Analyzing university spoken interaction: A CL/CA approach.

Steve Walsh, Tom Morton and Anne O'Keeffe
Newcastle University / Universidad Autónoma de Madrid / Mary Immaculate College,
University of Limerick

In this article, we consider how corpus linguistics (CL) and conversation analysis (CA) can be used together to provide enhanced descriptions of spoken interaction in the context of small group teaching in higher education. From our analysis of the data, we show how the two approaches can be combined in an iterative process to account for features of spoken discourse at both micro (word) and macro (text) levels. Beginning with CL and focusing largely on words and combinations of words, we then use CA to highlight pertinent interactional features. Our methodology follows an iterative process: from CL to CA, back to CL and so on. This approach to analysis provides powerful insights into the ways in which interactants establish understandings in educational settings and, in particular, highlights the inter-dependency of words, utterances and text in the co-construction of meaning.

1.   Introduction

In this study, we look at discourse in context by combining corpus linguistics (CL) with conversation analysis (CA) since corpus linguistics is unable to account for some of the features of spoken interaction which occur at the levels of utterance and turn. Our aim is to consider how corpus linguistics and conversation analysis can be used together to provide enhanced descriptions of spoken interaction. While CA and CL have both been used on their own to study spoken encounters, each has its limitations. CL, for example, largely ignores context and focuses on large scale analysis, whereas CA offers detailed descriptions but is unable to generalise to larger contexts. Using a combined CL and CA approach (henceforth, CLCA), we argue, cumulatively gives a more "up-close" description of spoken interactions in an educational setting than that offered 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. Using CL as a tool allows us to automatically search a large dataset, something which would have been impractical manually. However, while CL allows us to count frequencies and find keywords in micro-seconds, thus revealing patterns that we could not otherwise find, it does not allow us to explain the dynamics of these interactions. We use this as an exemplar to make the case for the complementarity of CL and CA.

Increasingly, CL is being applied to contexts and domains outside of the study of language itself where the focus is on the *use* of language in a given context. Such contexts include courtrooms and forensic linguistics (Cotterill 2010), the workplace (Koester 2006), educational contexts (O'Keeffe & Farr 2003, Walsh & O'Keeffe 2007), political discourse (Ädel 2010), the media (O'Keeffe 2006), among other areas. In all of these cases, CL is used as a tool and another approach, such as CA, discourse analysis or pragmatics, is drawn on as a framework. To call CL a methodological tool is not to denigrate it. None of the above studies could have achieved the same insights without CL. Essentially, we are distinguishing between "pure" CL research and research which *applies* CL. In the former, where the description of the language of the corpus is an end in itself (descriptive corpus research), it helps us find out more about the use of language in a certain context. The latter type of research, on the other hand, looks at the wider interactional context of language in use. In this case, the corpus and its

description is not an end in itself, the corpus is merely a means to the end of finding out more about a broader research question.

We also note the salience of our argument for research into spoken discourse. Perhaps it is no coincidence that as more and more modestly-sized specialised corpora emerge we see more "inter-marrying" between CL and other approaches to the analysis of discourse in context (see Santamaría-García 2010, this volume). As McCarthy & O'Keeffe (2010) point out, in the early days of CL, the aim was to have very large written corpora to serve the needs of lexicographers, whose focus was obviously on semantic and lexical patterning rather than on discourse context. As a result, large corpora were lexically rich but contextually poor. That is, when a researcher looks at a lexical item in a mostly written corpus of 100 million words or more, it is detached from its context. However, when the researcher records, transcribes, annotates and builds a small contextualised spoken corpus, a different landscape of possibilities opens up in areas beyond lexis to areas of use (especially issues of pragmatics, interaction and discourse).

Taking an overview of how we staged the combination of CL and CA in this study, the first layer of analysis (using CL) allowed us to scope out and quantify recurring linguistic features. By linking to the context of these recurring features in the corpus, we were led to contextual "patterns". The second layer of analysis (using CA) draws upon these contextual patterns in the quantitative analysis and investigates them more closely. For example, in the corpus exploration, there were interesting findings around the frequency and use of certain discourse markers, which clustered around specific contexts. This led us to a closer CA led investigation which, in turn, produced interesting findings above the level of turn and in relation to specific interactional features. The process was non-linear in that we sometimes used CL tools within the CA layer of analysis to quantify CA insights. We can say that our analysis progresses in an iterative manner: from CL to CA, back to CL and so on. There is an interdependence between the two modes of analysis.

