
Title	Tracing disciplinarity in primary and secondary teachers' talk: A corpus-based study
Author(s)	Paul Doyle
Source	<i>3rd Redesigning Pedagogy International Conference, Singapore, 1 - 3 June 2009</i>

This document may be used for private study or research purpose only. This document or any part of it may not be duplicated and/or distributed without permission of the copyright owner.

The Singapore Copyright Act applies to the use of this document.

TITLE OF PAPER:

TRACING DISCIPLINARITY IN PRIMARY AND SECONDARY TEACHERS'

TALK: A CORPUS-BASED STUDY

Paul Doyle

National Institute of Education, Nanyang Technological University

Singapore

Paper presented at the 3rd Redesigning Pedagogy International Conference

June 2009, Singapore

Abstract

In this paper, I focus on recurrent word sequences, or ‘lexical bundles’ (Biber et al. 1999), as markers of disciplinary variation in a corpus of primary and secondary teacher talk.

Frequently occurring lexical bundles can be classified using functional categories such as epistemic stance expressions, modality and topic related discourse organising expressions (ibid). However, in order to account for variation in lexical bundle distribution across disciplines, there is a need for an interpretative framework that relates to a specific community of language users operating in a single genre (Hyland, 2008). Classroom talk is a hybrid discourse (Biber, Conrad and Cortes, 2004) that exhibits both the characteristic interpersonal features of spoken language and ‘literate’ features of written language from textbooks, and that is especially rich in lexical bundles.

Using data from the Singapore Corpus of Research in Education (Doyle and Hong, 2009), I trace variations in discipline specific pedagogic practices as evidenced in teacher talk from English medium lessons in English Language, Mathematics and Science in Singapore classrooms. Frequent lexical bundles are classified using a framework adapted from Hyland’s (2008) taxonomy, and the distribution of the various categories is compared across the three school disciplines. The approach is evaluated for its ability to relate linguistic variation to significant disciplinary differences, and to highlight processes of knowledge construction in the classroom.

TRACING DISCIPLINARITY IN PRIMARY AND SECONDARY TEACHERS' TALK: A CORPUS-BASED STUDY

Introduction

In this paper, I attempt to trace disciplinarity in teacher talk using a corpus of some 600 classroom transcripts collected between 2003 and 2005 in Singapore Primary and Secondary schools. This corpus, the Singapore Corpus of Research in Education (Doyle & Hong, 2009; Hong, 2005), includes lessons in English, Mandarin, Malay and Tamil, but for the purposes of this paper I will focus on English medium instruction only in the following subjects: English Language, Mathematics, Science and Social Studies. Before looking at this data closely, I will define what I mean by 'disciplinarity' and I will explain how this can be traced in a corpus of transcripts. Then, I will briefly explain corpus-based methods for discourse analysis and highlight the benefits such an approach offers researchers interested in learning more about pedagogy as it is enacted in classrooms. Finally, I will explain the concept of a lexical bundle (Biber et al, 1999) which is the principle framework I use for investigating disciplinarity in the teacher talk.

Background

This section provides some theoretical background to the study in three areas: the concept of disciplinarity, corpus-based methods of discourse analysis, and the linguistic concept of lexical bundles.

Disciplinarity

What is disciplinarity? A sociological view would be that it is the distinctive epistemological configurations of knowledge claims, evidence bases and argument structures, and what is taken as a 'fact', which make up what we call 'science', 'humanities',

mathematics', and so on. From an applied linguistic view, it has been partially aligned with the concepts of register and genre, especially as those concepts have been developed by Michael Halliday and his followers (Halliday, 1994; Martin, 1990; Christie, 2000) within systemic functional linguistics. Recent work by followers of Bernstein and Halliday has evolved a rich discussion of disciplinarity in the context of pedagogy and language (Christie & Martin, 2007), with a particular focus on educational discourses. A key part of this, at least for systemic functional linguists and register/genre theorists is examination of classroom discourse (e.g. Christie, 2002; Schleppegrell, 2004).

Corpus-based methods of discourse analysis

In recent years in applied linguistic research, the power and value of the corpus as an explanatory tool has grown exponentially, informing first work in lexicography (such as the COBUILD project led by Sinclair, 1987) grammar, and language description (Biber et al, 1999), but subsequently penetrating many different sub fields of the discipline, including discourse analysis (Baker, 2008; Stubbs, 2006). One of the claims that corpus linguists make is that their work is empirical: it takes a scientific approach to the analysis of the language we speak and use, rather than the psycholinguistic appeal to 'intuition' and 'native speaker' inherent knowledge of language. When a corpus of sufficient size and diversity to be considered representative and balanced in terms of the language or language genres it is attempting to investigate is compiled, we can be confident that we are taking an empirical, scientific approach, not least because the studies we do will be replicable. Corpus-based studies of language quickly show us what is frequent and typical about language use in a way that is verifiable from the factual data: the words that people actually speak and write, rather than what they *think* they speak and write.

