Applying corpus linguistics and conversation analysis in the investigation of small group teaching in higher education.

Anne O'Keeffe, Mary Immaculate College, University of Limerick, Ireland

Steve Walsh, Newcastle University, U.K.

Abstract

In this paper, we consider how a combined corpus linguistics and conversation analysis methodology can reveal new insights into the relationship between interaction patterns, language use, and learning. The context of the paper is higher education small group teaching sessions and our data are drawn from a one million word corpus, the Limerick-Belfast Corpus of Academic Spoken English (LI-BEL CASE). Our methodology combines corpus linguistics (CL) and applied conversation analysis (CA), enabling quantitative findings to be elaborated by more close-up qualitative analysis of sequences of interaction. Using a combined CL and CA approach (henceforth CLCA) enables us to consider more closely the relationships between, for example, interaction patterns and lexical chunks and to evaluate the extent to which tutors create or prevent opportunities for learning. Using CLCA, we offer a more detailed description of the linguistic and interactional features of turns at talk and consider how these features combine in the joint enterprise of meaning-making. Put simply, CLCA enables us to characterize seminar talk in a more systematic way. Starting at the level of turn, and using CA to move to a higher level of discourse, we can identify specific features of the spoken interaction, such as sequential organisation and topic management. Similarly, starting again at the level of turn but using CL, we can move our analysis to lower levels of discourse, such as cluster and word patterns, in order to see how words combine. This CLCA approach offers a fuller, richer description of talk in small-group teaching than would be found using either CA or CL alone. Finally, we offer an evaluation of a CLCA methodology and consider its applications in other research settings.

1. Introduction

From an epistemological and ontological perspective, CA and CL have very different origins and research foci. CA takes as its starting point the sequence of turn-taking within interactions and looks at how interactants establish and maintain sequential order. By using a very small portion of naturally occurring textual data, the aim of the analyst is to show how interactants co-

construct meanings, repair breakdowns and orient to each other to maintain a sequential order to achieve the communicative purpose. Data are naturally occurring and the analyst aims to show what 'really happened' by asking the question 'why here, why now?' as connections are made to sequences of turns-at-talk. The ensuing rich, detailed and up-close commentary focuses on key features of the interaction which provide vital clues as to what is happening: these include pauses, overlapping speech, latched turns, 'smiley voice', laughter tokens, and so on. CL also draws upon naturally occurring data, but offers a very different type of analysis. Here, the aim of the analyst is to examine specific linguistic features of the data in terms of word frequency, concordances, multi-word units and keyness. Put simply, CL analyses are highly quantitative, use a large sample of data and set out to describe lexico-grammatical patterns and key linguistic features. CL allows for the (rapid) quantification of recurring linguistic features, which can be examined in their immediate linguistic contexts. Software programs enable accurate and consistent analysis carried out fast and without human bias.

On closer examination, however, we can see that both CA and CL have much in common. They both use empirical, naturally occurring data and refer to baseline comparisons with other types of interactions (sequential order in the case of CA, reference corpora in CL). They both look at language in its social context; an understanding of context lies at the heart of both approaches to analysis. For both CA and CL, turn level analysis is crucial to enhancing understandings of context. It is at this level that the most revealing insights can be found – we will come back to this point later in the article.

By way of illustration of the viability of a conjoint approach to the analysis of spoken discourse, we use Corpus Linguistics (CL) and Conversation Analysis (CA) in the investigation of a set of recordings which had been transcribed to form a *principled* collection, or corpus: the Limerick

2

Belfast Corpus of Academic Spoken English (hereafter LI-BEL). This corpus currently comprises almost one million words of recorded lectures, small group seminars and tutorials, laboratories and presentations. Following on from our previous research which looked at shorter segments of interaction in small groups interactions in higher education (see, for example, Walsh 2006, Walsh and O'Keeffe 2007), in this study, we are interested in looking at discourse in the context of longer stretches of text. To do this, we needed to combine CL with another approach since CL was unable to account for some of the features of spoken interaction which occurred at the 'higher levels' of utterance and turn (e.g. adjacency pairs). In order to conduct a detailed analysis at this level of the discourse, we were drawn towards conversation analysis (CA), an established and respected approach to providing detailed, micro-analytic descriptions of spoken interaction. This combined approach, using both CL and CA (henceforth, CLCA), we argue, cumulatively gives us a more 'up-close' description of spoken interactions in context (in this case, an educational setting) than could be gained by using either one on its own. From the analysis, we can gain powerful insights into the ways in which interactants establish understandings and observe how words, utterances and text combine in the co-construction of meaning.

In this paper, then, we talk about CL as a *methodological tool* which will help us investigate a corpus of small group interactions recorded in higher education. CL software enables us to automatically find word frequencies, collocations, keyness results, and so on (in this case, we used *Wordsmith Tools,* Scott 2009). CL software also allows us to rapidly retrieve textual information and from this to find patterns that we could not otherwise find, but it will not explain the dynamics of these interactions in the context. Essentially, we can say that CL offers us an overview of the landscape of the text we are analysing in much the same way we get an overall view of a landscape from an aeroplane flying at 20,000 feet. At this height, we can see the main

3

features of the landscape, but we are unable to pick out the detail. At 1,000, feet we can see the detail of the landscape but we need a different methodology to analyze the data; we need to use CA. CA allows us to look at the finer nuances of the interaction, at the sequential organisation of turns-at-talk, at adjacency pairs, at topic development and management, and so on. We believe that the two approaches, used together, are highly complementary by offering different, but mutually beneficial, perspectives on the same data-set.

CL is increasingly being applied to contexts and domains where other methodologies previously prevailed. Such contexts include courtrooms and forensic linguistics (Cotterill 2010), the discourse of the workplace (Koester 2006), the classroom and educational contexts (O'Keeffe and Farr 2003, Walsh and O'Keeffe 2007), the critical analysis of political discourse (Ädel 2010), advertising and the media (O'Keeffe 2006), among other areas. In all of these cases, CL is used alongside another complementary approach, including CA, discourse analysis, critical discourse analsysis and pragmatics. The use of CL with other complementary research methods is a testament to the strength insight that it can bring. None of these studies could have achieved the same insights without CL in addition to another approach. In all of these studies, and in the present one, CL is *applied* to achieve a particular goal rather than used to *describe* the language features of a given corpus. We are keen to draw the distinction between two types of corpus research, neither less valid than the other. In each of these types of research, the corpus is described differently (a parallel description is drawn between corpus-driven and corpus-based approaches in Tognini-Bonelli 2001):

1) *descriptive corpus research*: the corpus as an end in itself. The researcher looks 'into' the corpus so as to scrutinise the use of language and further our description of language patterns in a particular genre. An example of this type of research is Bednarek (2006). This is a corpus

4

study of evaluation in a corpus of newspapers. It tells us a lot about patterns of evaluation across and within the genre but it is not concerned with issues in a broader context of newspaper discourse, such as power, ideology, identity, and so on. This type of research makes available to the research community detailed findings which greatly enhance the description of the data type under scrutiny.

