

# Corpus Linguistics and Conversation Analysis at the Interface: Theoretical Perspectives, Practical Outcomes

Steve Walsh

## 1 Introduction

In this chapter, I offer a perspective on the merits and possible drawbacks of a combined corpus linguistics (CL) and conversation analysis (CA) methodology. The first part of the chapter provides a theoretical perspective of each methodology, considering their respective epistemological and ontological origins and traditions, before moving on to discuss how they might – in spite of their very different research positions – be used together, in combination. The broad argument for combining CL with CA is that CL is unable to account for some of the features of spoken interaction which occur at the levels of utterance and turn and largely ignores context, while CA is unable to identify linguistic patterns across larger corpora, limiting itself instead to detailed descriptions of small quantities of data. Each methodology, then, has its strengths and weaknesses – in combination, they have the potential to offer enhanced descriptions of spoken interaction. Using a combined CL and CA approach (henceforth, CLCA), cumulatively gives a more ‘up-close’ description of spoken interactions than that offered by using either one on its own. A CLCA analysis provides powerful insights into the ways in which interactants establish understandings and observe how words, utterances and text combine in the co-construction of meaning.

---

S. Walsh (✉)

School of Education, Newcastle University, King George VI Building,  
Newcastle Upon Tyne NE17RU, UK  
e-mail: steve.walsh@ncl.ac.uk

## 2 Corpus Linguistics: Epistemology and Ontology

One of the key methodological underpinnings common to both CL and CA is that they make use of corpora; their point of departure is always the building of a corpus. A corpus is a collection of texts that is stored on a computer; texts may be spoken or written, but for the purposes of this chapter, we are concerned only with spoken texts. Texts are examples of spoken discourse which have been recorded and transcribed and which include conversations, phone calls, university seminars, debates, etc. Essentially, any spoken discourse, produced in context and for a genuine purpose, can be regarded as a text. A corpus therefore is a collection of real language that people use in all types of situations.

The emergence of corpus linguistics goes back to the 1970s and 1980s when computers were being developed that were powerful enough to store and search large databases of stored texts. At this time, the main use of corpora was in the production of dictionaries – today, all major publishers producing dictionaries use corpora. The main advantage is clear: rather than relying on intuition, lexicographers were able to search very large databases to find examples of real language in use. The use of invented – or idealised – examples became a thing of the past. Today, computers can be used to search up to a billion words at any one time to identify examples and see how language is really used. Perhaps the most revolutionary work in the area of dictionary production at this time was the *Collins Birmingham University International Language Database* (COBUILD) project. This was set up at the University of Birmingham in 1980 under the direction of John Sinclair. From this database, 16 dictionaries have been produced to date, most notably the *Collins COBUILD English Language Dictionary* (1987, 2nd edition 1995, 3rd edition 2001, 4th edition 2003) and the *Collins Cobuild Grammar Patterns* series (1996; 1998).

While the main focus of the early CL work was lexicography, these studies also led to a focus of attention on grammar, and, in particular, heightened understandings of the relationship between words and grammar: *lexico-grammatical* features of language. What this focus of attention did was to direct attention towards the importance of words and chunks of words in grammatical relationships, rather than regarding grammar as the most important language system. Vocabulary suddenly became at least as important as grammar in our emerging understandings of language systems. Many grammatical relationships could also be linked much more to words.

Today, most grammar books of English are corpus-informed, a process which has many advantages. First, like lexicographers, grammarians no longer have to rely on their intuitions – examples can be derived from a corpus; more importantly, the ‘rules’ of grammar can also be derived from the corpus since patterns can be more easily established by looking at numerous examples. A second advantage is that it is now much easier to identify relationships across different text types and study how, for example, spoken grammars are different to written ones (Carter and McCarthy 2006), or how certain language structures are more common than others in some text-types (e.g. newspaper articles). Related to this, is the point that

corpus-based grammars can now make clearer claims about regional varieties such as differences between American and British or Irish English. Corpora also allow comparisons to be made over time, allowing us to comment on how certain grammatical features are more or less widespread; for example, *can I* is more common today in most contexts than *may I*.

If one of the main concerns of linguistics and, to some extent applied linguistics, is the study of patterns of use in language, CL has made the whole process much easier and faster. It is now possible to compare huge databases and make reliable claims about how language is *actually* used in context, rather than prescribing how it *should* be used. From a pedagogic perspective, the advantages of this are obvious and too numerous to mention here. CL, then, gives us, at a glance, an overview of how a particular word or grammatical structure is used across a range of contexts and text-types.