Thus, in this paper, I use the words spoken by teachers to investigate what they do when they teach their subject disciplines. The sample I have is representative: 455 transcripts,

representing 121 teachers at Primary 5 and Secondary 3 level, teaching full units of work in four core curriculum subjects.

Lexical bundles

The means by which I will attempt to investigate disciplinarity is through *lexical bundles*. A lexical bundle is a sequence of three, or four, or five, or more, contiguous words in the corpus. A detailed treatment of lexical bundles can be found in Biber et al (1999). Lexical bundles, also termed ‘multiword sequences’ (Stubbs, 2007) or ‘clusters’ (Scott, 2006; Mahlberg, 2007) or recurrent multiword combinations (Altenberg, 1998), are a feature of the fuzzy, conventional aspect of language, otherwise known as its ‘phraseology’. Example 1 shows a typical occurrence from the SCoRE data:

Example 1 I don’t know how¹

This example shows two features that are typical of lexical bundles. Firstly, they are syntactically incomplete. Lexical bundles typically cross phrase or clause boundaries. In Example 1, the first clause “I don’t know” is followed by “how” which is the beginning of a second dependent clause (for example, “how to answer that”).

The second feature of lexical bundles becomes apparent when we compare them with the similar category of fixed expressions. These include idioms and other phrasal language constructions. True idioms, however, are usually infrequent and are normally semantically opaque, that is, they have metaphorical meaning: e.g. “A bird in the hand is worth two in the bush”. Idioms are striking and memorable – but corpus studies show us they are really quite

¹ In this paper, lexical bundles are shown by underlining.

rare. Lexical bundles, in comparison, are quite unremarkable, but they are extremely frequent, especially in spoken discourse. In the SCoRE data, some 20 – 25% of the words in our corpus are found within lexical bundles comprising four words. Sinclair (2004) has claimed that as much as 80% of spoken text may be made up of phrasal language, and if this is so then it is clearly important to examine how this plays out in teachers' talk in classrooms.

Classifying lexical bundles

Lexical bundles can be classified in terms of their formal linguistic aspects and their functional role in texts. I will focus on the latter here, as I believe this has more efficacy for the exploration of disciplinarity, and fits better with the systemic functional linguistic approach to classroom discourse adopted by Christie (2002), Schleppegrell (2004) and others.

Biber, Conrad & Cortes (2004) identify three main discourse functions for lexical bundles: stance expressions, discourse organizers and referential expressions. Stance expressions carry epistemic evaluations or attitudinal / modality meanings: epistemic, desire, obligation, intention/prediction, and ability. Some examples of these categories from the SCoRE data are:

1. *Epistemic*

I don't know whether we have time to complete the discussion.

2. *Desire*

I want you to elaborate on each point.

3. *Obligation (directive)*

and you have to think what started the fire.

4. *Intention/prediction*

Today, class, *we are going to* learn about formation of, er, Malaysia.

5. *Ability*

You have to be able to identify the different parts.

Discourse organisers indicate overall discourse structure and to signal the informational status of statements: topic introductions, topic elaborations/clarifications, and identification/focus. Examples from the SCoRE data are:

1. *Topic introduction*

So, if you look at the main seed, you realise it's not like a bean seed.

2. *Topic elaboration/clarification*

What do you mean by do more with less?

3. *Identification/focus*

This is the one which is semi-permeable.

Finally, referential expressions identify an entity or single out some particular attribute of an entity as especially important and include: imprecision indicators, attribute specifiers, and expressions relating to time/place/text – deixis. Examples in the SCoRE data are:

1. *Imprecision*

This is probably some connection to the lungs or something like that lah

2. *Attribute specifiers*

(a) quantity:

How many of you are done with simple machines?

(b) tangible:

If I blow up the atom to the size of the stadium...

(c) intangible:

Okay, in the case of the African tulip seeds, what's the method of dispersal?

3. *Time/place/text deixis*

you are applying a force at the same time

A fourth category that I would like to add, which is particularly prevalent in this data are formulaic phrases with special conversational functions

Politeness formulae:

good morning class good (morning)

thank you class thank (you)

thank you very much

Significance

In recent years, a number of scholars have used lexical bundles as a means of exploring linguistic variation in corpora (Hyland, 2008a; Csomay, 2007, 2005; Biber & Barbieri, 2007; Mauranen, 2006; Scott & Tribble, 2006; Simpson-Vlach, 2006; Biber, Conrad & Cortes, 2003) and in particular the language of the classroom (Csomay, 2007; Thompson, 2006; Biber, Conrad & Cortes, 2004;) and student writing (Hyland, 2008b; Cortes, 2006, 2004).