2) *applied corpus research*: the corpus as a means to an end. The researcher looks beyond the corpus for both its research questions and its analysis. The corpus is a powerful methodological tool which leads to greater depth of analysis in combination with another theoretical framework. An example of this is O'Halloran (2010) whereby a corpus of British newspaper articles about immigrants is analysed within the framework of Critical Discourse Analysis and the result is an indepth analysis of the link between the use of language patterns and ideology. In this case, the researcher applies the detailed findings to another area of research as part of the overall study.

In this study, we could take a descriptive or corpus-driven approach to our data and, by comparison with other corpora, we could come to a description of the language of small group interactions in Higher Education settings. This in itself would be a valuable exercise but we want to address broader issues of interaction in these setting and so we do a narrow, fine-tuned analysis as we draw on a complementary framework within which to understand the patterns of language use in this context, namely CA.

Another factor that merits discussion, in the context of a CLCA approach is its appropriateness for research into spoken discourse. In the history of the development of corpora, large, written corpora were used to investigate lexical patterning rather than discourse context. Especially in the early days of corpus use, which was dominated by lexicographers, the aim was to have larger corpora so as to attain the greatest breadth of semantic coverage for lexical items. Getting behind the uses of the words in context was not the main focus of corpus researchers

5

(see McCarthy and O'Keeffe, 2010). In other words, large written corpora were lauded because they allowed lexicographers, in particular, to make important statements about lexical patterns, meanings and uses. Today, however, there is a growth in the use of smaller, specialised and highly context specific corpora which are often based on spoken interactions. Such corpora, we suggest, lend themselves very well to a combined approach (such as CL with CA) which allows the investigation of other features of the discourse. Where we are dealing with smaller corpora (say, 50,000 to one million words), the context is more clearly defined and it is therefore possible to investigate features of the discourse which operate beyond the level of sentence. In addition, the person who is analyzing interactional discourse is often the person who records and transcribes the data, and so, has an in-depth tacit knowledge of the interactional context. Hence, we find much more socio-contextual information about speakers, their roles and relationships and about the purpose of their interactions in such corpora. This level of detail would normally not be present in a large written corpus. In the present study, similar characteristics guide the analysis. We use around 50,000 words of spoken interaction (a sub-corpus of LIBEL) to describe and characterize both linguistic and interactional features by combining CL and CA.

CA

The main focus of CA is turn-taking, the nexus of turns is where it puts its interactional focus. Within this approach, emergent issues of topic, turn management, turn preference, sequential organisation, repair, etc. are explored in micro detail. To paraphrase Heritage (1997), it is the interaction, through the construction of turns, which is both *context-shaping* and *context-renewing.* Accomplishment through "talk-in-interaction" (Drew and Heritage, 1992: 3) is also a key concept where context is seen as dynamic and the observer a 'player' in interpreting the data. Essentially, a CA approach is based on the notion that interactants are constantly

6

conducting a social display to each other, and to analysts, of their cognitive states and of their understanding of each others' utterances by means of and by reference to the organisation of turn-taking, sequence and repair. CA analysis provides information about the microscopic details of the interaction such as pausing, overlapping speech, high- and low-rise intonation, soft voice, repair, back channels, and so on.   It is this detailed, microscopic analysis which demonstrates not only *what* understandings the interactants display to each other, but also *how* they do so. The interactional features which are identified and described reveal how understandings are accomplished; the entire process focuses on turns-at-talk and the place of one turn in relation to a subsequent or previous one.   The organisation of turn and sequence are mechanisms for displaying and checking mutual understanding and the organisation of repair is a mechanism for repairing breakdowns in mutual understanding.

CL

CL on the other hand starts with the 'data', taking it as representing an area of language use. For example one considers a corpus of newspapers or a corpus of spoken language rather than one newspaper article or one conversation. As such, this starting point is polar to that described above for CA. Corpus analysts move into the text via the analysis of single words, multi-word units and grammatical patterns, by looking at their frequencies relative to other baselines for frequencies (for example one might compare the frequencies of a newspaper corpus with those of a general corpus of written language). Key Word in Context (KWIC) searches also allow the analyst look at the word, unit or grammatical string as a vertical pattern. From this, patterns of use emerge. Patterns of use are a key concern for corpus linguists and software can automate the process by producing all of the statistically salient multi-word units in a corpus. From the search word, multi-word unit grammatical pattern, the analyst moves up context of use. When looking at spoken language, this means going to the source texts and looking at the turns and

turn sequences. Therefore, it is at the level of turn that CL meets CA. CL cannot go very far 'up' into the discourse without a framework within which to analyze the discourse and CA cannot go further 'down' into the transcript to look at patterns of use or to look at a large text or a number of texts. The merging of CL and CA thus creates a more powerful research model as Fig 1 represents:
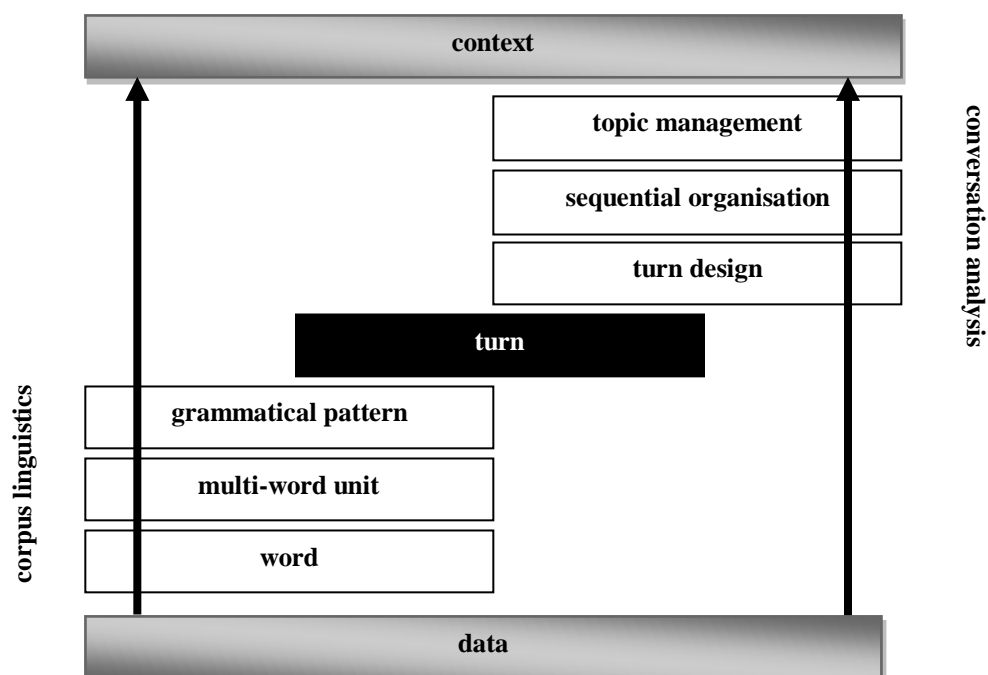

