Details of published version: Vaughan, E. 2010. How can teachers use a corpus for their own research? In: O'Keeffe, A & M. McCarthy (eds.), *The Routledge Handbook of Corpus Linguistics*. London: Routledge, pp. 471-484.

How can teachers use a corpus for their own research?

Elaine Vaughan

DRAFT

1. USING A CORPUS FOR YOUR OWN RESEARCH: BEING PROFESSIONALLY CURIOUS

Practising language teachers engage in corpus research and other types of research for a variety of reasons. The general stimuli and specific motivators for their research may lie in challenges and/or opportunities in the immediate teaching and learning environment, may be based on personal or academic interests, may be stimulated by interaction with colleagues or be in response to research findings which have been released into the public domain. However, a unifying feature of such research, and one which is crucial to the profession of teaching as a whole, is professional curiosity. This chapter is specifically aimed at practitioners who are interested in conducting their own professional or pedagogical research and would like to explore the possibilities of using a corpus in this regard. Some samples of the directions that corpus research in language teaching have taken are presented before we turn our attention particularly to the language of the wider professional context, specifically the language used by teachers. This is the site of a small but growing number of corpus-based studies, proof, if any were needed, that corpus methods can co-exist harmoniously with any number of paradigms of linguistic research (see, for example, Carter and McCarthy 2002 or Walsh and O'Keeffe 2007). An example of a corpus-based study into the interaction of English language teachers in meetings is presented at the end of the chapter as just one example of research inspired by professional curiosity and how a corpus can inform this type of work. The participation of practitioners and trainee teachers in their own corpus-based research is frequently advocated (e.g. O'Keeffe and Farr 2003; Tsui 2004; Römer 2006, 2009; O'Keeffe *et al.* 2007; McCarthy 2008; Breyer 2009) though this has not necessarily translated into the global provision of corpus analysis modules in teacher education and training programmes (see Farr, this volume; McCarthy 2008; Granath 2009). O'Keeffe *et al.* (2007: 246) further underline the

Details of published version: Vaughan, E. 2010. How can teachers use a corpus for their own research? In: O'Keeffe, A & M. McCarthy (eds.), *The Routledge Handbook of Corpus Linguistics*. London: Routledge, pp. 471-484.

need for reciprocity in the relationship between language researchers and language teachers, and they go as far as to claim that for the future '...research questions need to be driven by teachers, and indeed a more critical response to the findings of corpus linguistics needs to come from teachers.' As not all teacher education programmes offer training in corpus analysis, one of the decisive factors in making this happen will lie in the willingness of practising teachers to 'get their teeth into a corpus', to borrow Aston's (2002) evocative phrase.

A major benefit of corpus-based based research for language teachers lies is in its potential as a teacher development tool. Waters (2005) references the common distinctions made between teacher training, teacher education and teacher development. He uses teacher education as a superordinate term, but it is interesting in terms of our question and discussion here to look at what is generally understood to be contained in these distinctions. Teacher training concerns itself, for example, with the practical, classroom-based skills, teacher education has been seen as concerned with research and background knowledge, while teacher development has a focus on raising awareness of practices and fostering reflection and change (*ibid*: 211). Waters summarises the focus of these three interconnected aspects of 'teacher education' as *doing* (teacher training), *knowing* (teacher education) and *being* (teacher development). While the first two tend to be a mediated experience for teachers, the latter is different in that it is most often self-directed. Dörnyei (2007: 17) suggests that research excellence requires a number of essential characteristics in the researcher: genuine curiosity, common sense, good ideas and a blend of discipline, reliability (in the sense of thoroughness and systematicity displayed by the researcher him- or herself as opposed to the methodological concept) and social responsibility. Refreshingly, he also points out that research does not have to be the preserve of the elite few; the same argument applies to corpus-based research for language teachers. Corpus-based research can be applied as a means of investigating the *doing*, *knowing* and *being* of teaching and learning; what the practitioner-researcher needs is genuine, professional curiosity, a sense of how corpora are built and work and what sort of questions they have the potential to provide whole or partial answers for. We start with the kinds of issues that

DRAFT

Details of published version: Vaughan, E. 2010. How can teachers use a corpus for their own research? In: O'Keeffe, A & M. McCarthy (eds.), *The Routledge Handbook of Corpus Linguistics*. London: Routledge, pp. 471-484.

practitioner-researchers may have in consulting and using corpora in general and building their own corpus in particular. Then we look at the types of questions about language teaching that corpora have been used to pose and answer before taking the less travelled route of looking at the insights that a corpus-based analysis can generate about the profession of language teaching itself.

2. ISSUES IN USING AND BUILDING CORPORA

If we ask first of all what a corpus is, it provides a shortcut to what some of the issues you will face in using a corpus for your own research. Tognini-Bonelli (2001: 53) surveys definitions of what a corpus is and highlights that although the definitions she presents diverge, they agree in their basic assertion that a corpus is a collection of language text, 'though not necessarily texts' (*ibid.*) (see also Tognini-Bonelli, this volume). Aarts (1991: 45) suggests the criterion of 'running' text, Sinclair (1991: 171) integrates the idea of the text being 'naturally-occurring', while Francis (1982, 1992) introduces the term 'representative'. Crystal defines a corpus as 'a representative sample of language, compiled for the purpose of linguistic analysis' (1997: 414), and Biber *et al.* (1998: 4) characterise a corpus as 'a large and principled collection of natural texts'. Tognini-Bonelli provides her own, inclusive definition:

> A corpus is a computerised collection of authentic texts, amenable to automatic or semi-automatic processing or analysis. The texts are selected according to specific criteria in order to capture the regularities of a language, a language variety or sub-language.
>
> (2001: 55)

All corpora are collections of texts, but one could equally argue that the Web is a 'collection of texts', though clearly it is not a corpus in the conventional sense (see Lee, this volume for a further discussion on the Web as a corpus). So, what makes a corpus different? Sinclair (2001) provides an interesting answer for this, which is incredibly useful to bear in mind if you decide to create your own corpus. A collection of texts becomes a *corpus* when we treat it as such: the texts are gathered according to some kind of external

DRAFT

Details of published version: Vaughan, E. 2010. How can teachers use a corpus for their own research? In: O'Keeffe, A & M. McCarthy (eds.), *The Routledge Handbook of Corpus Linguistics*. London: Routledge, pp. 471-484.

