
11 Exploring teacher corpora

11.1 Introduction

This chapter is very different from all of the other chapters in this book from a number of perspectives. Up to this point, we have focused on what corpora can teach us about language in use and what, in turn, this tells us about language teaching. Here we are not looking at what we can learn about language use from a corpus, rather we are looking at what corpora can tell us about our own teaching and ourselves as part of a professional cohort. For example, we draw on corpora of classroom interactions and compare them with other question-driven institutionalised contexts, such as media interviews, to show what makes classroom interactions different. We also look at the specifics of teacher talk, for example we survey studies of teacher questioning strategies and wait-time (after questions have been asked) based on corpus data collected in the language classroom. The overall aim of this chapter is to make a case for the development of corpora and corpus skills as a tool for reflective practice within pre-service teacher education and ongoing in-career development.

Another reason why this chapter differs so much from other chapters is because here we do not see a teacher corpus as something which is 'off-the-shelf'. A teacher corpus is something small and evolving over time. In this chapter we look at very small amounts of data very closely, usually turn by turn. A corpus of teacher interactions is seen as developmental in that, like a portfolio, it grows over a teacher's career and also in the sense that it becomes a tool for development itself. By building up classroom extracts, a teacher can reflect closely on classroom practice. We are also interested here in looking beyond classroom practice. Though the classroom is the primary site for teacher interaction, there are other aspects of a teacher's working life which merit attention and understanding. These areas are steadily acquiring attention; for example, interactions outside of the classroom with colleagues in meetings, one-to-one teacher education feedback sessions or within professional development sessions. We will also look at a project in Hong Kong where a corpus resource service has been set up for teachers.

Looking at the language of a corpus does not necessarily always mean looking at other people's language. As we have argued, corpora can also be used by teachers as tools for reflective practice and professional development. In a practical sense this means that small corpora are created by teachers and analysed so as to reflect on, better understand and enhance their own professional practice. In the case of classroom practice, transcripts from classroom interactions can facilitate close inspection and build up sensitivity to the

language that we use so as to hone our judgements about what we say in the classroom. As Walsh (2006) notes, in a classroom context, where so much is happening at once, fine judgements can be difficult to make, and deciding to intervene or withdraw in the moment-by-moment construction of classroom interaction requires great sensitivity and awareness on the part of the teacher. Inevitably, teachers do not 'get it right' every time.

The overall aim of this chapter is to illustrate the growing application of corpora in teacher development and to provide frameworks within which teacher corpora can be used in different contexts. Looking at the language of the classroom is nothing new and many authors provide models for doing this (for example, Sinclair and Coulthard 1975; McCarthy 1991; Hatch 1992; McCarthy and Carter 1994; Johnson 1995; Riggensbach 1999; Celce-Murcia and Olshtain 2000; Hall and Verplaetse 2000; Hall and Walsh 2002; Mori 2002, 2004; Boxer and Cohen 2004; Kasper 2004; Markee 2004; Mondada and Pekarek Doehler 2004; Seedhouse 2004; Walsh 2006). Teacher educators will already be aware of commercially available video material which provides lessons for training and reflection in pedagogic practices. Here we are not arguing that these materials should be replaced by home-produced classroom corpora but we suggest that in-house teacher corpora can offer a valuable supplement to published training materials, especially in the area of methodological skills acquisition, because the practices of teaching must be interpreted within their contexts of realisation. In other words, socio-cultural and environmental factors which create and cast the lesson cannot easily be captured in their entirety by non-present third-party trainees in different educational and/or cultural surroundings. This is particularly true when the backgrounds, training conditions, and experience of trainees on teacher education programmes are socio-culturally at odds with that of the training materials available commercially. For instance, most teacher education videos are either British- or American-produced.

Another advantage of building and using a teacher corpus is that the transcript can then become a supplement to the video medium itself, or extracts from it can be examined as part of task-based activities on handouts. While a video clip could equally be used for this purpose, it is a far more ephemeral medium than the written transcript and does not allow for the same level of turn-by-turn analysis. For example, figure 1, overleaf, shows an example transcribed from a video clip, taken from O'Keeffe and Farr (2003) which, if played on video, involves less than 25 seconds of speech. However, when it is viewed as a transcript, it is frozen for turn-by-turn analysis.

With the advent of digital recording facilities, it is also possible to design such materials for teacher education whereby the audiovisual clip can be aligned with the transcript.

Figure 1: Sample material for awareness-raising in relation to teaching new vocabulary (O’Keeffe and Farr 2003: 401)

Student:	What’s the difference between ‘collaborate’ and ‘cooperate’?
Trainee:	Well ‘collaborate’ is generally used for something which is negative and ‘cooperate’ is more positive.
Student:	So can I say ‘I am cooperating with Maria on this project’? Collaborate would be wrong here?
Trainee:	Well yes, no, mm I’m not too sure. What does the dictionary say? Let’s check.

- a) Use a dictionary to find the differences in meaning between these two words.
- b) Use any large corpus from the electronic library to establish how these near-synonyms differ in terms of use and lexical patterns.
- c) Redesign the part of the lesson in the extract above to make it more effective.

11.2 Classroom discourse

Once a classroom corpus is created (see chapter 1 on building your own small corpus), the next step is to build up strategies and frameworks for its use. For the most part, classroom corpora will be used qualitatively; that is, extracts will be read and analysed manually. While applications such as concordances and word frequency list software will be used to search for certain words, phrases or discourse patterns, turn-by-turn analysis will be the main focus. Therefore, the corpus in this context is a large electronic resource that can be searched automatically to find extracts to suit one’s pedagogical goal in a teacher education and professional development context, and it may be used very effectively as a supplement to existing video resources, as we noted above.

McCarthy and Walsh (2003) note that, for language teachers, understanding the discourse of the classroom itself is crucial. We teach *through* discourse with our learners; language teaching is unique in that language is both the medium and the content of teaching. In many parts of the world, the main exposure to discourse in the target language that learners will have is in the classroom itself, via the teacher. A number of studies have compared the discourse of the classroom with ‘real’ communication (e.g. Nunan 1987). But, as van Lier tells us (1988: 267), ‘the classroom is part of the real world, just as much as the airport, the interviewing room, the chemical laboratory, the beach and so on.’ A teacher corpus is therefore a resource of real-world interactions from the classroom and other sites of teacher interaction, and this database needs to be interpreted within a framework which will help us best understand the structure of the discourse that we find within it (see below).

11.3 Frameworks for the analysis of classroom language

We feel that there is no point in collecting classroom data without having an awareness of the main analytical models within which these data can be interpreted and understood.

We now survey three models, none of which is directly corpus-related but all of which offer powerful models for analysing classroom corpus data: Discourse Analysis (DA), particularly its concept of 'exchange structure', Conversation Analysis (CA) and Socio-cultural Theory (SCT). Data could be analysed using any one or even none of these models. However, we hope to show that by applying these models to actual data a triangulation of the three perspectives can offer a very rich insight for teachers. As we present each of these perspectives, we will also provide illustrations of the type of insights that they have brought to our understanding of language teaching and classroom discourse. Generally these are not corpus related, but they give a sense of how these models can be applied in a general sense.