Fig. 1 CL and CL as a combined framework for analysis of spoken language

Figure 1 illustrates the many levels at which we can begin to look at real data from CL and CA perspectives. They both start from the data and can bring us to understandings of context of use. With CL, the tools give us easy access to the study of lexis, multi-word units and grammatical patterns. CL studies also look at data from turn level (see for example, Tao 2003; Csomay 2007; Buttery and McCarthy in press). Results from analysis at any of these levels can bring us insights about context of use. With CA, analysts focus on turn design, turn sequence and topic management to show how social actions are accomplished in a particular context.. In a combined model, the analyst has more analytical tools at his/her disposal and so greater

8

insight can be gained. While CA offers an emic, close-up perspective of the data, CL complements this by providing specific information about the 'bigger picture', about which language features occur, where and in what numbers, and how certain words combine in particular ways with other words. Both start from the data and work towards context.

Having established a position on the suitability of a combined CLCA methodology, we now turn to a consideration of the context in which the study took place.

2.   Context

The focus of this study is small group teaching (henceforth SGT) in higher education contexts. SGT, such as seminars and tutorials, is used to support monologic class sessions (lectures) by allowing tutors and students to engage in discussion and debate. While small group teaching is generally highly valued by both instructors and students, there are some sources of dissatisfaction, with departments identifying problems relating to student engagement and tutor skills and a lack of time (see, for example, Bennet, Howe and Truswell 2002). In our study, we were interested in identifying some of the reasons for students' apparent dissatisfaction and for a lack of engagement, especially in relation to tutor skills in managing the interaction, their 'interactional competence' (see Walsh 2006, Young 2008).

From the perspective of corpus linguistics, much influential work on spoken interaction in higher education is based on two corpora, the Michigan Corpus of Academic Spoken English (MICASE) (Simpson, Briggs, Ovens, Swales 2002) and the TOEFL 2000 Spoken and Written Academic Language (T2K-SWAL) Corpus (see Biber *et al* 2001; Biber *et al* 2002; Biber 2006, among others). MICASE corpus comprises data from across a range of speech events in higher education. It includes contexts relevant to the study reported here, such as classroom discussions, seminars, lab work and advising sessions. Studies based on the MICASE corpus

9

have explored a wide range of phenomena in academic spoken interaction, such as metadiscourse in lectures (Lorés 2006), the use of conditionals (Louwerse, Crossley, and Jeuniauxa 2008), and, of more direct relevance to this study, the effect of class size on lecture discourse (Lee 2009). The T2KSWAL corpus project involved the large-scale collection written and spoken data to 'accurately represent the linguistic charactieristics of spoken and written academic registers' (Biber *et al* 2004: v). Ultimately, these data were used to evaluate the representativeness of the ESL/EFL materials and assessment instruments in relation to the international TOEFL exams. Within the dataset, there were large collections of classroom-based recordings. Biber (2006) details how the corpus allows for the complete complement of the typical situations in which students might find themselves. From a classroom perspective, this ranged from lectures, seminars and tutorials. Biber (2006) provides an account of these as part of the linguistic description of the range of university registers which pertain.

Outside corpus linguistics, there is quite a long history of research into spoken interaction in higher education. Some of this research has taken as its main focus spoken interaction in SGT situations and used CA as a research methodology. More recent research on talk-in-interaction in SGT in higher education has uncovered important aspects of the processes or 'machinery' by which seminars and tutorials 'get done', for example, by focusing on cues and signals used to manage interaction and participant roles (Viechnicki 1997), sequential organisation and negotiation of meaning (Basturkmen 2002), and the issue of 'topicality' in small group discussion (Stokoe 2000; Gibson, Hall and Callery 2006). Other research has explored the formulation and uptake of tasks and resistance to 'academic' identities (Benwell and Stokoe 2002).

10

Much of the more recent work on talk in SGT (particularly that of Benwell and Stokoe) draws on perspectives from ethnomethodology, conversation analysis and discursive psychology. In these perspectives, human social activity such as small group seminars or tutorials, are seen as locally produced accomplishments in which participants display their own understandings of the unfolding context. Participants take actions to further their own goals and agendas and display their orientations to others' actions. In SGT contexts, tutors will demonstrably orient to the accomplishment of pedagogical goals and tasks, and students may accept or resist these actions (Benwell and Stokoe 2002).

In the present study, we were interested in the ways in which tutors and students manage the complex relationship between pedagogic goals and the talk used to realise them. In SGT settings, as in most educational contexts, there is a strong relationship between pedagogic goals and pedagogic actions and the language used to achieve them (Seedhouse, 2004). Understanding this relationship, and the ways in which tutors and students engage in tightly organised and intricate negotiations of a set of pedagogic agendas, lies at the heart of any enterprise which sets out to improve teaching and learning in higher education. We adopt the strong position taken by others that interaction and learning are inextricably linked (c.f van Lier, 1996, 'interaction IS the learning'). Any attempt to enhance learning in SGT should therefore, we suggest, begin by gaining a closer understanding of the interactions taking place. By using a combination of CL and CA, we are able to provide a more realistic description of the relationship between pedagogic actions and the language used to achieve those actions in classroom discourse (Walsh, 2006), thus offering a greater understanding of the finer interactional adjustments and variations which exist in SGT interaction. We can then use these insights as a means to the end of addressing the problem identified above, that of the relationship between student engagement and tutor interactional skills in SGT in one higher education context.

11

3. Data and analysis

The study is based on data from the Limerick Belfast Corpus of Academic Spoken English (hereafter LI-BEL), which currently comprises 500,000 words of recorded lectures, small group seminars and tutorials, laboratories and presentations. These data were collected in two universities on the island of Ireland: Limerick and Belfast, across common disciplinary sites within the participating universities: Arts and Humanities, Social Sciences, Science, Engineering and Informatics and Business. From the main corpus, a sub-corpus of 50,000 was created by identifying all the instances of small groups teaching. We define these as sessions comprising between 15 and 25 students and where there was evidence of sustained interaction either between the instructor and the students or the students alone. It is perhaps significant that we were only able to identify 50,000 words (or 10% of the corpus) in which there was evidence of extended interactions. This, in itself, is indicative of the current state-of-play of SGT in the two universities under investigation; it is apparent that tutorials and seminars are functioning more as extensions of lectures than offering opportunities for interaction and sustained debate.