(cf. Clear 1992) criteria (teacher-student interaction or soap opera discourse, say) and we expect 'an investigation into the patterns of the language used will be fruitful and linguistically illuminating' (*ibid*: xi). The crucial idea here is 'patterns': rather than evaluating a text (which of course can be spoken or written) as an object from beginning to end, we are considering its parts, not the sum of its parts, but the parts themselves. Using corpus software, we can search for patterns in the language of the texts and find how these are similar or different to patterns of language use in other corpora and contexts. Corpus-based analysis can be used in tandem with complementary discourse analytic methods, thus exploiting to the full the corpus contents.

The issue of how large a corpus should be, or even how small a corpus *can* be, has been a bone of contention. There does not appear to be an upper limit on corpora: the British National Corpus (BNC) contains 100 million words, the American National Corpus (ANC) (currently 22 million words) when complete will also have 100 million words and the COBUILD Bank of English stands at a massive 450 million words. The Cambridge International Corpus (CIC) is even more of a behemoth at over one billion words. These corpora represent a mixture of spoken and written texts, though not necessarily in equal proportions, for example, the BNC consists of 90% written text and only 10% spoken text. Representativeness within a corpus, 'or the extent to which a sample includes the full range of variability in a population' (Biber 1993: 243), is probably the more salient issue. Language data have proven resistant to standard approaches to statistical sampling (Clear 1992: 21) and sampling frames. However, Biber proposes strata and sampling frames for representative corpus design based on *register*, or situationally-defined text categories such as 'fiction', 'news article' etc., and linguistically-defined text types, such as various written or spoken modes. With regard to sample size, his previous research on 1,000-word samples from the London-Lund and Lancaster/Oslo/Bergen corpus concluded that these relatively small samples yielded similar functional and grammatical findings (Biber 1990). The register approach taken by Biber *et al.* (1999), for example, has meant that, amongst other things, it is possible to compare and contrast how language is used in different contexts.

DRAFT

Details of published version: Vaughan, E. 2010. How can teachers use a corpus for their own research? In: O'Keeffe, A & M. McCarthy (eds.), *The Routledge Handbook of Corpus Linguistics*. London: Routledge, pp. 471-484.

In terms of consulting an existing corpus, there are a number of issues for the novice corpus researcher (see chapters by Tribble and Evison, this volume). One that is not always emphasised across a literature that, on the whole, assumes its audience is the university-based language teacher/pre-service trainee, is access to resources such as computer labs and corpus software. While these are a pre-requisite for *hands on* activities with language learners, the lack of these types of facilities in your teaching environment does not preclude engagement with corpus-based research and pedagogical activity. There is much that can be done with access to a personal computer, the internet and a printer. As Conrad (1999) has pointed out, without any computer facilities at all, the findings of corpus-based studies can still be (and have still been) of use to language teachers; for the purpose of this chapter, we will assume that the minimum of PC, internet access and printer is available to the reader. Practical concerns aside, there are a number of essential skills practitioners new to using corpora need to develop. Frankenberg-Garcia (forthcoming, 2009) summarises these and suggests that teachers need to know the following:

1. What corpora are available;
2. How to formulate a corpus query;
3. How to interpret the results of this query.

Firstly, teachers need to know what corpora are freely and commercially available (see Lee, this volume); more importantly, it is essential to develop an awareness of what these available corpora can and cannot offer. Being able to critically evaluate what is available and make an informed choice in relation to whether an existing corpus is appropriate for the investigation of a research question, or whether a new, more specialised, corpus is required is the first step for the budding teacher-researcher (see Reppen, this volume). For example, the Michigan Corpus of Academic Spoken English (MICASE) is an excellent resource if you wish to investigate the kind of language it will be necessary to teach your students so that they will be able to operate in the (North American) academic domain, but it will not provide much information about how friends interact with one another in casual conversation. Secondly, once an appropriate corpus resource has been chosen or created, it

DRAFT

Details of published version: Vaughan, E. 2010. How can teachers use a corpus for their own research? In: O'Keeffe, A & M. McCarthy (eds.), *The Routledge Handbook of Corpus Linguistics*. London: Routledge, pp. 471-484.

is necessary to be able to adequately frame corpus queries, or ask relevant questions. For the teacher embarking on corpus-based research, this is possibly less of a difficulty as the likelihood is that you have already isolated a particular genre of discourse or particular language feature that you want to investigate. Again, it is worth emphasising that choosing an appropriate corpus is crucial – if you want to analyse occurrences of the item *it*, for example, a very large corpus may contain far more information than it is feasible to analyse and a smaller corpus size may be more than adequate (cf. Biber 1990). Finally, once these queries have yielded results, the teacher-researcher needs to be able to interpret these results and take into account considerations such as corpus size and composition and their impact on the type of data that is generated. In other words, an ability to interpret what corpus data 'means', what variables are impacting on the corpus results being returned and what follow-up queries may be required in order to explore them are skills, or as Mukharjee (2002: 179) puts it, 'corpus literacy', that teachers need to cultivate. Given these complexities, on-line professional development courses for practising teachers such as *An Introduction to Corpora in English Language Teaching* (McCarthy *et al.* 2007) are invaluable. This course provides modules on how to use corpora to investigate and teach grammar and vocabulary and the implications of the corpus evidence in skills-based teaching. Evison (this volume) gives a thorough introduction to how to 'get your teeth into' a corpus from exploring word frequencies to exploring discourse, and in the further reading section at the end of this chapter, other texts that give accessible and practical introductions are recommended.

