33

How can teachers use a corpus for their own research?

Elaine Vaughan

1 Using a corpus for your own research: being professionally curious

This chapter is aimed at teachers who are interested in conducting their own professional and/or pedagogical research and want to explore the possibilities of using a corpus approach in this regard. Practising language teachers engage in and/or with research (cf. Borg 2013), including corpus research, for a wide variety of reasons. Motivation for this engagement may arise from challenges or opportunities in the immediate teaching and learning environment, personal or academic interests, interaction with colleagues or as 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. The potential range of research that curiosity might generate is hard to predict, given that there is much to be done, and not all of what is being done finds its way into the public domain. However, professional engagement in research (of whatever type) is critical to quality research of benefit to the practice and profession of second language teaching and learning. The nexus of teaching and research has at different times been problematised as imbalanced and top-down, unsustainable or bearing little practical value for the language classroom (see Medgyes 2017; Paran 2017), although the classroom is not necessarily, nor should it be, the sole end of any "teaching-related" research. Discussion of the relationship between teaching and research does often pivot around a perceived nexus of roles, hierarchies and priorities (e.g. Xerri 2017). There are deeper issues of access to research (Sato and Loewen 2019), its intrinsic usefulness and whether or not it actually reflects or informs the profession or practice (McKinley 2019). Early advocates of corpus linguistics in second language pedagogy emphasised integration of corpus-based research into teachers' professional repertoires and practices (e.g. Fligelstone 1993; Coniam 1997; Leech 1997). In tandem with this, *corpus literacy*, the ability to access, exploit and interpret corpus output, has been advocated (Mukherjee 2006; Frankenberg-Garcia 2012; Callies 2019). The importance of novice teachers' and experienced practitioners' participation in corpus-based research is also a consistent refrain over time (e.g. O'Keeffe and Farr 2003; Tsui 2005; McCarthy 2008; Breyer 2009; Römer 2009, 2011; Farr 2011; Timmis 2015).

DOI: 10.4324/9780367076399-33

However, this has not necessarily translated into widespread provision of corpus linguistics in language teacher education and training programmes (O'Keeffe and Farr 2003; McCarthy 2008; Granath 2009; O'Keeffe and Farr 2019; and Chapter 32, this volume). Nor, indeed, has it translated into global awareness of corpus research and methods within language teaching as a profession. Clearly, despite regular discussion of the "potential" of corpus linguistics vis-à-vis language teaching and learning (see Timmis 2015: 7–8), language teaching professionals' ownership of and engagement in corpus research is still a relatively elusive factor. The selection of work surveyed later in this chapter focuses broadly on corpus research that intersects with key areas of second language teaching and learning and then zones in on professional interaction on the spoken to written spectrum, including work which focuses on language use *inside* the classroom, the context that most pervades the published literature. *Outside* the classroom, the lion's share of work takes place in developmental contexts such as language teacher education (LTE). This chapter advocates work that considers professional language in the workplace, a site of research that language teaching professionals are uniquely positioned to conduct as individuals and teams. This can be achieved through a combination of professional engagement and a desire to stand back and survey the evidence objectively – or professional curiosity allied with a research perspective.

There is no scarcity of published work which deals specifically with the relationship between corpora, corpus linguistics, language research, pedagogy and education (see McEnery and Xiao 2011 or Vaughan and McCarthy 2017). This work defines and describes types of corpora, demonstrating the direct and indirect applications corpora may have for the classroom and language-based research in the sphere of second language teaching and learning (see Flowerdew 2012 for a book-length treatment). The present chapter offers in its contribution to this plethora an outline of the various contexts of corpus research in second language teaching and learning and contributes a perspective on how teachers, researchers and teacher-researchers can and have been using corpora for their own research. One of the major benefits of corpus-based based research for language teachers lies in its potential as a teacher development tool (e.g. O'Keeffe and Farr 2003; O'Keeffe *et al.* 2007; Farr 2011). To conceptualise how, Waters's (2005) characterisation of the tripartite elements within his preferred superordinate term "teacher education" is useful: teacher *training*, teacher *education* and teacher *development*. Teacher training, or perhaps more appropriately "preparation" (see Mann 2005), concerns itself with the practical, for example, classroom-based skills. Teacher education is more connected to research and background knowledge, while teacher development has a focus on raising awareness of practices and fostering reflection and change (Waters 2005: 211). He characterises these three interconnected aspects 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. Corpus-based research can be a means of investigating the *doing*, *knowing* and *being* of teaching and learning. In this regard, 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. This chapter also suggests that those of us engaged in researching professional contexts and practices reconsider the limits of our purview and what teaching-related research might logically encompass. It suggests one instance of where we might extend these boundaries, presenting an excerpt from a case study of teacher professional life, not within the classroom or language education context, but within the professional