Exchange structure

This approach to discourse analysis stems from a highly influential study by Sinclair and Coulthard (1975). Based on the analysis of recorded classroom interactions, Sinclair and Coulthard produced a model for understanding classroom discourse, which has subsequently been applied to the study of other contexts, for example doctor-patient interactions (Coulthard and Ashby 1975). In their analysis, Sinclair and Coulthard found that teachers divided their lessons into different phases of activity (called 'transactions'). Discourse markers (see chapter 8 for a detailed treatment) typically marked the beginnings and ends of transactions, along with intonational cues. These marking devices are termed 'frames' and are generally limited to items such as *okay, well, right, now, good*, uttered with strong stress, high falling intonation and followed by a short pause. It was noted that teachers frequently followed a frame (indicating the beginning of a transaction) with a 'focus', that is, a metastatement about the upcoming transaction. Here is an example from an EFL class where the teacher is setting up a task. The discourse markers *right, alright* and *okay* operate as frames and are followed by a focus, which functions as a signalling statement:

(11.1)

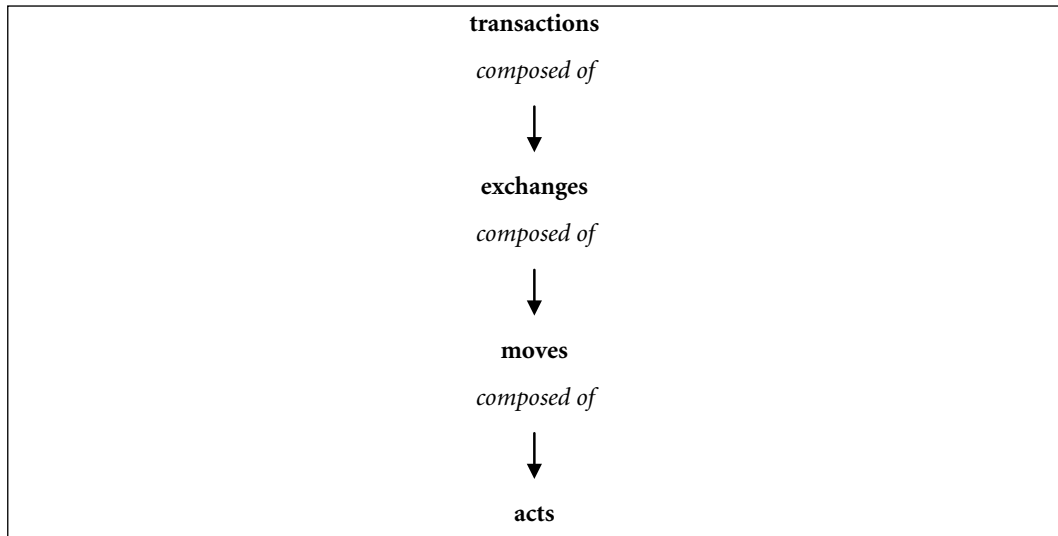
Teacher: **Right** so what I'm going to do is I'm going to give you amm a thing. **Right?** I'm going to give you the thing an object **alright?** And I want you to decide what it is cos it may not be a hundred percent clear when you see the object what it is. **Alright?** You have to decide what it is. You decide what the selling points are and then we have to present it.

(LIBEL)

Sinclair and Coulthard's (1975) model for the structure of a lesson involves a hierarchy consisting of levels, each composed of elements from the level below it (figure 2).

At the level of 'exchange', Sinclair and Coulthard observed the following as characterising classroom interactions:

- (1) question-and-answer sequences
- (2) pupils responding to teachers' directions
- (3) pupils listening to the teacher giving information

Figure 2: Levels of Sinclair and Coulthard's (1975) hierarchical structure of a lesson

The question-and-answer sequence receives most attention. As a sequence, it consists of a minimum of three elements (often referred to as IRF):

- (1) the question (or *Initiation*)
- (2) the answer (or *Response*)
- (3) the teacher's feedback (or *Follow-up*)

Here is an example from Sinclair and Coulthard (1975):

Teacher: . . . What else will cut the piece of wood?	<i>Initiation</i> (I)
Student: Saw.	<i>Response</i> (R)
Teacher: The saw yes.	<i>Follow-up</i> (F)

Note, in this example from Walsh (2001), the use of the discourse marker *so* whereby the teacher marks the new phase of activity. Here we see that the IRF sequence is repeated:

Teacher: So, can you read question two, Junya.	(I)
Junya: [Reading from book] Where was Sabina when this happened?	(R)
Teacher: Right, yes, where was Sabina?	(F)
In Unit 10, where was she?	(I)
Junya: Er, go out . . .	(R)
Teacher: She went out, yes.	(F)

Typically the teacher's follow-up evaluates the learner's answer (*right, yes*); such feedback is important to the learner. This is one of the distinguishing features of classroom discourse. Coulthard (1977) notes that the three-part exchange structure was suggested as the norm for classroom discourse for two reasons: firstly, answers directed at the teacher can be difficult for others to hear and so need repetition. Secondly, and more importantly, a distinguishing feature of classroom discourse is that the questions which a teacher asks are ones to which she already knows the answer (referred to as 'display questions', see below).

Often answers which are correct in terms of the question are not the ones the teacher is seeking and therefore it is essential for him/her to provide feedback indicating whether a particular answer is the one (s)he is looking for. For example:

- Teacher: What does the food give you? (I)
 Student: Strength (R)
 Teacher: Not only strength we have another word for it. (F)
 Student: Energy (R)
 Teacher: Good girl, energy, yes. (F)

(adapted from Coulthard 1977: 125)

IRF exchanges are also found in everyday conversation, but the follow-up element is not normally evaluative, for example:

(11.2)

- S1: What's the last day of the month? (I)
 S2: Friday. (R)
 S1: Friday. (F)
 We'll invoice you on Friday. (I)
 S2: That would be brilliant. (R)
 S1: And fax it over to you. (I)
 S2: Er, well I'll come and get it. (R)
 S1: Okay. (F)

(CANCODE. See also McCarthy and Walsh 2003: 176)

Very often in casual conversation, the response to an initiation involves tokens such as *great*, *brilliant*, *excellent*, *sure*. As we have discussed in chapter 7, these have a relational rather than an evaluative function, for example to show interest, surprise, shock and so on. For example, here they mark agreement between friends:

(11.3)

- S1: . . . it just goes to show you can't take people at face value.
 S2: No.
 S1: And you don't know what's going on either.
 S2: **Exactly.**

(LCIE)

The powerful nature of the three-part exchange as a classroom structure is illustrated by Coulthard (1977: 125) in this next example, where he notes that the absence of the feedback move signals to the student that the answer is wrong.