When CL was in its infancy and being used mainly in the production of dictionaries, the main focus was on building large corpora: the bigger the better. The reason for this is both to ensure that as many examples as possible are available, and also to ensure that rarer words, or words which are less commonly used, could also be studied with the same degree of reliability. Essentially, the larger the sample, the more accurate are the claims which can be made about a particular feature. The trend of aiming for large or very large corpora has, to a large degree, come to an end. There has been a shift in recent years towards using smaller, more context-specific and locally derived corpora in order to highlight specific examples of language use in spheres such as business, medicine, science, classrooms or everyday conversations. These more specific corpora may be used, for example, by translators and materials designers. For a translator working on a medical document, a small corpus (for example five lakh to one million words) of medical articles, is more useful than a general corpus of ten million words. Equally, an author of business text books could find out a lot more from one million words of business language than from a much larger general corpus. Smaller still are corpora used for research; it would, for example, be quite feasible to conduct a small-scale research project using a corpus of 100,000 words, providing that it was designed with a specific context in mind.

In this chapter, then, CL is presented as a *methodological tool* which can be used to investigate, for example, 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 key words in micro-seconds, thus revealing patterns that we could not otherwise find, it does not allow us to explain the dynamics of these interactions. One of the main reasons for using a combined CLCA methodology is that CA does allow us to reveal in some detail which is actually 'happening' in interactions. We return to this below.

To return to the arguments made above about the importance of CL in the study of language *use*, it is probably fair to say that CL is being increasingly 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, educational contexts (O'Keeffe and Farr 2003; Walsh and 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. Under this ‘applied’ view of CL, language in use is the prime focus and the research endeavour is to uncover, using a complementary methodology, the broader interactional context in order to gain understandings of ‘what is really happening’. The interest lies less in the linguistic features *per se* and more in what is being accomplished through their use. So, for example, we might be interested in studying the ways in which discourse markers are used in an educational context (cf. Yang 2013), or the use of modal verbs in transactional encounters. In both instances, the corpus and its description is not an end in itself, but a means to finding out more about a broader research question.

One of the consequences of the recent shift towards smaller corpora (O’Keeffe et al. 2007) is that there has been a corresponding movement towards combining CL with other methodologies, particularly when the focus is on spoken discourse. As McCarthy and 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). We can say, then, that there has been some ‘meeting of the ways’ between CL and CA approaches: both CL and CA highlight the importance of context, albeit in different ways, and CL has recently started to recognise the value of smaller, context-specific corpora.

Before considering in more detail the relative merits (and shortcomings) of a combined CLCA methodology, I offer an overview of the origins and research traditions of CA.

### 3 Conversation Analysis: Epistemology and Ontology

The origins of conversation analysis (CA) lie in sociology, not linguistics or applied linguistics. The original interest arose out of a perceived need to study ordinary conversation as social action; CA’s underlying philosophy is that social contexts are not static but are constantly being formed by participants through their use of language and the ways in which turn-taking, openings and closures, sequencing of acts, and so on are locally managed (Sacks et al. 1974). Interaction is examined in relation to meaning and context; the way in which actions are sequenced is central to the process. In the words of Heritage (1997, p. 162):

In fact, CA embodies a theory which argues that sequences of actions are a major part of what we mean by context, that the meaning of an action is heavily shaped by the sequence of previous actions from which it emerges, and that social context is a dynamically created thing that is expressed in and through the sequential organisation of interaction.

According to this view, interaction is *context-shaped* and *context-renewing*; that is, one contribution, or ‘turn-at-talk’ is dependent on a previous one and subsequent contributions create a new context for later actions. Context is “both a project and a product of the participants’ actions” (Heritage 1997, p. 163). According to Sidnell (2010, p. 1), CA aims to “describe, analyse, and understand talk as a basic and constitutive feature of human social life”. In its early days, CA focused on describing conversations between friends; only later did it look at institutional settings (see below).

According to Seedhouse (2005, pp. 166–67), the basic principles which CA adopts are:

There is order at all points in any interaction: talk- in-interaction is systematically organised, deeply ordered and methodic.

Contributions to interaction are context-shaped and context-renewing (see above).

No order of detail can be dismissed as disorderly, accidental, or irrelevant (cf. Heritage 1984): CA has a detailed transcription system, and a highly empirical orientation.