Biber, Conrad and Cortes (2004) studied the occurrence lexical bundles in university registers: tutorials and textbooks, and found some interesting similarities and differences. One of their key findings was that classroom talk (here referring to university tutorials) contains more lexical bundles than either lectures or textbooks alone, and that this can be explained by the fact that lexical bundles characteristic of these spoken and written pedagogical modes both occur in classroom talk. Lexical bundles found in spoken registers are typically centred on pronouns and clause fragments (a verb) – for example, you want me to, whereas those found in written registers tend to be centred on nouns – for example, the nature of the and as a result of. Biber, Conrad and Cortes (2004) found that classroom discourse contains *both* types of lexical bundle. They argue that classroom teaching displays

similarities to textbook language, at least at the university level, by containing lexical bundles more typical of academic written registers, such as referential expressions (or something like that), as well as containing lexical bundles with discourse organising functions not found in textbooks or conversation, such as want to talk about.

Nesi and Basturkmen (2006) used the BASE corpus to look at the discourse signalling role of lexical bundles. They found broad similarities with Biber, Conrad and Cortes (2004) in terms of the top twenty lexical bundles occurring in lectures, and ample use of ‘topic introduction/focus’ lexical bundles.

Hyland (2008) has also looked at the role in academic discourse. His particular focus was on written academic registers and the extent to which lexical bundles varied by discipline, in a corpus containing research articles, PhD theses and MA/MSc dissertations from four disciplines: electrical engineering, microbiology, business studies and applied linguistics. He found “considerable differences” (Hyland, 2008a: 11) in the occurrence and frequency order of the top 50 four-word lexical bundles across the four disciplines: “over half the items in each (frequency) list do not occur at all in any other discipline and only 30% of the strings in each discipline are found in two other fields” (*ibid*: 12). In addition, he found only 5 of the four-word lexical bundles occurred in all four disciplines and only 14 in three disciplines (Hyland, 2008a: 12-13).

A key question for us, then, is to examine to what extent primary and secondary school academic spoken discourse shares these characteristics with tertiary level registers.

Methodology

To find the most frequent lexical bundles in the SCoRE data, *Wordsmith Tools* software (Scott, 2006) was used to produce a word list index for all 455 transcripts. This index enabled the use of the cluster tool in *Wordsmith Tools*, which allowed me to determine

the length of the lexical bundles, the minimum frequency for inclusion in the resulting list of lexical bundles, and a method for determining whether a lexical bundle would be ‘delimited’ in any way – this was set to ‘sentence breaks’ for the purpose of this study. I created lists of 3, 4 and 5 word-long lexical bundles for each subject sub-corpus: English Language, Mathematics, Sciences. and Social Studies. These lists were then exported as *Microsoft Excel* format files, so that I could use that program’s statistical, graphical and list management tools to manipulate and then analyse the lists of lexical bundles.

In addition to these lists of lexical bundles, I also carried out a lexical bundles analysis using Wordsmith Tools’ keywords feature. A keyword is a word that has a significantly higher frequency in a target corpus or text when compared to a reference corpus. Frequencies for the item in both corpora are compared and a ‘keyness value’ is calculated using a statistical procedure known as Dunning’s Log likelihood function (Dunning, 1993; Oakes, 1998:172). The significance of keyword comparisons is that the items revealed as key are those most typical of the target text, “what the text is really about, avoiding trivia and insignificant detail”. (Scott & Tribble, 2006: 56). In this case, the keyword comparison was conducted on the frequency lists for the lexical bundles, following Mauranen (2006: 279), producing a list of key lexical bundles for each subject. In each case, the reference corpus was the combined transcripts from the other three subjects. Thus, this analysis reveals what is distinctive, in terms of lexical bundles, about each subject.

Following Biber, Cortes and Conrad (2004) and Hyland (2008), it was decided to focus on the four-word lexical bundles. Others have argued for look at longer sequences, especially in written texts (Mahlberg, 2007), but it was felt that the interactional nature of classroom discourse would mitigate against success here. The longest lexical bundle in SCoRE, other than counting sequences, contains nine words: now this is what I want you to do. It has a frequency of 10 occurrences in just 7 transcripts.

Case study: zooming in and out of the data

Our analysis also showed us that one of the most frequent lexical bundles in the SCoRE data is the sequence one two three four – a counting sequence –which occurs 350 times. Furthermore, this sequence occurs in 135 transcripts out of 455 (29.7%) and across all four subjects. The discussion below shows how this lexical bundle was further analysed using the Concordance tool in *Wordsmith Tools*.

To explore this lexical bundle further, we use the concordance (see Figure 3 in Appendix 2) which shows all the occurrences in all the transcripts listed with partial context. Clicking on one of these lines in the concordance display will take us to the full context, i.e. the transcript itself, and allow us to analyse the function of that instance of the lexical bundle. Two examples, one from a Mathematics lesson (Figure 1) and one from an English Language lesson (Figure 2) are shown below.