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: Small group teaching in higher education

In many higher education settings, small group teaching (henceforth SGT) contexts such as seminars and tutorials are used to support lectures by allowing tutors and students to engage in discussion and debate. To take the example of one subject, Psychology, SGT can account for around 40% of the contact time of first and second year undergraduates and up to 75% of final year and post-graduate students (Bennet et al. 2002). From the perspective of corpus linguistics, much influential work on spoken interaction in higher education is based on the Michigan Corpus of Academic Spoken English or MICASE (Simpson et al. 2002). This 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 have explored a wide range of phenomena in academic spoken interaction, such as metadiscourse in lectures (Lorés 2006), the use of conditionals (Louwerse et al. 2008), and, of more direct relevance to this study, the effect of class size on lecture discourse (Lee 2009).

Outside corpus linguistics, 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". Such work has focused on cues and signals used to manage interaction and participant roles (Viechnicki 1997), sequential organisation and negotiation of meaning (Basturkmen 2002), the issue of 'topicality' in small group discussion (Stokoe 2000, Gibson et al. 2006), and the formulation and uptake of tasks and resistance to "academic" identities (Benwell & Stokoe 2002). 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, SGT sessions are seen as locally produced accomplishments in which participants take actions to further their own goals and agendas and display their orientations to others' actions and make relevant certain identities. In SGT contexts, tutors will demonstrably orient to the accomplishment of pedagogical goals and tasks, and students may accept or resist these actions (Benwell & Stokoe 2002). At all times during interaction in these SGT contexts, as in other educational contexts, there is a complex relationship between pedagogic goals and the talk used to realise them. By looking closely at the interactions taking place in SGT settings, we show that tutors and students engage in tightly organised and intricate negotiations of a set of pedagogic agendas, and in doing so, use as tools both the

machinery of interaction (Levinson 2006) and specific linguistic features, such as discourse markers, to achieve their goals.

3. Corpus analysis

The study is based on data from the *Limerick-Belfast Corpus of Academic Spoken English* (hereafter LI-BEL), which currently comprises circa 1,000,000 words of recorded lectures, small group seminars and tutorials, laboratories and presentations, circa 500,000 words of which are transcribed. 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 words was created by identifying all the instances of SGT, sessions labelled in the corpus as "tutorial" or "lab" and comprising up to 25 students.

Our analytic procedure went from CL to CA: our broad goal was to first identify the high frequency and key lexical features of this type of interaction, in comparison with a reference corpus. These features would be salient in the sense that they would be signalling what is quantitatively distinctive about this type of interaction from a lexical perspective using frequency lists and keywords analyses (as detailed below). The 'keyness' of these items was also explored qualitatively by going back to the data to find out what it was they were "doing" in the interactions. Plot dispersions aided this process as discussed and exemplified below. In other words, these items were seen as salient in a qualitative sense in that they cluster at certain stages in the interaction, thus signalling different functions, and often pedagogical goals. The CL analysis thus provides us with a broad picture of how participants were going about doing SGT sessions. We then use a CA methodology to explore sequences of interaction in greater depth where there was a clustering of these statistically significant, or salient, features.

We began our analysis by using *WordSmith Tools* (Scott 2008) to identify key words and word frequencies for both single words and multi-word units (henceforth, MWU). By MWU, we refer to units of two or more words sometimes referred to as 'n-grams', 'lexical bundles', 'lexical phrases', 'clusters', 'chunks', though with slightly varying definitions (see Greaves & Warren 2010). We used the one-million word *Limerick Corpus of Irish English* (LCIE) as a reference corpus for our keyword analysis. This contains over one million words of everyday conversation (see Farr et al. 2004). LCIE was chosen as the reference corpus firstly because it is a spoken corpus but

also because, like the data in the LI-BEL sub-corpus, it comprises Irish English speakers (the LI-BEL sub-corpus was recorded in Limerick and so primarily comprises Irish English speakers). Therefore, LCIE is the most suitable point of comparison. Table 1 illustrates the top 20 key words:

Table 1. Top 20 key words from LI-BEL sub-corpus using LCIE as a reference corpus

1	okay	11	any
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

Further analysis into the context using concordance lines revealed differences in the functioning of these key words. For example, *if* when used in 'first conditional' type structures had three main functions:

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

Another step that we took at this point was to look at the single-word frequency lists in comparison to another corpus, in this case the BNC academic component. Figure 1 illustrates the items of the key words from Table 1 in comparison with the BNC (per million word results). The rationale for doing this is simply to get another angle on the results, a triangulation process. It also brings to light frequency differences between spoken and written academic data. For example, one thing that came to light in this comparison of word lists was the frequency differences between the semi-modal/lexical verb *need* (not differentiated in the search). It is the most frequent of the modal/semi-