The analysis is bottom-up and data driven: researchers should approach the data without prejudice or bias and adopt CA’s principle of ‘unmotivated looking’.

One of the main concerns of CA is turn-taking in talk-in-interaction (Hutchby and Wooffitt 2008). Adjacency pairs, repair, and preference are the other main foci of attention. In CA, the basic unit of analysis is a Turn Constructional Unit (TCU), approximately the same as a single utterance which carries meaning. A single turn may comprise several TCUs and any single TCU may indicate the end of a turn, marked by a transition relevance place (TRP), at which point any other speaker may take the floor, or the original speaker may retain his or her turn. This basic turn-taking mechanism underpins all CA research, which adopts the ‘next turn proof procedure’ (REF) as an indicator of the robustness of the method. Essentially, any one turn-at-talk can be related to any other turn in a logical and systematic way so that analysts view the interaction in the same way as participants.

Apart from turn-taking, another area of interest for CA is adjacency pairs, based on the premise that much human communication proceeds through paired utterances; greeting/greeting, question/response, invitation/acceptance, etc. An understanding of adjacency pairs entails a realisation that there are preferred and dispreferred second pair-parts. So, for example, the preferred second-pair part of invitation is acceptance. Space precludes a fuller treatment of adjacency pairs and preference structure, but see, for example Schegloff (2007) and Hutchby and Wooffitt (2008).

The final system which is of concern to CA is repair, defined as “the treatment of trouble occurring in interactive language use” (Seedhouse 2004, p. 34). Repair is essential for intersubjectivity, or mutual meaning-making, and interactants constantly make use of a range of repair strategies in order to understand and be understood. There is no limit to what can be repaired in spoken interaction, making it a key method for interactants to achieve mutual understanding.

Although the original focus of CA was naturally occurring conversation, it is perhaps in specific institutional settings, where the goals and actions of participants are clearly determined, that the value of CA approaches can be most vividly realised.

The discussion turns briefly to an institutional discourse perspective before looking specifically at CA in the L2 classroom.

An institutional discourse CA methodology takes as its starting-point the centrality of talk to many work tasks: quite simply, the majority of work-related tasks are completed through what is essentially conversation, or “talk-in-interaction” (Drew and Heritage 1992, p. 3); many interactions (for example, doctor-patient interviews, court-room examinations of a witness, classrooms) are completed through the exchange of talk between specialist and non-specialists (*ibid.*):

Talk-in-interaction is the principal means through which lay persons pursue various practical goals and the central medium through which the daily lives of many professionals and organizational representatives are conducted.

The purpose of a CA methodology in an institutional setting is to account for the ways in which context is created for and by the participants in relation to the goal-oriented activity in which they are engaged (Heritage 1997, p. 163). All institutions have an over-riding goal or purpose which constrains both the actions and interactional contributions of the participants according to the business in hand, giving each institution a unique interactional “fingerprint” (Heritage and Greatbatch 1991, pp. 95–6). Thus, the interactional patterning (or “fingerprint”) which is typical of, for example, a travel agent will be different from that of a classroom and different again from that of a doctor’s surgery. In each context, there are well-defined roles and expectations which, to some extent, determine what is said.

By examining specific features in the institutional interaction, an understanding can be gained of the ways in which context is both constructed and sustained; features which can be usefully examined include turn-taking organisation, turn design, sequence organisation, lexical choice and asymmetry of roles (Heritage 1997). The second language classroom is, of course, a clear example of an institutional setting with asymmetrical roles, goal-oriented activities and a context which is constantly being created for and by participants through the classroom interaction. While the discourse of L2 classrooms does not and should not be interpreted as having any resemblance to conversation, there are nonetheless good reasons for using a CA methodology (Edwards and Westgate 1994, p. 116):

The point is not that classroom talk ‘should’ resemble conversation, since most of the time for practical purposes it cannot, but that institutionalised talk [...] shows a heightened use of procedures which have their ‘base’ in ordinary conversation and are more clearly understood through comparison with it.

The relevance of a CA approach to the L2 classroom context is not difficult to perceive. CA attempts to account for the practices at work which enable participants in a conversation to make sense of the interaction and contribute to it. There are clear parallels: classroom talk is made up of many participants; it involves turn-taking, -ceding, -holding and -gaining; there have to be smooth transitions and clearly defined expectations if meanings are to be made explicit. Possibly the most significant role of CA is to **interpret** from the data rather than **impose** pre-determined categories.