CL analysis

Using *WordSmith Tools* (Scott, 2009) key words and word frequency lists for both single words and multi-word units were generated. The one-million word Limerick Corpus of Irish English (LCIE) was used as a reference corpus (Farr et al 2002). Table 1 illustrates the top 20 key words.

1	okay	11	any

12

2	ye[1]	12	exactly
3	alright	13	different
4	you	14	include
5	et cetera	15	if
6	so	16	this
7	that	17	can
8	what	18	about
9	of	19	next
10	your	20	literally

Table 1 – top 20 key words from LI-BEL sub-corpus compared reference corpus (LCIE)

Through concordance and source text analysis via Wordsmith differences in the functioning of these higher frequency words was brought into relief. For example, *if* when used in 'first conditional' type structures had three main functions:

- pedagogic illustration of 'general truths/facts' *if John Kerry takes Texas, …he takes every vote…*;
- projecting, 'meaning when you find yourself in this situation' *if you are on TP and you have a class that…*;
- demonstrating, *if you click the mouse and then click…*

Figure 2 illustrates the most salient items when we looked at the LI-BEL sub-corpus frequencies using LCIE as a baseline for comparison.

---

[1] *Ye* is the plural form of *you* used in Irish English. Even though it is prevalent in LCIE, it operates as a key word in the LI-BEL sub-corpus along with the standard for *you*.
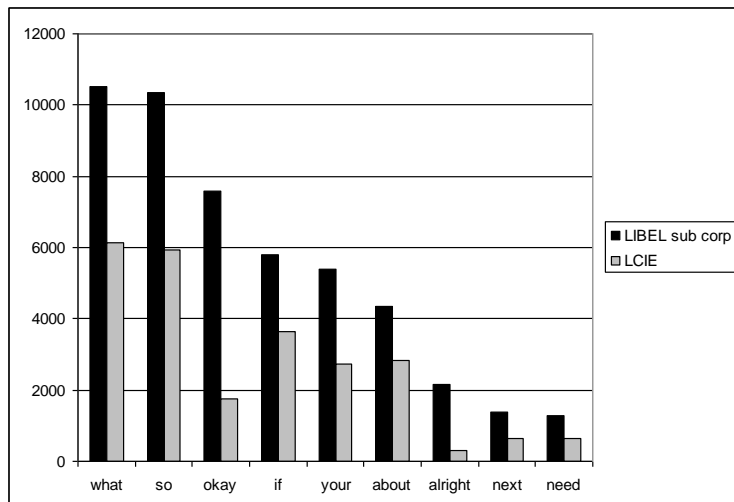
Fig 2 – Single word frequencies in LI-BEL sub-corpus and LCIE, results normalised per million words

As Fig 2 illustrates, nine words single-word items were found to be significantly different in frequency when compared to the reference corpus (the Limerick Corpus of Irish English, LCIE). Some of these are context-specific, for example, the prevalence of the interrogative pronoun *what,* discourse markers *so, okay, alright*, deictic *next* (as in *next week, next semester, next lecture*), modality (*need to*), and so on. These initial corpus results, when examined in KWIK searches point some core actions in classroom contexts, as eliciting information (with *what*), signposting the discourse and locating learning (using discourse markers and deixis) and directing learners to perform certain actions and carry out tasks (*what I need you to do, you need to*).

Having scoped out the word frequencies and word patterns related to these (using concordance searches), we next moved to multi-word unit (MWU) level of analysis. We identified over 128 MWUs in all and these further illuminated the earlier results for key word and the single word frequencies. This resulted in the emergence of clear categories into which the words and their patterns could be divided. We found that like the single word items referred to above, the most frequently used MWUs in this context have a broad discourse marking function. They signpost,

14

manage, demonstrate, sequence, set up activities/groups and they mark out shared and new

knowledge (Carter and Fung 2007), as table 2 illustrates, based on our results.

Category	Examples
Elicitation from individuals and groups	*any ideas of how you could, do you have any idea, do you think that would, did anyone else come up with, did ye come up with anything, any ideas of how, anyone have any idea, anything else to add, do you have any, do you think, give me an example of,*
Feedback on elicitations	*amm okay so, okay so, very good, a very good point, yeah yeah yeah,*
managing tasks and activities	*do you think you could, I want ye to, I want you to, you're going to, we're going to talk about what did your group have? anything else? talk to the person next to*
demonstrating and sequencing	*do ye know how to, I'm going to, and then, what you can do is put, we're going to what we're going to do*
Relating to past or future references within the lecture or course	*do you remember, next week, as I was saying, the next*
Discourse markers of shared space	*you know what I mean, you know the way you, so you know, then you know, like you know, you know that, you know a, you know ah, I mean, I think, I suppose*

Table 2 – Broad functional categorisation of significant multi-word units in LIBEL sub-corpus

This is the point to which CL brings us. We could go back 'down' into any one of the high

frequency items (words or multi-word units) and do more indepth analyses in the context of the

data but if we want to move 'up' in our more general understanding of the data, we need to employ a discourse framework. Hence we move to the CA phase of the analysis.

CA analysis

Our CL analysis clearly identified a number of key linguistic features whose distribution was in some way marked in terms of frequency. In order to gain a deeper understanding of spoken interaction in this context, we needed to see how these high frequency and / or key features actually operated in speakers' turns and in longer sequences of interaction. In our qualitative analysis (see below), we examined the corpus using CA, building 'collections' (Hutchby and Woofitt, 2008) of similar instances of stretches of interaction where there was both a clustering of the linguistic features identified in the corpus analysis in addition to specific patterns of sequential organisation (Schegloff, 2007).

Moving on to longer stretches of discourse which our CL analysis had helped us to identify (because these stretches of text were marked in some way), our focus shifted now to an analysis of the sequential organisation of turns at talk. Our CA analysis enabled us to identify a number of specific interactional features of the discourse which we were then able to consider alongside the linguistic features previously identified in the CL analysis. The dialectic between CA and CL thus allowed us to better understand why certain items were clustering at certain points. We now look at the most salient contexts in which high frequency items clustered. For the sake of convenience, we label each context according to its predominant pedagogic function.

