the same level of close reading. Schleppegrell (2007, p. 143) writes that in the register of mathematics, function words such as *more*, *less* and *as many as* can entail non-obvious concepts that need to be learned explicitly.

Complex and dense noun phrases have been noted in the mathematics register by a number of writers (e.g., Wilkinson, 2018; Schleppegrell, 2007). These sometimes express specialised concepts such as *area under a curve* (Wilkinson, 2018, p. 170). Schleppegrell (2007) gives examples of how complex noun phrases may then be part of longer clauses, which have to be carefully unpacked, such as *Sides of the triangle that are in the same positions are corresponding sides of a triangle* (2007, p. 144). Schleppegrell (2007) also notes conjunctions such as *if* and *when*, used in different ways from everyday life. In Candarli et al.'s (2019) keywords comparison of KS2 and KS3 mathematics corpora, *if* was key in the KS3 corpus (using KS2 as reference corpus). Examination of concordance lines in the KS3 mathematics corpus showed that the majority of occurrences were in problems that had been set using templates such as in the following examples:

(3) If one rectangle has twice the area of the other, find the length of the smaller rectangle. (Y8 worksheet)
(4) If there are 28 chairs in the classroom, how many tables are there? (Y7 worksheet)

These examples point to syntactic features of the mathematics register that may be challenging for students, and which teachers may not perceive, given that they involve high-frequency words.

### *Vocabulary*

Thompson and Rubenstein (2000) developed a detailed list of 12 potential pitfalls in learning the vocabulary of mathematics, which has been

widely used in studies of learning mathematics in various contexts, such as work by Riccomini et al. (2015). For most of the categories of pitfall, they give examples from five core areas of mathematics: number, algebra, geometry, statistics/probability and discrete mathematics. They note that the categories are not mutually exclusive, and a number might be found in a single lesson. Their final three categories concern technological terms, translation into languages other than English and abbreviations, which are less relevant to our research. The remaining nine categories are shown in Table 7.1, with some of Thompson and Rubenstein's examples (2000, p. 569).

It can be seen that categories 1, 2, 4, 7 and 8 all concern polysemy and homonymy in some way. Category 5 also concerns differences in meaning, as related to collocation. Polysemy was discussed in Chapters 2 and 6 of this book, and it features centrally in other studies of the language of mathematics, for example, in writing by Schleppegrell (2007), Wilkinson (2018, 2019) and Powell et al. (2017). It is also of concern to teachers, as seen, for example, in a blog post for teachers, in which Quigley (2021) discusses the number of synonyms that are used in schools and the wider world for talking about mathematical operations. He gives five alternative, nearly synonymous terms for *subtract/ion*: *minus*, *take away*, *take off*, *decrease* and *reduce*, which students need to be able to understand and switch between.

The notion of tiers of vocabulary (Beck et al., 2002) was discussed in Chapter 2 and also arose in Chapter 6. In their study of children's use of mathematical language, Powell et al. do not use the term 'tier', but the notion can be traced in their division of the vocabulary of mathematics (2017, p. 23). They describe three types of word, plus a fourth group, symbolic vocabulary, words such as *zero* and *equal*, which verbalise mathematics symbols. Their first lexical group is technical words, which have a meaning in mathematics only, such as *numerator*. This group corresponds to Beck et al.'s Tier 3. The second group is termed 'sub-technical', a term that is often used synonymously to Beck et al.'s Tier 2, especially in Higher Education. However, Powell et al. (2017) include polysemy over and above the usual understanding of Tier 2, writing that sub-technical words have a meaning in mathematics and a meaning in everyday language. Their examples include *round* and *regroup*. Their third group comprises words 'from everyday language that students encounter in mathematics (e.g., *more, longest*)' (2017, p. 23).

As was the case for science, reported in Chapter 6, while there have been a number of descriptions of isolated features of the language of school mathematics, there has not yet been a systematic, corpus-based study. The study reported in this chapter attempts to fill that gap, focusing specifically on the language around the transition.

In the second half of this chapter, we describe a series of studies that aimed to address the following research questions:

*Table 7.1* 'Vocabulary issues and examples', from Thompson and Rubenstein (2000, p. 569), with categories 11–12 and some examples omitted.

| | *Category of potential pitfall* | *Examples* |
|---|---|---|
| 1 | Some words are shared by mathematics and English, but they have distinct meanings. | *number*: power, prime, factor<br>*algebra*: origin, function, domain, radical, imaginary<br>*geometry*: volume, leg, right<br>*statistics/probability*: mode, event, combination<br>*discrete mathematics*: tree |
| 2 | Some mathematics words are shared with English and have comparable meanings, but the mathematical meaning is more precise. | *number*: divide, equivalent, even<br>*algebra*: continuous, limit, amplitude<br>*geometry*: similar, reflection |
| 3 | Some mathematical terms are found only in a mathematical context. | *number*: quotient, decimal, denominator, algorithm<br>*statistics/probability*: outlier, permutation |
| 4 | Some words have more than one mathematical meaning. | *number*: inverse, round<br>*algebra*: square, range, base, inverse, degree<br>*geometry*: square, round, dimensions, median, base, degree, vortex<br>*statistics/probability*: median, range<br>*discrete mathematics*: dimensions, inverse, vortex |
| 5 | Modifiers may change mathematical meanings in important ways. | *algebra*: root or square root, prime or relatively prime<br>*geometry*: polygon or regular polygon, bisector or perpendicular bisector |
| 6 | Some mathematical phrases must be learned and understood in their entirety. | *number*: at most, at least<br>*geometry*: if-then, if-and-only-if<br>*statistics/probability*: stem-and-leaf |
| 7 | Some words shared with science have different technical meanings in the two disciplines. | *number*: divide, density<br>*algebra*: solution, radical, variable<br>*geometry*: prism, degree, image, radian<br>*statistics/probability*: simulation, experiment<br>*discrete mathematics*: matrix, element, cell, tree |
| 8 | Some mathematical terms sound like everyday English words. | *algebra*: sine or sign<br>*geometry*: pi or pie, plane or plain |
| 9 | Some mathematical words are related, but students confuse their distinct meanings. | *number*: factor and multiple, hundreds and hundredths<br>*geometry*: theorem and theory |

1. What are the key lexico-grammatical features in KS3 mathematics in comparison with KS2 mathematics?
2. Which words are significantly more frequent in KS3 mathematics than in KS2 mathematics? To what extent, if any, do the functions and meanings of the top 30 keywords change in KS3 mathematics in comparison with KS2 mathematics?

## Method

### *The corpora*

In this study, we analysed the mathematics sub-corpus, which comprises all the written sub-registers of mathematics, and the spoken corpus of teacher talk in mathematics, as described in Chapter 3. As shown in Table 7.2, this sub-corpus is comprised of 1138 texts of 538,912 tokens in total. 'Text length' refers to tokens and SD stands for standard deviation. The number of texts and tokens is higher in the KS3 sub-corpus than in the KS2 corpus, and this difference is taken into account when we interpret the findings of this chapter.

### *Key feature analysis*

Key feature analysis is a relatively new corpus technique, proposed by Biber and Egbert (2018) as complementary to multidimensional (MD) analysis (described in Chapter 3 and used in Chapter 4). Key feature analysis focuses on grammatical features rather than individual word types or lemmas. Most previous studies in corpus linguistics have used keyness analysis to examine keywords (Gabrielatos, 2018), although key part-of-speech categories and key semantic domains have also been investigated (e.g., Culpeper, 2009). Biber and Egbert (2018) proposed the use of Cohen's *d* formula (see Biber & Egbert, 2018, for the calculation) to measure keyness for grammatical features since such features occur much more frequently than word types in corpora. The present study employs key feature analysis, utilising Cohen's *d* formula to investigate the positive and negative key lexico-grammatical features in the KS3 mathematics corpus in comparison with the KS2 mathematics corpus.

As we reported in Chapter 4, MD analysis did not capture any linguistic variation in the registers of mathematics across the key stages apart from two statistically significant findings, both concerning the discourse of the assessment sub-registers, which becomes (1) more explicit at KS3 than at KS2 and (2) more non-impersonal at KS3 than at KS2. However, it is possible that there is other linguistic variation that was not identified by the MD analysis. We therefore decided to conduct a key feature analysis of the two corpora to determine if

*Table 7.2* Sub-corpus of mathematics across the key stages.

| | *Key Stage 2* | | | | *Key Stage 3* | | | |
|---|---|---|---|---|---|---|---|---|
| | *Number of texts* | *Tokens* | *Mean text length* | *SD of text length* | *Number of texts* | *Tokens* | *Mean text length* | *SD of text length* |
| Written | 415 | 160,012 | 386 | 1086 | 694 | 245,698 | 354 | 373 |
| Spoken | 18 | 82,031 | 4557 | 2225 | 11 | 51,171 | 4652 | 1467 |
| Total | 433 | 242,043 | | | 705 | 296,869 | | |

there are positive and negative key grammatical features in KS3 mathematics in comparison with KS2 mathematics. For positive keyness, we used the threshold of $d \geq 0.20$; and for negative keyness, $d \leq -0.20$, for small effect size, following Cohen (1988). The grammatical features included both 67 linguistic features used in Biber's work (see 1988) and part-of-speech tags used in the Penn Treebank project (see Santorini, 1991). This combination allowed us to examine a wide range of features, including symbols and comparative and superlative adjectives. We used both the Multi-dimensional Analysis Tagger (MAT) v.1.3.2 (Nini, 2019), which utilises the Stanford parser (Toutanova et al., 2003), and #LancsBox 6.0 (Brezina et al., 2020), which uses Tree Tagger in order to tag grammatical features in our sub-corpus of mathematics. The latter tool has 96.36% average accuracy in tagging English corpora (Schmid, 1994). Linguistic features that require punctuation boundaries in order to be identified, including sentence relatives (Biber, 2019, p. 102), were removed from our analysis because the transcripts in our spoken sub-corpus were not punctuated, apart from question marks.