Trn0517 spk	I want two pieces out of five, how many pieces must I take away?
Trn0518 spk2	[...] Two! three!
Trn0519 spk8	Two!
Trn0520 spk7	Okay! Out of five pieces of cake, I want two. I give away three. Alright! Right. So three take out! Okay. Out of eight pieces, I only want three. How many do I take away?
Trn0521 spk2	[...] Five!
Trn0522 spk7	<u>One, two, three, four</u> , five. So is three eighth bigger or two fifth bigger?
Trn0523 spk8	Two fifth.
Trn0524 spk9	Two fifth bigger.

Figure 1: Lexical bundles in a Maths lesson (SCoRE transcript: score_maths_012).

Trn0086	spk12	Now there are twenty seven, there are twenty seven of you. I'm going to divide the group in to three and I'm going to give you a number so that number one will go first, number two and so on okay.
Trn0087	spk15	Sit with Rafie.
Trn0088	spk12	Cannot. Why you in love with him?
Trn0089	spk15	Kurubu.
Trn0090	spk16	Kurubu.
Trn0091	spk12	No Kurubu, no nothing. I divide you, Kurubu.
Trn0092	spk14	Kurubu.
Trn0093	spk12	<u>One, two, three, four</u> , five, six, seven, eight, nine. Remember your numbers. You are team "A", Abu ah Abu.

Figure 2 Lexical bundle in an English Language lesson (SCoRE transcript: score_eng_018).

What these two examples show is that the lexical bundle serves different purposes in different contexts. In the maths lesson, it is used as part of the explanation of the subtraction procedure in a discussion of fractions; in the English language lesson, it is part of the organisational language for setting up groups. Similar variation can be found in Science and Social Studies lessons. Yet there is a common function to all of these examples: the sequence clearly has a ‘counting’ function.

Findings

A search for four-word lexical bundles in SCoRE located 119,418 instances with a frequency of 10 or higher, and 5,477 different four-word bundles with a frequency of 10 or higher, in 2.28 millions words. This means that 2,396 different lexical bundles occurred per million words. In comparison, Nesi and Basturkmen (2006), examining lectures from the BASE corpus and the Michigan Corpus of Spoken Academic English (MICASE), found just 996 different four-word lexical bundles per million words. This difference is evidence for a greater degree of formulaic language used in classroom interactions between teacher and students as compared to monologic lectures (over 2.4 times more).

One reason for more formulaic language is likely to be that classroom discourse comprises, at least in part, conventional ways of doing things that teachers repeat across lessons and subjects. The data in SCoRE bears this out. Looked at as a percentage of the total words in the SCoRE corpus, the total words found in four-words lexical bundles comprised almost 21% of the corpus. If three-word lexical bundles were also taken into account, the figure would be greater still: most studies show that three-word lexical bundles are about five times more frequent than four-word lexical bundles in corpora. In SCoRE, there were 451,384 three-word lexical bundles found with a frequency greater than 10, and 14,873 different three-word lexical bundles: these three-word lexical bundles make up nearly 60% of the whole corpus in terms of words².

How does this compare with other corpora of spoken academic discourse? Looking at lectures in BASE, Thompson (2006) noted the prevalence of what he termed discourse organising four-word ‘clusters’ (i.e. lexical bundles) in that corpus. Table 1 shows the top ten four-word lexical bundles from the British Academic Spoken English (BASE) corpus and the top ten lexical bundles in the teachers’ talk in SCoRE (Thompson, 2006).

Table 1 Top 10 Lexical Bundles (4-word) in BASE and SCoRE.

Rank	BASE	SCoRE
1	<u>we are going to</u>	<u>I want you to</u>
2	<u>I am going to</u>	<u>how many of you</u>
3	<u>you are going to</u>	<u>you are going to</u>
4	<u>I do not know</u>	<u>if you want to</u>
5	<u>you have got a</u>	<u>we are going to</u>
6	<u>the end of the</u>	<u>what do you think</u>
7	<u>is going to be</u>	<u>I don’t want to</u>
8	<u>at the end of</u>	<u>you look at the</u>
9	<u>it’s going to be</u>	<u>one two three four</u>
10	<u>it is going to</u>	<u>what do you mean</u>

² It should be noted that many four-word lexical bundles subsume three-word lexical bundles, so the percentage figures for three-word and four-word lexical bundles cannot simply be added together here.

Only two lexical bundles, we are going to and you are going to are common to both sets, and there are some differences of interest. The school classroom discourse represented in SCoRE appears to have a more interactional emphasis rather than an organisational one. There is evidence of question forms (how many of you, what do you think, and what do you mean) as well as modal stance (if you want to, I want you to and I don't want to) related to getting people to do things.