3. BUILDING YOUR OWN CORPUS

In terms of building and using a corpus for your own research, there are some points to consider which benefit from further discussion (for a more extensive coverage, see Reppen, this volume). Broadly, these are:

- Access and consent

DRAFT

Details of published version: Vaughan, E. 2010. How can teachers use a corpus for their own research? In: O'Keeffe, A & M. McCarthy (eds.), *The Routledge Handbook of Corpus Linguistics*. London: Routledge, pp. 471-484.

- Recording and transcribing spoken data/compiling and storing written data
- Your position in the research

Practising teachers have a significant advantage when it comes to investigating the language of language teaching as their position in teaching institutions means that the data required is within immediate proximity. Once you have identified the language you require, the next step is securing consent to record / collect and use that data. This means that you will need to approach your students or your colleagues and ask their permission to record (in the case of spoken data) or compile (in the case of written data) language material that 'belongs' to them. Most learners and colleagues are very cooperative if you are upfront about why you want to research a particular sphere of language use and what you will do with it. This is even more the case if the results of your analysis are going to address a practical problem or highlight good practices. Always obtain consent in written form and do so at the beginning of the research. The consent form should assure the learners or colleagues who cooperate in the research that you will anonymise the data and treat it ethically. It should also make clear what the data may be used for: it is best to cover all the possibilities in this regard as it may not always be possible to predict the research paths collecting the data will bring you down. For example, you may end up publishing your research, using extracts in conferences presentation or other professional meetings or creating/contributing to a larger project etc. A thorough consent form will cover all these possibilities. Make sure that your corpus design is replicable and you keep as much contextual information about it as possible. The first of these considerations is crucial. As a rule of thumb, once you have finalised your corpus design, another researcher should be able to add to your corpus, or build a companion corpus, by adhering to your design principles. More importantly, you will be able to maintain and add to your corpus so that it becomes an organic entity.

If your corpus is to contain spoken texts, then you will need to have recording equipment and a means of transcribing and storing the audio files (see Adolphs and Knight, this

DRAFT

Details of published version: Vaughan, E. 2010. How can teachers use a corpus for their own research? In: O'Keeffe, A & M. McCarthy (eds.), *The Routledge Handbook of Corpus Linguistics*. London: Routledge, pp. 471-484.

volume). Sound quality is essential for spoken data collection, and the more speakers that will be present, the more important this is. As long as the quality of the recording it allows is good, any recording device can be used. However, a digital recorder is the better investment as it is possible to download free transcription software (two examples are: *Express Scribe* and *SoundScriber*) which allows you, amongst other things, to control the playback of your audio recordings using 'hot' keys on the keyboard and slow down the playback (a very valuable aid to accurate transcription). Transcription is a slow process; McCarthy (2008: 571) estimates that one hour of talk can take 12-15 hours to transcribe. Despite its labour-intensive nature, the transcription of spoken data is an excellent skill to hone. One of the benefits of collecting and transcribing spoken data yourself is that your familiarity with the texts that your corpus contains grows exponentially. In addition, as you transcribe, you may identify particular language features which will be interesting to look into in greater detail when you do your preliminary analyses. Whether or not the same is true with written texts depends on your sources. If you are using material from the Internet, your job may be to gather on principle rather than physically input the data (while observing copyright restrictions). If, however, you are creating a corpus of, for example, student essays or other written work, you will more than likely be typing these using a word processing programme and thus gaining the same sort of familiarity and insight mentioned above in relation to spoken data transcription. Many books and studies that deal with transcribed data provide transcription conventions that you could potentially use and modify (see, for example, Eggins and Slade 1997: 1-5; O'Keeffe *et al.* 2007: 6).

Two final points should be made on the collection of spoken corpora which are implicated in the third of the issues outlined above, the position of the researcher. When teachers create their own spoken corpora, they are more often than not amongst the participants in the interaction they are recording; for example, a corpus of students performing interactive classroom tasks may include the instructions given by the class teacher, who is also the corpus compiler. Being involved in an authentic situation as a participant and as an analyst can be extraordinarily positive but it is also important to acknowledge, and be conscious of, the potential biases that this dual role may bring. In setting out to record authentic, spoken

DRAFT

Details of published version: Vaughan, E. 2010. How can teachers use a corpus for their own research? In: O'Keeffe, A & M. McCarthy (eds.), *The Routledge Handbook of Corpus Linguistics*. London: Routledge, pp. 471-484.

interaction, an oft-mentioned Catch-22-like situation occurs: we inevitably introduce a degree of artificiality into a previously authentic situation, which is physically present as the microphone. The impact that recording has can be mitigated in the early stages of your study by making sure that your learners and/or colleagues are clear about why you are collecting the data and that they are confident you will treat it ethically (e.g. by anonymising it). A short time into recording, most participants forget that the microphone is there and so the 'microphone effect' is also mitigated in this way. It is also good practice to ensure that the participants have access to the corpus, or at least the transcribed texts, particularly if you are dealing with data that is sensitive in some way.

4. WHAT CAN BE DONE WITH CORPORA INSIDE AND OUTSIDE THE CLASSROOM?

The only possible answer to the question 'how can teachers use a corpus for their own research?' is, unfortunately, 'it depends on the question', so rather than second-guessing the multitude of questions one could potentially ask, an admittedly highly selective view of the types of questions language teachers have been asking and the kind of uses corpora have been put to can be presented. We can ask instead how corpora have been deployed in the *doing*, *knowing* and *being* of language teaching. Despite a frosty initial reception, corpora have become 'part of the pedagogical landscape' according to Sinclair (2004: 2). Johansson (2009: 40) suggests that the primary areas of relevance to the 'pedagogical landscape', which corpora have permeated to a greater or lesser degree, are in the production of dictionaries, grammars, textbooks and teaching materials as well as syllabus design, classroom activities, testing and basic research. Some areas, such as compiling dictionaries (e.g. the Collins COBUILD dictionary) or the production of grammar reference books (e.g. Biber *et al.* 1999; Mindt 2000; Carter and McCarthy 2006), are more firmly established, while others, particularly the use of corpora in the classroom, as products of teacher choice and discretion are less widespread. Mindt (1996) and Römer (2004a, 2004b, 2005) have identified a mismatch between corpus evidence based on authentic language use and the content of English language textbooks. A notable exception in these terms is the entirely

DRAFT

Details of published version: Vaughan, E. 2010. How can teachers use a corpus for their own research? In: O'Keeffe, A & M. McCarthy (eds.), *The Routledge Handbook of Corpus Linguistics*. London: Routledge, pp. 471-484.

corpus-informed textbook series *Touchstone* (McCarthy *et al.* 2005a, 2005b; see also McCarten, this volume).