16

Organisational talk

Much of what goes on in SGT entails tutors organising learning in some way, often temporally or spatially. Here, tutors' pedagogic goals are to inform students about different procedural matters (the date and time of an examination, the materials to bring to the next session, and so on). Consider extract 1 below, where the tutor makes frequent reference to time (*the next day, week nine*, and so on). The primary purpose here is to alert students to upcoming tasks and activities, and to the overall organisation of modules and courses. Note too the use of *okay* at the end of this sequence as a marker of a transition to the next stage of the SGT session. *Okay* was the third most statistically 'key' lexical item in the corpus, occurring very frequently in stretches of organisational talk.

Extract 1

1   T: I'll talk to ye about those *the next day* (.) that's

2   *week nine* (.) this is only *week two* isn't it? so amm

3   we're okay with regards to that (.) you see now amm

4   okay

The interactional organisation of 'organisational talk' is characterised by long turns by one participant (normally the tutor), while the other participants produce short responses or no responses at all. It is here that the tutor may use discourse markers such as 'okay?' to check understanding. Often tutors do not wait for a verbal response (presumably relying on visual information to monitor the state of comprehension of the other participants). In our data, it was

17

very obvious that the tutor may also perform the role of both questioner and answerer, as

evidenced below in extract 2. Here we see the tutor producing both the first and second-pair

parts of a question, with no pause between them to indicate a turn transition relevance place,

showing that no response is expected:

Extract 2

1   T: so if I told you I wanted horizontal lines in that

2   first box (.) what would your first thing to do be

3   (.) well would you kind of come in here like this and

4   go like that? no you wouldn't because free hand you

5   don't need to have the sheet attached fully to the

6   desk

Instructional talk

Much of the interaction of SGT was found to be reminiscent of more traditional classroom

discourse (see Cazden 1988), dominated by display questions, Initiation-Response-Follow-up

(IRF) exchanges (see Sinclair and Coulthard 1975), short utterances from students, and so on.

In what we are calling 'instructional talk, the discourse is highly controlled, with the main

responsibility for managing the interaction firmly in the hands of the tutor. Turn-taking is tightly

controlled by the tutor, who manages both next turn allocation and questions addressed to

individual participants, thus making the respondee's provision of the second pair-part strongly

relevant. In terms of corresponding linguistic features, the most obvious example is found in concordance searches of the pattern *tell me*:

```
 class what would you tell me your aims might be?

  that's it. Can anyone tell me what type of type of objective that is. Students will be

         . I want you to tell me come back to me and tell me do you think there's

ns. And ahh can anybody tell me why that's not the case? Why is it that you know there

. Fine out. And can you tell me what type of an objective that is? They're labelling a

ot      . I want you to tell me come back to me and tell me do you think there's anythin

the groups saying right tell me what your list is tell me what your list and that's your

the groups saying right tell me what your list is tell me what your list and that's your

                   So tell me what ye'd put down instead so? If you were talking about

he sense of?      . And tell me have you any ideas of how you could possibly include

self I've now told you. Tell me anyone again how could you change that objective to mak
```

Fig 3 – Extracts from concordance lines of 'tell me'

Another example is *I want you / ye to*:

```
 right then over here what I want you to do is take any line in the middle make a wavy line like

single group in the class. I want ye to look at that lesson plan and I want you to critique under

o minutes. I've got     . I want you to tell me come back to me and tell me do you think there's

izontal like this. So what I want you to practice is roughly around maybe five eight mils apart.

with your setsquares. What I want you to do is do a bit of printing in there. Lovely lovely plan

in    your  pairs  again  I want you to look at this lesson plan. First of all I want you to

an and in your pairs again I want you to look at this lesson plan. First of all I want you to

sson plan okay? Number two I want you to decide amm would you be happy if I told you in the next

e you okay for two minutes I want you to discuss that in the sense of is there anything missing
```

Fig. 4 Concordance extracts of the pattern 'I want you to'

19

Consider extract 3 below, which comes from a teacher education seminar, and which makes extensive use of the multi-word units (MWUs) found in the corpus and used for eliciting information (as in lines 7, 9-10, 12):

Extract 3

1   T: say now it's we'll give a a topic example again

2   (.)ahh let me see what would you be doing at all? the

3   same genetics again we'll go back to that (.) amm (.)

4   you're teaching your fifth year class genetics and

5   you say pupils should in will be enabled to?

6   S: (find out?)

7   T: perfect find out (.) and *can you tell me* what type of

8   an objective that is? they're labelling a diagram (.)

9   what type of an objective is that? *anyone have any*

10   *idea*? remember I was telling you that there's three

11   types of objectives (.) what type of objective is

12   that? *can anyone name* any one of the three? oh this

13   is fun at nine isn't it? kind of you know twenty

14   questions

20

In this extract, the tutor performs a lot of interactional work in the pre-sequence (lines 1-5) which precede the actual elicitation: 'you say pupils should in will be enabled to?' (line 5) said with a rising intonation to mark it as a question. The student's second pair-part response 'find out' at line 6 is met with an assessment by the tutor ('perfect find out', line 7). In the remainder of the extract, the tutor offers a series of first-part question-answer adjacency pairs in an attempt to elicit 'what type of an objective that is' (lines 7-8). Despite the tutor's prompting ('anyone have any idea?' (lines 9-10); 'remember I was telling you that there's three types of objectives (.)' (line 11); 'can anyone name any one of the three?' (line 12), these elicitations are met by silence from students. Eventually, the silence provokes a meta-comment on the discourse at lines 12-14: 'oh this is fun at nine isn't it? kind of you know twenty questions'.

Discursive talk

We are using the term 'discursive talk' to describe those interactions where the main pedagogic goal of the tutor is to get students to describe personal experiences, reflect on their feelings or attitudes towards a particular phenomenon, relate personal feelings, emotions and experiences to an aspect of the course they are taking. One of the most important indicators of success in any educational discourse, arguably, is a tutor's ability to create shared space where learning can take place. This is particularly true in a higher education context, where students must feel able and willing to participate and contribute to the discussion. In this study, we were interested

21

in the ways in which tutors, through their choice of linguistic and interactional features, created 'space for learning': interactional space in which students could become involved, engaged, and willing to take risks in the discussion (see, for example, Csomay 2005, Walsh 2006). Our quantitative analysis showed quite clearly that what we called 'discourse markers of shared space' occurred frequently in this context, labelled here 'discursive talk'. The main characteristic of discursive talk are student accounts of experiences that they are having as part of the course, often accompanied by assessments of situations and behaviour. The tutor accepts and builds on these accounts, converting them into pedagogical material in the form of reflective statements about appropriate behaviour, roles and identities in the professional practice of the discipline. Agreement to assessments is favoured (there is a lack of dispreferred responses) and there is frequent use of interpersonal discourse markers to provide supportive responses to the speaker (*yeah*) and to mark shared knowledge (*you know*; *you see*).