### *Keyword analysis*

Keyword analysis (Rayson, 2019) was used to identify words that are significantly more frequent in KS3 mathematics in comparison with KS2 mathematics, and vice versa, in the same way as was done in Chapters 5 and 6 for English and science. Then, through qualitative analysis, we aimed to find out whether the meanings and functions of these words are different between KS2 and KS3. As in Chapters 5 and 6, we used the lemma as our unit of analysis, so here the term 'keyword' refers to lemma. As previously, we used Cohen's *d* statistic because it takes into account dispersion (Brezina, 2018), with the threshold value for keyness set at > ± 0.20.

As in previous chapters, we also used the dispersion measure normalised 'deviation of proportions' ($DP_{norm}$) (Lijffijt & Gries 2012), to take into consideration to what extent words in our corpus are (un)evenly distributed across the individual files. There is a growing consensus that both keyness statistic and dispersion values need to be taken into account to produce a robust keyword analysis (Brezina, 2018; Egbert & Biber, 2018; Gries, 2021). $DP_{norm}$ generates a value from 0 (perfectly even distribution) to 1 (extremely uneven dispersion) and corrects for corpus parts that are different in terms of size (Lijffijt & Gries 2012), which is the case for all of our sub-corpora. We determined the threshold of $\leq 0.95$ for $DP_{norm}$ to filter out keywords that had an extremely uneven distribution, as for the studies in Chapters 5 and 6.

As in Chapter 5, #LancsBox v.6.0 (Brezina et al., 2020) was used to identify keywords and calculate Cohen's *d* and $DP_{norm}$ values. In Chapters 5 and 6, we removed all words in the top 200 *New GSL* (Brezina & Gablasova, 2015) in order to focus on topic-specific words. We did not take this step for the mathematics analysis. This was because the literature, as discussed earlier, suggested a high level of polysemy between mathematics

and everyday words including some very frequent words, which could be missed if the *GSL* words were discounted. We therefore removed a narrower group of words from our analysis, just symbols, cardinal numbers, list markers (e.g., (a.)) and function words (prepositions, pronouns etc). Symbols were removed because our key feature analysis in this chapter had already focused on them. We removed cardinal numbers, list markers and function words because we judged that their analysis would not give further insights into the language demands of mathematics. We do note that function words may offer useful insights into discourse analysis (McEnery, 2006). On the initial, unfiltered keyword list, there was a range of pronouns that were significantly more frequent in KS2 mathematics registers than in KS3 registers, but our key feature analysis had already captured this finding.

### *Concordance and collocational analysis*

The next step in our analysis involved reading the concordance lines of each keyword and identifying the meanings and functions of keywords in their context based on our reading of their discourse patterns in the expanded concordance lines. Previous studies have shown the prevalence of polysemy in the language of mathematics (e.g., Thompson & Rubenstein, 2000), and the keyword analysis alone would not help us to identify this. We therefore followed the keyword analysis with qualitative analyses of concordance lines of all the KS3 keywords that met the threshold for effect size, and the top 30 KS2 keywords, in order to identify their meanings and functions in context, as was done in Chapters 5 and 6. We used the KWIC (Key Words in Context) tool in #LancsBox 6.0 for this. We then examined collocational networks, using GraphColl in #LancsBox 6.0, because, as noted in Chapter 6, collocation is associated with polysemy. Collocation was also noted as a factor in one of Thompson and Rubenstein's (2000) potential pitfalls in mathematics vocabulary, discussed above.

## Findings

### *Key feature analysis*

Figure 7.1 shows the positive and negative key grammatical features of the KS3 corpus in comparison with the KS2 corpus.

As the last three rows in Figure 7.1 show, three grammatical features were represented more frequently in KS3 than KS2: nominalisations, present participle or gerund verb form and symbols. Examples of nominalisations in KS3 are *product*, *factorisation* and *probability*, shown below:

(5) Every positive integer can be uniquely expressed as a product of primes. (Y7 worksheet)

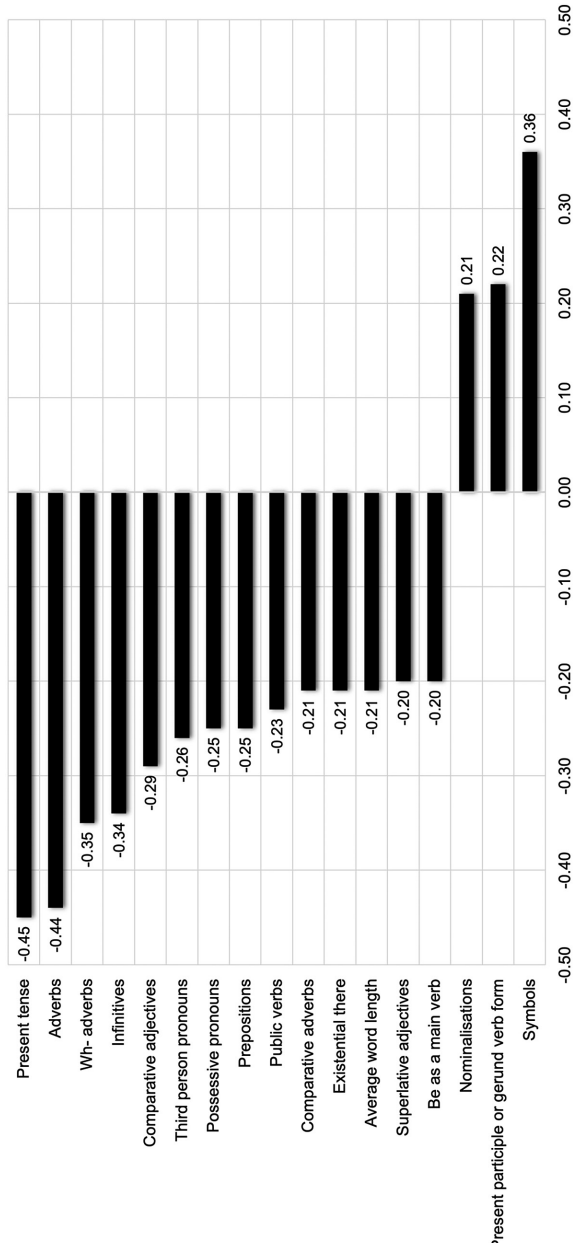



*Figure 7.1* Key grammatical features for the KS3 mathematics registers.

(6) Use prime factorisation to find the highest common factor and the lowest common factor of two numbers. (Y8 assessment)
(7) … calculate theoretical probability for events with equally likely outcomes. (Y8 presentation)

Cruz Neri and Retelsdorf's (2022) systematic analysis found an association between nominalisations and lower comprehension and performance in mathematics. The following example from a KS3 mathematics worksheet at KS3 shows the other two frequent features: a present participle verb form, *connecting*, and symbols *x, y* and =.

(8) A quantity y is inversely proportional to the square of a quantity x, and when x = 5, y = 4. If there is more information connecting the two quantities the constant of proportionality can be calculated… (Y7 worksheet)

The frequent use of symbols at KS3 is a reflection of the multisemiotic nature of mathematics at the secondary school level (Wilkinson, 2019) and the programme of study at KS3 since students learn operations and algebraic representations at KS3 (DfE, 2013b). The present participle verb form *connecting* is part of the complex noun phrase *information connecting the two quantities*. Students have to unpack the meaning of a complex syntactical construction to understand and then solve the exercise.

The top 14 rows in Figure 7.1 show negative key grammatical features, that is, features that are less frequent in the KS3 corpus than in the KS2 corpus. The less frequent occurrence of the features present tense, adverbs, infinitives, third person pronouns, prepositions, public verbs, existential *there* and *be* as a main verb together suggest that the language of KS2 mathematics is much more clausal than that of KS3. These clausal features and nominalisations occur in complementary distribution (see Biber & Egbert, 2018). That is, the use of nominalisations in KS3 mathematics language compresses information that might have been found in clauses in KS2. As Veel (2005, p. 184) writes, 'Nominalisation is the process by which events, qualities and relationships come to be represented not as verbs, adverbs or conjunctions, but as things, nouns'.