workplace. This is a context that is relatively unexplored by comparison, but can yield rich insights on one of the (many) 'occluded genres' (Swales 1996: 46) of the profession of English language teaching.

2 Issues in using corpora for educational research

The provenance of corpus linguistics as a research field is mapped out more comprehensively by McCarthy and O'Keeffe (2010 and Chapter 1, this volume), but anywhere an initial definition is presented, the preliminary step is usually to define what a corpus is (and, of course, indicate the clunky plural form, *corpora*). This is because corpus linguistics is 'ultimately about *finding out about the nature and usage of language*' (McEnery and Hardie 2012: 228; italics in original), via the construction/investigation of language corpora using a range of corpus analysis techniques (see Vaughan and O'Keeffe 2015 and Chapter 9, this volume). All corpora are collections of texts of various types, but are characterised by purpose and design; the use of corpus software to explore features and patterning; and, critically, comparison with other corpora. Most descriptions of defining characteristics of a corpus make mention of corpus size. Indeed, the issue of how large a corpus should be, or even how small a corpus *can* be, has been a recurrent theme (see Vaughan and Clancy 2013). There has never been an upper limit on corpora, particularly those designed to inform grammar reference works or for use in lexicography. The *Cambridge International Corpus* (CIC), contains over a billion words of spoken and written texts and has been used to inform Cambridge dictionaries, but also English language course books, vocabulary builders and grammar reference works, such as *The Cambridge Grammar of English* (Carter and McCarthy 2006). Many of the corpora available via the online corpus management platform *Sketch Engine* run to billions of words. Clearly, to inform reference works, a corpus should be as large as possible, and the same holds true for corpora that are designed to represent national varieties. However, capturing the totality of any language variety, of whatever specific type (see Chapter 6, this volume), is a virtual impossibility: that would mean capturing all of language use, all the time, everywhere. The challenge, then, is to sample from an impossible-to-define population, an appropriate, *representative* amount.

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 a more salient issue than corpus size where corpora are designed to produce generalisable findings (see Chapters 2–6, this volume). Language data have proven resistant to standard approaches to statistical sampling (Clear 1992: 21) and sampling frames, and so Biber proposes strata and sampling frames suitable for language 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 (see Chapters 2–6, this volume). With regard to sample size, Biber's previous research on 1,000-word samples from the *London-Lund and Lancaster/Oslo/Bergen* corpus had concluded that these relatively small samples yielded similar functional and grammatical findings (Biber 1990). These connected points raise questions that are important for teachers planning to actually build corpora for their own research, though it should be emphasised that it is not necessarily essential to build a corpus. Whether or not that happens is entirely dependent on the nature of the research being planned. For many, it may be possible to do independent research on particular language features using existing corpora, some of which are accessible via user-friendly online search platforms, such as the *Corpus of*

Contemporary American English (COCA) (see Davies 2010), and produce more generalisable results. For most *small corpus studies*, the best way to describe many of those discussed later, results are not strictly generalisable, and while that should always be acknowledged, in many ways they actually do not need to be, given the equal value placed on generating context-dependent knowledge in the broader fields of language description. It is more important to be able to establish whether or not a broad or specific finding that emerges from a small corpus study might be particular to the very specific context it describes or hold across other examples of that context type via comparison with published results, for example.