- Teacher: Can you think why I changed 'mat' to 'rug'? (I)
 Student: Mat's got two vowels in it. (R)
 Teacher: (F)
 Teacher: Which are they? What are they? (I)
 Student: 'a' and 't' (R)

Teacher:	(F)
Teacher: Is 't' a vowel?	(I)
Student: No.	(R)
Teacher: No.	(F)

(Coulthard 1977: 125)

However, the IRF routine in classroom interaction has been seen by many as unproductive as an interactional format, especially as a model for spoken interaction outside of the classroom. The argument put forward is that the IRF exchange is a poor model for learning pragmatics and discourse norms of the target language since it differs from everyday interaction (as the above examples show). IRF exchanges, it is argued, fail to give opportunities for tackling the complex demands of everyday conversation, especially since teachers usually exercise the follow-up role, while learners often remain in passive, respondent roles. Ohta (2001), for example, finds that the overwhelming majority of classroom follow-up moves are spoken by the teacher; learners get few opportunities to use typical listener follow-ups and only experience the teacher's moves as peripheral participants. Peer-to-peer interaction, Ohta argues, can provide the best opportunities for learners to produce appropriate listener responses (this ties in with the joint-production model of *confluence* that we discuss in chapter 7).

Walsh (2002), in his analysis of different modes of teacher talk, illustrates how these may hinder or optimise learner contributions. Kasper (2001), however, argues that the negative reputation of the IRF exchange may not be entirely warranted and that what really matters is the kind of interactional status assigned by the teacher to individual learners. Teachers can help their learners become actively involved in interaction, even within the typical IRF pattern, she argues. Exposure to the teacher's use of follow-up moves, along with explicit guidance on the use of responsive moves, can help students gradually move towards more productive use in peer-to-peer speaking activities.

Conversation analysis (CA)

CA gives us a framework for looking at 'local' aspects of interaction in detail, especially how participants in a conversation work hard to make it successful (see Pomerantz and Fehr 1997). CA focuses on how speakers decide when to speak during conversation, i.e. the rules governing 'turn-taking' (see Sacks, Schegloff and Jefferson 1974), and how they show they are listening (by using response tokens such as *umhm*, *yeah*, *right*, see chapter 7). It also deals with how speaker turns can be related to each other in sequence and might be said to go together as 'adjacency pairs', for example, complain + denial, greeting + greetings, or, as in Figure 3, yes/no question + yes/no answer:

Figure 3: Concordance line examples of adjacency pairs from CANCODE

1	Did you know that?	<\$2>	No I didn't.
2	Did you find them?	<\$3>	No I didn't.
3	Did you knock?	<\$1>	No I didn't.
4	Did you see that one?	<\$1>	No I didn't.

Or in this example, from CANCODE:

(11.4)

[Speaker 2 has been relating how she was stung by a wasp while asleep]

S1: **Well perhaps it was nosing around minding its own business and you frightened it.**

S2: **Oh I see. It's my fault is it!**

S1: Well.

S2: He can never see my side.

S3: [laughs]

S1: Wasps don't sting unless threatened.

(CANCODE)

Not all second pairs have the same significance; therefore, there is said to be 'preference organisation', whereby some second-pair-parts are preferred and some are dispreferred (see Pomerantz 1984). When the two pair-parts do not fit, speakers have to work hard to repair potential problems, for example an invitation anticipates acceptance rather than rejection or hesitation. Compare the following:

S1: Would you like a cup of tea Ursula? S2: Ooh I'd love one

(preferred response)

versus

S2: (pause) You know I just don't know

(invented dispreferred response).

(CANCODE)

Another important focus of CA is how turns are organised in their local *sequential* context at any given point in an interaction and the systematicity of these sequences of utterances (see Schegloff 1982). For example, one can talk about the sequentiality of greeting or leave-taking routines in different situations (as discussed in Chapter 8). CA also places great importance on how seemingly minor changes in *placement* within utterances and across turns are organised and meaningful, for example, the difference between whether a vocative is placed at the beginning, mid or end point of an utterance (see Jefferson 1973). Other concerns of CA include openings and closings of conversations (Schegloff and Sacks, 1973), and topic management (i.e. how speakers launch new topics, change the subject, decide what to talk about, etc.; see Gardner, 1987).

McCarthy and Walsh (2003) note that CA has brought a number of key insights for language teaching, including how teachers and learners have to deal with the special turn-taking circumstances of the classroom (only teachers normally select the next speaker, it is difficult to interrupt the teacher, teachers often do not wait long enough for students to answer, etc.). Pedagogically, CA insights suggest that some adjacency pairs will be easy to learn (e.g. the ritualised ones like greeting–greeting, offer–accept), but that dispreferred sequences will require skill and practice (see Dörnyei and Thurrell 1994). There has been growing support for CA as a means of understanding and improving speaking in pedagogical contexts in recent years (see Boxer and Cohen 2004). Mori (2002) uses CA to analyse a speaking activity in a class of non-native-speaking learners of Japanese, where students

exchanged experiences and opinions with Japanese native speakers invited to the class. The resulting interaction resembled an interview, with a succession of questions by the students and answers from the native-speaker guests. Interestingly, more natural discussion came about when students made spontaneous utterances and when they seemed to be attending more to the moment-by-moment unfolding of the talk.

Wong (2000) notes that CA illuminates how local choices unfold in interaction and can focus on aspects of talk which are relevant for the participants themselves. A number of important studies into second language acquisition have been undertaken using CA (Hall and Verplaetse 2000; Markee 2000, 2004; Mori 2002, 2004; Hall and Walsh 2002; Lazaraton 2002; Seedhouse 2004; Kasper 2004; Mondada and Pekarek Doehler 2004, among others). Ducharme and Bernard (2001) look at learners of French, using micro-analyses of videotaped interactions and retrospective interviews to gain insights into the perspectives of participants. Mondada and Pekarek Doehler (2004) also look at the French second language classroom, providing an empirically based perspective on the contribution of CA and socio-cultural theory (see below) to our understanding of learners' second language practices. Mori (2004) focuses on a peer interactive task in a Japanese as a foreign language classroom. Through close observation of vocal and non-vocal conduct, Mori demonstrates how the students transform, moment by moment, their converging or diverging orientations towards varying types of learning and learning opportunities. Kasper (2004) examines a dyadic learning context in a German class between a native speaker and a beginning learner. Weiyun He (2004) appraises the 'uses and non-uses' of CA in the context of Chinese language learning. While she sees numerous applications of CA to teaching and research, such as in oral language assessment, she concedes that CA does not address introspective matters that may be important to language learning, and it is not designed to document learning longitudinally. Also pointing to the shortcomings of CA, Rampton et al. (2002) warn of the lack of a 'learning' dimension. Because CA is a very local kind of analysis, they argue, it lends itself less easily to providing evidence of actual development of language ability over time.