Our comparative corpus analyses into such items show telling differences as illustrated in fig. 5 in the comparison of *you see* and *you know.*
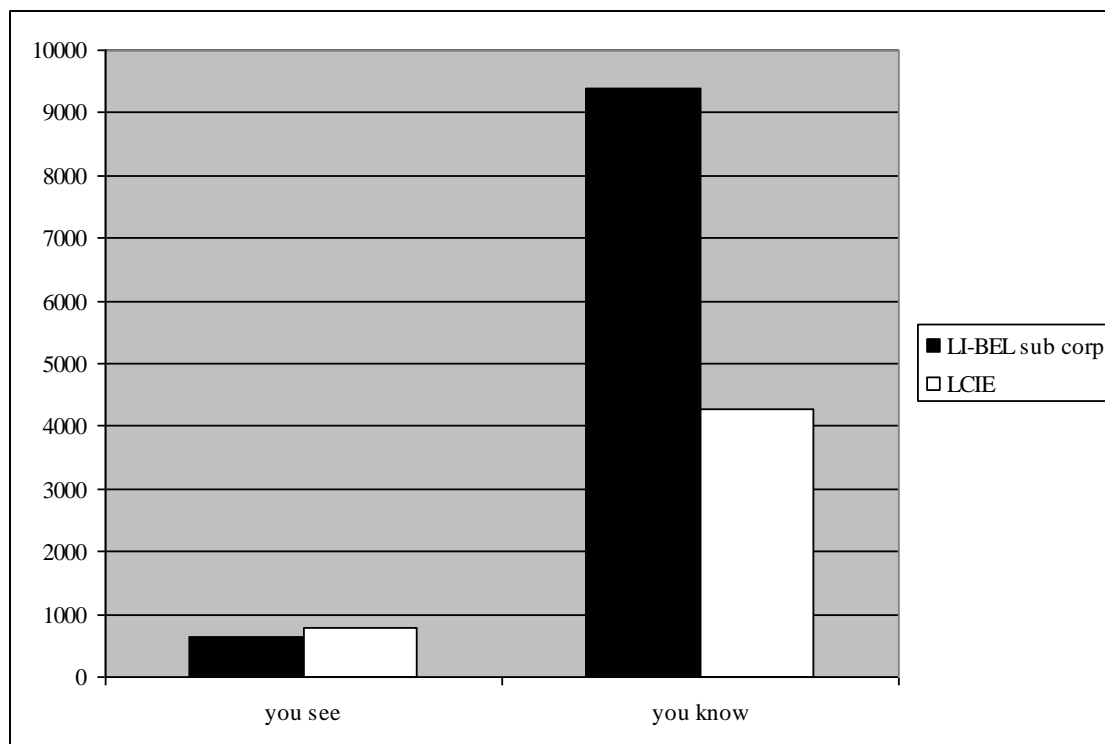
Fig 5 – Comparison of frequencies of 'you see' and 'you know' in LI-BEL sub-corpus and LCIE (normalised per million words)

You see usually marks new information while *you know* generally marks shared information (Carter and McCarthy 2006). It is telling that in a corpus recorded in higher education classrooms that we find an exceptional number *you knows* (marking shared information) but we find more or less the same amount of *you see*s (marking new information). The priority to build on and appeal to shared knowledge and 'shared space' is central to both the pedagogic and interactional process.

The interactional features of this kind of talk show that there is considerable symmetry; tutor and students adopt almost equal interactional roles and it may not be immediately obvious who the tutor is. Typically, turns are evenly distributed and often managed by students themselves, in a

23

way which closely resembles everyday conversation. Tutors may initiate exchanges as a form of open invitation to produce accounts of experience, as in extract 4, taken from a film studies seminar.

Extract 4

1   T: how are you getting on with your other ahh module (.)

2   ahh the the the filming one

3   S3:   we're filming a scene at the moment we're editing and

4   it's crazy

5   T: yeah you see it is crazy isn't it (.) this week now

6   is going to be unbelievable

7   S3:   it's just music and we're just putting it together

8   you see now (.) you know you've all the footage but

9   you're there trying

10   S?: (   )

11   S3   we have so much footage and it's just like you some

12   people have to accept that some of it

13   S? (   )

14   T: yeah well it's like essays isn't it (.) I mean you

15   can't write from the   middle you know=

16   S4   =that was my idea so we can't lose that and you're

17   like going=

18   T: =who who's the director?

24

19	S3	I wish I was the director
20	S2	in our in my group John C is
21	T:	okay yeah you see that's the thing like you know I
22		mean like really it does all come down to the
23		director and the people should respect that
24		immediately you know (.) that doesn't happen that
25		often (.) you know what I mean it can get the roles
26		can get dispersed

This interaction appears more symmetrical and conversational than the more traditional asymmetry which is associated with classroom discourse. In response to the teacher's opening turn, one student (S3) produces an account of a group's experiences of making a film, including an assessment of the situation (it's crazy), to which the tutor offers a preferred (agreeing) response with the discourse marker 'yeah' and the repetition of the assessment, before building on this to project what experiences will be like in the future. It seems apparent that participants can express feelings such as frustration with aspects of the course, or in the case above, with other students' behaviour. In lines 11 and 12, S3 indicates that 'some people' may have problems in accepting that material has to be cut, and in line 19, seems to be expressing frustration either about the existing director, or the lack of a director's role in the group.

The role of the tutor here is to 'take a back seat', listen to what students contribute, take their experiences and feelings and build on them, and so on. The pedagogic goal is to reinforce appropriate behaviours and identities, especially in a context where professional practice is

25

important, as in the one above. However, there may be a tension between the establishment of a more 'equal' turn-taking system, with the freedom to express feelings, and the need for tutors to convert this into pedagogically useable material. This can be seen in the tutor's last turn in the extract (lines 21-26), in which 'okay' marks a switch in orientation, and the content about appropriate roles and behaviours is prefaced with a lengthy string of hedges, indicating pragmatic work in switching roles from an empathic listener to a 'reflexive judge' (Baumgart, 1976). This tutor does quite a lot of interactional work in order to change footing ('okay yeah you see that's the thing like you know I mean like really'); his stance after this preface is that of teacher again, giving instruction and passing on new knowledge. The interactional work is apparently needed in order to change from equal interactant to tutor, to move from a position of role symmetry to one of role asymmetry.

Extract 4 again highlights the value of using a combined CLCA methodology. Here, we have seen how the use of CA across longer stretches of discourse allows analysts to make inferences about roles, about the ways in which specific social actions are accomplished and about the ways in which interactants perform multiple identities. CL, on the other hand, highlighted a clustering of specific linguistic features (you know and you see) which helped us to locate the parts of the corpus where 'identity work' was taking place.