One of the biggest influences on CA-led classroom-based research was the call of Firth and Wagner (1997) for greater sensitivity towards contextual and interactional aspects of language use by focusing more on the participants in SLA research and less on cognitive processes. Since the late 1990s, these studies have highlighted the ways in which learning and interactional competence can be approached and described through a micro-analytic mode of inquiry (see, for example, Hellermann 2008; Markee 2008). From this body of research has emerged the field now known as CA-SLA or CA-for-SLA: Conversation Analysis for Second Language Acquisition. By focusing on micro-details of video- or audio-recorded interaction, CA-for-SLA aims to document micro-moments of learning and understanding by drawing upon participants' own understanding of the ongoing interaction, from an emic perspective. This perspective is revealed through a detailed analysis of vocal (words and grammar, suprasegmentals, pace of talk, etc.) and non-vocal (silence, body language, embodiment of surrounding artefacts, etc.) resources within the sequential development of talk. CA-for-SLA studies have succeeded in demonstrating 'good' examples of 'interactional competence' and/or understanding of certain information by students by using interactionally and pedagogically fruitful instances of talk; for instance through the use of repair sequences (e.g. Hellermann 2009, 2011).

To summarise this necessarily brief overview of the use of CA for the study of classroom discourse, we can make a number of claims concerning its appropriateness. Firstly, under CA, there is no preconceived set of descriptive categories at the outset. The aim of CA is to account for the structural organisation of the interaction as determined by the participants. That is, there should be no attempt to 'fit' the data to preconceived categories; evidence that such categories exist and are utilized by the participants must be demonstrated by reference to and examples from the data. Thus, the approach is strictly empirical. Secondly, there is a recognition that the context is not static and fixed, but dynamic and variable. A dynamic perspective on context allows for variability; contexts are not fixed entities which operate across a lesson, but dynamic and changing processes which vary from one stage of a lesson to another (Cullen 1998). A CA methodology is better-equipped to take variations in linguistic and pedagogic purpose into account since one contribution is dependent on another. Third, the approach recognises that all spoken interactions are goal-oriented. Under institutional discourse, the behaviour and discourse of the participants are goal-oriented in that they are striving towards some overall objective related to the institution. In a language classroom, for example, the discourse is influenced by the fact that all participants are focusing on some pre-determined aim, learning a second language. Different participants, depending on their own agenda may have different individual objectives; nonetheless, the discourse which is jointly constructed is dependent on both the goals and the related expectations of the participants. Finally, CA offers a multi-layered perspective on classroom discourse. Because no one utterance is categorised in isolation and because contributions are examined in sequence, a CA methodology is much better-equipped to interpret and account for the multi-layered structure of classroom interaction.

## 4 A CLCA Methodology

In light of the different research traditions of CL and CA outlined in the preceding sections, the reader might be forgiven for coming to the conclusion that the two methodologies are incompatible and that there is little point in pursuing the enterprise of CLCA. In this section, therefore, I present a practical example to demonstrate how this methodology was utilised in a recent study (see Walsh et al. 2011). The study reported here took place in a higher education, small group teaching (henceforth SGT) context, where seminars and tutorials are used to support larger lectures. These sessions are important in that they are designed to allow tutors and students to engage in debate and discussion. They account for up to 40 % of the time of undergraduate students and up to as much as 75 % of the time of postgraduate students (Bennett et al. 2002). The 2010 study used a corpus of 500,000 words taken from two universities in Ireland, one in the north, the other in the south.

Previous CL studies on spoken interaction in higher education have arisen principally from 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).

From a CA perspective, 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 and Stokoe 2002). In most of these studies, 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 and 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, the aim of Walsh et al.’s 2010 study was to demonstrate how tutors and students engage in tightly organised and intricate negotiations of a set of pedagogic agendas, using both interactional and linguistic resources to achieve their goals.

A CLCA methodology essentially entails looking at the same data-set through two different lenses: one CL, the other CA. Thus, the same text is subjected to two



treatments, each offering a unique but complementary perspective on the data. A useful starting point is to use CL in the first layer of analysis as a means of scoping out and quantifying recurring linguistic features. This analysis enables the identification of recurring patterns, each specific to the context. 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 2010 study, there were interesting findings around the frequency and use of certain discourse markers, which clustered around specific contexts. This led 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 adopted an iterative approach to analysis, from CL to CA, back to CL and so on. Key to this is the interdependence between the two modes of analysis, which was non-linear in that, for example, CL tools were sometimes used within the CA layer of analysis to quantify CA insights.