In terms of testing and assessment, Taylor and Barker (2008) review the contribution of corpora from the early 1990s on (see also Barker, this volume). Cambridge ESOL, for example, has used information from the BNC and COBUILD corpus as well as the purpose-built Cambridge Learner Corpus to revise the examinations it administers amongst other things. In the USA, the TOEFL 2000 Spoken and Written Academic Language Corpus (T2K-SWAL) corpus has been used to identify patterns of language use in these spoken and written academic contexts and whether the listening and reading components of the TOEFL examination reflected these uses (see Biber *et al.* 2004). Learner corpora are also used in the interrogating the theoretical bases of foreign language teaching. Nesselhauf (2004) and Granger (2009) provide extensive overviews of the potential of learner corpora for the fields of language teaching and second language acquisition research (see chapters by also Guilquin and Granger and Xiaofei Lu, this volume).

Over the last decade or so, the number of studies addressing themselves to the use of corpora in language teaching within and beyond the classroom has proliferated. The studies themselves could usefully be divided into those which deal with data-driven learning (Johns 1986, 1991) involving the direct or mediated use of corpora in the classroom, and studies that use corpora to interrogate the content, theoretical bases and practices of language teaching. In terms of using corpora as resources for language learners, Aston (2001) reports on using the BNC with advanced learners of English as a reference tool when learners come across unfamiliar vocabulary items, such as *blunder* or *hamfisted*, using concordance lines to deduce how and when such items are used, and what contexts they tend to be used in. Bernardini (2001) provides an insightful view of the learner experience of using large corpora for autonomous language learning. Gavioli (2001) discusses how to equip language learners with the skills they need to develop in order to actually benefit from using corpora to learn about language. She also gives examples of the type of graded corpus-based activities she has created using lonely hearts columns, which are very interesting (see also chapters by Gilquin and Granger, Sripicharn, Tribble, this volume). McCarthy (2002) uses

DRAFT

Details of published version: Vaughan, E. 2010. How can teachers use a corpus for their own research? In: O'Keeffe, A & M. McCarthy (eds.), *The Routledge Handbook of Corpus Linguistics*. London: Routledge, pp. 471-484.

5-million-word samples of spoken and written language (from the Cambridge and Nottingham Corpus of Discourse in English and the Cambridge International Corpus) to address the question of what an advanced level vocabulary for English language students might contain and Coxhead (2002) uses an English academic corpus of over 3.5 million words to develop an academic wordlist in response to the needs of learners preparing for academic study (see Coxhead, this volume). That is not to say that large corpora like these are a prerequisite for corpus-based studies, Tribble (2001) uses a 'micro-corpus' of approximately 14,000 words to explore how a particular writing genre is organised and uses this information to develop the linguistic resources that learners will need to exploit in order to write successfully in this genre (see also Tribble, this volume). The genre represented in the corpus is university promotional material for MA programmes in Applied Linguistics and was compiled using the web search engine *Alta Vista*. From the point of view of how teachers can use corpora as a resource to develop their own language awareness, Tsui (2004, 2005) discusses *TeleNex*, a Hong Kong-based website which provides advice for teachers from language specialists and their peers on queries on a range of grammatical and lexical issues, such as discrepancies between prescriptive grammars and authentic usage and queries regarding explaining the usage of synonymous items such as *tall* and *high* to their students. The final aspect of how teachers have been using corpora to carry out research mentioned above, the investigation of language teaching practices and professional research is one that is ripe for expansion: namely, the investigation of teachers' professional language, particularly that which occurs outside the classroom.

The aspects of teacher language that have been prioritised in the existing (not exclusively corpus-based) research is centred, on the whole, on understanding teacher language in connection to how classrooms work, and how the profession considers its practices within them reflexively. In terms of teacher language within the classroom in the L2 context, Walsh (2006) reviews and summarises its major features as follows:

- teachers control patterns of communication in the classroom

DRAFT

Details of published version: Vaughan, E. 2010. How can teachers use a corpus for their own research? In: O'Keeffe, A & M. McCarthy (eds.), *The Routledge Handbook of Corpus Linguistics*. London: Routledge, pp. 471-484.

- the classroom is dominated by question and answer routines
- 'repair' or correction of learner errors is a prerogative of the teacher
- teachers typically modify their speech to accommodate learners

Walsh posits a framework (*Self-Evaluation of Teacher Talk*, or *SETT*) to aid teachers in their description of language used in the L2 classroom context and as a conduit for understanding the complex interactional processes that occur within it (*ibid*: 62-92) (see further reading) and this provides a useful framework for teachers considering using a corpus to research their linguistic practices within the classroom.