It is unsurprising that comparative adjectives and adverbs and superlative adjectives are negative key features in KS3 compared with KS2 since comparisons are one of the subject matters at KS2 (DfE, 2013a). Average word length is also a negative key feature of KS3 mathematics. This may be traced back to the highly frequent short symbols and notations, which might have reduced the average word length in the KS3 sub-corpus.

The clausal nature of KS2 language may also arise from its tendency to express problems in terms of real-world actions and events, discussed above. The following extract from a worksheet at KS2 includes present tense verbs (e.g., *buys*, *pays*), third person pronouns, and an adverb (*how*).

(9)

Sasha buys 5 lollies.
She pays with a £2 coin.
She receives 85p change.
How much does one lolly cost? (Y6 worksheet)

The occurrence of third person pronouns and active present tense verbs contribute to an imageable discourse that can be associated with everyday life experiences of students. Walkington et al. (2018) noted an association between the use of third person pronouns and higher performance in mathematics story problems. The frequent use of such linguistic features realises person-oriented, concrete[1] descriptions in KS2 mathematics, which is intended to provide relatable input to primary school students, as well as to prepare them for the KS2 SATs, which also present mathematics problems in such terms.

Taken together, our key feature analysis showed that there was a shift away from a clausal style, typical of conversation registers, in KS2, marked by key features, including present tense verbs, public verbs, *be* as a main verb, to, in KS3, a phrasal style, typical of academic prose and characterised by nominalisations, symbols and present participle or gerund verb forms. This finding suggests an increase in phrasal complexity of mathematics registers at KS3 in comparison with KS2. Phrasal complexity in academic written registers is associated with dense informational packaging (Biber et al., 2020). Nevertheless, the effect size of all the key features remains small in this study (Cohen, 1988). Our analysis of the negative key features, including third person pronouns and possessive pronouns in the KS3 corpus, together with the more frequent use of symbols in KS3 than in KS2, are indicative of a change from concrete, interpersonal discourse in KS2 to abstract discourse at the level of both lexis and content in KS3.

### *Results of keyword analysis*

Table 7.3 shows the keywords in KS3 mathematics, using the KS2 mathematics corpus as the reference corpus. Normalised frequency is per 1000 words. In the final column, we show our interpretation of the meaning and/or use of each lemma in the KS3 corpus, based on our study of concordance lines.

In the keyword analysis of KS2 mathematics using KS3 as a reference corpus, more than 30 keywords met the threshold for effect size, but as in Chapters 5 and 6, we only studied the top 30 to keep the qualitative analysis of meaning and use manageable. Keywords are shown in Table 7.4, ranked by Cohen's *d*. The final column shows our analysis of the meaning and/or use of each lemma in the KS2 corpus based on our study of extended corpus lines.

### *Discourse functions of keywords*

In KS3 mathematics, half of the keywords serve as discourse organisers, taking their primary function into account. This means that they are concerned

*Table 7.3* Keywords in KS3 mathematics with reference to KS2 mathematics, ranked by Cohen's *d*.

| *Rank* | *Lemma* | *Raw frequency* | *Normalised frequency* | $DP_{norm}$ | *Cohen's d* | *Meaning/use* |
|---|---|---|---|---|---|---|
| 1 | able | 350 | 1.18 | 0.77 | 0.38 | Discourse organisation: objective setting |
| 2 | apply | 93 | 0.31 | 0.90 | 0.37 | Self-assessment |
| 3 | learning | 77 | 0.26 | 0.90 | 0.36 | Discourse organisation: objective setting |
| 4 | extension | 126 | 0.42 | 0.90 | 0.33 | Discourse organisation: sequencing |
| 5 | mind | 54 | 0.18 | 0.89 | 0.32 | Discourse organisation: objective setting |
| 6 | expression | 230 | 0.77 | 0.89 | 0.32 | Topic-specific: algebraic thinking |
| 7 | unit | 319 | 1.07 | 0.85 | 0.31 | Topic-specific: measurement; Discourse organisation: sequencing |
| 8 | calculator | 177 | 0.60 | 0.86 | 0.31 | Discourse organisation: directive discourse |
| 9 | step | 517 | 1.74 | 0.85 | 0.31 | Discourse organisation: sequencing |
| 10 | solution | 151 | 0.51 | 0.89 | 0.27 | Topic-specific: problem solving |
| 11 | simplify | 261 | 0.88 | 0.86 | 0.26 | Topic-specific: algebraic thinking |
| 12 | example | 301 | 1.01 | 0.75 | 0.26 | Discourse organisation: exemplification |
| 13 | down | 305 | 1.03 | 0.68 | 0.25 | Discourse organisation: directive discourse |
| 14 | type | 106 | 0.36 | 0.88 | 0.24 | Topic-specific: geometry and data visualisation |
| 15 | learn | 98 | 0.33 | 0.86 | 0.24 | Discourse organisation: objective setting |
| 16 | happen | 82 | 0.28 | 0.87 | 0.23 | Topic-specific: reasoning and probability |
| 17 | student | 309 | 1.04 | 0.84 | 0.22 | Self-assessment |

(*Continued*)

*Table 7.3* (Continued)

| *Rank* | *Lemma* | *Raw frequency* | *Normalised frequency* | $DP_{norm}$ | *Cohen's* ***d*** | *Meaning/use* |
|---|---|---|---|---|---|---|
| 18 | term | 740 | 2.49 | 0.86 | 0.21 | Topic-specific: algebraic thinking |
| 19 | end | 205 | 0.69 | 0.75 | 0.21 | Discourse organisation: objective setting; topic-specific |
| 20 | worksheet | 99 | 0.33 | 0.85 | 0.21 | Discourse organisation: directive discourse |
| 21 | around | 36 | 0.12 | 0.84 | 0.21 | Topic-specific: estimation |
| 22 | next | 269 | 0.91 | 0.73 | 0.21 | Discourse organisation: sequencing |
| 23 | sum | 245 | 0.83 | 0.86 | 0.21 | Topic-specific: calculation |
| 24 | rule | 433 | 1.46 | 0.84 | 0.21 | Topic-specific: generalisation |
| 25 | section | 140 | 0.47 | 0.89 | 0.20 | 1. Discourse organisation: sequencing; 2. topic-specific: data visualisation |
| 26 | negative | 246 | 0.83 | 0.88 | 0.20 | Topic-specific: comparison and ordering |

with the procedures and expectations of the classes and their activities. The following example illustrates the discourse organisation function of *able* in objective setting.

(10) By the end of today's lesson, I will be *able* to use algebraic notation. (Y7 presentation)

In KS2 mathematics, keywords that functioned as discourse organisers constituted only a quarter of the overall keywords. The specific function was mostly to give students directions, as shown in the following example of *look*:

(11) *Look* at the shapes below. Do any of the shapes have the same area? (Y6 worksheet)

This comparative analysis suggests that there was a change in the overall communicative functions of the keywords at the discourse level from KS2

*Table 7.4* Keywords in KS2 mathematics with reference to KS3 mathematics ranked by Cohen's *d*.

| *Rank* | *Lemma* | *Raw frequency* | *Normalised frequency* | $DP_{norm}$ | *Cohen's d* | *Meaning/use* |
|---|---|---|---|---|---|---|
| 1 | how | 1739 | 7.18 | 0.37 | 0.44 | Topic-specific: counting and calculation |
| 2 | look | 253 | 1.05 | 0.63 | 0.42 | Discourse organisation: directive discourse |
| 3 | box | 338 | 1.40 | 0.67 | 0.41 | Topic-specific: concrete word problems; discourse organisation: directive discourse |
| 4 | answer | 965 | 3.99 | 0.54 | 0.40 | Assessment-oriented |
| 5 | explain | 269 | 1.11 | 0.68 | 0.39 | Topic-specific: reasoning |
| 6 | many | 925 | 3.82 | 0.47 | 0.39 | Topic-specific: counting and calculation |
| 7 | model | 168 | 0.69 | 0.80 | 0.38 | Topic-specific: problem solving |
| 8 | challenge | 170 | 0.70 | 0.81 | 0.38 | Discourse organisation: sequencing |
| 9 | mark | 819 | 3.38 | 0.85 | 0.38 | Assessment-oriented |
| 10 | correct | 383 | 1.58 | 0.61 | 0.33 | Assessment-oriented |
| 11 | much | 244 | 1.01 | 0.62 | 0.32 | Topic-specific: counting and calculation |
| 12 | complete | 289 | 1.19 | 0.74 | 0.31 | Discourse organisation: directive discourse |
| 13 | method | 348 | 1.44 | 0.65 | 0.31 | Topic-specific: calculation |
| 14 | buy | 159 | 0.66 | 0.81 | 0.30 | Topic-specific: concrete word problems |
| 15 | child | 172 | 0.71 | 0.84 | 0.30 | Topic-specific: concrete word problems |
| 16 | cost | 150 | 0.62 | 0.84 | 0.30 | Topic-specific: calculation |
| 17 | sheet | 250 | 1.03 | 0.72 | 0.29 | Discourse organisation: directive discourse |
| 18 | here | 474 | 1.96 | 0.51 | 0.28 | Discourse organisation: introduction |
| 19 | more | 349 | 1.44 | 0.53 | 0.28 | Topic-specific: counting and calculation |
| 20 | number | 2352 | 9.72 | 0.46 | 0.27 | Topic-specific: calculation |
| 21 | bar | 316 | 1.31 | 0.79 | 0.27 | Topic-specific: problem solving |
| 22 | calculation | 216 | 0.89 | 0.71 | 0.27 | Topic-specific: calculation |