Sociocultural theory (SCT)

Sociocultural theories of learning focus on the social nature of the classroom interaction. Learners collectively construct their own knowledge and understanding by making connections, building mental schemata and concepts through collaborative meaning-making (Walsh 2006). Within this view, learners are seen as interacting with the 'expert' adult teacher 'in a context of social interactions leading to understanding' (Röhler and Cantlon 1996: 2). This notion has its origins in the work of Vygotsky (1962, 1978), a Russian psychologist who developed the sociocultural theory of mind. Lantolf and Appel (1994b), Lantolf (2000) and Lantolf and Thorne (2006) have been very influential in applying Vygotskian theory to language pedagogy. The concepts of 'scaffolding' and 'the zone of proximal development' (ZPD) are of central importance to this perspective. Scaffolding is the cognitive support provided by an adult or other guiding person to aid a learner, and is realised in dialogue so that the learner can come to make sense of difficult tasks. Scaffolded support is given up to the point where a learner can 'internalise external knowledge and convert it into

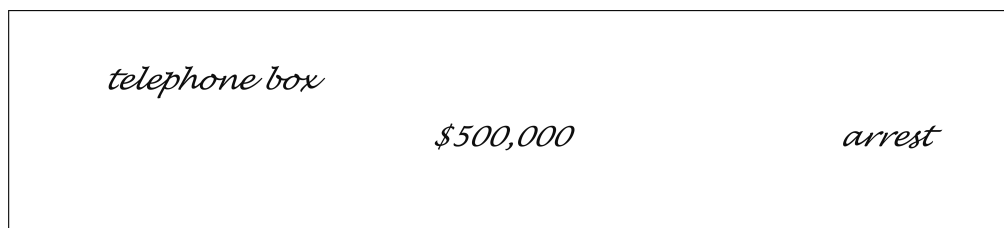
a tool for conscious control' (Bruner 1990: 25). The ZPD is the distance between where the learner is developmentally and what (s)he can potentially achieve in interaction with adults or more capable peers (Vygotsky 1978: 86). According to Lantolf (2000: 17), the ZPD should be regarded as 'a metaphor for observing and understanding how mediated means are appropriated and internalized.' In the Vygotskian paradigm, instructors (or peers) and their pupils interactively co-construct the arena for development, it is not pre-determined and has no lock-step limits or ceiling. Meaning is created in dialogue (including dialogue with the self, often manifested in 'private speech') during goal-directed activities.

Walsh (2006) notes that central to the notion of scaffolding are the polar concepts of challenge and support. He points out that learners are led to an understanding of a task by, on the one hand, a teacher's provision of appropriate amounts of challenge to maintain interest and involvement, and, on the other, support to ensure understanding. Johnstone (1989) presents scaffolding as a strategy used by learners and teachers to overcome 'shortcomings' in the learner's interlanguage, while Anton (1999) advocates the use of careful and particular error correction as a means of assisting learners through the ZPD. Machado (2000) demonstrates how peer-to-peer scaffolding in the preparatory phases of spoken classroom tasks (mutual help with the interpretation of the tasks and the wording of meanings) is reflected in evidence of internalisation of such help in the performance phases of the same tasks. Machado suggests that peer-to-peer scaffolding may be just as important as expert-novice scaffolding (see also Kasper 2001; Ko et al. 2003).

11.4 Applying the frameworks to a corpus of classroom data

Bringing together the three frameworks that we have surveyed above, we will now consider some of their key insights and concerns in the context of actual corpus data. Figure 4 and example (11.5) are taken from an extract from an EFL class (from the LIBEL corpus, see appendix 1) where the teacher is trying to build a schema (or cognitive outline) for a newspaper text that the students are going to read as part of a reading lesson. She puts three vocabulary items on the blackboard. We begin the extract as she finishes writing the last two items:

Figure 4: Extract from an EFL class



(11.5)

[the numbers on the left refer to turn numbers]

- 1 Teacher: ... ok ah so five hundred thousand dollars and arrest those are three things three items from a newspaper story. You can ask me yes no

questions that means I can only answer yes no or no okay? amm to find out a little bit more about the story. Now the dollar sign gives you a clue when asking the questions.

- 2 Student 1: Is it a fin=
- 3 Teacher: Is it a fine? No no it's not a fine.
- 4 Student 2: It's a robbery
- 5 Teacher: Yes yes a robbery umhm.
- 6 Student 3: Is it a re=
- 7 Teacher: A what? a reward? Sorry reward am no no that's not a reward no.
- 8 Student 3: Is it a phone=
- 9 Teacher: A coin box yeah
- 10 Student 3: [five syllables unintelligible] one phonebox.
- 11 Teacher: Not from one box. Not from one box from several boxes. Many boxes all right the five hundred thousand dollars came from many boxes yep ok. Anything else you can find out?

(LIBEL)

DA and CA: turn-taking in the classroom

The issue of the controlled or institutionalised nature of classroom discourse comes to the fore particularly in DA and CA models. Teachers have rights to initiation and evaluative feedback. Or in CA terms, there is a turn pre-allocation which assigns the questioning and evaluative role to the teacher, who is the holder of institutional power in a classroom context. Using DA and CA to examine extract (11.5) closely, we can make the following general observations about its turn structure:

discourse analysis	conversation analysis
<ul style="list-style-type: none"> • The teacher's move in turn 1 sets up the students as the initiators by getting them to ask the questions. • This seems to change the usual IRF structure by giving the students the right to initiate. • On closer examination, this is not so. Turn 1 is an initiation, turn 2, albeit a question from a student, is actually the response to the initiation in turn 1 by the teacher. • Turn 3 on the surface seems to be the teacher's response to turn 2, but it is in fact the teacher's evaluative feedback on turn 2. 	<ul style="list-style-type: none"> • The teacher is normally in the role of questioner, but in turn 1 she sequentially allocates this role to the students. • However, while the teacher attempts to redress the teacher-centred turn pre-allocation of classroom discourse (i.e. where the teacher gets to ask all the questions), she merely replaces it with another turn pre-allocation (where students <i>have to</i> ask the questions). That is, students are normally pre-allocated the role of answerer; now they are pre-allocated the role of questioner.

discourse analysis	conversation analysis
<ul style="list-style-type: none"> • The exchange pattern, therefore, comprises the classic IRF structure, controlled by the teacher. • However, the teacher has decentralised the questioning role within the classic IRF structure so that the students are asking questions. She is not always answering the students' questions, in fact she sometimes responds with another question or gives feedback on theirs. • Students do not have the right to make evaluative comments on the teacher's questions. 	<ul style="list-style-type: none"> • In reality, however, the teacher does not really change the turn pre-allocation or sequentiality of classroom discourse here: (1) she still usually selects the next speaker, (2) she manages and steers the topic by virtue of her responses, (3) she interrupts the students but they do not interrupt her, (4) she does not allow wait time between question and answer, (5) her responses to the students' questions are evaluative, and (6) on a number of occasions she does not adhere to the adjacency pairings of question + answer; instead she answers a question with another question.