Findings from the field of language teacher education (LTE) are uniquely illuminating in terms of the professional concerns of language teachers. Farr (2005a) analyses trainer-trainee interaction in LTE in terms of the types of interventions used in teaching practice feedback and this focus on the discourse of teacher training is obviously an interesting dimension in language teaching research as it provides us with a very important locale of teacher language (see also Farr, this volume). In fact, what we are seeing in the unique interaction that studies such as Farr (2003, 2005a, 2005b), work by Reppen and Vásquez (2007) and Vásquez and Reppen (2007) is trainee teachers captured in the process of *becoming* teachers. The studies referred to here focus on a specific language event in the life of the trainee, feedback meetings on trainee's observed classes, an event which is inherently face-threatening (Reppen and Vásquez 2007: 16) and necessitates deft interpersonal and linguistic negotiation. Vásquez and Reppen's (2007) report on collecting a corpus of post-observation meetings is especially interesting as it illustrates how corpus-based studies can inform and, in the case of this study, transform practices. In their workplace, on an intensive English programme on which their MA in TESL student teachers gain practical ESL teaching experience, the reflective rather than evaluative model of post-observation feedback is favoured. With this in mind, the supervisors/mentors approached the feedback meeting as a discursive space for trainees to reflect on their teaching practices (*ibid*: 159). However, corpus-based analyses of participation patterns

DRAFT

Details of published version: Vaughan, E. 2010. How can teachers use a corpus for their own research? In: O'Keeffe, A & M. McCarthy (eds.), *The Routledge Handbook of Corpus Linguistics*. London: Routledge, pp. 471-484.

indicated that, in practice, the supervisors/mentors did more of the talking than the trainees. This insight led to a change in practices for the supervisors/mentors involved: by increasing the number of questions they asked the trainees, they were able to turn the floor over to the trainees more effectively and give them the tools to use the discursive space for reflection.

5. WHAT CAN A CORPUS TELL US ABOUT THE PROFESSION OF LANGUAGE TEACHING?

The case of C-MELT

This small Corpus of Meetings of English Language Teachers (C-MELT) consists of six meetings in two language teaching institutions, in two different countries, México and Ireland. In all approximately 3.5 hours of interaction was recorded (c. 40,000 words). Its defining characteristic is as the situated language and practices of two local communities of teachers who form part of a (hypothesised) larger, global community. Its underlying purpose was to put teachers in the frame, the rationale being that practices outside the classroom are at least as interesting and just as deserving of research attention as those that occur within it. At the meetings that make up C-MELT, teachers talk about the day-to-day business of teaching: placing students according to ability, examinations/assessment, student attendance and motivation, administrative issues, workplace frustrations etc. Placing C-MELT in the larger teacher language context, if we look at the range of interaction that teachers engage in, it is possible to divide them into interaction that occurs *inside* and *outside* the classroom (see Figure 1 below for some dimensions of teacher language that focus on the face-to-face spoken mode) or characterise them in terms of whether they occur in the professional *frontstage* (classroom) or *backstage* (outside the classroom) (Goffman 1971). Most research on teacher language is conducted with an eye to the classroom, but this research is about what happens unrehearsed in the staffroom when the 'gloves are off'.

DRAFT

Details of published version: Vaughan, E. 2010. How can teachers use a corpus for their own research? In: O'Keeffe, A & M. McCarthy (eds.), *The Routledge Handbook of Corpus Linguistics*. London: Routledge, pp. 471-484.

INSERT FIGURE 1 HERE

Quantitative analyses using the wordlist and concord functions of *WordSmith Tools* (Scott 1999) were carried out and the results were viewed through the prism of the community of practice (after Lave and Wenger 1991; Wenger 1998). It was possible then to isolate and analyse in turn a range of linguistic markers of community and explore them in depth. The wordlist function made it possible to explore aspects of the teachers' shared linguistic and professional repertoire (see Vaughan 2007: 179). Particular shorthand for talking about student ability and negotiating the practice of placing students became evident; extract 1 below shows an example of how teachers placing students (taken from a meeting in an Irish language school) shows how this practice is negotiated.

Extract 1: Student placement meeting

(1) **Siobhán:** **He's not strong**.
(2) **Sally:** Now **he's he's weak in it you know**.
(3) **Siobhán:** Hm.
(4) **Sally:** **The others would be all stronger than him.**
(5) **Niall:** Ali? I had him on Friday.
(6) **Sally:** Yeah did **how did you find him he'd be weak now in that class.**
(7) **Niall:** Yeah I would then I'd suggest maybe.
(8) **Aoife:** **Switch**.
(9) **Niall:** **Swapping** the two of them.

One of the most interesting features of the terms that have become part of the community's shared repertoire is the fact that the language itself is neither highly esoteric in form, nor complex in basic linguistic meaning, but rather encodes highly detailed and entailed professional knowledge.

Information such as laughter, sighing or any other marked extralinguistic behaviour was included in the transcription and the wordlist revealed an unexpected frequency in *laughs*, *laughing* and *laughter*. Humour turned out to be a highly salient, multifunctional, marker of community (Vaughan 2008). It is used to invoke shared knowledge, create and maintain

DRAFT

Details of published version: Vaughan, E. 2010. How can teachers use a corpus for their own research? In: O'Keeffe, A & M. McCarthy (eds.), *The Routledge Handbook of Corpus Linguistics*. London: Routledge, pp. 471-484.

solidarity, resist authority in a socially sanctioned way and downtone potentially face-threatening acts. When the teachers wish to vent frustrations, resist institutional strictures or criticise (or mock) students, the humorous frame provides a way of doing so that will not contravene the professional code. Extract 2 below is a quite typical example of how the teachers in C-MELT use humour.

Extract 2: Student placement meeting
[Note: *ye* is the Irish English form for *you* plural]

(1) **Ciarán:** So anyone to go down?
(2) **Michaela:** No but it would cheer us up a lot if you could tell us when Juan is leaving.
<$E> laughter </$E>
(3) **Siobhán:** That's exactly what I wanted to know.
(4) **Ciarán:** **<$E> laughing </$E>** I'll check that out for ye.
(5) **Siobhán:** Please do **<$E> laughs </$E>**.
(6) **Michaela:** It would make it worth the time.
(7) **Siobhán:** Oh he's unbearable. He's unbearable.

These are just two aspects of community that corpus-based analysis prompted and revealed. Many types of practices, communicative or otherwise, make up the work of language teachers. It would be interesting to see further research of naturally-occurring language, for example, from blogs or informal conversation, which pertain to the backstage practices of being a teacher. This would, to some extent, extend the purview of language teaching beyond the classroom, and provide some life and colour for the picture of the liminal spaces in the language teaching professionals' life. These are the places where the professional mask that is presented to students is put to one side and a new one, used to do the hidden work of teaching, assumed; the backstage spaces where professional successes and failures are discussed, critiqued and laughed about and bonds of community and professional identity forged. More corpus-based research that pushes the boundaries of what linguistic genres are taken to represent the profession of English language teaching would be most welcome and could be a fruitful starting point for practising teachers who want to use a corpus for their own research.