If we look at the SCoRE data from a disciplinary perspective, however, differences become apparent within this interactional discourse. The percentage of words found in lexical bundles compared to the total number of words in the Maths sub-corpus is almost 25% (24.96), whereas the other subjects have lesser percentages: English language being just under 21% (20.75), with Science and Social Studies both around 14% (14.57 and 14.14 respectively). One implication to draw from this is that Mathematics teaching displays a greater reliance on formulaic phrases than the other subjects.

Moreover, if we compare the top ten four-word lexical bundles across the three subject areas in SCoRE, we find there is little difference among the lists. Table 2 shows these lists for Maths, EL and Science.

Table 2 Top ten four-word lexical bundles in three subjects in SCoRE.

Rank	English Language	Maths	Science
1	<u>I want you to</u>	<u>I want you to</u>	<u>I want you to</u>
2	<u>if you want to</u>	<u>how many of you</u>	<u>you look at the</u>
3	<u>you are going to</u>	<u>you know how to</u>	<u>what do you think</u>
4	<u>how many of you</u>	<u>you are supposed to</u>	<u>we are going to</u>
5	<u>what do you think</u>	<u>we are going to</u>	<u>you are going to</u>
6	<u>do you want to</u>	<u>one two three four</u>	<u>if you look at</u>
7	<u>we are going to</u>	<u>know how to do</u>	<u>I don't want to</u>
8	<u>I don't want to</u>	<u>what do you mean</u>	<u>how many of you</u>
9	<u>you don't have to</u>	<u>if you want to</u>	<u>one two three four</u>
10	<u>don't know how to</u>	<u>can you tell me</u>	<u>if you want to</u>

What the table shows is that, in terms of the most frequent lexical bundles found, the three curriculum subjects have more in common with each other than with the lectures. This is, in part, a confirmation that we are looking at different genres, and of the differences between monologic and dialogic³ academic spoken discourses.

However, if we compare those lexical bundles that occur in at least 20% of all the transcripts within each subject sub corpus, we find again interesting differences (the reason for using a cut-off figure of 20% of all texts is to eliminate the idiosyncratic use of phrases by one or two teachers). Appendix 1 gives a list of these high frequency and widely distributed lexical bundles by each subject. It also shows which lexical bundles are common to all four subjects, which lexical bundles are unique to one subject, and which subjects ‘avoid’ lexical bundles that the other three subjects all share.

As noted above, Mathematics appears to depend more on lexical bundles than the other subjects. Not only this, it has the greatest number of ‘unique’ lexical bundles. When we examine the lexical bundles generated by the keyword comparison procedure in Wordsmith Tools, this becomes more apparent, as will be described in the next section.

Distribution of categories across the four school disciplines

Analysing the data shown in Appendix 1 reveals some interesting aspects subject-specific classroom discourse. Subjects **share** many common ‘pedagogically-oriented’ bundles (e.g. how many of you) related to question and answer routines (IRFs), basic classroom directives and procedures (e.g. go back to your) and the dominant role of the teacher (e.g. I want you to). However, subjects also **differ** in which bundles are most used and vary in their lexical bundle ‘profiles’:

³ I use ‘dialogic’ here in a purely neutral sense to contrast with monologic discourse and not as an echo of Alexander’s (2006) notion of dialogic classroom teaching.

- Maths teachers seem to use lexical bundles that frame metacognitive knowledge of procedures (e.g. how do you get, how do you know) significantly more than Science, English Language and Social Studies teachers.
- Science teachers appear to use lexical bundles related to ‘events’ in class, around which they can form a narrative: now I want you, and which are source of visual evidence: (e.g. can you see that, now look at the)
- English teachers frequently ask for opinions about third party motivations: (e.g. do you think they).

Thus, I would argue, there is evidence here for variations in discipline specific pedagogic practices in core curriculum subjects. While the differences are not as striking as Hyland (2008) found in written academic registers, perhaps partly obscured by the generic aspects of classroom practices such as regulatory talk which are common to all subjects, it is possible to link the differences found to disciplinary orientations.

Key lexical bundles

Examining the lexical bundles produced by the keyword comparison procedure reveals clearer evidence of disciplinary variation. The top 10 ‘key’ lexical bundles in English Language, Mathematics, Sciences and Social Studies transcripts are shown in Table 3. The analysis produces a somewhat skewed result: for Mathematics, there are 118 key lexical bundles and for English Language, there are 88, but for Science and Social Studies there are only 9 and 11 key lexical bundles respectively. This is partly explained by the smaller size of the Science and Social Studies sub corpora, but the figures are also a clear conformation of the earlier finding that Mathematics has a much higher dependence on lexical bundles than the other subjects.