modal verbs in the sub-corpus but in relative terms it is used less frequently in the BNC data.
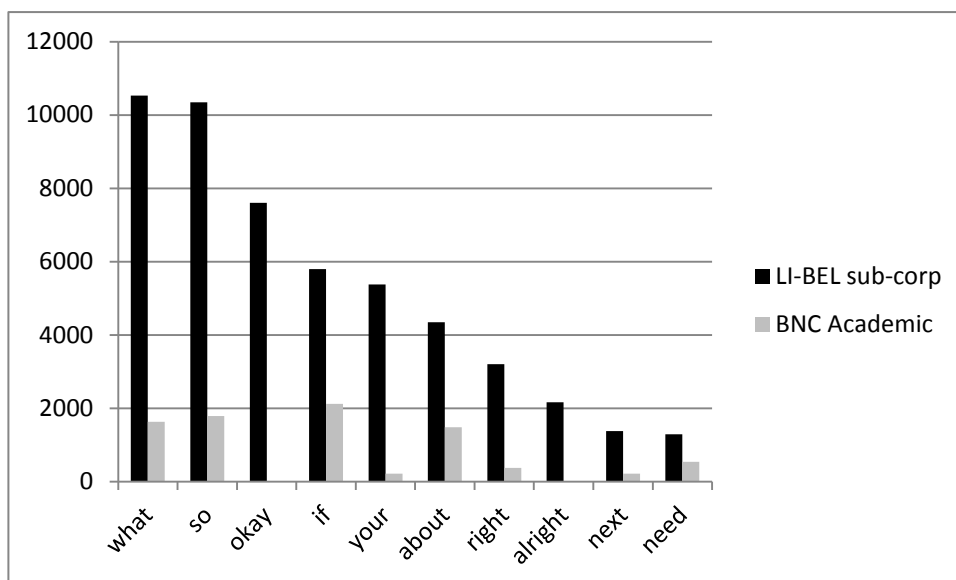

Figure 1. Single-word frequencies LI-BEL sub-corpus versus BNC Academic, per million words

In relation to items on both Table 1 and Figure 1, the context of SGT accounts for the presence of some of them quite readily through concordance line analysis, for example, the prevalence of the interrogative pronoun *what* (e.g. *What do you think of it?*), discourse markers *so, okay, alright*, deictic *next* (as in *next week, next semester, next lecture*). Concordancing also showed that the relatively high frequency of *need* in comparison to the BNC is relate to the speech act of giving instructions (*what I need you to do, you need to,* etc.).

At this lexical level therefore, the corpus data is pointing us to certain contexts such as eliciting information, signposting the discourse, locating learning and teaching in time and giving instructions to learners to perform certain actions and carry out tasks. However, these are just pointers that are emerging as hypotheses as a result of keywords, frequency counts, concordance searches. The next move for a corpus analyst is to look at patterns.

Two- to six-word multi-word units were generated with a cut-off frequency of four occurrences. These were then examined through concordance searches to lead to 128

items which were seen as most salient to the SGT context. Many items were eliminated at this point, for example the five-word unit *seventeen eighty to eighteen forty* occurs five times but on analysis of concordance lines, we find that it is uttered by one lecturer and all uses are in the context of the same history seminar. Many such "teaching content" examples were identified. At this stage, we have word frequency lists, key word lists, concordances and a list of MWUs. Having synthesised these, we were able to begin to classify them according to broad functions (see Table 2). We were also able to draw on plot dispersions to see where in one class transcript they clustered (as exemplified further below). Many of these items, we accept, can have multiple functions, but by looking at them closely in context, we were able to identify strong functional patterns into which most of the items broadly fall. We stress that these functions are by no means the only possible functions of these items in the data. Rather, they are the most prevalent functions. A similar approach is taken in McCarthy & Carter 2002 when they generate multi-word cluster lists based on the five-million word Cambridge and Nottingham Corpus of Discourse in English and come up with a list of broad functions of MWUs in spoken interaction. Most notable in this area is the work of Biber and his associates who look at lexical bundles in spoken and written academic contexts (see, for example, Biber & Barbieri 2007, Biber et al. 2002, Biber et al. 2004).

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

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*

Apart from the first category ("Elicitation"), the items broadly function as markers of discourse on the part of the speaker in order to orient the listener. They signpost, manage, demonstrate, sequence, set up activities/groups and they mark out shared and new knowledge (Fung & Carter 2007).