FURTHER READING

DRAFT

Details of published version: Vaughan, E. 2010. How can teachers use a corpus for their own research? In: O'Keeffe, A & M. McCarthy (eds.), *The Routledge Handbook of Corpus Linguistics*. London: Routledge, pp. 471-484.

Aijmer, K. (ed.) (2009) *Corpora and Language Teaching*, Amsterdam: John Benjamins. (The papers in this volume are all geared towards the practical issues involved in using corpora in the classroom and for pedagogical research. It looks at new types of corpora, corpora and second language acquisition and direct and indirect approaches in integrating corpus research into the language classroom.)

O'Keeffe, A., McCarthy, M.J. and Carter, R. (2007) *From Corpus to Classroom: Language Use and Language Teaching*, Cambridge: Cambridge University Press. (This book provides a practical introduction to the discipline of corpus linguistics and its relationship to language teaching using numerous practical examples of the types of linguistic information a corpus can provide. It also contains extensive overviews of previous studies and illustrates corpus insights from a variety of linguistic genres. With regard to the potential of corpus research in relation to teacher language, Chapter 11, *Exploring Teacher Corpora* is of particular interest.)

Sinclair, J.M. (2004) *How to Use Corpora in Language Teaching*, Amsterdam: John Benjamins. (This collection of papers is divided into sections which investigate how teachers can use corpora in the classroom, as a source of information about how language works - including answering tricky student questions – as a resource, for research and processing and programming resources available for working with corpora.)

Walsh, S. (2006) *Investigating Classroom Discourse*, London: Routledge. (This provides an extensive overview of the history and dimensions of classroom language research and thus a solid grounding in the literature of a particularly dense field of study. In addition, it outlines an extremely useful and user-friendly framework within which to analyse teacher talk in the classroom, Self-Evaluation of Teacher Talk (SETT), which has valuable application to corpus-based research into classroom discourse.)

DRAFT

Details of published version: Vaughan, E. 2010. How can teachers use a corpus for their own research? In: O'Keeffe, A & M. McCarthy (eds.), *The Routledge Handbook of Corpus Linguistics*. London: Routledge, pp. 471-484.

REFERENCES

Aarts, J. (1991) "Intuition-based and observation-based grammars," in K. Aijmer and B. Altenberg (eds.), *English Corpus Linguistics. Studies in Honour of Jan Svartvik*, London: Longman, pp. 44-62.

Aston, G. (2002) "Getting one's teeth into a corpus," in M. Tan (ed.) *Corpus Studies in the Language Classroom*, Bangkok: IELE Press, pp. 131-143.

Biber, D. (1990) "Methodological issues regarding corpus-based analyses of linguistic variation," *Literary and Linguistic Computing*, 5(4): 257-269.

Biber, D. (1993) "Representativeness in corpus design," *Literary and Linguistic Computing*, 8(4): 243-257.

Biber, D., Conrad, S. and Reppen, R. (1998) *Corpus Linguistics: Investigating Language Structure and Use*, Cambridge: Cambridge University Press.

Biber, D., Johansson, S., Leech, G., Conrad, S. and E. Finegan (1999) *Longman Grammar of Spoken and Written English*, Harlow: Pearson.

Biber, D., Conrad, S., Reppen, R., Byrd, P., Helt, M., Clark, V., Cortes, V., Csomay, E. and Urzua, A. (2004) "Representing Language Use in the University: Analysis of the TOEFL 2000 Spoken and Written Academic Language Corpus," *Report Number: RM-04-03, Supplemental Report Number: TOEFL-MS-25*, Princeton, NJ.: Educational Testing Service. (Available online at URL: https://www.ets.org/Media/Research/pdf/RM-04-03.pdf)

DRAFT

Details of published version: Vaughan, E. 2010. How can teachers use a corpus for their own research? In: O'Keeffe, A & M. McCarthy (eds.), *The Routledge Handbook of Corpus Linguistics*. London: Routledge, pp. 471-484.

Bernardini, S. (2001) "'Spoilt for choice': A learner explores general language corpora," in G. Aston (ed.) *Learning with Corpora*, Houston: Athelstan, pp. 220-249.

Breyer, Y. (2009) "Learning and teaching with corpora: Reflections by student teachers," *Computer Assisted Language Learning*, 22(2): 153-172.

Carter, R. and McCarthy, M.J. (2002) "From conversation to corpus: A dual analysis of a broadcast political interview," in A. Sanchez-Macarro (ed.).*Windows on the World: Media Discourse in English*, Valencia: University of Valencia Press, pp. 15-39.

Carter, R. and McCarthy, M.J. (2006) *Cambridge Grammar of English*, Cambridge: Cambridge University Press.

Clear, J. (1992) "Corpus sampling," in G. Leitner (ed.) *New Directions in English Language Corpus Methodology*, Berlin: Mouton de Gruyter, pp. 21-31.

Conrad, S.M. (1999) "The importance of corpus-based research for language teachers," *System*, 27(1): 1-18.

Coxhead, A. (2002) "The academic word list: A corpus-based word list for academic purposes," in B. Kettemann and G. Marko (eds.) *Teaching and Learning by Doing Corpus Analysis*, Amsterdam: Rodopi, pp. 73-89.

Crystal, D. (1997) *The Cambridge Encyclopedia of Language*, Cambridge: Cambridge University Press.

Dörnyei, Z. (2007) *Research Methods in Applied Linguistics*, Oxford: Oxford University Press.

DRAFT

Details of published version: Vaughan, E. 2010. How can teachers use a corpus for their own research? In: O'Keeffe, A & M. McCarthy (eds.), *The Routledge Handbook of Corpus Linguistics*. London: Routledge, pp. 471-484.