(*Continued*)

*Table 7.4* (Continued)

| *Rank* | *Lemma* | *Raw frequency* | *Normalised frequency* | $DP_{norm}$ | *Cohen's d* | *Meaning/use* |
|---|---|---|---|---|---|---|
| 23 | leave | 251 | 1.04 | 0.59 | 0.27 | Topic-specific: concrete word problems |
| 24 | show | 404 | 1.67 | 0.58 | 0.27 | Topic-specific: data visualisation |
| 25 | say | 374 | 1.55 | 0.49 | 0.27 | Discourse organisation: explanation, exemplification |
| 26 | near | 186 | 0.77 | 0.81 | 0.26 | Topic-specific: rounding |
| 27 | half | 200 | 0.83 | 0.70 | 0.25 | Topic-specific: calculation |
| 28 | money | 170 | 0.70 | 0.83 | 0.23 | Topic-specific: concrete word problems |
| 29 | use | 910 | 3.76 | 0.41 | 0.20 | Discourse organisation: directive discourse; topic-specific concrete word problems |
| 30 | pay | 63 | 0.26 | 0.83 | 0.20 | Topic-specific concrete in word problems |

to KS3, and that there were different phases and activities in KS3 texts and lessons to a greater extent than in KS2. This is further evidenced by keywords indicating sequencing, including *step* and *next*, in KS3 mathematics. This difference seems unlikely to pose comprehension challenges for secondary school students, but it is an indicator of the number of activities that teachers fit into a one-hour lesson, and hence the pace and intensity of KS3 lessons.

There were only three assessment-oriented keywords – *answer*, *mark*, *correct* – in KS2 mathematics despite the emphasis on SATs preparations in lessons and written resources, and in KS3 there were two: *apply* and *student*. An example from KS2 is as follows:

(12) Husna's number is 306,042. She adds 5,000 to her number. What is her new number? Circle the *correct* answer. (Y6 worksheet)

There was an important difference in the meaning of these assessment-oriented keywords between KS2 and KS3. Whereas such keywords focused on receiving marks for correct answers in exercises and practice tests in KS2, probably due to the washback effect of SATs, the assessment-oriented keywords *student* and *apply* at KS3 were concerned with self-assessment and application of learning and reasoning to other contexts.

### *Patterns of meanings of keywords*

We analysed extended concordance lines for the content-specific keywords and found five groups of differences as follows: (1) part-of-speech categories; (2) concrete versus abstract keywords; (3) polysemous keywords in KS3 mathematics that were rare and/ or have different meanings in KS2 mathematics; and (4) collocational networks of keywords in two corpora. Each group is now discussed.

#### *Part-of-speech categories*

The first difference is concerned with the part-of-speech categories of keywords in KS2 and KS3 mathematics; 65% keywords ($n$ = 17) in KS3 mathematics were nouns, such as *solution*, compared with 47% ($n$ = 14) in KS2; 27% keywords ($n$ = 8) in KS2 mathematics were verbs, such as *buy* compared with 15% ($n$ = 4) in KS3. Nouns are typically found in academic prose and are one of the features of grammatical complexity in academic writing (e.g., Biber & Gray, 2016). This means that the language of KS3 mathematics may pose greater grammatical complexity than that of KS2 mathematics.

#### *Concrete and abstract keywords*

The second difference between KS2 and KS3 mathematics language is the use of concrete versus abstract keywords. In KS2, many more keywords are topic-specific, compared with KS3. In KS2, the majority of these topic-specific keywords referred to counting or calculation, or to concrete objects and processes in KS2 mathematics registers, as seen in the following example:

(13) Chen and Megan each buy a sandwich. Chen gets 5p change from £2. Megan gets £2.25 change from £5. How much more does Megan pay than Chen? (Y6 worksheet)

Topic-specific keywords in KS3 tend to be more abstract and are less likely to refer to everyday concepts, as seen in the following example, which includes the topic-specific keywords *simplify* and *expression*.

(14) Step 1: discover laws of indices for multiplication and division.
Step 2: discover laws of indices for powers and the zero index.
Step 3: simplify expressions using laws of indices. (Y8, presentation)

We note that *simplify* and *expression* are surrounded by other technical and abstract vocabulary, which may potentially create challenges for KS3 students in decoding (see Schleppegrell, 2007; Wilkinson, 2019). As for other language features that we have discussed, this indicates a shift from words for concrete objects and processes, such as *child*, *box*, *money*, *pay* and *buy*, in everyday contexts at KS2, to abstract concepts, processes and thinking at KS3.

*Polysemy*

The third difference in topic-specific keywords between KS2 and KS3 mathematics concerns polysemy. We found a number of polysemous words that had domain-specific meanings in the discipline of mathematics, including *expression*, *unit*, *solution*, *type*, *term* and *negative* in the KS3 mathematics corpus. This is consistent with Thompson and Rubenstein's first 'potential pitfall' for students 'Some words are shared by mathematics and English, but they have distinct meanings' (2000: 569, cited above). Cruz Neri and Retelsdorf's (2022) systematic review of the role of linguistic features provided strong evidence for a link between the occurrence of polysemous words and lower comprehension and performance in mathematics. With the exception of *unit* and *term*, these polysemous words were very infrequent in KS2.

The following example shows a typical use of *expression* in KS3 mathematics:

(15) Write an *expression* for the total area of the two congruent rectangles. (Year 7 presentation)

Because polysemous topic-specific keywords occurred very rarely in the KS2 mathematics corpus, we consulted the general corpus, BNCBM, which was described in Chapter 5, to identify other meanings of the *expression*. The most frequent meaning of the lemma in the BNCBM is a look on people's faces, occurring mainly in the fiction sub-corpus, followed by an idiom or turn of phrase, occurring most often in the speech sub-corpus. Examples are as follows:

(16) He had a strange expression on his face, half sad and half wild. (BNCBM, fiction)
(17) They [olives] don't have stones in. Is the expression 'pitted'? (BNCBM, speech)

There are no examples of *expression* in the sense of 'mathematical symbol' in the BNCBM, and it is very infrequent in the KS2 mathematics corpus. This suggests that this meaning is rare outside of secondary school and more advanced mathematics registers.

Occasionally, a topic-specific keyword identified in KS3 mathematics corpus tends to be used with an everyday, general meaning in KS2. The following example of *around* from KS3 shows the word used with the sense of 'approximately':

(18) Supposing that the population of the UK is around 60 million … (Y7 worksheet)

In the following KS2 extract, *around* is used with its everyday meaning, as an adverb denoting direction:

| *line* | *speaker* | *utterance* |
|---|---|---|
| 43 | T062 | … the value is what? six what? six what? turn around and tell your friend six what? (.) six what? (.) which column is the six in? what is the value of it? (…) okay year five (.) four three two one okay <name M> what's the value of the six in that number? |

*Extract 7.4,* Year 5 mathematics lesson recording, Teacher 062.

Usage-based approaches to language acquisition hold that users of a language need to be exposed to vocabulary and other linguistic features frequently in order to develop form-meaning relationships (e.g., Ellis & Wulff, 2015). Our corpus analysis suggests that students are unlikely to encounter the domain-specific meanings of most of these polysemous words until they start KS3, and it therefore seems likely that they will present a challenge.

### *Collocation*

The fourth difference between KS3 and KS2 mathematics language concerns collocation. Collocational networks are visual representations of 'networks of words that collocate with each other' (Brezina et al., 2015, p. 139), which help the analyst to detect the lexico-grammatical relationships in a corpus. Our collocational analyses of the top 30 keywords in KS2 mathematics suggested that such keywords collocated with a more restricted repertoire of words in KS2 than in KS3, despite their occurring significantly more frequently in KS2 than in KS3 mathematics. Due to space limitations, we will illustrate this by giving an example of *how*, which was the most frequent keyword in KS2 mathematics. We used the statistic LogDice with a value of at least seven, collocation frequency of at least five and a span of three words to the left and right of the node word, using #Lancsbox 6.0 (Brezina et al., 2020). As in Chapter 6, the LogDice measure, which identifies associations between words 'without the low-frequency bias' was selected because it is a standardised measure that does not depend on the corpus size (Gablasova et al., 2017, p. 165).

Figure 7.2 and Figure 7.3 illustrate the first-order collocational networks of *how* in KS2 and KS3 mathematics, respectively. The shade of the collocate indicates the frequency – darker shades indicating more frequent collocations than lighter shades. Distance between the node word and collocates indicates strength – shorter distances showing a stronger collocational relationship than longer distances (Brezina et al., 2020).