Argumentative talk

A key aim of higher education is to foster criticality and promote individualised thinking. Most tutors would be delighted if students would engage with their discipline, discuss, debate and argue about new concepts, challenge existing principles and offer new ideas of their own.

26

Unfortunately, all too often, this does not occur, with students playing more of a passive role and not really engaging with the content of their discipline, apparently disinterested and only motivated by information which well help them pass the course or achieve success in an assignment. In our data, there were instances of what we are calling 'argumentative talk' where there was some kind of discussion or debate, even argument.

Typically, and based on our quantitative CL analysis, argumentative talk occurred most frequently when there was a preponderance of discourse markers of shared space. Accompanying these discourse markers, we found a heavy use of frequent examples of negation or adversative items such as 'but', as exemplified in Figure 6 below:

```
hurling and      laughing        you know. But this is it. This is our oral tradition.

        settle on that.        Exactly amm but what about like I totally agree with that

 symbol of ahh well regression I suppose. But what did happen in Limerick was the amm

aying about his intro= or you know        but your not allowed say that. What's your

          That's a problem.       Yeah no but you'll know this it's it's the      . It

at they have it kind of      . Yeah okay. But you're saying that maybe that they okay so

          understand them at all. Okay? But a sense of territory is there. Okay?

        because that was my fault. Okay? But the next all the rest of the classes do ut

   and then you come back and       right but you're wrong     . What do you think?
```

Fig 6 – Sample concordance lines of 'but' preceded by a discourse marker in argumentative contexts

From an interactional perspective, contexts in which argumentative talk could be found were characterized by a symmetrical speech exchange system, with 'give and take' in the interaction as tutor and students collaboratively negotiate meanings and co-construct understandings. There can be quite rapid exchanges of assertions, with frequent occurrence of dispreferred

27

options such as straight rebuttals, and there is a high frequency of latched turns and a relative lack of pauses at transition relevance places. Extract 5, which is from a politics seminar, is a clear example of argumentative talk in action:

Extract 5

1	S5	are we are we defining (.) ethnicity or nationalism
2	T	they're blurring (.) for the purposes of this class
3		they're blurred (.) oh no they're not amm (.) no
4		we're ethnicity is what we're doing=
5	S5	=but it's it's very close to nationalism when I see
6		territory people and and its its ((unintelligible))
7		you have a nation (.) ahh it's one of the=
8	T	=not for the people who claim they're ethnically
9		different but a people within the nation=
10	S5	=yeah I suppose=
11	T	=and that that's kind of the point is that (.) yes
12		you'd you'd think on the outer we'd have a nation but
13		if it worked out like that well then we wouldn't have
14		ethnic conflict (.) do you get my point (.) okay so
15		(.)any anybody else think there's anything else there
16		that should be up there

Extract 5 opens with an apparent challenge from S5 (are we are we defining (.) ethnicity or nationalism) followed by an uncertain response from the tutor (lines 2-4). Note the frequent use of pausing (.) which may indicate hesitation or uncertainty. In line 5, the same student appears to be dissatisfied with the tutor's previous response and interrupts (indicated =) with a further challenge. S5 also appears to show some uncertainty in line 8 (a pause (.) followed by *aggh*), allowing the tutor an opportunity to interrupt again in line 11. The tutor succeeds in holding the floor from lines 11-16 and, despite some obvious transition relevance places (marked (.), nobody challenges his explanation further. Indeed, he even closes down space in 14 (*do you get my point (.) okay*). The discourse marker 'okay' here seems to show a degree of finality to the discussion, pointing to a transition in this stage of the seminar and a time to move on.

The pedagogic orientation appears to be towards an open and dialogic exploration of disciplinary knowledge, similar to Barnes' (1992) and Mercer's (2000) 'exploratory talk'. However, this micro-context actually shows characteristics of 'disputational talk' (Mercer, 1995, 2000) in which participants, rather than interacting to build knowledge together, dispute each other's meanings in ways which may not move the discussion forward.

4. Discussion

In terms of our actual findings within the study, we can claim that it has several implications. First, there is a need for further research to consider more carefully the relationship between language use, interaction and learning in small group teaching sessions. At present, we only have a partial understanding of the complex relationships between language, pedagogy, interaction, learning and knowledge. The linguistic and interactional features identified in our data, we suggest, perform a central role in co-constructing meaning, in promoting criticality and in engaging learners in academic debate. There is more work needed to promote an understanding of the ways in which these features assist in the creation of space for learning.

Second, we would argue that there is a need for tutors to develop greater interactional competence in order to facilitate the kind of 'whole class interactive teaching' such as that currently being advocated in the national literacy strategy in secondary classrooms. Classroom interactional competence (Walsh, 2011) refers to the specific interactional strategies that tutors use to help learners express new ideas, discuss key concepts, question accepted knowledge, and articulate emerging understandings. By helping tutors gain greater interactional competence, we suggest, the overall quality of learning can be enhanced.

Third, there is a need to look more closely at ways of including and involving students more fully in the discourse of SGT sessions, raising students' interactional competence and facilitating a more interactive, engaging learning environment. We suggest that much can be done to improve the learning experience of students by helping them to consider how they can become better interactants, more able to articulate complex ideas or take a particular stance in relation to an idea, concept or theory.

Finally, further research is needed to evaluate and assess the extent to which the micro-contexts we have identified in this study stand up to closer scrutiny when extended to other contexts in which SGT takes place. This study was carried out in one national context (Ireland) using a relatively small corpus. Further studies using larger corpora across a range of contexts in higher education would be likely to reveal the robustness of the framework for understanding interaction in these contexts.

30

6. Conclusion

In terms of the overarching focus of this paper, namely the proposition that CL and CA are suitable bedfellows, we have shown in our exemplar not only how they can be mutually beneficial, but how they can actually synergize. Through an over and back iterative process between the extended text and the corpus results, a methodological dialectic, we were able to identify and verify four distinct micro-contexts which emerged through a combination of the tutors' and students' orientations to certain pedagogic goals, and the speech exchange systems set up to produce this knowledge as an interactional accomplishment between them. Implicated in, and indexical of, these micro-contexts, is the use of high frequency items in the corpus at particular points in the interactions. Had we used CL on its own, we would have achieved an interesting insights but it would not have brought us anywhere near the depth of understanding of the interactions compared with what a CA framework could explain. Had we looked at the data purely from a CA perspective, we may have established the four micro-contexts, but we wouldn't have been able to support the fact that the words and patterns which are contained therein were actually high frequency items (i.e. key words, high frequency words and multi-word units). In addition, by drawing on quantitative methods within CL, we were able to reference our findings against another dataset. By way of final reflection on what CL can gain from CL, we see the narrowness of the CA transcription as something that needs to be accommodated more into the transcription of spoken corpora and the need to align recordings with transcripts is also something which will hopefully become more a reality as the next generation of spoken corpora emerge (see Adolphs and Knight, this volume).