Using *WordSmith Tools* (Scott 2008) key words and word frequencies were identified for both single words and multi-word units (henceforth, MWU), units of two or more words sometimes referred to as *lexical bundles*, *lexical phrases*, *clusters*, *chunks*, though with slightly varying definitions (see Greaves and Warren 2010). 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’ *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...*

Other features which were identified through concordance line analysis include 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* is related 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 pointed towards 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 key words, frequency counts, concordance searches. When the analysis was extended to patterns (2–6-word MWUs), concordance searches produced a total of 128 items which were salient to the SGT context. These items were then categorised according to their approximate functions in the discourse. The analysis, at this stage, was moving towards looking at longer stretches of discourse at the level of turn and longer sequences. At this point, the main focus switched to CA.

An initial CA analysis showed that the 128 items identified in the corpus as being salient played an important role as resources for participants’ courses of action or ‘interactional projects’. Schegloff (2007) describes interactional projects as a form

of interactional organisation 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.” (p. 244). 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, p. 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, four such speech exchange systems were identified, each with distinguishing interactional features and clear pedagogic goals (cf. Walsh 2006):

- (a) Procedural talk, with a focus on organising learning and comprising long tutor turns and correspondingly little participation by students. Specific MWUs such as ‘what I want you to do is’ were also found in high frequency.
- (b) Didactic talk, with a focus on eliciting information or giving feedback. The MWU *tell me* is prominent in this micro-context, while turn-taking is controlled tightly by the tutor. Display questions prevail and the three-part exchange structure IRF dominates. (Tutor Initiates, student Responds, tutor gives Feedback)
- (c) Empathic talk. Here, students have more space and manage the floor, producing ‘tellings’ or accounts of personal experiences. There is more equality in turn-taking and roles are more symmetrical. Discourse markers play a key part in this micro-context, especially *you know* and *you see* which function to create ‘shared space’ for learning.
- (d) Argumentational talk. This micro-context was found to occur when there was shared space, but the discussion was more combative, with a focus on agreeing and disagreeing. Words like *but* and *maybe* were used frequently to show disagreement or indicate stance.

## 5 Discussion

This aim of this chapter was to demonstrate the appropriateness of CL and CA in providing enhanced descriptions of spoken interactions in higher education small group settings. Four speech exchange systems (micro-contexts) were identified in the data, each with distinctive interactional, linguistic and pedagogic features or ‘fingerprints’ (Drew and Heritage 1992, p. 26). The four speech exchange systems 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. Using a CLCA methodology, I suggest, allows useful comparisons to be made both across and within these micro-contexts. For example, a comparison of didactic and empathic talk reveals very different profiles or 'fingerprints'. 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.

When the CL analysis is related more closely to the CA findings, the single words and MWUs identified as being salient are found across all micro-contexts; more importantly, they are found to do different interactional work in relation to the particular agenda of the moment. Indeed, it is striking that the participants in this study used single words and MWUs to carry 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 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.

It seems evident from the study presented here that there is much to be gained from using a combined CLCA methodology. First, the methodology allows two (at least) perspectives on the same dataset: one (using CL) offering an overview of the data and a profile of the most important recurring linguistic features in specific contexts of use; the other (using CA) offering a fine-grained, up-close view of the same data and highlighting the ways in which meanings are co-constructed. This dual perspective on the same dataset, I would suggest, facilitates a closer understanding of what linguistic and interactional resources are used to create meaning. Specifically, there is an opportunity for the analyst to examine in some detail the ways in which linguistic, interactional and textual features combine in any communicative encounter. Second, the methodology allows enhanced understandings of specific features of spoken discourse in a particular context. Arguably, it allows the analyst to focus more on language *use* (what we do with language) and less on language *usage* (what language is); the issue of what language does rather than what language is has been taxing applied linguists for many years (ref). Third, this methodology goes some way at least in compensating for the deficiencies of each method when used alone: CA, which is unable to extend its findings beyond the relatively small sample of data it typically utilises; CL which is only able to make general observations on the data, without offering the kind of interactional detail which CA provides. A CLCA methodology compensates for all these deficiencies and allows analysts to provide both greater depth and coverage in their findings.