The graphs show that *how* co-occurs with *many* and *much* more strongly than the other collocates in both KS2 and KS3 mathematics. The other shared collocates are 'money', 'know' and a number of function words, such

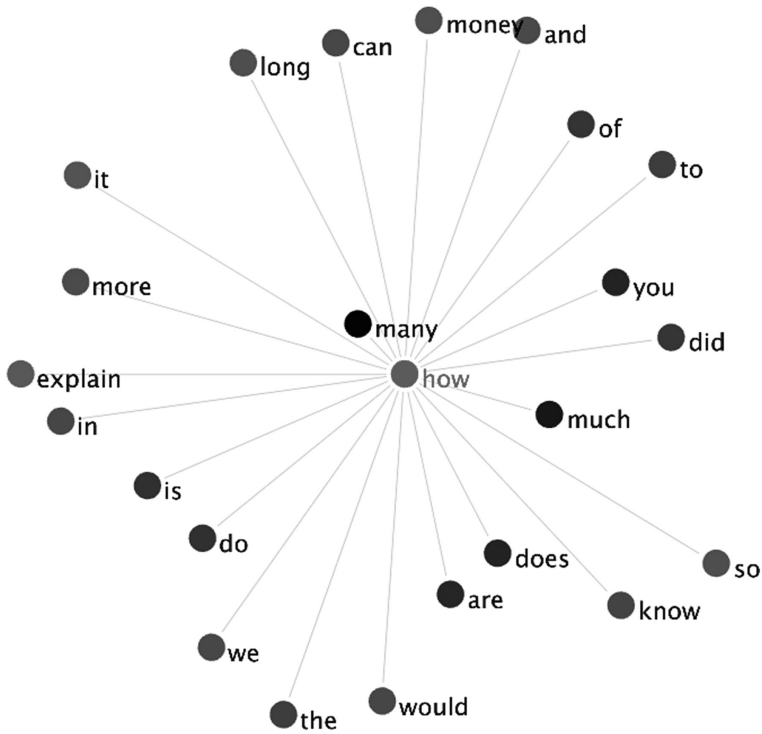



*Figure 7.2* The collocational network of 'how' in KS2 mathematics registers.

as 'of', 'you', 'do', 'does' and 'would'. Collocational networks are insightful in revealing both similarities and differences in discourses between the two corpora, suggesting differences within the similarity in this example. When we look at the differences, *how* co-occurs with a greater number of content words in KS3 mathematics, including *different*, *degrees*, *people*, *work*, *find*, *understand* and *objective*. At the discourse level, these collocates suggest that there was a focus on differences, objectives and understanding of mathematical concepts, processes and reasoning in KS3 mathematics, as example 17 illustrates. In KS2 mathematics, on the other hand, there was only one collocate, *explain,* that suggested a focus on mathematical reasoning, as can be seen in example (20).

(19) Objective:
Understand *how* to find missing term or pattern in sequences. (Y7 presentation)

(20) Amy completes the calculation 145 ÷ 6. She gets a remainder of 7. Explain *how* you know Amy is incorrect. (Y6 assessment)

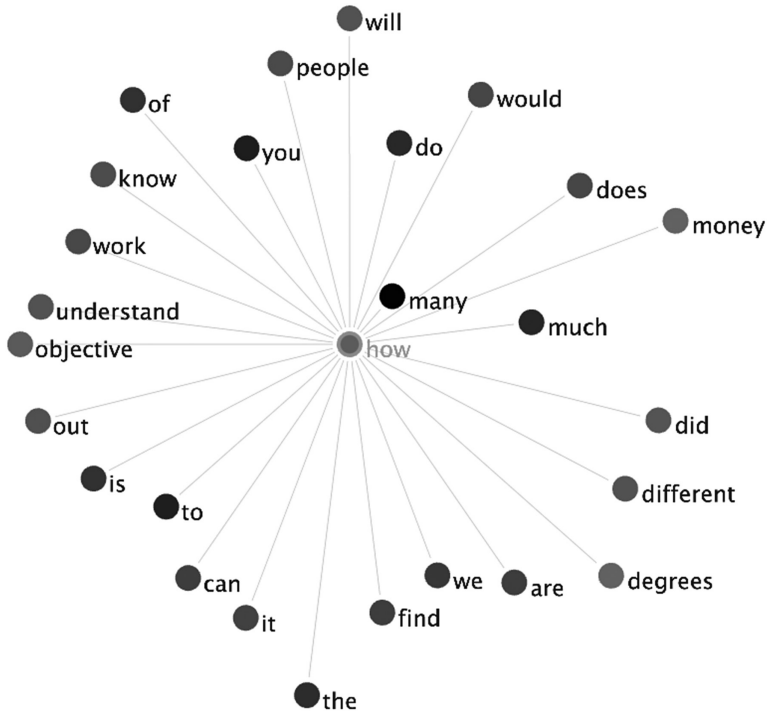



*Figure 7.3* The collocational network of 'how' in KS3 mathematics registers.

Examining collocates also demonstrates differences between the language of school and everyday language, even where the meaning difference is very subtle. *Happen** is a keyword in KS3 mathematics with reference to KS2 mathematics. When we examined its collocates using LogDice, with a value of at least 7.0, collocation frequency of at least five and span of five words to left and right, we identified five collocates. Table 7.5, produced using the GraphColl facility in #LancsBox 6.0, shows these.

*Table 7.5* Collocates of the lemma *happen* in the KS3 mathematics corpus, ranked by LogDice.

| *Rank* | *Position (L/R)* | *Collocate* | *LogDice* | *Freq as collocate* | *Freq in corpus* |
|---|---|---|---|---|---|
| 1 | L | event_n | 11.81 | 10 | 69 |
| 2 | M | probability_n | 9.05 | 6 | 349 |
| 3 | L | will_v | 8.33 | 7 | 692 |
| 4 | L | that_other | 7.3 | 5 | 1019 |
| 5 | L | can_v | 7.05 | 5 | 1211 |

A typical concordance line is:

(21) The probability that an event would *happen* is a number between 0 and 1. (Y8 presentation)

In the KS2 mathematics, there are only nine occurrences of *happen**, too few to calculate collocations. Nonetheless, the examples suggest a mathematical use, as in the following example:

(22) What has *happened* to the numerator? What do we notice? (Y6 presentation)

We used GraphColl to identify collocates of the lemma *happen* in the BNCBM, using the same parameters as for the search in KS3 mathematics. It is to be expected that in a much larger, general corpus, there would be a wider range of collocates that meet the thresholds set, as shown in Table 7.6.

We note that the top two collocates of *happen* in KS3 mathematics, *probability* and *event*, do not appear in the corresponding list for the BNCBM.

*Table 7.6* Collocates of the lemma *happen* in the BNCBM, ordered by LogDice.

| *Rank* | *Left/right* | *Collocate* | *LogDice* | *Freq as collocate* | *Freq in corpus* |
|---|---|---|---|---|---|
| 1 | L | gonna_v | 8.82 | 37 | 2151 |
| 2 | R | happen_v | 8.63 | 29 | 1871 |
| 3 | L | thing_n | 8.20 | 59 | 6037 |
| 4 | L | that_adv | 8.11 | 8 | 418 |
| 5 | L | wait_v | 7.79 | 13 | 1393 |
| 6 | L | will_v | 7.78 | 60 | 8397 |
| 7 | L | what_pron | 7.77 | 96 | 13862 |
| 8 | L | something_n | 7.77 | 29 | 3822 |
| 9 | L | would_v | 7.71 | 60 | 8865 |
| 10 | L | likely_adj | 7.6 | 5 | 315 |
| 11 | R | again_adv | 7.49 | 19 | 2928 |
| 12 | L | anything_n | 7.46 | 13 | 1883 |
| 13 | L | bad_adj | 7.43 | 12 | 1750 |
| 14 | L | exactly_adv | 7.42 | 8 | 997 |
| 15 | L | suppose_v | 7.35 | 7 | 874 |
| 16 | R | often_adv | 7.34 | 6 | 679 |
| 17 | L | sure_adv | 7.29 | 5 | 514 |
| 18 | L | let_v | 7.22 | 10 | 1668 |
| 19 | L | these_other | 7.22 | 16 | 2993 |
| 20 | L | allow_v | 7.21 | 6 | 793 |
| 21 | L | ever_adv | 7.18 | 10 | 1724 |
| 22 | R | if_con | 7.17 | 55 | 11978 |
| 23 | L | make_v | 7.16 | 39 | 8434 |
| 24 | M | funny_adj | 7.15 | 6 | 848 |
| 25 | L | never_adv | 7.11 | 15 | 3020 |
| 26 | L | this_other | 7.06 | 75 | 17934 |

Both lists show modality, but the KS3 list includes *can*, which is not in the list for the BNCBM. In the BNCBM, the frequency of markers of modality including *gonna*, *will*, *would*, *likely* and *if*, and the words *thing*, *something* and *anything*, can be traced to lines such as the following:

(23) It's just not *gonna happen*, is it if we're honest. (Speech)
(24) *If anything happens* to me, he should give it to them. (Fiction)

The adjectives *bad* and *funny* occur in examples such as:

(25) … something quite *funny happened* at work with that. (Speech)
(26) I'm only really here in case *something bad happens*. (e-language, SMS)

In the BNCBM, events that happen are generally unexpected, while in the KS3 mathematics corpus, they are predicted, in association with the study of probability. This is a very subtle difference in meaning, and unlikely to be problematic in itself. However, for KS3 students, it will be one of tens or hundreds of words that are used in a slightly different way from the one that they are familiar with on a daily basis in school.