The use of dispersion plot graphs is another CL tool which we drew on. By way of example, compare the sample plot graphs for the high frequency items *last week, next week* and *okay* in the data (cf. Figures 2-4). These show us whether these items cluster at certain points and in which files (i.e. which interactions/classes). References to *last week* and *next week* prevail at the beginning and end of interactions whereas *okay* is more dense at the beginning of interactions but is used throughout as well, with "clusterings" around certain phases.
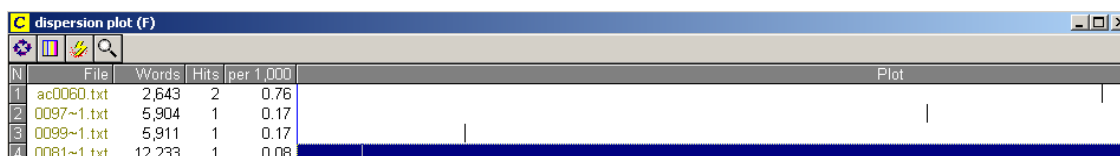

Figure 2. Sample dispersion plot graph of *last week* in LI-BEL sub-corpus
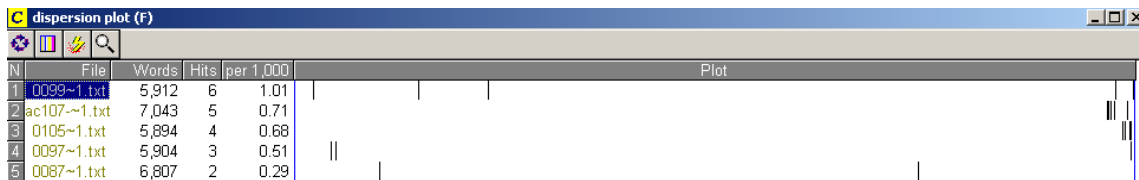
Figure 3. Sample dispersion plot graph of *next week* in LI-BEL sub-corpus
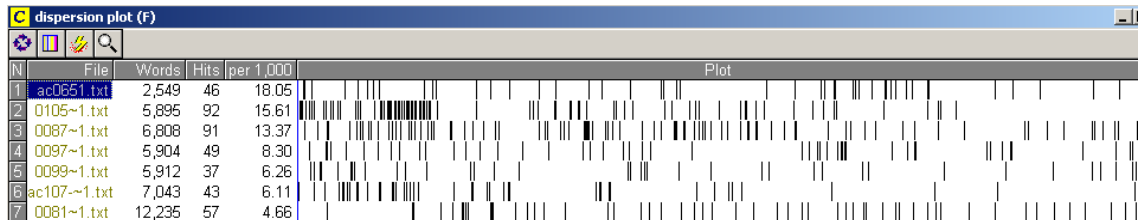

Figure 4. Sample dispersion plot graph of *okay* in LI-BEL sub-corpus

```
  in your pairs again I want you to look at this lesson plan. First of all I want you to
okay for two minutes I want you to discuss that in the sense of is there anything
                     I want you to tell me come back to me and tell me do you think
                     I want you to look at that lesson plan and I want you to critique
    okay? Number two I want you to decide amm would you be happy if I told you in the
 then over here what I want you to do is take any line in the middle make a wavy line
Yeah get them up not I want you all kind of sitting beside each as you go around okay?
l like this. So what I want you to practice is roughly around maybe five eight mils
   setsquares. What I want you to do is do a bit of printing in there. Lovely lovely
```

Figure 5. Sample concordance lines of *I want you/ye to*

However, in order to gain a deeper understanding of spoken interaction in this context, we needed to see how the salient features we identified actually operated in speakers' turns and in longer sequences of interaction. It is at this point that we need to bring in another framework. By looking at micro-contexts within a CA framework, we were able to bring their interactional and pedagogical relevance into relief (as we detail below). 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.

4. CLCA analysis

When we looked at how the single words and multi-word items which turned out to be salient in the sub-corpus were used over phases of interaction, we found that they played an important role as resources for participants' courses of action or 'interactional projects'. Schegloff (2007: 244) describes interactional projects as a form of

interactional organization in which a course of conduct "is developed over a span of time (not necessarily in consecutive sequences) to which co-participants may become sensitive, which may begin to inform their inspection of any next sequence start to see whether or how it relates to the suspected project, theme, stance, etc.". These interactional projects are less tightly bound than the kinds of sequences or 'sequences of sequences' built up out of adjacency pairs, although they can themselves include such sequences, but they do set up specific types of identifiable speech exchange systems within SGT sessions.

In producing these speech exchange systems participants use the different 'organizations of practice' (Schegloff 2007: xiv) such as turn design, turn-taking, orientation to actions such as requesting and telling, building coherent sequences through adjacency pairs, repairing trouble, word selection and overall structuring of the interaction, in specific ways. In SGT interaction, in common with other types of pedagogical interaction, it is the tutor's interactional project to pursue pedagogical goals, and this leads to a reflexive relationship between such goals and the "shape" of the interaction (Seedhouse 2004). In the dataset, we identified four such speech exchange systems, and in this section we describe their characteristics and illustrate them with data extracts. Our aim is to show the items identified as salient in the corpus findings in their interactional environments, and to show how they are used alongside some of the other organizations of practice to build four main speech exchange systems through which SGT sessions are talked into being. We leave for a further report how such systems combine in the overall structuring of SGT sessions. In the following we / In Sections 4.1 -4.4 below we detail the four main speech exchange systems.