Table 3 Top 15 key lexical bundles by school subject in SCoRE.

English Language	Mathematics	Science	Social Studies
sit on the floor	is that clear to	side of the heart	and so on and
what did he do	that clear to you	zero point six metre	so you find that
what's the meaning of	three hundred and sixty	the shape of the	on and so forth
you want to write	hundred and eighty minus	then you tell me	trying to tell you
in front of the	is the same as	the red blood cells	you do at home
you can talk about	how do you get	thirty five point five	is the purpose of
to come up with	how to find the	so you can see	so on and so
you want to be	is it clear to	can you see that	if you look at
look at the picture	it clear to you	take a look at	what is the purpose
know how to spell	one hundred and eighty		and things like that
do you think he	the ratio of the		a lot of people
want you to read	i want to find		
the two of you	what do you get		
can anyone tell me	they are the same		
come up with the	want to find the		

What the keyword comparison in Table 3 clearly shows is the lexical bundles most characteristic of each subject. An inspection of the full list⁴ shows that Mathematics has a number of items centred on the word “clear” (is that clear to, that clear to you, is it clear to, it clear to you) which are part of longer bundles (is that clear to you, is it clear to you) and which express a focus on checking understanding. The other subjects have no equivalent common lexical bundles, which suggest that this is a priority with Maths teachers. Similarly,

⁴ The full list is not reproduced here for reasons of space but is available from the author on request.

Science has so you can see, can you see the, take a look at which all suggest the use of demonstration or experience as a means of building learning. English, on the other hand, has a large number of lexical bundles that contain ‘think’ – do you think he, you think they are, do you think they – which are clearly related to third persons (i.e characters in narratives) and which elicit students’ opinion. In fact the lexical bundle what do you think has a negative keyness value for Mathematics. That is, it is found statistically less frequently in Mathematics classroom talk than it is in the other subjects.

Evaluation

What the analyses above show is that, using the corpus-based approach we can identify meaningful variation among lexical bundles across subject disciplines. Furthermore, there is some evidence that these variation relate to different strategies for knowledge building in the classroom, although such a claim needs to be supported by additional, qualitative analyses of these lexical bundles in the context of their transcripts.