## Conclusion

This chapter has identified some differences in the language of KS2 and KS3 mathematics. We found differences at the lexico-grammatical level, manifested through the reliance on nominalisations in KS3 mathematics, at the multisemiotic level, where the symbols are one of the key features in KS3 mathematics, at the semantic level, in polysemous words identified as keywords in KS3 mathematics and used with their specialised meanings, and at the discourse level, seen in the broad discourse functions of keywords in KS2 and KS3. At the level of lexis, we found a greater occurrence of abstract keywords in KS3 than in KS2 mathematics and showed how collocational networks of keywords point to differences in the broader discourses of KS2 and KS3 mathematics. These new insights into the language of mathematics at the transition stage suggest that there is a leap from KS2 to KS3 mathematics at multiple levels, which may pose challenges for students in terms of comprehension and performance in mathematics (see Cruz Neri & Retelsdorf, 2022). Increased awareness of these language demands in mathematics on the part of both students and teachers may support students' transition from primary to secondary school.

## Note

1 We thank Robbie Love for his preliminary analysis and finding on the greater concrete nature of KS2 mathematics than KS3 mathematics (see Candarli et al., 2019).

## References


Allen, R., & Sims, S. (2018). *How do shortages of maths teachers affect the within-school allocation of maths teachers to pupils?* Report for the Nuffield Foundation. https://www.nuffieldfoundation.org/wp-content/uploads/2018/06/Within-school-allocations-of-maths-teachers-to-pupils_v_FINAL21.pdf

Beck, I., McKeown, M. G., & Kucan, L. (2002). *Bringing words to life: Robust vocabulary instruction*. Guilford Press.

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

Biber, D., & Conrad, S. (2019). *Register, genre, and style* (2nd ed.). Cambridge University Press.

Biber, D., & Egbert, J. (2018). *Register variation online*. Cambridge University Press. https://doi.org/10.1017/9781316388228

Biber, D., & Gray, B. (2016). *Grammatical complexity in academic English: Linguistic change in writing*. Cambridge University Press.

Biber, D., Gray, B., Staples, S., & Egbert, J. (2020). Investigating grammatical complexity in L2 English writing research: Linguistic description versus predictive measurement. *Journal of English for Academic Purposes*, *46*, 100869. https://doi.org/10.1016/j.jeap.2020.100869

Brezina, V. (2018). *Statistics in corpus linguistics: A practical guide*. Cambridge University Press.

Brezina, V., & Gablasova, D. (2015). Is there a core general vocabulary? Introducing the new general service list. *Applied Linguistics*, *36*(1), 1–22. https://doi.org/10.1093/applin/amt018

Brezina, V., McEnery, T., & Wattam, S. (2015). Collocations in context: A new perspective on collocation networks. *International Journal of Corpus Linguistics*, *20*(2), 139–173. https://doi.org/10.1075/ijcl.20.2.01bre

Brezina, V., Weill-Tessier, P., & McEnery, A. (2020). #LancsBox v. 6.0. [software]. http://corpora.lancs.ac.uk/lancsbox

Byrne, J., & Pendeville, P. (2020). Does a child's mathematical language improve when they engage in cooperative group work in mathematics? *Education 3-13*, *48*(6), 627–642. https://doi.org/10.1080/03004279.2019.1636109

Candarli, D., Love, R., & Deignan, A. (2019, July). *Using corpora to explore the language challenges of the transition from primary to secondary school* [Paper presentation]. The 10th International Corpus Linguistics Conference 2019, Cardiff, UK.

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences*. Erlbaum.

Cruz Neri, N., & Retelsdorf, J. (2022). The role of linguistic features in science and math comprehension and performance: A systematic review and desiderata for future research. *Educational Research Review*, *36*(May), 100460. https://doi.org/10.1016/j.edurev.2022.100460

Culpeper, J. (2009). Keyness: Words, parts-of-speech and semantic categories in the character-talk of Shakespeare's Romeo and Juliet. *International Journal of Corpus Linguistics*, *14*(1), 29–59. https://doi.org/10.1075/ijcl.14.1.03cul

Deieso, D., & Fraser, B. J. (2019). Learning environment, attitudes and anxiety across the transition from primary to secondary school mathematics. *Learning Environments Research*, *22*, 133–152. https://doi.org/10.1007/s10984-018-9261-5

Department for Education (DfE). (2013a). *Mathematics programmes of study: Key stages 1 and 2*. Crown copyright. https://www.gov.uk/government/publications/national-curriculum-in-england-mathematics-programmes-of-study

Department for Education (DfE). (2013b). *Mathematics programmes of study: Key stage 3*. Crown copyright. https://www.gov.uk/government/publications/national-curriculum-in-england-mathematics-programmes-of-study

Dowker, A., Sarkar, A., & Looi, C. Y. (2016). Mathematics anxiety: What have we learned in 60 years? *Frontiers in Psychology, 7*(508), 1–16.

Egbert, J., & Biber, D. (2018). Incorporating text dispersion into keyword analyses. *Corpora*, *14*(1), 77–104. https://doi.org/10.3366/cor.2019.0162

Ellis, N. C., & Wulff, S. (2015). Usage-based approaches to SLA. In B. VanPatten, & J. Williams (Eds.), *Theories in second language acquisition: An introduction* (pp. 75–93). Routledge.

Evans, D., Borriello, G. A., & Field, A. P. (2018). A review of the academic and psychological impact of the transition to secondary education. *Frontiers in Psychology*, *9*(1482), 1–18. https://doi.org/10.3389/fpsyg.2018.01482

Evans, D., & Field, A. P. (2020a). Maths attitudes, school affect and teacher characteristics as predictors of maths attainment trajectories in primary and secondary education. *Royal Society Open Science*, *7*(September), 200975. https://doi.org/10.1098/rsos.200975

Evans, D., & Field, A. P. (2020b). Predictors of mathematical attainment trajectories across the primary-to-secondary education transition: Parental factors and the home environment. *Royal Society Open Science*, *7*(June) 200422. https://doi.org/10.1098/rsos.200422

Gablasova, D., Brezina, V., & McEnery, T. (2017). Collocations in corpus-based language learning research: Identifying, comparing, and interpreting the evidence. *Language Learning*, *67*(June), 155–179. https://doi.org/10.1111/lang.12225

Gabrielatos, C. (2018). Keyness analysis: Nature, metrics and techniques. In Taylor, C., & Marchi, A. (Eds.), *Corpus approaches to discourse: A critical review* (pp. 225–258). Routledge.

Gries, S. T. (2021). A new approach to (key) keywords analysis: Using frequency, and now also dispersion. *Research in Corpus Linguistics*, *9*(2), 1–33. https://doi.org/10.32714/ricl.09.02.02

Hutchings, M. (2015). *Exam factories: The impact of accountability measures on children and young people*. Report commissioned by the National Union of Teachers. https://schoolsweek.co.uk/wp-content/uploads/2015/07/Exam-factories.pdf

Leung, C. (2005). Mathematical vocabulary: Fixers of knowledge or points of exploration? *Language and Education*, *19*(2), 126–134. https://doi.org/10.1080/09500780508668668

Lijffijt, J., & Gries, S. T. (2012). Correction to Stefan Th. Gries' "Dispersions and adjusted frequencies in corpora", *International Journal of Corpus Linguistics*. *International Journal of Corpus Linguistics*, *17*(1), 147–149. https://doi.org/10.1075/ijcl.17.1.08lij

Litster, J. (2013). *The impact of poor literacy skills on adults: Research review*. Report for the National Research and Development Centre for Adult Literacy and Numeracy. https://maths4us.files.wordpress.com/2013/08/nrdc_impacts-of-numeracy-review_june13-m4u.pdf

Madjar, N., Zalsman, G., Weizman, A., Lev-Ran, S., & Shoval, G. (2018). Predictors of developing mathematics anxiety in school students: A 2-year prospective study. *International Journal of Psychology*, *53*(6), 426–432. https://doi.org/10.1002/ijop.12403

McEnery, T. (2006). *Swearing in English: Bad language, purity and power from 1586 to the present*. Routledge.

Neumann, E. (2021). Setting by numbers: Datafication processes and ability grouping in an English secondary school. *Journal of Educational Policy*, *36*(1), 1–23. https://doi.org/10.1080/02680939.2019.1646322

Nini, A. (2019). The multi-dimensional analysis tagger. In T. Berber Sardinha & M. Veirano Pinto (Eds.), *Multi-dimensional analysis: Research methods and current issues* (pp. 67–96). Bloomsbury Academic. https://doi.org/10.5040/9781350023857.0012

Noyes, A. (2012). It matters which class you are in: Student-centred teaching and the enjoyment of learning mathematics. *Research in Mathematics Education*, *14*(3), 273–290. https://doi.org/10.1080.14794802.2012.734974

O'Meara, N., Prendergast, M., Cantley, I., Harbison, L., & O'Hara, C. (2020). Teachers' self-perceptions of mathematical knowledge for teaching at the transition between primary and post-primary school. *International Journal of Mathematical Education in Science and Technology*, *51*(4), 497–519.