4.1 Procedural talk

Throughout the corpus there are sustained sequences of interaction where participants orient to the interactional project of informing, and being informed, about different types of procedural matters. We labelled this speech exchange system 'procedural talk'. These stretches of interaction often involve very long turns by the tutor, with minimal verbal contributions from the students. For example, the tutor may perform the role of both questioner and answerer as he/she talks through a procedure.

(1)

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

2	in that first box (.)
3	what would your first thing to do be (.)
4	well would you kind of come in here like this
5	and go like that?
6	no you wouldn't
7	because free hand you don't need to have
8	the sheet attached fully to the desk

At lines 1-5 the tutor produces two consecutive first-pair parts of question-answer adjacency pairs. At line 6, he supplies the second-pair part to the second question, in doing so cancelling out the conditional relevance of the second pair part of the first adjacency pair. There are no gaps for the other party in the interaction (the students) to supply second-pair parts. Micro-pauses such as the one at line 3 are not hearable as a handing over of the floor in this speech exchange system[2]. Orientation to the other party is carried on through the use of the discourse marker *okay*:

(2)

1	T:	but say for example
2		I want to do an example
3		more like this (.) okay? (.)
4		again this time
5		though I want to
6		actually reuse the cell okay?

Even though *okay* is used with a rising intonation, and can be seen as having the function of checking understanding, it is not hearable as an invitation for the students to verbally confirm understanding.

Figure 5. Sample concordance lines of *I want you/ye to*

```
in your pairs again I want you to look at this lesson plan. First of all I want you to
okay for two minutes I want you to discuss that in the sense of is there anything
                     I want you to tell me come back to me and tell me do you think
                     I want you to look at that lesson plan and I want you to critique
   okay? Number two I want you to decide amm would you be happy if I told you in the
 then over here what I want you to do is take any line in the middle make a wavy line
Yeah get them up not I want you all kind of sitting beside each as you go around okay?
l like this. So what I want you to practice is roughly around maybe five eight mils
    setsquares. What I want you to do is do a bit of printing in there. Lovely lovely
```

Procedural talk is indexed in the corpus findings by the use of MWUs which have the function of indicating actions which the tutor wants the students to do, such as the four word MWU *I want you to* (see concordance lines in Figure 5) or, less frequently, the six-word MWU *what I want you to do*, as in this example:

(3)

1	T:	what I want you to do
2		is take any line in the middle
3		make a wavy line like so
4		and then practise the bit of parallel drawing once again
5		offset that a certain distance
6		and draw your lines offset perfectly from that.

In this extract, the interactional project and associated pedagogical goal is the demonstration of a procedure. Thus, the actions the tutor wants the students to do are likely to be more immediate in time, or may even be carried out by them as the tutor speaks. However, procedural talk can refer to actions more remote in time:

(4)

1	T:	I'll talk to ye about those
2		the next day (.)
3		that's week nine (.)
4		this is only week two isn't it?
5		so erm we're okay with regards to that (.)
6		you see now erm okay

In this extract we see the temporal orientation to future events at lines 2 and 3. These references tend to occur where tutors are detailing programme requirements, such as the dates for handing in assignments.

4.2 Didactic talk

When we looked at the MWUs associated with the function of eliciting information from individuals or groups, or giving evaluative feedback, we also saw that they were indicative of a specific speech exchange system. One of the most frequent was *tell me*, examples of which can be seen in the concordance lines in Figure 6.

```
    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
                      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
```

Figure 6. Extracts from concordance lines of *tell me*

In this speech exchange system, turn-taking is tightly controlled by the tutor, with next turn allocation firmly in his/her hands and questions addressed to individual participants, thus making the respondee's provision of the second pair-part strongly relevant. The interaction is reminiscent of classic descriptions of classroom discourse (use of Initiation Response Feedback, IRF, exchanges, teachers' use of display questions, students nominated to provide answers, short utterances from students). We labelled this context 'didactic talk'.

Extract 5, from a teacher education seminar, shows *tell me* as part of the four-word MWU *can you tell me* (line 12) and two other MWUs which have the function of elicitation in this interactional context (lines 16 and 21):

(5)

1	T:	say now it's -
2		we'll give a a topic example again
3		(.) ah let me see
4		what would you be doing at all?
5		the same genetics again
6		we'll go back to that (.) erm (.)
7		you're teaching your fifth year class genetics
8		and you say pupils should in

9		will be enabled to?
10	S:	(find out?)
11	T:	perfect find out (.)
12		and can you tell me what type of
13		an objective that is?
14		they're labelling a diagram (.)
15		what type of an objective is that?
16		anyone have any idea?
17		remember I was telling you
18		that there's three types of objectives (.)
19		what type of objective is that?
20		(0.5)
21		can anyone name any one of the three?
22		(0.5)
23		oh this is fun at nine isn't it?
24		kind of you know twenty questions