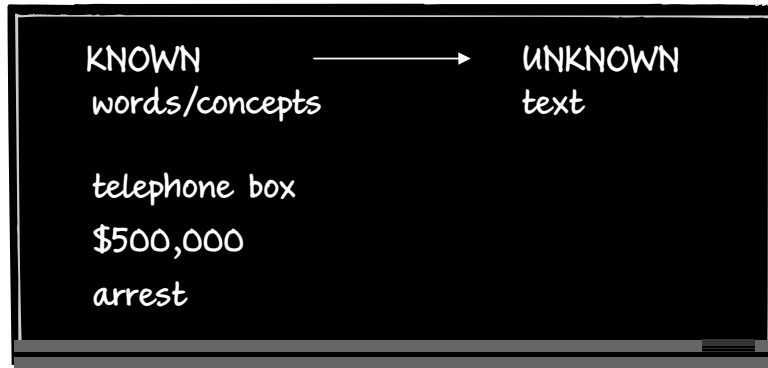
Some of the pedagogical reflections from this close analysis of the extract are:

positive	negative
<ul style="list-style-type: none"> • Getting students to take on the role of questioner is a good idea because it is normally monopolised by the teacher. • By getting the students to ask the questions, the teacher decentralised the lesson. • As the students are asking the questions, the teacher has the opportunity to assess how much vocabulary they already know in relation to the text that they are going to read and to appraise the amount of new vocabulary which will have to be presented. • Students can learn from each other by listening to each other's questions and the teacher's responses to these. This sets up a peer-peer interaction as well as a student-teacher interaction. 	<ul style="list-style-type: none"> • While the turn structure is devolved, the exchange is still highly controlled. • It would have been better to allow more wait time while the questions were being answered (see below). • The teacher interrupted the students in three out of five of their responses. • The teacher should have resisted reverting to the control position so soon. By turn 11, only after five contributions from the students, she intervenes. • In the teacher's initiation of the task, she says that the students must ask the questions and that she can only answer 'yes' or 'no'. However, she does not adhere to this arrangement and so never really hands control over to the students.

Socio-cultural theory: scaffolding and the ZPD

Extract (11.5) is an interesting one from the perspective of scaffolding. The teacher is preparing the students for a reading task. She needs to guide them through the ZPD by bridging the gap between what is known and unknown (figure 5). She does this by trying to build up the schema, or conceptual outline, of the story. The way in which she achieves this is interesting. Though it is teacher-led, it draws on peer-to-peer scaffolding. The teacher sets it up by giving three key words/concepts that she is confident the students will know.

Figure 5: Moving from the known to the unknown



Peer-to-peer scaffolding is set up through her yes/no question routine. Students have to listen to each other's questions carefully so as to collaboratively increment the collective understanding of the schema of the text. Learning takes place interactively between teacher and student, as well as between students.

The issue of the amount of scaffolding provided by the teacher is interesting to consider here. She provides the following scaffolds:

Table 1: Teacher scaffolding, a turn-by-turn analysis

student	teacher scaffold	type of scaffold
Turn 2 student says <i>fin</i> =	Turn 3 teacher provides <i>fine</i>	lexical
Turn 6 student says <i>re</i> =	Turn 7 teacher provides <i>reward</i>	lexical
Turn 8 student says <i>phone</i> =	Turn 9 teacher provides <i>coin box</i>	lexical and schematic ¹
Turn 10 student suggests <i>one phone box</i> [as far as can be established]	Turn 11 teacher provides the information that <i>the five hundred thousand dollars came from many boxes</i>	schematic

¹ On one level the teacher is giving an alternative lexical item to phone box, but at a schematic or conceptual level she is helping to add to the outline of the overall story by focusing on the phone box as a key factor in the story.

The following comments could be made about the teacher's approach, some positive and some more critical:

- She keeps the momentum of the guessing phase going by incrementing the new information at a steady pace, rather than letting it slow, so as to elicit the full or extended utterance from any one of the students. This sustains a high level of interest.
- She moves from lexical to schematic or conceptual scaffolds, building up key vocabulary before introducing schematic (or conceptual) information.
- She intervenes too soon in turns 3 and 7, for example, even before the students have had a chance to finish the words they are trying to construct.
- She provides too much scaffolding overall and should allow the students to engage in more guesswork for longer. This would promote more peer-to-peer scaffolding. Providing additional wait time would assist in this.
- By turn 11 when she provides the key information about there being many phone boxes, she has only had questions from two students at that stage.
- This could be counteracted by saying that the teacher knows the class and their level of need best and her goal is to build up a schema for the main task of the lesson, the newspaper story that they are going to read. She works at a pace that she knows will suit the class.

11.5 Looking at questioning in the classroom

Following on from this three-way analysis above, it is clear that questions have a central role in the classroom. Even when the teacher tried to hand over the questioning role to her students, she struggled with it, and that perhaps reflects the link between questioning and control. Classrooms, like a number of other institutional contexts such as political interviews, doctor-patient exchanges and courtroom interactions, are typified by a pervasion of questions. Raising an awareness of questions, how they are phrased, how many of them are asked, who they are asked to and how long the teacher waits for an answer are key issues to consider in teacher education and practice. Close scrutiny of classroom data can help considerably here. CA research tells us that the speaker who has high contextual status (e.g. lawyer in a courtroom, teacher in a classroom) normally controls the development of the discourse through questioning (see Coulthard and Ashby 1975; Sinclair and Coulthard 1975; Blum-Kulka 1983; Drew 1985; Fisher and Groce 1990; Heritage and Greatbatch 1991, among others). Hutchby and Wooffitt (1998) point out that institutional formats typically involve chains of question-answer sequences, in which the institutional figure asks the questions and the witness, pupil or interviewee is expected to provide the answers. This format is pre-established and normative rules operate, which means that participants can be constrained to stay within the boundaries of the question-answer framework.

In contrast, in casual conversation, roles are not restricted to those of questioner and answerer, and the type and order of turns in an interaction may vary freely. In this extract

from a casual conversation, for example, we see how questions meander from speaker to speaker as the conversation evolves in real-time, without any pre-allocation of questioning turns or chains of question-answer sequences:

(11.6)

[Twix and Snickers are chocolate bar brand names]

S1: I remember when I was in France ages ago when people were calling Twix Radars.

S2: Radars?

S1: Do you remember when Snickers were called Marathon?

S2: Yeah.

S1: And Twix were called Radars.

S3: Were they called Radars? I never knew that.

S2: Yeah the way they change the names of things like films.

S1: They just translate them

S2: No they don't 'Analyse This' right, they called it 'Mafia Blues'. It was an English word why change the name?

S1: They probably didn't know what analyse meant or something.

S3: Yeah do you know the 'Runaway Bride' is that what it is called?

S2: Yeah.

S1: Yeah.

S3: Am in France it was called 'Just married'

S2: 'Just married' that was it

S1: What?

S3: It was in English like.

S2: Yeah you used to see it on buses and it was like 'Just Married' and I was like that's 'Runaway Bride'. And I was like 'oh my god'.

S3: I wouldn't mind if they translated it into a French word but it was in English as well.

(LCIE)

Though many institutional interactions are question-laden, the pattern of how they are used is not necessarily homogenous. It can be instructive to compare classroom transcripts with data from other settings. Here we consider how classroom interaction compares and contrasts with media interviews.