Farr, F. (2003) "Engaged listenership in spoken academic discourse," *Journal of English for Academic Purposes*, 2(1): 67-85.

Farr, F. (2005a) "Reflecting on reflections: the spoken word as a professional development tool in language teacher education," in R. Hughes (ed.) *Spoken English, Applied Linguistics and TESOL: Challenges for Theory and Practice*, Hampshire: Palgrave, Macmillan, pp. 182-215.

Farr, F. (2005b) "Relational strategies in the discourse of professional performance review in an Irish academic environment: the case of language teacher education," in K. P. Schneider and A. Barron (eds.) *Variational Pragmatics: The Case of English in Ireland*, Berlin: Mouton de Gruyter, pp. 203-234.

Francis, W.N. (1982) "Problems of assembling and computerising large corpora," in S. Johansson (ed.) *Computer Corpora in English Language Research*, Bergen: Norwegian Computing Centre for the Humanities, pp. 7-24.

Francis, W.N. (1992) "Language corpora B.C.," in J. Svartvik (ed.) *Directions in Corpus Linguistics. Proceedings of Nobel Symposium 82, Stockholm 4-8 August 1991*, Berlin: Mouton de Gruyter, pp. 17-32.

Frankenberg, A. (forthcoming, 2009) "Raising Teachers' Awareness to Corpora," in N. Kübler (ed.) *Corpora, Language, Teaching and Resources: From Theory to Practice*, Frankfurt: Peter Lang.

Gavioli, L. (2001) "The learner as researcher: Introducing corpus concordancing in the classroom," in G. Aston (ed.) *Learning with Corpora*, Houston: Athelstan, pp. 108-137.

Goffman, E. (1971) *The Presentation of Self in Everyday Life*, Harmondsworth: Penguin.

DRAFT

Details of published version: Vaughan, E. 2010. How can teachers use a corpus for their own research? In: O'Keeffe, A & M. McCarthy (eds.), *The Routledge Handbook of Corpus Linguistics*. London: Routledge, pp. 471-484.

Granath, S. (2009) "Who benefits from learning how to use corpora?" in K. Aijmer (ed.) *Corpora and Language Teaching*, Amsterdam: John Benjamins, pp. 47-65.

Granger, S. (2009) "The contribution of learner corpora to second language acquisition and foreign language teaching," in K. Aijmer (ed.) *Corpora and Language Teaching*, Amsterdam: John Benjamins, pp. 13-32.

Johansson, S. (2009) "Some thoughts on corpora and second language acquisition," in K. Aijmer (ed.) *Corpora and Language Teaching*, Amsterdam: John Benjamins, pp. 33-44.

Johns, T. (1986) "Microconcord: A language learner's research tool," *System*, 14(2): 151-162.

Johns, T. 1991. "Should you be persuaded – two samples of data-driven learning materials," in T. Johns and P. King (eds.) *Classroom Concordancing*, English Language Research 4, University of Birmingham, pp. 1-16.

Lave, J. and Wenger, E. (1991) *Situated Learning: Legitimate Peripheral Participation*, Cambridge: Cambridge University Press.

McCarthy, M.J. (2002) "What is an advanced level vocabulary?," in M. Tan (ed.) *Corpus Studies in the Language Classroom*, Bangkok: IELE Press, pp. 15-29.

McCarthy, M.J. (2008) "Accessing and interpreting corpus information in the teacher education context," *Language Teaching*, 41(4): 563-574.

McCarthy, M.J., McCarten, J. and Sandiford, H. (2005a) *Touchstone. Student's Book 1*, Cambridge: Cambridge University Press.

DRAFT

Details of published version: Vaughan, E. 2010. How can teachers use a corpus for their own research? In: O'Keeffe, A & M. McCarthy (eds.), *The Routledge Handbook of Corpus Linguistics*. London: Routledge, pp. 471-484.

McCarthy, M.J., McCarten, J. and Sandiford, H. (2005b) *Touchstone. Student's Book 2*, Cambridge: Cambridge University Press.

McCarthy, M.J., O'Keeffe, A. and Walsh, S. (2007) *An Introduction to Corpora in English Language Teaching. ELT Advantage: Online courses for teachers*, Boston: Heinle Cengage, http://eltadvantage.ed2go.com/eltadvantage/index.html

Mindt, D. (1996) "English corpus linguistics and the foreign language teaching syllabus," in J. Thomas and M. Short (eds.) *Using Corpora for Language Research*, London: Longman, pp. 232-247.

Mindt, D. (2000) *An Empirical Grammar of the English Verb System*, Berlin: Cornelsen.

Mukharjee, J. (2002) *Korpuslinguistik und Englischunterricht: Eine Einführung (Sprache im Kontext series 14)*, Frankfurt: Peter Lang.

O'Keeffe, A. and Farr, F. (2003) "Using language corpora in initial teacher education: Pedagogic issues and practical applications," *TESOL Quarterly*, 37(3): 389-418.

Nesselhauf, N. (2004) "Learner corpora and their potential for language teaching," in J.M. Sinclair (ed.) *How to Use Corpora in Language Teaching*, Amsterdam: John Benjamins, pp. 125-152.

O'Keeffe, A., McCarthy, M.J. and Carter, R. (2007) *From Corpus to Classroom: Language Use and Language Teaching*, Cambridge: Cambridge University Press.

Römer, U. (2004a) "Comparing real and ideal language learner input: The use of an EFL textbook corpus in corpus linguistics and language teaching," in G. Aston, S. Bernardini and D. Stewart (eds.) *Corpora and Language Learners*, Amsterdam: John Benjamins, pp. 151-168.

DRAFT

Details of published version: Vaughan, E. 2010. How can teachers use a corpus for their own research? In: O'Keeffe, A & M. McCarthy (eds.), *The Routledge Handbook of Corpus Linguistics*. London: Routledge, pp. 471-484.