The speech exchange system here is a 'sequence of sequences', more exactly a 'question series' (Schegloff 2007: 207). That is, it consists of a series in which the same sequence type (in this case question-answer) is done repeatedly with the same parties as first pair part and second pair part speaker. In terms of turn design, tutors in this context work hard to ensure that the first pair part of the question-answer adjacency pair will meet with a preferred response. In Extract 5 we see how the tutor builds up to the production of the base first pair part by filling in contextual information. At lines 1-7 the tutor does this "pre" work before delivering the first pair part at line 8. After delivering the first pair part of the next question in the series at lines 12 and 13, the tutor backtracks to remind the students of information they may need to answer the question. It is as if 'pre-expansions' (Schegloff 2007) instead of being done interactionally as preliminary adjacency pairs, are "packed into" the tutor's turns, presumably with the same aim of pre-expansion, that is to try to ensure the successful delivery of the base second pair part. Interestingly, there is evidence that the nature of the speech exchange system itself is salient to at least one of the participants, the tutor, as can be seen in his metadiscursive comment at lines 23 and 24.

4.3 Empathic talk

The discourse markers which had the function of indicating "shared space" between participants were indexical of a type of interactional project and resulting speech exchange system that we labelled 'empathic talk'. In orienting to this interactional project, (mainly) students produce "tellings" - accounts of experiences that they are having as part of the course, often accompanying these tellings with 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/monitor shared knowledge (*you know*; *you see,* see Carter & McCarthy 2006, Andersen 2001). The overall pattern, then, is another type of action sequence, but this time more akin to reciprocal or exchange sequences (Schegloff 2007: 195-207), in which the same sequence (telling followed by assessment/reflection) is repeated but with different speakers occupying the relevant slots.

In terms of the turn-taking practices in this speech exchange system, the overall pattern is more symmetrical than either procedural or didactic talk. However, turn allocation is still controlled by the tutor at a crucial point – that is, the interactional project is launched by the teacher eliciting tellings from the students, as in this example:

(6)

1	T:	how are you getting on with
2		your other uh module (.)
3		uh the the the filming one
4	S3:	we're filming a scene at the moment
5		we're editing and it's crazy
6	T:	yeah you see it is crazy isn't it (.)
7		this week now is going to be unbelievable
8	S3:	it's just music and we're just putting it together
9		you see now (.) you know you've all the footage
10		but you're there trying
11	S:	( )

12	S3:	we have so much footage
13		and it's just like you
14		some people have to accept that some of it
15	S:	( )
16	T:	yeah well it's like essays isn't it (.)
17		I mean you can't write from the middle you know=
18	S4:	=that was my idea
19		so we can't lose that
20		and you're like going=
21	T:	=who who's the director?
22	S3:	I wish I was the director
23	S2:	in our in my group John C is
24	T:	okay yeah you see that's the thing like you know
25		I mean like really
26		it does all come down to the director
27		and the people should respect that immediately
28		you know (.)that doesn't happen that often (.)
29		you know what I mean
30		it can get- the roles can get dispersed

In response to the tutor's elicitation (lines 1-3), S3 launches a telling which is an account of a group's experiences of making a film. The turn includes an assessment at line 5 (*it's crazy*), to which the tutor offers a preferred (agreeing) response with the acknowledgment marker *yeah*, the discourse marker of shared space *you see* and the repetition of the assessment. At least two interactional projects seem to be in operation here. For the students, it is an opportunity to express feelings such as frustration with aspects of the course or with other students' behaviour. At line 14, S3 indicates that *some people* may have problems in accepting that material has to be cut, and at line 22, seems to be expressing frustration either about the existing director, or the lack of a director's role in the group.

The tutor orients to an interactional project in which they have to strike a balance between accepting students' experiences and the pedagogic goal of reinforcing appropriate behaviours and identities in the context of professional practice. This can be seen in the tutor's last turn in the extract (lines 24-30), 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 switch identity orientation from that of equal interactant to that of tutor, to move from a position of role symmetry to one of role asymmetry.