31

References

Ädel, A. (2010) How to use corpus linguistics in the study of political discourse. In A. O'Keeffe and M. J. McCarthy (eds) *The Routledge Handbook of Corpus Linguistics.* London: Routledge, pp. 591-604.

Basturkmen, H. (2002). Negotiating meaning in seminar-type discussions and EAP. *English for Specific Purposes, 21*(1), 233-242.

Baumgart, N. (1976). Verbal interaction in university tutorials. *Higher Education, 5*, 301-317.

Bednarek, M. (2006) *Evaluation in Media Discourse. Analysis of a Newspaper Corpus*. London/New York: Continuum

Bennett, C., Howe, C., & Truswell, E. (2002). *Small group teaching and learning in psychology*. York: LTSN Psychology University of York.

Benwell, B. M., & Stokoe, E. H. (2002). Constructing discussion tasks in university tutorials: shifting dynamics and identities. *Discourse Studies, 4*(4), 429-453.

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

Biber, D., Reppen, R., Clark, V. and Walter, J. (2001) 'Representing Spoken Language in University Settings: The Design and Construction of the Spoken Component of the T2K-SWAL Corpus'. In R. Simpson and J. Swales (eds) *Corpus Linguistics in North America*. Ann Arbor, MI: Michigan University Press, pp. 48–57.

Biber, D., Conrad, S., Reppen, R., Byrd, P., and Helt, M. (2002) 'Speaking and writing in the university: A multi-dimensional comparison'. *TESOL Quarterly*, 36, 9-48.

Buttery, P. and McCarthy, M.J. (in press) "Lexis in spoken discourse". In J. Gee and M. Handford (eds) *The Routledge Handbook of Discourse Analysis*. London: Routledge.

Cazden, C. (1988) *Classroom Discourse: The Language of Teaching and Learning*, Heinemann, Portsmouth, NH.

Carter, R.A. & Fung, L. (2007). Discourse markers and spoken English: Native and non-native use in pedagogic settings. *Applied Linguistics*, 28(3), 410-439.

Cotterill (2010) How to use corpus linguistics in Forensic Linguistics? In O'Keeffe, A and M. J. McCarthy (eds) *The Routledge Handbook of Corpus Linguistics*. London: Routledge, pp 578-590.

Csomay, E. (2007) 'A Corpus-Based Look at Linguistic Variation in Classroom Interaction: Teacher Talk' *Journal of English for Academic Purposes*, 6(4): 336- 355.

Drew, P., & Heritage, J. (Eds). (1992). *Talk at Work*. Cambridge and New York: Cambridge University Press.

Farr, F., Murphy, B. and O'Keeffe, A. (2002). The Limerick Corpus of Irish English: design, description and application. *Teanga,* 21: 5-29.

Gibson, W., Hall, A., & Callery, P. (2006). Topicality and the structure of interactive talk in face-to-face seminar discussions: implications for research in distributed learning media. *British Educational Research Journal, 32*(1), 77–94.

33

Hopper, R. and Drummond, K. (1992) 'Accomplishing interpersonal relationship: the telephone openings of strangers and intimates'. *Western Journal of Communication* 56: 185-199.

Heritage, J. (1984). *Garfinkel and Ethnomethodology*. Cambridge: Polity Press.

Heritage, J. (2004). Conversation analysis and institutional talk: analysing data. In D. Silverman (Ed.), *Qualitative research: Theory, method and practice* (2nd ed., pp. 222-245). London, Thousand Oaks, CA and New Delhi: Sage.

Koester, A. (2006). *Investigating workplace discourse*. London: Routledge.

McCarthy, M.J. and O'Keeffe, A. (2010) Historical perspective: what are corpora and how have they evolved? *The Routledge Handbook of Corpus Linguistics.* London: Routledge,3-13.

Mercer, N. (2000). *Words and Minds: How We Use Language to Think Together*. London: Routledge.

O'Halloran, K. (2010) How to use corpus linguistics in the study of media discourse. In A. O'Keeffe & M. McCarthy (eds) *The Routledge Handbook of Corpus Linguistics.* London: Routledge, 563 – 577.

O'Keeffe, A. (2006). *Investigating Media Discourse*. London: Routledge.

O'Keeffe, A. and F. Farr 2003. "Using Language Corpora in Language Teacher Education: pedagogic, linguistic and cultural insights", TESOL Quarterly 37(3): 389-418.

Richards, K. (2005). Introduction. In K. Richards & P. Seedhouse (Eds.), *Applying Conversation Analysis.* Basingstoke: Palgrave Macmillan.

Scott, M. (2009) *WordSmith Tools Version 5.0*. Liverpool: Lexical Analysis Software Ltd.

Seedhouse, P. (2004). *The interactional architecture of the language classroom: A conversation analysis perspective*. Oxford: Blackwell.

Sinclair, J. McH (2004) *Trust the Text: Language Corpus and Discourse*. London: Routledge.

Sinclair, J. McH and Coulthard, M. (1975) *Towards an Analysis of Discourse*. Oxford: Oxford University Press.

Stokoe, E. H. (2000). Constructing topicality in university students' small-group discussion: a conversation analytic approach. *Language & Education, 14*(3), 184-203.

ten Have, P. (*2007). Doing conversation analysis*. 2nd ed. London: Sage

Tao, H. (2003) Turn initiators in spoken English: A corpus-based approach to interaction and grammar. In Leistyna, P. and Meyer, C. F. (eds.) *Corpus Analysis: Language Structure and Language Use.* Amsterdam: Rodopi, 187-207.

Tognini-Bonelli, E. (2001) *Corpus Linguistics at Work*. Amsterdam: John Benjamins.

Viechnicki, G. B. (1997). An empirical analysis of participant intentions: discourse in a graduate seminar. *Language and Communication, 17*(2), 103-131.

Walsh, S. (2006). *Investigating Classroom Discourse*. London: Routledge.

Walsh, S., and O'Keeffe, A. (2007). Applying CA to a modes analysis of third-level spoken academic discourse. In H. Bowles & P. Seedhouse (Eds.), *Conversation analysis and languages for specific purposes*. Bern: Peter Lang.

Walsh, S. (2011) *Exploring classroom discourse: language in action.* London and New York: Routledge.