There are, naturally, also some shortcomings to this methodology. The first is that there is a presupposition in the arguments put forward here that researchers are able to use both CL and CA. That is rarely the case since the two research traditions are, by definition, mutually exclusive. It would be unusual, but not unheard of, for a conversation analyst to use a CL methodology and the same is true in reverse. One way round this is for conversation analysts to work with corpus linguists in a spirit of shared expertise (cf. Walsh et al. 2010). A second shortcoming is that the methodology, while following an iterative process, is somewhat imprecise in terms of which steps should be taken and when. Should one, for example, commence with CL and then do CA, or vice versa? What precise steps should be taken once the first analysis has been completed and in what sequence? There are no exact answers to these issues; I would only say that with a little trial and error, it is possible to make effective use of the two methodologies.

## 6 Conclusion

This chapter set out with the proposition that CL and CA can be usefully combined in the analysis of spoken data. I have suggested how, in spite of their ontological and epistemological differences, these two research methodologies can be combined and offer a surprisingly rich and comprehensive perspective on a corpus. This combined CLCA approach has the potential to provide far more detailed analysis than that offered when each is used in isolation. In the study reported here, for example, detailed descriptions of the same corpus of academic spoken English were given from at least three perspectives: linguistic (portraying the use of high frequency items, key words, MWUs, discourse markers, question forms and so on), interactional (focusing on turn-taking and turn design, sequential organization) and pedagogic (looking at specific pedagogic functions at a given moment to include eliciting, explaining, instructing and so on). Arguably, a CLCA approach allows for a much more detailed description of a particular context (for example, small group teaching in higher education), offering insights into the ways in which language is used to mean, convey information and establish joint understandings. The approach, above all, underlines the centrality of joint enterprise in any spoken encounter: people establish understandings together and share equal responsibility for that goal in most cases.

While each methodology has its own merits, it also has significant shortcomings as outlined above. CL on its own, for example, may provide interesting lists of high frequency items which can then be explained functionally, but its perspective is a surface level one; a CA perspective, on the other hand, enables us to identify particular exchanges and sequence organisations, but misses the fact that particular linguistic features may occur in each exchange structure. Essentially, there is much to commend this combined methodology and the future is likely to

show further evidence of the power and potential of the two methodologies. Future research is likely to result in a narrowing of the perceived gap which currently exists between each approach: for example, there have already been moves to look more quantitatively at turn openings and closings using CL (refs), while there has been a corresponding prediction that CA will become more quantitative in the future (ref). By looking more at specific interactional features (such as discourse markers), it is not inconceivable that CL will begin to offer turn-level analyses which have relevance for CA. In short, we can predict that a combined CLCA methodology is here to stay and that we'll be witnessing a growth in its adoption in coming years.

## References

- Ädel, A. 2010. How to use corpus linguistics in the study of political discourse. In *The Routledge handbook of corpus linguistics*, ed. A. O'Keeffe and M.J. McCarthy, 591–604. London: Routledge.
- Basturkmen, H. 2002. Negotiating meaning in seminar-type discussions and EAP. *English for Specific Purposes* 21(1): 233–242.
- Bennett, C., C. Howe, and E. Truswell. 2002. *Small group teaching and learning in psychology*. York: LTSN Psychology University of York.
- Benwell, B.M., and E.H. Stokoe. 2002. Constructing discussion tasks in university tutorials: Shifting dynamics and identities. *Discourse Studies* 4(4): 429–453.
- Carter, R., and M.J. McCarthy. 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 *The Routledge handbook of corpus linguistics*, ed. A. O'Keeffe and M.J. McCarthy, 578–590. London: Routledge.
- Cullen, R. 1998. Teacher talk and the classroom context. *English Language Teaching Journal* 52(3): 179–187.
- Drew, P., and J. Heritage. 1992. Analyzing talk at work: An introduction. In *Talk at work: Interaction in institutional settings*, ed. P. Drew and J. Heritage, 3–65. Cambridge: Cambridge University Press.
- Edwards, A., and D. Westgate. 1994. *Investigating classroom talk*. London: Falmer.
- Farr, F., B. Murphy, and A. O'Keeffe. 2004. The Limerick corpus of Irish English: Design, description and application. *Teanga* 21: 5–29.
- Firth, A., and J. Wagner. 1997. On discourse, communication, and (some) fundamental concepts in SLA research. *The Modern Language Journal* 81: 285–300.
- Gibson, W., A. Hall, and P. Callery. 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., and M. Warren. 2010. What can a corpus tell us about multi-word units? In *The Routledge handbook of corpus linguistic*, ed. A. O'Keeffe and M.J. McCarthy, 212–226. London: Routledge.
- Hellermann, J. 2008. *Social actions for classroom language learning*. Clevedon: Multilingual Matters.