4.4 Argumentational talk

The fourth type of interactional project was one in which there was also a shared cognitive space, but this time it is disputed rather than used to build empathy. This speech exchange system, which we labelled 'argumentational talk', was indexed in the corpus data by the appearance of *but* as a "tying" element at the beginning of turns (Tao 2003), sometimes preceded by an agreeing or acknowledging token such as *yeah* or *okay* (see Figure 7 for examples from the corpus).

```
hurling and        laughing        you know. But this is it. This is our oral tradition.
          symbol of ahh well regression I suppose. But what did happen in Limerick was
the amm
aying about his intro= or you know         but you're not allowed say that. What's your
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?
```
Figure 7. Sample concordance lines of *but* preceded by a discourse marker in argumentational talk

Extract 7, which is from a politics seminar, is a clear example of argumentational talk in action, with an example of turn-initial *but* at line 5, and turn-initial negation at line 9:

(7)

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
6           when I see territory people
7           and and its its ((unintelligible))
8           you have a nation (.) ahh it's one of the=
9    T:     =not for the people
10          who claim they're ethnically different
11          but a people within the nation=
12   S5:    =yeah I suppose=
13   T:     =and that that's kind of the point is that (.) yes
14          you'd you'd think on the outer we'd have a nation
15          but if it worked out like that
16          well then we wouldn't have ethnic conflict (.)
17          do you get my point (.)
18          okay so (.) any anybody else think
19          there's anything else there
20          that should be up there
```

In line 1 we see evidence of greater scope for student agency in that the first pair part of a question-answer sequence is launched by a student. The tutor's rather equivocal answer, with the self-repair at lines 3 and 4, is met by the student's quite strongly disaffiliative *but*. This is met at line 9 by the teacher's use of a turn-initial negative to rebut the student's assertion. At line 12, the student produces a more strongly affiliative acknowledgement token *yeah* (Drummond & Hopper 1993), but it is immediately mitigated by the addition of *I suppose*. Having secured some agreement, the tutor moves the sequence towards its close by producing what could be analysed as a "post-completion musing" (Schegloff 2007: 143) in which he offers an analysis or assessment of the prior sequence as *the point* before going on to open a new sequence at lines 18-20. This extract is an illustration of something seen in other examples of argumentational talk in the corpus: it can be a risky strategy as it may expose the tutor's epistemological authority by allowing for assertions and counter assertions. Thus, in this case, and in others throughout the corpus, tutors often quite hastily moved to close down this speech exchange system by using sequence-closing third turns such as assessments or post-completion musings.

5.   Discussion

This paper set out to use corpus linguistics (CL) and conversation analysis (CA) to provide enhanced descriptions of spoken interaction in a small group teaching higher education context. From the data and subsequent CLCA analysis, we were able to characterise four speech-exchange systems, each with distinctive interactional, linguistic and pedagogic features or 'fingerprints' (Drew & Heritage 1992: 26). The four speech exchange systems we identified are robust throughout the data. That is, at any point one or other will be operating, whether for long spates of interaction or for shorter bursts.

Our results allowed us to make comparisons both within and across these interactional contexts. For example, when we compare didactic and empathic talk, very different profiles or 'fingerprints' emerge. The former is characterised by short learner turns, tightly controlled turn-taking, evidence of IRF exchange structures, extensive use of the MWUs *tell me* and *can you tell me* and the main pedagogic function of eliciting. The main focus of empathic talk, on the other hand, is "show and tell": the tutor's pedagogic goal is to promote debate and discussion and create a safe environment for that to take place.

In linking the corpus findings to the CA analysis, we observe that in all four speech exchange systems, participants use the single words and MWUs which we found to be salient in the data for carrying out specific actions that move forward their interactional projects. Thus they are helpful both to participants and analysts in solving what Schegloff (2007) describes as the 'action-formation' problem: that is, how language formations are designed to be recognizable by interlocutors as particular actions, such as requesting, telling, eliciting, etc. Not only are these units used by participants to carry out specific acts, but they function as indices, both for participants and for ourselves as analysts, of the current speech exchange system one is in. For this reason, they are bound up with the interactional competence displayed by participants in SGT sessions as they move forward their particular agendas and respond appropriately at any moment in the interaction.

6.   Conclusion

Although there have been many attempts to characterise spoken interaction in educational settings by focusing on micro-contexts (see, for example Seedhouse 2004; Walsh 2006), none, as far as we know, offer the same level of detail as the present

study. From our CLCA analysis, we are able to provide detailed descriptions of the interaction from three perspectives: linguistic (portraying the use of high frequency items, keywords, MWUs, discourse markers, question forms and so on), interactional (focusing on turn-taking and turn design, sequential organization, etc.) and pedagogic (looking at specific pedagogic functions at a given moment to include eliciting, explaining, instructing and so on).