In media interviews, interviewers and interviewees generally confine themselves to a question-answer sequence, respectively. The power-role holder does not normally engage in a wider range of feedback responses (Greatbatch 1988). For example, (11.7) is an extract from the BBC TV programme *Breakfast with Frost* in which the host, David Frost, interviews the then Secretary of State for Education, Ruth Kelly:

(11.7)

[Speaker 1 = David Frost, Speaker 2 = Ruth Kelly]

S1: And would you like to see, I gather
between the line you would, would
you like to see more foundation

- schools and more specialist schools
as soon as can be managed? *Initiation*
- S2: I think the idea of a specialist school
is an extremely important one. A school
that has its own mission and ethos. A
school that is strong and autonomous.
And they have really a very important
role to play in the future . . . *Response*
- S1: Will the 160 or so grammar schools survive
under your system, under your aegis? *Initiation*
- S2: Well, as long as parents want them in the
way they are, that's right. But I don't want
to see more selection in the process. What
I do want to see is really good state schools,
strong and autonomous, who want to co-operate
in the best interests of their students. *Response*

(*Breakfast with Frost*, BBC TV, 23 January 2005)

Statements are often made by both interviewer and teacher as a follow up to a response. When an interviewer uses a statement, it normally refers forward as a preface to or as part of the next question (Greatbatch 1988), whereas when a teacher makes a statement it is typically referring back to the student's response in an evaluative way (as discussed above):

(11.8)

[In this extract from the BBC programme *Newsnight*, presenter Jeremy Paxman is interviewing Richard Caborn, then British Minister for Sports and Tourism, about the British government's intentions to liberalise licensing laws in relation to extending the hours within which alcohol can be legally sold. Speaker 1 = Richard Caborn Speaker 2 = Jeremy Paxman]

- S1: . . . We have evidence to show where we have
relaxed in England on Sundays, in Scotland
when we allowed the opening hours to extend,
there was a reduction in the problems related
to nuisance through drink. Also you can cite
many other countries that you don't get those
problems on the Continent. *Response*
- S2: But we're not on the Continent. This is a north
European and Anglo-Saxon problem. *Statement as Initiation*
- S1: France and Germany are north Europe. When
they come over and go to a show at the Barbican
and they can't get a drink after 11.00, they look
at us bemused. *Response*
- S2: So we're doing it to placate French and German
tourists. *Statement as Initiation*

S1: Jeremy, when you're walking in Derbyshire and you can't get a drink at 4pm in the afternoon, because of the licensing laws, you get a little annoyed. *Response*

S2: So we're doing it to placate French and German tourists and walkers in Derbyshire. *Statement as Initiation*

S1: Plenty of other people who'd want of an evening to go and relax having a drink. *Response*

(*Newsnight*, BBC TV Tuesday, 8 July, 2003, Full transcript
<http://news.bbc.co.uk/1/hi/programmes/newsnight/3055548.stm>)

The goal of the media interview is primarily to elicit information whereas the classroom goal is to facilitate learning, and so the teacher's questions and responses must increment knowledge rather than assume it. Many of the teacher's questions and responses serve to build up shared knowledge. Notice in extract (11.9) how the teacher stages her responses and questions so as to repeat what has been said for the benefit of others in the class. She gradually builds new information and extends vocabulary by repeating and recycling the students' responses.

(11.9)

[In this language classroom extract, the teacher is introducing a newspaper article on healthy eating for university students. They are discussing what constitutes a healthy lunch.]

Teacher: What do you think they might mean by a healthy lunch then?

Student: Having something else to ah eat.

Teacher: So what might they eat normally? Maybe.

Student: Chips, burger.

Teacher: Okay. Fries fries burger.

Student: Drinks.

Teacher: What kind of drinks? All right fizzy drinks?

[laughter]

Teacher: You know the expression fizzy drinks. Have you come across 'fizzy'?

Students: Yeah.

Teacher: What, Sebastian very kindly came in showing us there and what you just finished there. Is a fizzy drink am coke fanta fizzy drinks po= we also use the word pop am there tends to be a lot of chemicals in these drinks . . . So burgers pop what else might they eat normally?

Student: Eat sandwich.

Teacher: Yeah.

Student: Sweets.

Teacher: Yeah chocolate. Yeah cake. The food we like unfortunately. So what might be a healthy option?

Student: Vegetables.

Teacher: Vegetables okay what else?

[Three turns later]:

Student: Yogurt.

Teacher: Yeah yogurt am maybe water or if they don't like water and they don't like milk what else could they drink that's not fizzy?

Student: Juice.

Teacher: Orange juice apple juice . . . what system do we have in England and in Ireland for school lunches for kids in schools?

(LIBEL)

The classroom context differs greatly from the media interview in that there is a constant dialectic between student responses and pedagogic goals. In the media interview, as noted by Carter and McCarthy (2006), the interviewer typically does not follow up on responses in the same way that the teacher does; instead the listener or viewer is usually left to make his/her own evaluation of the interviewee's answer. The goal of the interviewer is to elicit information and to entertain rather than to teach the interviewee or the audience.

Something that the media interview and the classroom interaction have in common is the use of display questions. These are typically questions to which the questioner already knows the answer. As Carter and McCarthy (2006: 717) note, they are common in contexts such as classrooms, quiz shows and other tests of knowledge, and media interviews. The purpose of a display question is to put knowledge or information on public display. In the classroom, this is an important way of transmitting and testing knowledge for teachers and students. In these display question situations such as classrooms and quizzes, the questioner follows up the answer by stating whether it is the correct one or not. However, in media interviews, as we have noted, the follow up is very often left to the listener or viewer. We will now take a close look at other types of questions, including display questions, and the impact that they may have on the course of classroom interaction.

Questioning and question types

Questions are broadly defined as utterances which require a verbal response from the addressee and there are a number of types, based on a variety of structural patterns. Carter and McCarthy (2006: 715–727) distinguish between the following forms which function as questions:

- 1 Yes-no questions: these are one of the most common question types. The anticipated response is either *yes* or *no*.

Do you know what a freebie is?

(LIBEL)

- 2 *Wh*-questions: questions with *what*, *when*, *where*, *which*, *who(m)*, *whose*, *why*, *how* request specific information concerning persons and things, and the circumstances surrounding actions and events (e.g. time, manner, place, etc.). The anticipated response to such questions is not *yes* or *no*, but information which provides the missing content of the *wh*-word.

What adjective would you use to describe someone who says 'hi how are you I'm it's nice to meet you'?

(LIBEL)

- 3 Alternative questions: these questions give the answerer a choice between two or more items contained in the question which are linked by *or*. Alternative questions may be *yes-no* interrogatives or *wh*-interrogatives. An alternative question may offer the recipient the choice of one or all of the alternatives.

Is this is this a word, a phrase or a clause?

(LIBEL)

- 4 Declarative questions: not all *yes-no* questions have interrogative form, and a declarative clause may function in context as a question. The intonation is typically rising (↗) (asking for confirmation) or falling (↘) (strongly assuming something).

↗ *You are sick today?*

(LIBEL)

S1: ↘ *So you're going to be here about quarter past?*

S2: *Yeah quarter past, twenty past, yeah.*

S1: *That's fine.*

(CANCODE)

- 5 Tag questions: questions may include a tag after a declarative clause. Tag questions are highly interactive in that they may constrain the range of possible or desired responses from the addressee. Some patterns are more constraining than others.

*You've worked hard **haven't you?***

(CANCODE)

- 6 Echo and checking questions: echo questions repeat part of the previous speaker's utterance, usually because some part of it has not been fully understood. They often have declarative word order and a clause-final *wh*- word.