Powell, S. R., Driver, M. K., Roberts, G., & Fall, A.-M. (2017). An analysis of the mathematics vocabulary knowledge of third- and fifth-grade students: Connections to general vocabulary and mathematics computation. *Learning and Individual Differences*, *57*, 22–32.

Quigley, A. (2021). *Does reading *really* matter in mathematics?* https://www.theconfidentteacher.com/2021/04/does-reading-really-matter-in-mathematics/

Rayson, P. (2019). Corpus analysis of key words. In Chapelle, C. A. (Ed.), *The concise encyclopaedia of applied linguistics* (pp. 320–326). Wiley.

Riccomini, P. J., Smith, G. W., Hughes, E. M., & Fries, K. M. (2015). The language of mathematics: The importance of teaching and learning mathematical vocabulary. *Reading and Writing Quarterly*, *31*(3), 235–252. https://doi.org/10.1080/10573569.2015.1030995

Rice, J. K. (2001). Explaining the negative impact of the transition from middle to high school on student performance in mathematics and science. *Educational Administration Quarterly*, *37*, 372–400. https://doi.org/10.1177/00131610121969352

Santorini, B. (1991). *Part-of-speech tagging guidelines for the Penn Treebank Project*. Department of Computer and Information Science, University of Pennsylvania.

Schleppegrell, M. (2007). The linguistic challenges of mathematics teaching and learning: A research review. *Reading and Writing Quarterly*, *53*(2), 139–159. https://doi.org/10.1080/10573560601158461

Schmid, H. (1994). Probabilistic part-of-speech tagging using decision trees. In *Proceedings of the international conference on new methods in language processing*. https://www.cis.lmu.de/~schmid/tools/TreeTagger/data/tree-tagger1.pdf

Shanahan, T., Shanahan, C., & Misischia, C. (2011). Analysis of expert readers in three disciplines: History, mathematics and chemistry. *Journal of Literacy Research*, *43*(4), 393–429. https://doi.org/10.1177/1086296X11424071

Simpson, A., & Cole, M. W. (2015). More than words: A literature review of language of mathematics research. *Educational Review*, *67*(4), 369–384. https://doi.org/10.1080/00131911.2014.971714

Stone. (n.d.). *Maths timetabled hours*. https://drstonemaths.wordpress.com/2020/08/18/maths-timetabled-hours-part-2/

Thompson, D. R., & Rubenstein, R. N. (2000). Learning mathematics vocabulary: Potential pitfalls and instructional strategies. *The Mathematics Teacher*, *93*(7), 568–574.

Toutanova, K., Klein, D., Manning, C., & Singer, Y. (2003). Feature-rich part-of-speech tagging with a cyclic dependency network. In *Proceedings of the 2003 conference of the North American chapter of the association for computational linguistics on human language technology* (pp. 252–259). Association for Computational Linguistics.

Turan, E., & Smedt, B. De. (2022). Mathematical language and mathematical abilities in preschool: A systematic literature review. *Educational Research Review*, *36*(April), 100457. https://doi.org/10.1016/j.edurev.2022.100457

Veel, R. (2005). Learning how to mean-- scientifically speaking: Apprenticeship into scientific discourse in the secondary school. In Christie, F., & Martin, J. R. (Eds.), *Genre and institutions: Social processes in the workplace and school* (pp. 161–195). Bloomsbury Publishing Plc.

Walkington, C., Clinton, V., & Shivraj, P. (2018). How readability factors are differentially associated with performance for students from different backgrounds when solving mathematics word problems. *American Educational Research Journal*, *55*(2), 362–414. https://doi.org/10.3102/0002831217737028

Wilkinson, L. C. (2018). Teaching the language of mathematics: What the research tells us teachers need to know and do. *Journal of Mathematical Behavior*, *51*, 167–174. https://doi.org/10.1016/j.jmathb.2018.05.001

Wilkinson, L. C. (2019). Learning language and mathematics: A perspective from linguistics and education. *Linguistics and Education*, *49*, 86–95. https://doi.org/10.1016/j.linged.2018.03.005

# 8 Conclusion

*Alice Deignan, Duygu Candarli and Florence Oxley*

## Introduction

In this book, we have explored the language demands of the transition from primary and secondary school using corpus linguistics techniques. Our corpus-based analyses have indicated notable differences between the language of KS2 and KS3 registers at multiple levels, creating potential challenges for students at the beginning of secondary school. Throughout, our approach has been purely descriptive. As linguists, it is not our role to evaluate the language used by teachers, who have expertise in how children learn and in their subjects and disciplinary language. This chapter briefly reviews, contextualises and reflects on our key findings.

## Key issues and findings

### *The move from generalist to specialist teachers*

For the vast majority of school students in England, Year 7 is the first time that they will encounter subject specialist teachers. Most primary school teachers, while usually having a background in a particular subject, will work closely with one group of students for every subject across the timetable. This means that most primary school teachers may have more insight into what language is familiar to their pupils and what language might prove challenging than their secondary school colleagues. Secondary school teachers, by contrast, are responsible for a much larger number of students, whom they see for a much smaller proportion of their school lives. This limits opportunities for secondary school teachers to get to know each student and learn about what parts of the curricula they might find challenging. Additionally, there may be limited opportunities for secondary school teachers to discuss subject content and language with colleagues from other departments. There is an awareness of the need for primary-secondary and cross-disciplinary collaboration and cooperation, but this work requires funding, resources and time (Quigley, 2018, 2020).

DOI: 10.4324/9781003081890-8

### *Register features*

In Chapter 4, we presented a multi-dimensional (MD) analysis of the language in our KS2 and KS3 sub-corpora for the core disciplines, English, mathematics and science. While disciplinary language showed some similar features, this MD analysis revealed significant differences in the language that children encounter in each of these three disciplines. This aligns with well-established research and theory, which has posited that different academic disciplines have distinct registers (e.g., Biber & Conrad, 2019). In this study and those reported in Chapters 5–7, marked differences were also identified between the language used in KS2 and that used in KS3. For example, we found that science registers overall involved significantly more informational discourse at KS3 than KS2, suggesting an increasing phrasal complexity. Furthermore, in Chapter 4, some significant differences were found even within disciplines, at the level of the sub-registers. These sub-registers refer to the different types of written and spoken teaching material that students were exposed to within each subject and included resources like lesson presentations and worksheets. The following sections will outline our findings about some of the features of language that have emerged as reportedly or potentially challenging for transitioning students.

### *Polysemy*

One key issue that has surfaced in the language data in this book is polysemy. Across all the subjects that we studied in detail, we found that a central language issue was that vocabulary that students had encountered previously takes on new meanings at KS3. These may be more specialist, narrow and subtle, and in some cases, very different. Educated adults can perceive metaphorical relationships between KS2 and KS3 uses, and between everyday and KS3 uses. We would argue that the relationships are motivated but not predictable, say between *device* meaning a mobile phone and meaning a literary tool such as figurative language.

The keyword analysis presented in Chapter 5 compared the language used in KS2 English teaching, KS3 English teaching and a corpus representing everyday English and identified four specific patterns of polysemy. First, some words that were used frequently in specific contexts in KS2 and KS3 English teaching were used similarly frequently in everyday language but restricted to different contexts such as *feature*. Second, some words that were used frequently in KS3 English teaching and less frequently in KS2 English teaching were also used in everyday language, but only rarely, and more often with different meanings, such as *technique*. Third, some words carried more precise or nuanced meanings in KS3 English teaching than in KS2 English teaching or everyday language, such as *explore*. Lastly, in some cases, the meaning of a polysemous word was context-dependent, with accurate interpretation relying on the readers' ability to use collocates to

select the most contextually appropriate meaning, as in the case of *develop*. It seems likely that many KS3 students would lack the language and reading experience to do this accurately.

In the science study that we reported in Chapter 6, we found similar issues. We found five types of polysemy: (1) contextual differences; (2) fine-grained differences in use; (3) meaning differences; (4) lexico-grammatical differences; and (5) frequency differences. The first three groups are sections on a cline with, at one end, words that have meanings whose difference is barely perceptible, and, arguably, a product of context alone. This does not mean that students will find these unproblematic, however. Typical context primes us to expect particular words and interpretations (Hoey, 2005), so words that are encountered in an unfamiliar context may need additional processing time. At the other end of this cline are words whose meanings are very clearly distinct, and in the middle group, more finely split differences. The allocation of a specific word to one of these groups is less important than an understanding of the range and extent of the issue. Our exploration of meaning in the science corpus convinced us that far from being a marked phenomenon seen in a few interesting cases, difference in word meaning between science and everyday registers is the norm.

We found a similar pattern in our study of mathematics – reported in Chapter 7. Our concordance and collocation analyses showed numerous examples of polysemy, which ranged from words which take on specialist meaning in a mathematical context through to words which have very different meanings in mathematics in KS2 and everyday life.