Had we used CL on its own we would have achieved interesting lists of high frequency items which we could have explained functionally but it would not have brought us anywhere near the depth of understanding compared with what a CA framework could explain. Had we looked at the data purely from a CA perspective, we would have possibly identified the four main speech exchange systems but we would not have been able to back up the fact that the words and patterns they contain were actually high frequency items (that is, keywords, 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 (in this case LCIE). We can therefore safely assert that CL and CA are "well met".

In this paper, we have looked at the organization of practice at the level of action orientation and the building of sequences. Further research could look at how the different sequences (i.e. the four types of talk) combine and cohere in the overall sequential organization of SGT sessions. This indeed should turn out to be a fruitful line of enquiry, and one in which the combined CLCA method used in this study would be extremely appropriate.

Notes

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.*

2. In line with Sidnell (2010: ix), the CA notation (.) indicates 'a "micropause", hearable, but not readily measurable without instrumentation, ordinarily less than 0.2 of a second.'

References

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

Andersen, G. 2001. *Pragmatic Markers and Sociolinguistic Variation. A Relevance-theoretic Approach to the Language of Adolescents*. Amsterdam/Philadelphia: John Benjamins.

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.

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. & Barbieri, F. 2007. "Lexical bundles in university spoken and written registers". *English for Specific Purposes*, 26 (3), 263-286.

Biber, D., Conrad, S. & Cortes, V. 2004. "*If you look at…*: Lexical bundles in university teaching and textbooks". *Applied Linguistics*, 25 (3), 371-405.

Biber, D., Conrad, S., Reppen, R., Byrd, P. & Helt, M. 2002. "Speaking and writing in the university: A multidimensional comparison". *TESOL Quarterly*, 36 (1), 9-48.

Carter, R. & McCarthy, M. J. 2006. *Cambridge Grammar of English. A Comprehensive Guide to Spoken and Written Grammar and Usage*. Cambridge: Cambridge University Press.

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

Drew, P. & Heritage, J. 1992. "Analyzing talk at work: An introduction". In P. Drew & J. Heritage (Eds.), *Talk at Work: Interaction in Institutional Settings*. Cambridge: Cambridge University Press, 3-65.

Drummond, K. & Hopper, R. 1993. "Back channels revisited: Acknowledgment Tokens and speakership incipiency". *Research on Language & Social Interaction,* 26 (2), 157-177.

Farr, F., Murphy, B. & O'Keeffe, A. 2004. "The Limerick Corpus of Irish English: Design, description and application". *Teanga,* 21, 5-29.

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

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.

Greaves, C. & Warren, M. 2010. "What can a corpus tell us about multi-word units?". In A. O'Keeffe & M. J. McCarthy (Eds.), *The Routledge Handbook of Corpus Linguistic.* London: Routledge, 212-226.

Koester, A. 2006. *Investigating Workplace Discourse*. London: Routledge.

Lee, J. 2009. "Size matters: An exploratory comparison of small- and large-class university lecture introductions". *English for Specific Purposes,* 28 (1), 42-57.

Levinson, S. 2006. "Cognition at the heart of human interaction". *Discourse Studies,* 8 (1), 85-93.

Lorés, R. 2006. "The referential function of metadiscourse: Thing(s) and idea(s) in academic lectures". In A. Hornero, M. Luzón & S. Murillo (Eds.), *Corpus Linguistics: Applications for the Study of English*. Bern: Peter Lang, 315-334.

Louwerse, M., Crossley, S. & Jeuniauxa, P. 2008. "What if? Conditionals in educational registers". *Linguistics and Education,* 19 (1), 56-69.

McCarthy, M. J. & Carter, R. A. 2002. "*This, that and the other*: Multi-word clusters in spoken English as visible patterns of interaction". *Teanga,* 21, 30-52.

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

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

O'Keeffe, A. & Farr, F. 2003. "Using language corpora in language teacher education: Pedagogic, linguistic and cultural insights". *TESOL Quarterly*, 37 (3), 389-418.

Scott, M. 2008. *WordSmith Tools. Version 5*. Liverpool: Lexical Analysis Software.

Schegloff, E. A. 2007. *Sequence Organization in Interaction: A Primer in Conversation Analysis, Vol. 1*. Cambridge: Cambridge University Press.

Seedhouse, P. 2004. *The Interactional Architecture of the Language Classroom: A Conversation Analysis Perspective*. Oxford: Blackwell.

Simpson, R. C., Briggs, S. L., Ovens, J. & Swales., J. M. 2002. *The Michigan Corpus of Academic Spoken English*. Ann Arbor, MI: The Regents of the University of Michigan.

Sidnell, J. 2010. *Conversation Analysis: An Introduction*. Chichester: Wiley-Blackwell.

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

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

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. & 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.

Authors' addresses

Steve Walsh

School of Education, Communication and Language Sciences

King George VI Building

Newcastle University

Queen Victoria Road

Newcastle upon Tyne

NE1 7RU

United Kingdom

steve.walsh@newcastle.ac.uk

Anne O'Keeffe

Department of English Language and Literature

Mary Immaculate College

University of Limerick

South Circular Road

Limerick

Ireland

anne.okeeffe@mic.ul.ie

Tom Morton

Departamento de Filología Inglesa

Universidad Autónoma de Madrid

28049-Madrid

Spain

tom.morton@uam.es