- Hellermann, J. 2009. Looking for evidence of language learning in practices for repair: A case study of self-initiated self-repair by an adult learner of English. *Scandinavian Journal of Educational Research* 53(2): 113–132.
- Hellermann, J. 2011. 'Members' methods, members' competencies: Looking for evidence of language learning in longitudinal investigations of other-initiated repair'. In *L2 Interactional competence and development*, ed. J.K. Hall, J. Hellermann, and S. Pekarek Doehler, 147–172. Bristol: Multilingual Matters.
- Heritage, J. 1984. A change-of-state token and aspects of its sequential placement. In *Structures of social action*, ed. J.M. Atkinson and J. Heritage, 299–345. Cambridge: Cambridge University Press.
- Heritage, J. 1997. Conversational analysis and institutional talk: Analysing data. In *Qualitative research: Theory, method and practice*, ed. D. Silverman. London: Sage Publications.
- Heritage, J., and D. Greatbatch. 1991. On the institutional character of institutional talk: The case of news interviews. In *Talk and social structure: Studies in ethnomethodology and conversation analysis*, ed. D. Boden and D.H. Zimmerman. Berkeley: University of California Press.
- Hutchby, I., and R. Wooffitt. 2008. *Conversation analysis*, 2nd ed. Cambridge: Polity Press.
- Lee, J. 2009. Size matters: An exploratory comparison of small- and large-class university lecture introductions. *English for Specific Purposes* 28(1): 42–57.
- Lorés, R. 2006. The referential function of metadiscourse: thing(s) and idea(s) in academic lectures. In *Corpus linguistics: Applications for the study of English*, ed. A. Hornero, M. Luzón, and S. Murillo, 315–334. Bern: Peter Lang.
- Louwerse, M., S. Crossley, and P. Jeuniaux. 2008. What if? Conditionals in educational registers. *Linguistics and Education* 19(1): 56–69.
- Markee, N. 2008. Toward a learning behavior tracking methodology for CA-for-SLA. *Applied Linguistics* 29: 404–427.
- McCarthy, M., and A. O'Keeffe. 2010. Historical perspective: What are corpora and how have they evolved? In *The Routledge handbook of corpus linguistics*, ed. A. O'Keeffe and M.J. McCarthy, 3–13. London: Routledge.
- 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.
- O'Keeffe, A., M. McCarthy, and R. Carter. 2007. *From Corpus to classroom*. Cambridge: Cambridge University Press.
- Sacks, H., E.A. Schegloff, and G. Jefferson. 1974. A simplest systematics for the organization of turn-taking for conversation. *Language* 50(4): 696–735.
- Schegloff, E.A. 2007. *Sequence organization in interaction: A primer in conversation analysis*, vol. 1. Cambridge: Cambridge University Press.
- Scott, M. 2008. *WordSmith tools (version 5)*. Liverpool: Lexical Analysis Software.
- Seedhouse, P. 2004. *The interactional architecture of the language classroom: A conversation analysis perspective*. Oxford: Blackwell.
- Seedhouse, P. 2005. Conversation analysis and language learning. *Language Teaching* 38(4): 165–187.
- Sidnell, J. 2010. *Conversation analysis- an introduction*. West Sussex: Wiley-Blackwell.
- Simpson, R.C., S.L. Briggs, J. Ovens, and J.M. Swales. 2002. *The Michigan Corpus of Academic Spoken English*. Ann Arbor: The Regents of the University of Michigan.
- Stokoe, E.H. 2000. Constructing topicality in university students' small-group discussion: A conversation analytic approach. *Language and Education* 14(3): 184–203.
- 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 A. O’Keeffe. 2007. Applying CA to a modes analysis of third-level spoken academic discourse. In *Conversation analysis and languages for specific purposes*, ed. H. Bowles and P. Seedhouse. Bern: Peter Lang.
- Walsh, S., T. Morton, and A. O’Keeffe. 2011. Space for learning: Language use, interaction and orientation to knowledge in small group teaching in higher education. *International Journal of Corpus Linguistics*.
- Yang, S. 2013. Unpublished Ph.D. thesis, University of Newcastle.