S1: *He's called Oliver.*

S2: *He's called **what?***

S1: *Oliver.*

S1: *Steve was singing with the group.*

S2: ***Who** was singing, sorry? (stressed)*

S1: *Steve, Steve Jones.*

S2: *Oh.*

(CANCODE)

A corpus of classroom interactions provides a very good starting point for reflecting on teacher questioning strategies and how these affect the classroom interaction, and ultimately the learning outcome. Farr (2002) looked at the questions in a corpus of

classroom interactions of five pre-service teachers who were undertaking a language teacher education course. In these EFL classes, the teachers were working with advanced level students. Her research showed that declarative questions produced the longest answers:

Table 2: Question types and answer length (Farr 2002)

question type	average number of words per answer
yes-no	7.36
wh-	10.51
alternative	9.33
declarative	18.33

Research into classroom questions also uses a functional categorisation including display questions, as mentioned above, and referential questions (see Banbrook and Skehan 1989; Farr 2002):

- 1 **Referential questions:** genuine questions to which the teacher does not already know the answer

Teacher: So how long have you studied English Jong?

(LIBEL)

- 2 **Display questions:** questions to which the teacher already knows the answer

Narrow display questions: display questions to which there is only one anticipated response in terms of either content or form

Teacher: What do you call that what they're wearing?

Student: Uniform.

(LIBEL)

Broad display questions: display questions to which there is a range of possible answers in terms of content or form from a range of possibilities already known to the teacher

Teacher: Marie can you tell me what did you find in the third paragraph?

(LIBEL)

Farr (2002) also looked at functional questioning strategies in her corpus of pre-service teachers and she found the following breakdown:

Table 3: Breakdown of functional questioning strategies (Farr 2002)

question type	total
referential	13
narrow display	38
broad display	74

Pica and Long (1986) examined the difference in linguistic performance between experienced and inexperienced teachers in Philadelphia. In terms of questioning, they found that, among inexperienced teachers:

- more display questions were employed in classroom talk than in informal conversation.
- almost four times as many display questions were asked as referential questions (see also Long and Sato 1983).

In another study, Brock (1986) examined the effect of using more referential questions in the language classroom. She found that by increasing the frequency of referential questions, students produced longer and more syntactically complex responses. While display questions produced an average answer length of 4.2 words, referential questions produced an average of ten-word answers. Farr (2002) found the following correlation between question type and length of answer in her corpus-based study:

Table 4: Question type and average length of student reply (Farr 2002)

question type	total occurrences	average number of words per reply in student answers
referential	13	17.92
narrow display	38	3.34
broad display	74	12.44

Another important factor in classroom questioning strategies that has arisen from research is the amount of time that the teacher pauses after asking the question; that is, the ‘wait time’ after asking a question before the teacher added a new or re-formulated question. White and Lightbown (1984) found that teachers rarely waited longer than two seconds for a reply from their students. Farr (2002) calculated that only 27% of all the questions that she looked at allowed any wait time. O’Keeffe and Farr (2003) suggest how a corpus of classroom interactions can be used to focus on questions and questioning strategies so as to promote teacher awareness and reflection.

11.6 Teacher corpora in professional development

Adolphs et al. (2004) look at communication in the professional context of health care in a corpus-informed study of staged telephone conversations between callers and advisers in the UK’s NHS Direct health advisory service. They make a case for applied clinical linguistics, which involves the synergy of those involved in the health services, educators and corpus linguists. By looking at the communicative events within the profession empirically, they argue, a better understanding of the interaction can be reached and this can lead to better practice. This model lends itself even more readily to the broader professional

context of language teaching since as a professional group we are more linguistically equipped to reflect on our own language use. Within this model, contexts beyond the classroom would be included so as to examine, for example, how we communicate with colleagues, trainees and administrators in non-classroom contexts such as meetings, staffrooms, offices, which are part of the wider situational matrix of teaching.

As noted by Sarangi (2002: 106), the primary focus of classroom-based teacher–pupil interaction is at the expense of looking at what happens outside the classroom. Corpora are beginning to have applications to teacher talk outside of the classroom, particularly in the broadening model of teacher observation. Two corpora have been independently developed to focus on this type of interaction and to learn from it (see Farr 2003, 2005; Vásquez and Reppen 2004; Vásquez 2004, 2005). Farr, working with the Post Observation Teacher Training Interactions (POTTI) corpus of over 80,000 words, looks at the interaction of trainers and trainees on an Irish postgraduate teacher education programme (see also chapter 6). Her work gives many insights into the post-observation interaction, including the role of relational strategies such as inclusive pronoun use when advising, so as to draw on professional solidarity, the use of first name vocatives, hedged directives, shared socio-cultural references as well as engaged listenership (responses, overlaps, interruptions) and small talk. Extract (11.10) is an example from Farr (2005: 214), where at the beginning of a post-observation session small talk is used as a relational strategy by the trainer to mitigate forthcoming criticism (the trainee had made a major organisational mistake in her teaching practice by preparing the wrong lesson). The small talk extends for 19 turns in all:

(11.10)

- Trainer: . . . are you feeling okay now cos you were you weren't feeling great earlier you said?
- Trainee: Em not any better I can tell you actually+
- Trainer: Really?
- Trainee: +I'm very tired and em I think I've an ear infection or something every time I talk I can it's like major feedback in my ear+
- Trainer: Oh
- Trainee: +yeah I'll need to get to the doctor or something.
- Trainer: You need to be careful with that.

(Farr 2005: 214)

Vásquez and Reppen's work draws on a corpus of language teachers and their mentors in a longitudinal, action research study in an American university intensive English programme. Post-observation meetings between mentors and teachers were recorded and transcribed over a period of two years. The authors were involved as mentors in these interactions and their initial findings showed that they were responsible for the majority of the talk in the meetings and that teachers tended to be passive. Based on this, changes were made to their practice with the goal of eliciting more talk from teachers. Focusing primarily on interactional data from four teacher/mentor pairs collected over two semesters, Vásquez and Reppen (in press) describe how this study enabled mentors to become aware

of the linguistic and interactional subtleties of their existing practices. They illustrate how mentors were able to successfully change the meeting dynamics from mentor-centered to more teacher-centered through changes in the distribution of talk among participants. Important changes came about, for example, as a result of the ways that teachers were positioned by mentors in the openings of meetings. As in Farr's work, Vásquez and Reppen have created their own corpus to look at their own professional practices in context.

Vaughan (in press) looks at a corpus of English language teacher meetings in which she participated. She applies Goffman's (1959) dramaturgical metaphor of *frontstage* and *backstage* to teacher discourse. She contrasts the teachers' highly regulated and formalised frontstage talk in the classroom with their less organised backstage identity. Somewhere between this highly regulated and formalised frontstage and less organised backstage lies the area of mediated interaction which has as its goal the facilitation of professional development (e.g. Edge 1992, 2002) and reflective practice (e.g. Walsh 2002, 2003). Vaughan argues that, while the frontstage interaction has been considered the most significant type of discourse that teachers engage in, interaction outside the classroom, the teacher's *backstage* (teacher to teacher) discourse, is equally significant and has not thus far received as much attention as it merits. Vaughan, working with a corpus of over 40,000 words of teacher staff meetings, looks at how characteristics of this Community of Practice (after Lave and Wenger 1991; Wenger 1998) may be realised in linguistic features, and how these features together comprise a 'badge of identity'. She finds, for example, that the type of vague language used by the teachers is specific to their practices and that humour is key to the establishment of a shared communicative space. She also highlights the creation of this space through the construction of in- and out-groups.