References

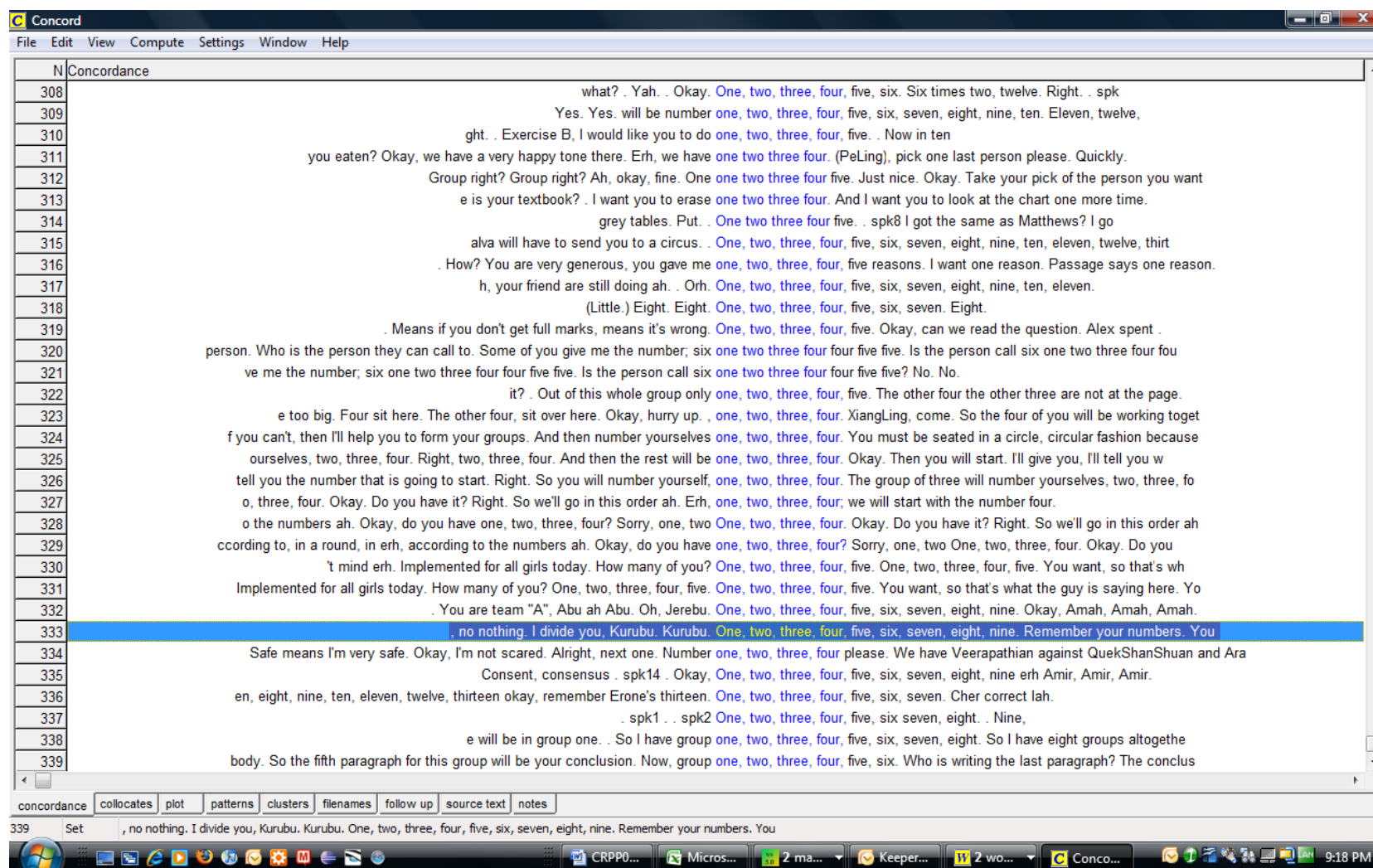
- Alexander, R. 2006. *Towards dialogic teaching. Rethinking classroom talk*. Third edition. York, UK: Dialogos.
- Altenberg, B. 1998. On the phraseology of spoken English: the evidence of recurrent word combinations. In A. P. Cowie (Ed.), *Phraseology. Theory, analysis and applications* (pp. 101-122). Oxford: Oxford University Press.
- Biber, D. 2006. *University language. A corpus-based study of spoken and written registers*. Amsterdam: John Benjamins.
- Biber, D. 2004. Lexical bundles in academic speech and writing. In B. Lewandowska-Tomaszczyk (Ed.), *Practical applications in language and computers, PALC 2003* (pp. 165-178). Łódź studies in language, 9. Frankfurt am Main: Peter Lang.
- Biber, D. and Barbieri, F. 2007. Lexical bundles in university spoken and written registers. *English for Specific Purposes*, 26, 263-286.
- Biber, D. and Conrad, S. 1999. Lexical bundles in conversation and academic prose. In H. Hasselgard & S. Oksefjell (Eds.), *Out of corpora: Studies in honour of Stig Johansson* (pp. 181-189). Amsterdam: Rodopi.
- Biber, D., Conrad, S. and Cortes, V. 2004. 'If you look at...': Lexical Bundles in University Teaching and Textbooks. *Applied Linguistics*, 25 (3), 371-405.
- Biber, D., Johansson, S., Leech, G., Conrad, S. and Finegan, E. 1999. *Longman grammar of spoken and written English*. London: Longman, Pearson Education Ltd.
- Christie, F. 2002. *Classroom discourse analysis. A functional perspective*. London: Continuum.
- Christie, F. and Martin, J. (Eds.) 2007. *Language, knowledge and pedagogy. Functional linguistic and sociological perspectives*. London: Continuum.

- Cortes, V. 2006. Teaching lexical bundles in the disciplines: An example from a writing intensive history class. *Linguistics and Education*, 17, 391-406.
- Cortes, V. 2004. Lexical bundles in published and student disciplinary writing: Examples from history and biology. *English for Specific Purposes*, 23, 397-423.
- Csomay, E. 2007. A corpus-based look at linguistic variation in classroom interaction: Teacher talk versus student talk in American University classes. *Journal of English for Academic Purposes*, 6, 336-355.
- Csomay, E. 2005. Linguistic variation within university classroom talk: A corpus-based perspective. *Linguistics and Education*, 15, 243-274.
- Doyle, P. and Hong, H. 2009. Compiling a multimodal corpus of educational discourse in Singapore schools. Unpublished report. Singapore: Centre for Research in Pedagogy and Practice.
- Dunning, T. 1993. Accurate methods for the statistics of surprise and coincidence. *Computational Linguistics*, 19, 1: 61-74.
- Hyland, K. 2008a. As can be seen: Lexical bundles and disciplinary variation. *English for Specific Purposes*, 27 (1), 4-21.
- Hyland, K. 2008b. Academic clusters: text patterning in published and postgraduate writing. *International Journal of Applied Linguistics*, 18, 1: 41-62.
- Mahlberg, M. 2007. Corpus stylistics: bridging the gap between linguistic and literary studies. In M. Hoey, M. Mahlberg, M. Stubbs and W. Teubert (Eds.), *Text, discourse and corpora. Theory and analysis* (pp. 219-246). London: Continuum.
- Mauranen, A. 2006. Speaking the Discipline: Discourse and socialisation. In K. Hyland & M. Bondi (Eds.), *Academic discourse across disciplines* (pp. 271-294). Linguistic Insights, 42. Bern: Peter Lang.