### *Other language issues*

Each of the three disciplines that we studied presented some unique problems. In English, there was a marked change in the most frequent words, which we traced to a shift in orientation. KS2 focused on language analysis, and a view of text as an object to be understood. The most frequent words in KS3 show a very different approach to texts, as artefacts that have been created for a purpose, for an audience and which have intended and actual effects.

Our study of science discourse showed a stronger continuity of approach across KS2 and KS3, but a significant increase in volume of material. This is reflected linguistically in the volume of new vocabulary that students face in Year 7 by comparison with Years 5 and 6, indicated by the numbers of frequent word types and keywords in our KS3 science corpus. This is far higher than comparable findings for English and mathematics. As the corpora for different subjects are not exactly the same sizes, it is not possible to be precise about the exact extent of the difference, and further research with more calibrated corpora would be revealing.

Our comparison of KS3 and KS2 mathematics language began with key feature analysis, which showed us that KS2 is more clausal than KS3;

explanations and problems are expressed using features such as present tense verbs, adverbs and third person pronouns. KS3 contains comparatively more symbols, present participle verb forms – found in dense nominal groups – and nominalisations, all indicating a dense, non-clausal style, which is likely to be new to many students.

## Context

### *Awareness of the linguistic challenges of transition*

Meston et al. (2020) have written about how teachers and students have different understandings of the purpose of academic language and conversation, meaning that children may often not understand why teachers use academic language in the ways that they do. Nagy and Townsend (2012, p. 93) have argued that, more than simply acquiring new technical vocabulary, becoming proficient with academic language involves learning about new and more complex jobs that language can do and how to do them. This requires some quite sophisticated reorganising of children's knowledge of language in order to accommodate new concepts, like grammatical metaphor, and relationships between objects and ideas, for example, taxonomic relationships in science. While some features of academic language may seem obvious or intuitive to adult teaching staff, who are already proficient users of academic language, these same linguistic features may not be transparent or easy to grasp for children, whose conceptual understanding and academic language skills and knowledge are still developing.

By investigating changes in academic language and how it is used during the transition, we hope to increase awareness of how and why some children might struggle to access learning and stay engaged, with a view to informing how students can be supported in the future. Findings from this project have already been, and continue to be, shared with teachers and school leaders, who are enthusiastic about the topic, recognising the issues we highlight. It thus represents an application of corpus linguistic techniques to a societal issue, adding to studies that contribute to areas such as healthcare (Semino et al. 2018) and university-level education (Nesi & Gardner, 2012).

### *Academic language and home learning environment*

Throughout this book, we have alluded to issues of social justice. We have repeatedly found language challenges that, it seems to us, are likely to be greater for students from less educated and literate family backgrounds. Some children are disproportionately disadvantaged by the ways that academic language is used in teaching. Serbin et al.'s (2013) quantitative study identified parental support as a factor in academic success at transition, which helped students overcome otherwise disadvantaged backgrounds. They also suggested that girls' relative success over boys could be linked to

different parenting styles towards girls and boys. They argue that parental support might not be available due to work schedules and family issues (p. 1344), in such cases, tailored support such as homework help or after-school tutoring could help. Such questions are beyond the scope of our descriptive linguistic analysis, but we hope that our findings will help to focus support for students who struggle with academic registers.

### *Understanding the purpose of academic language*

Phillips Galloway et al. (2015) conducted interviews about academic language with 23 fourth to eighth grade students (Years 5–9 in England and Wales). The students tended to make value judgements about academic language, and relating it to social norms, in phrases such as 'more proper' and 'finer words' (p. 228). In this and a related study, the researchers found almost no reference to academic language being useful for communicating complex ideas or for communicating school subject matter. Rather, 'the intents behind the use of academic registers were associated with portraying a positive image both cognitively ("smart") and socially ("nice")' (2015, p. 230). Meston et al. (2020) coded interviews with teachers and students about academic conversation and examined how both groups perceived the purpose. They found a number of divergences, the biggest being comments that they classified 'practising social norms', which was mentioned significantly more by students than by teachers. This is consistent with the interviews that we described in Chapter 1, in which students told us repeatedly about the need to use 'good' words and 'upgrade' their vocabulary to more formal, 'posher' words. These findings suggest that many students do not understand the connection between academic language and its purpose, perhaps having an impoverished view of it as ornamental and serving a purely social function. This will not help them with the subtle and specialised meanings that these registers can convey.

## Research on school language and transition

Our main contribution in this book has been to bring multiple corpus linguistics techniques, including multi-dimensional analysis, keywords, key feature, concordance and collocation analysis, to research the variation within school language registers in a systematic way. This approach allowed us to capture variation between KS2 and KS3 registers, which would be impossible to trace by using a single method. Our second contribution has been to examine school language registers from a 'transition' angle and describe discipline-specific registers in a more fine-grained manner than has been possible in the past. Mathematics registers, for instance, are characterised by technical vocabulary and their multisemiotic nature in the literature (e.g., Wilkinson, 2019). However, we found that this characterisation applies to KS3 mathematics rather than KS2 mathematics. We further noted

important variations within different sub-registers of a single discipline. Finally, we have been able to compare the challenges of three core academic subjects and have found that while they have issues in common, notably polysemy, there are specific challenges for each.

## Future research and ways forward

This project has aimed to gather knowledge and information about when and how the language that children encounter in school changes over the course of the transition.

In conducting our research, we have constructed a large and versatile corpus of spoken and written school language data, comprising disciplinary sub-corpora and student and teacher interviews, teacher talk in lesson recordings, worksheets, lesson presentations, reading extracts, textbooks, assessment documents, poems, plays and fictional books. As mentioned in Chapter 3, to our knowledge, ours is the first corpus constructed using such a diverse range of source material and disciplines and containing such a high token count. Alongside the studies presented in this book, numerous other studies are currently being conducted, examining the language of the science and lesson presentation sub-corpora and pupils' own self-reported experiences of school language during transition, among others. Our corpus is limited to school contexts in England; however, the potential language challenges of the transition from primary to secondary data may be applicable to other countries given that 'international data are consistent in revealing a "dip" in attainment following transfer to secondary school' (West et al., 2010, p. 24). Therefore, further research on school language registers is needed at the transition from primary to secondary school in different countries.

The work of the project with school practitioners is ongoing and aims to achieve more widespread awareness of the language issues that face all children, but especially those whose first language is not English, those from lower SES backgrounds and those with additional needs. We are attempting to move awareness from a word-list approach, and towards an understanding that often the issue is not new words, but unfamiliar uses of known words in new structures and contexts. As noted earlier, discussions of genre and register with school students suggested that they have a simplistic understanding, sometimes formulated in terms of 'posh' words versus words to be used in everyday contexts. We aim to promote a much more nuanced and non-evaluative understanding of genre and register. We were struck early on in the research by the sheer volume of language that students encounter in every hour of every day when they arrive at secondary school. Making sense of each new use of a word or lexico-grammatical structure, each unexpected approach to information organisation or interactional demand might be straightforward in itself, but multiplied and added to the emotional strains of the transition, the language of secondary school presents a significant challenge.

## References

Biber, D., & Conrad, S. (2019). *Register, genre, and style* (2nd ed.). Cambridge University Press.

Hoey, M. (2005). *Lexical priming: A new theory of words and language*. Routledge.

Meston, H. M., Phillips Galloway, E., & Brown McClain, J. (2020). 'They're the ones who hold the answers': Exploring educators' and students' conceptions of academic conversation. *Journal of Adolescent and Adult Literacy*, *64*(4), 409–491.

Nagy, W., & Townsend, D. (2012). Words as tools: Learning academic vocabulary as language acquisition. *Reading Research Quarterly*, *47*(1), 91–108. https://doi.org/10.1002/RRQ.011

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

Phillips Galloway, E., Stude, J., & Uccelli, P. (2015). Adolescents' metalinguistic reflections on the academic register in speech and writing. *Linguistics and Education*, *31*, 221–237.

Quigley, A. (2018). *The "Reading Gap" between primary and secondary school*. https://www.theconfidentteacher.com/2018/03/the-reading-gap-between-primary-and-secondary-school/.

Quigley, A. (2020). *Thinking about school transition*. https://www.theconfidentteacher.com/2020/05/thinking-about-school-transition/

Semino, E., Demjen, Z., Hardie, A., Payne, S. A., & Rayson, P. E. (2018). *Metaphor, cancer and the end of life: A corpus-based study*. Routledge.

Serbin, L. A., Stack, D. M., & Kingdon, D. (2013). Academic success across the transition from primary to secondary schooling among lower-income adolescents: Understanding the effects of family resources and gender. *Journal of Youth and Adolescence*, *42*(9), 1331–1347. https://doi.org/10.1007/s10964-013-9987-4

West, P., Sweeting, H., & Young, R. (2010). Transition matters: Pupils' experiences of the primary–secondary school transition in the West of Scotland and consequences for well-being and attainment. *Research Papers in Education*, *25*(1), 21–50.

Wilkinson, L. C. (2019). Learning language and mathematics: A perspective from linguistics and education. *Linguistics and Education*, *49*, 86–95. https://doi.org/10.1016/j.linged.2018.03.005

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

# Index