Corpora also have great potential as a linguistic resource for teachers who wish to either improve their own language awareness or want to find out more about a specific structure in a language that comes up for them in the classroom. A number of studies illustrate the role of using a corpus in developing teachers' linguistic awareness both in pre-service education and in-service development and support (see Hunston 1995; Allen 1999; Conrad 1999; O'Keeffe and Farr 2003; Tsui 2004, 2005).

Allan (1999) and Tsui (2004, 2005) provide details of an exciting Hong Kong-based corpus facility which supports English teachers' grammar queries online. The website, *TeleNex*, was set up in 1993 to provide professional support to English language teachers in Hong Kong schools (see Tsui 2004). It is supported by a team of language specialists at the Teachers of English Language Education Centre (TELEC) of the Faculty of Education, The University of Hong Kong (see Tsui 1996; Tsui and Ki 2002). The website is designed to include a conference area in which a number of discussion corners have been set up, including one on the English language. Within this 'corner', teachers send questions seeking help and advice on language issues. The questions are responded to by both school teachers and language specialists in TELEC, some of whom are full-time staff specifically recruited to support the website and some are academic staff in the Faculty of Education. The service has evolved so that teachers can now learn to use the corpus resources independently as well as avail themselves of the support team's responses, and obviously they

can respond to each other's queries. In a period of eight years, more than one thousand questions were submitted (Tsui 2005).

When answering teachers' questions, corpus data is consulted for evidence of language structure and use. What is interesting is that this is done from both a local and an international context of use. Internationally, mostly British and American English corpora are used (the BNC and COBUILD Direct). Locally, the team has amassed data of considerable size to reflect how forms are used by successful users of English in Hong Kong. These include the Modern English Corpus (see Tsui 2005), a five-million-word native speaker collection consisting of one million words of spoken texts from radio phone-ins, panel discussions, casual conversations and lectures and two million words of literary and academic texts, and two million words from feature articles in the *South China Morning Post*, and the TeleCorpora, which includes a 20-million-word sub-corpus of articles from the *South China Morning Post* and a learner corpus of more than two million words. TeleCorpora is now available for on-line access by registered users of the TeleNex website (<http://www.telenex.hku.hk>). Reflecting on the project, Tsui (2005) believes that the process has led to many existing concepts about language being challenged (she provides a number of examples, including a query on whether *because* can be used to begin a sentence or turn). This offers an example of how a corpus can become an end in itself rather than just a means to an end. It can offer a tool for awareness-raising at all stages of professional development.

Meanwhile, at the Pennsylvania State University in the USA, a website is available to which teachers can upload their own data of any kind and gain assistance in coding and analysing it using the site's own online software, which, when fully developed, will include capabilities for measuring features such as lexical density and variation, as well as the more conventional tools of frequency lists and concordances, all linked to sophisticated databases. The site also encourages and enables data-sharing among practitioners, an invaluable step in the creation of a community of corpus-aware professionals. The website is under the aegis of the CALPER project (Centre for Advanced Language Proficiency Education and Research; see <http://calper.la.psu.edu/>).

11.7 Conclusions and considerations

A corpus as a complementary resource

As we have stressed here, we are not advocating a corpus of classroom interactions as a replacement for video resources, but rather we are saying that the one complements the other. A video offers the opportunity to look at the classroom interaction in close detail, its transcription allows us to look even closer (and commercially available videos often include transcripts, for example Bampfield et al. 1997). A teacher-made corpus of classroom interactions adds to this kind of resource because it comes from a local context, reflects local teaching conditions and can be viewed with local insights. It is something that can be built up gradually over time and not something that needs to be of a certain size before it can be of any use. Even one hour of recording can offer many reflective opportunities. As we have

seen here, most is to be gained by looking at short extracts. In this way, a teacher corpus is one from which much can be gained qualitatively, where the corpus is an end in itself. In other chapters in this book, we sometimes used corpora as a means to an end, to help us identify lexical frequencies and language patterns, for example, which will inform *what* we teach. A corpus of teacher interactions, on the other hand, informs us about *how* we teach and interact in the classroom and with colleagues. Here, we have been concerned not so much with what can be gained *from* a corpus as what can be gained *by* it.

A teacher-made corpus provides a mirror for our own practice which we can hold up to ourselves and learn from what we see. In the future, the optimum situation will certainly be to have digital audio-visual corpora, thus merging image and transcript (the BASE corpus has already achieved this for the majority of its data; see appendix 1). The further down the line we go with audio-visual corpora, the more challenges we face. For example, how best should we code the visual aspects of non-verbal communications? How many cameras would be needed to capture a classroom interaction? Classroom interactions, like most social interactions, are *multi-modal* in nature, combining both verbal and non-verbal components and units (Saferstein 2005). If we are to properly transcribe the audio-visual interaction, should we transcribe and align teacher and student gestures and other non-verbal components such as position of teacher, direction of gaze, movement of hand and so on? Current research at the University of Nottingham, for example, is looking at ways of building an audio-visual corpus so that ultimately concordance lines can be generated with the visual as well as verbal (Carter et al. 2006; Adolphs and Carter (forthcoming)).² At a technical level this poses many challenges. A number of projects are underway to this end, for example see Pea (in press).

From turn to theory

Teaching and learning do not just happen. They are part of an interactional process built around teaching goals, learning styles, individual differences and classroom conditions, among other things. By extracting actual classroom interactions from a corpus and breaking them down turn by turn, we have been able to explore this interactional process very closely. However, to do so we have needed to draw on some existing frameworks. The importance of teacher awareness of frameworks for analysing discourse is something we see as fundamental since they help us interpret our practice. This also points to a wider issue in corpus linguistics: the question as to whether corpus linguistics is a theory or a method (see Tognini-Bonelli 2001). For us, a corpus is a database and the processes of corpus linguistics offer a powerful methodological tool. The interpretation of the results that we generate from either qualitative or quantitative analyses need to be interpreted within existing applied linguistic frameworks, as well as enabling us to refine those frameworks and generate novel ones, in the classic dialectical process. Here we have used three frameworks: DA (discourse analysis), CA (conversational analysis) and Sociocultural theory, but there are many others including CDA (critical discourse analysis) (Fairclough 1989, 1992, 1995),

² See <http://www.nottingham.ac.uk/english/research/cral/projects.html>

Language Identity, Language Socialization and many Second Language Acquisition models that could have been applied (see McCarthy 1991; Hatch 1992; McCarthy and Carter 1994; Johnson 1995; Riggenbach 1999; Celce-Murcia and Olshtain 2000; Boxer and Cohen 2004; Seedhouse 2004; Walsh 2006).

Throughout this book we have drawn on frameworks to interpret what we find in language corpora and these frameworks often lead us to new insights which, in turn, suggest new ways of exploiting corpora. This process is unlikely ever to come to a finite end. Nor should it, for corpora are endlessly fascinating treasure-houses which always have something new to offer. There is no such thing as a used up, worn-out corpus.