- Nesi, H. and Basturkmen, H. 2006. Lexical bundles and discourse signalling in academic lectures. *International Journal of Corpus Linguistics*, 11 (3), 283-304.
- Oakes, M. 1998. *Statistics for corpus linguistics*. Edinburgh: Edinburgh University Press.
- Partington, A. and Morley, J. 2004. From frequency to ideology: investigating word and cluster/bundle frequency in political debate. In B. Lewandowska-Tomaszczyk (Ed.), *Practical applications in language and computers, PALC 2003* (pp. 179-192). Łódź studies in language, 9. Frankfurt am Main: Peter Lang.
- Schleppegrell, M. 2004. *The language of schooling. A functional linguistic perspective*. Mahwah, New Jersey: Lawrence Erlbaum Associates.
- Scott, M. 1996. *Wordsmith Tools*. Oxford: Oxford University Press.
- Scott, M. and Tribble, C. 2006. *Textual Patterns. Key words and corpus analysis in language education*. Amsterdam: John Benjamins.
- Simpson-Vlach, R. 2006. Academic speech across disciplines: Lexical and phraseological differences. In K. Hyland & M. Bondi (Eds.), *Academic discourse across disciplines* (pp. 295-316). Linguistic Insights, 42. Bern: Peter Lang.
- Stubbs, M. 2007. Quantitative data on multi-word sequences in English: the case of the word 'world'. In M. Hoey, M. Mahlberg, M. Stubbs and W. Teubert (Eds.), *Text, discourse and corpora. Theory and analysis* (pp.163-190). London: Continuum.
- Thompson, P. 2006. A corpus perspective on the lexis of lectures, with a focus on Economics lectures. In K. Hyland & M. Bondi (Eds.), *Academic discourse across disciplines* (pp. 253-270). Bern: Peter Lang.

SECTION BREAK – DON'T DELETE !

Appendix 1 Concordance display, Wordsmith Tools.

Figure 3 Concordance of lexical bundle one two three four

Appendix 2

High Frequency Lexical Bundles in SCoRE: text coverage > 20% in each Subject

<i>English Language</i>	<i>Mathematics</i>	<i>Science</i>	<i>Social Studies</i>
are going to do are you going to at the end of do you know what do you think the do you think they do you want to don't know how to go back to your going to give you	are you going to can you see that can you see the can you tell me do you know what do you mean by do you want to don't know how to don't want you to go back to your how do you get how do you know	at the same time can you see the can you tell me don't know how to how many of you	a lot of people are some of the can you tell me do you think the do you think they do you want to going to give you how many of you
how many of you	how many of you	how many of you	how many of you
I ask you to I don't know how I don't want to	how to find the I am going to I ask you to I don't want to I don't want you	I don't want to	I am going to I don't want to
I want to see I want you to I will give you	I want to see I want you to I will give you	I want to see I want you to	I just want to I need you to I want to see I want you to I will give you if you look at if you want to
if you want to	if you want to	if you want to	if you want to
let's look at the many of you have one two three four	is the same as know how to do now I want you okay this is the one two three four should be able to	know how to do now I want you one two three four	let's take a look okay so this is one two three four
so that you can so what do you	so this is the so what do you so what is the	so this is the	so this is the

so you have to	so you have to	take a look at	take a look at
the end of the	tell me what is	tell me what is	
the rest of the	thank you very much		
the rest of you	the rest of you	the rest of the	the rest of you
those of you who	they are the same		
to look at the	those of you who	to look at the	
two three four five	two three four five	two three four five	two three four five
want you to do	want to find the	want you to do	
we are going to	want you to do	we are going to	
	we are going to		
	what do you do		
	what do you get		
what do you mean	what do you mean	what do you mean	what do you mean
what do you think	what do you think	what do you think	what do you think
	what is the answer		
what I want you			
why do you think		why do you think	why do you think
you are going to		you are going to	you are going to
	you are supposed to		you are supposed to
	you can do it		
you don't have to	you don't have to	you don't have to	you don't have to
you have to do	you have to do		
	you know how to	you know how to	
you look at the	you look at the	you look at the	you look at the
	you look at this		
	you tell me what	you look at your	
		you to look at	
you want to go	you want me to		
	you want to find		
Total: 41	57	31	32
Unique: 8	21	4	5
Avoided: 2	1	2	

Lexical bundles common to all subjects, and thus not indicative of disciplinarity, are shown in the shaded rows. Lexical bundles ‘unique’ to a subject are shown in bold. For lexical bundles that are found in all but one subject, the empty subject slot is shaded. For example, “can you tell me” is found in Mathematics, Science and Social Studies, but not in English Language.