



RiUPTC

Repositorio Institucional
UPTC

repositorio.uptc@uptc.edu.co

Analysing university spoken interaction: a CL/CA approach

Steve Walsh

Professor of Applied Linguistics

Newcastle University, UK

Reino Unido

Introduction.

In this paper, I consider how corpus linguistics (CL) and conversation analysis (CA) can be used together to provide enhanced understandings of spoken interaction in the context of small group teaching in higher education. From an analysis of the data, I 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, CA is then used to highlight pertinent interactional features. The 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. The aim is to consider how corpus linguistics (CL) and conversation analysis (CA) can be used together to generate fine-grained 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, CL is regarded 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

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. Increasingly, more and more modestly-sized specialised corpora emerge and we see more 'inter-marrying' between CL and other approaches to the analysis of discourse in context. As McCarthy and O'Keeffe (2010, pp 3-13) 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).

In this study, the first layer of analysis (using CL) was designed to scope out and quantify recurring linguistic features, or 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; the 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.

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, Howe and Truswell 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, Briggs, Ovens, Swales 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, pp. 315-334), the use of conditionals (Louwerse, Crossley, and Jeuniaux 2008, pp. 56-69), and, of more direct relevance to this study, the effect of class size on lecture discourse (Lee, 2009, pp. 42-57).

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, pp. 103-131), sequential organisation and negotiation of meaning (Basturkmen 2002, pp. 233-242), the issue of ‘topicality’ in small group discussion (Stokoe 2000, pp. 184-203; Gibson, Hall and Callery 2006, pp. 77–94), and the formulation and uptake of tasks and resistance to ‘academic’ identities (Benwell and 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 and Stokoe 2002, pp. 429-453). 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, pp. 85-93) and specific linguistic features, such as discourse markers, to achieve their goals.

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 is 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.

WordSmith Tools (Scott 2008) was used to identify key words and word frequencies for both single words and multi-word units (henceforth, MWU: units of two or more words). The one-million word Limerick Corpus of Irish English (LCIE) was used as a reference corpus for our key word analysis. This contains over one million words of everyday conversation (LCIE, see Farr et al 2004, pp. 5-29). 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:

Insert Table 1 Here.

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

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

¹Ye 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*.

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 key words, 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. At this stage, the CL analysis has produced word frequency lists, key word lists, concordances and a list of MWUs. These patterns were then grouped according to their key functions.

In order to gain a deeper understanding of spoken interaction in this context, CA was then used to see how salient features actually operated in speakers' turns and in longer sequences of interaction.

CLCA analysis.

In considering the ways in which single words and multi-word items were used over phases of interaction, I found that they played an important role as resources for participants' courses of action or 'interactional projects'. 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 the data, four exchange structures were identified, each with distinct interactional features and pedagogic goals:

Procedural talk, where the aim is to organize and manage learning.

Didactic talk, with a focus on eliciting information or giving feedback.

Empathic talk, where opinions and feelings are discussed.

Argumentational talk, where ideas are discussed and debated.

Conclusion.

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, four speech-exchange systems were identified, each with distinctive interactional, linguistic and pedagogic features or ‘fingerprints’ (Drew and Heritage 1992, p. 26). The analysis enables comparisons to be made 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.

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 I know, offer the same level of detail as the present study. From the CLCA analysis, it is possible to provide detailed descriptions of the interaction from 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, etc) and pedagogic (looking at specific pedagogic functions at a given moment to include eliciting, explaining, instructing and so on).

Had CL been used on its own, it would have been possible to produce lists of high frequency items which might have been explained functionally; this analysis 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, key words, high frequency words and multi-word units). In addition, by drawing on quantitative methods within CL, we were able to reference our findings against another dataset (in this case LCIE). We can therefore safely assert that CL and CA are ‘well met’.

References.

- Basturkmen, H. (2002). *Negotiating meaning in seminar-type discussions and EAP*. In: English for Specific Purposes, 21(1).
- 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*. In: Discourse Studies, 4(4).
- 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.
- Farr, F., Murphy, B. & O’Keeffe, A. (2004). *The Limerick Corpus of Irish English: design, description and application*. In: *Teanga*, 21.
- 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*. In: *British Educational Research Journal*, 32(1).
- Lee, J. (2009). *Size matters: an exploratory comparison of small- and large-class university lecture introductions*. In: *English for Specific Purposes*, 28(1).
- Levinson, S. (2006). *Cognition at the heart of human interaction*. In: *Discourse Studies*, 8(1).
- 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.
- Louwerse, M., Crossley, S. & Jeuniaux, P. (2008). *What if? Conditionals in educational registers*. In: *Linguistics and Education*, 19(1).
- 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.

- 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.
- Stokoe, E. H. (2000). *Constructing topicality in university students' small-group discussion: a conversation analytic approach*. In: *Language & Education*, 14(3).
- Viechnicki, G. B. (1997). *An empirical analysis of participant intentions: discourse in a graduate seminar*. In: *Language and Communication*, 17(2).
- Walsh, S. (2006). *Investigating Classroom Discourse*. London: Routledge.