Römer, U. (2004b) "A corpus-driven approach to modal auxiliaries and their didactics," in J.M. Sinclair (ed.) *How to Use Corpora in Language Teaching*, Amsterdam: John Benjamins, pp. 185-199.

Römer, U. (2006) "Pedagogical applications of corpora: Some reflections on the current scope and a wish for future developments," *Zeitschrift für Anglistik und Amerikanistik*, 54(2): 121-134.

Römer, U. (2009) "Corpus research and practice," in K. Aijmer (ed.) *Corpora and Language Teaching*, Amsterdam: John Benjamins, pp. 83-98.

Reppen, R. and C. Vásquez, (2007) "Using corpus linguistics to investigate the language of teacher training," in J. Waliński, K. Kredens and S. Goźdź-Roszkowski (eds.), *Corpora and ICT in Language Studies*, PALC 2005, Frankfurt am Main: Peter Lang, pp. 13-29.

Scott, M. (1999) *WordSmith Tools Version 3*, Oxford: Oxford University Press.

Sinclair, J.M. (1991) *Corpus, Concordance, Collocation*, Oxford: Oxford University Press.

Sinclair, J.M. (2001) "Preface," to M. Ghadessy, A. Henry and R.L. Roseberry (eds.), *Small Corpus Studies and ELT*, Amsterdam: John Benjamins, pp. vii-xv.

Taylor, L. and Barker, F. (2008) "Using corpora for language assessment," in E. Shohamy and N.H. Hornberger (eds.) *Encyclopedia of Language and Education (2nd Edition)*, (available online at URL:
http://www.springerlink.com/content/x1434l3131507322/fulltext.pdf)

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

DRAFT

Details of published version: Vaughan, E. 2010. How can teachers use a corpus for their own research? In: O'Keeffe, A & M. McCarthy (eds.), *The Routledge Handbook of Corpus Linguistics*. London: Routledge, pp. 471-484.

Tsui, A.B.M. (2004) "What teachers have always wanted to know – and how corpora can help them," in J.M. Sinclair (ed.) *How to Use Corpora in Language Teaching*, Amsterdam: John Benjamins, pp. 39-61.

Tsui, A.B.M. (2005) "ESL teachers' questions and corpus evidence," *International Journal of Corpus Linguistics*, 10(3): 335-356.

Vásquez, C. and Reppen, R. (2007) "Transforming practice: Changing patterns of interaction in post-observation meetings," *Language Awareness*, 16(3): 153-172.

Vaughan, E. (2007) "I think we should just accept our horrible lowly status': Analysing teacher-teacher talk in the context of community of practice," *Language Awareness*, 16(3): 173-189.

Vaughan, E. (2008) "Got a date or something?: An analysis of the role of humour and laughter in the workplace," in A. Ädel and R. Reppen (eds.) *Corpora and Discourse*, Amsterdam: John Benjamins, pp. 95-115.

Walsh, S. and O'Keeffe, A. (2007) "Applying CA to a modes analysis of higher education spoken academic discourse," in H. Bowles and P. Seedhouse (eds.) *Conversation Analysis and Language for Specific Purposes*, Berlin: Peter Lang, pp. 101-139.

Waters, A. (2005) "Expertise in Teacher education: Helping teachers to learn," in K. Johnson (ed.) *Expertise in Second Language Learning and Teaching*, Basingstoke: Palgrave Macmillan, pp. 210-229.

Wenger, E. (1998) *Communities of Practice: Learning, Meaning, and Identity*, Cambridge: Cambridge University Press.

DRAFT

Details of published version: Vaughan, E. 2010. How can teachers use a corpus for their own research? In: O'Keeffe, A & M. McCarthy (eds.), *The Routledge Handbook of Corpus Linguistics*. London: Routledge, pp. 471-484.

DRAFT

Details of published version: Vaughan, E. 2010. How can teachers use a corpus for their own research? In: O'Keeffe, A & M. McCarthy (eds.), *The Routledge Handbook of Corpus Linguistics*. London: Routledge, pp. 471-484.

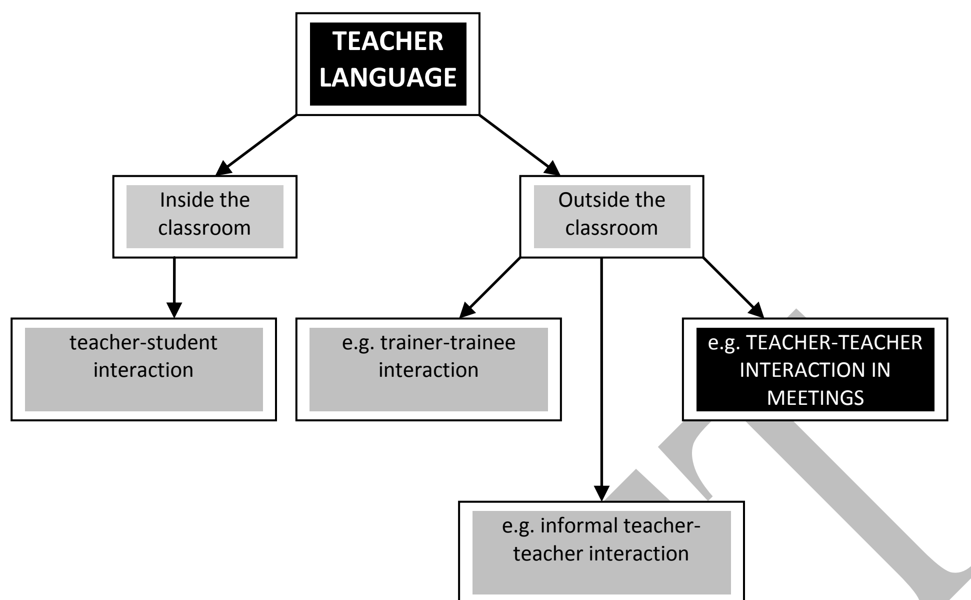

Figure 1 Dimensions of teacher language

DRAFT