In terms of consulting existing corpora, one obvious obstacle is corpus literacy (discussed later). Other obstacles may be access to resources such as basic corpus training, online corpus platforms, free-standing corpus software and computer labs for use with students, amongst others. These are rarely emphasised across a literature that, on the whole, assumes its audience is the university-based language teacher/pre-service trainee (Frankenberg-Garcia 2016). While some of these resources are a prerequisite for any form of research and hands-on activities with language learners, the lack of some of these types of facilities in the immediate environment does not preclude engagement with corpus-based research for pedagogical/professional activity. These concerns aside, there are a number of essential skills practitioners new to using corpora need to develop. These include having a broad overview of the range and types of corpora that exist and are freely and commercially accessible (see Chapter 8, this volume) in order to develop awareness of what these corpora can, and cannot, offer. The vast majority of accessible, large, general-purpose corpora will be useful for generating examples of particular structures and features that may be the language focus in a class. This is a common reason for corpus consultation, which might not be considered as formal research per se but is a good way of becoming familiar with corpora.

Being able to critically evaluate 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 early-stage practitioner-researcher. Corpus literacy, or the ability to generate and frame corpus queries, as well as interpret results is essential. From a research perspective, understanding how corpora are compiled, formulating appropriate corpus queries and being aware of the impact of corpus size and composition on output related to these queries are skills as critical as mastering the mechanics of using specific software to explore corpora. For the mechanics and beyond, the Lancaster University MOOC *Corpus Linguistics: Method, Analysis, Interpretation*, hosted by *Future Learn* at the time of writing, is an excellent resource. Investigating language use questions, for example, establishing the lexical characteristics of written academic texts like journal articles, or queries around most frequent phrasal verbs in informal conversation, is likely to be possible using existing corpora. However, teachers about to embark upon corpus research – the readers of this chapter – are likely to have framed research questions linked to their professional practices or domains, for which there may not be an existing corpus that will answer these questions adequately. Therefore, the next section deals specifically with research that is predicated on gathering naturally occurring samples of professional language use and focuses on aspects of building a bespoke corpus of language data for this purpose that this entails.

3 Building your own corpus based on your practice

Among the many advantages that practising teachers have when it comes to investigating professional and pedagogical questions is that their position in teaching institutions means that the relevant participants and data are within immediate proximity. Adopting a corpus approach is, naturally, only one of the many methodological approaches that might be appropriate to the investigation of questions immediate to the profession and practice of language teaching. It is particularly useful if the research is based on the investigation of spoken or written interaction, which can be captured and contained electronically. Corpus methods can coexist harmoniously with many paradigms of research and can be used in tandem with complementary discourse analytic methods, such as conversation analysis (e.g. Walsh *et al.* 2011). This means that the contents of a small corpus can be explored in a more detailed way once corpus analytical procedures have yielded an overview of patterns in the data (see Chapters 2–4, this volume). There are some specific points to consider, related to the macro- and micro-ethics of doing research in the second language teaching and learning environment (cf. Kubanyiova 2008), especially one's own, which should be mentioned here. The macro-ethics are procedural, usually codified by institutions and organisations. These generally outline professional conduct around approaching access to sites of research, ensuring informed consent is gained, and provide guidelines around using and protecting data. Micro-ethics are of particular concern for situated research of the type that we move towards here. These are closer, context-specific issues, such as our own position vis-à-vis the research, our research values and other aspects of working ethically with the participants involved, where the participants are typically our colleagues and/or students.

Once access has been negotiated, the next step is securing consent to record/collect, use and safely store data. This means that students' or colleagues' 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 more likely to engage if teacher-researchers are up-front about why the interaction or language materials are necessary for the research and are clear about the purpose of the research itself. This is even more the case if it is going to address a practical problem or highlight good practices (see Healy and Onderdonk 2013). Informed consent should be obtained at the beginning of the research, and the learners or colleagues who cooperate in the research should be confident that their data will be treated ethically, understand how the data will be used and disseminated, for example, in presentations publications or as a contribution to a larger project (a thorough consent form will cover all these possibilities). Where the research community at large is concerned, the corpus design should be carefully considered, fully described and its contextual information preserved in order to ensure that it is replicable (information about the data is commonly referred to as "metadata"; see Chapters 2–4, this volume). As a rule of thumb, another researcher should be able to add to the corpus or build a companion corpus by adhering to the design principles. This creates the possibility for what is originally a context-specific corpus research project to become an organic, community endeavour.

If the corpus is to contain spoken texts, then sensitive enough recording devices, a consistent system of transcription and secure storage of sound files will be important (see Chapter 3, this volume). Transcription is a slow process; McCarthy (2008: 571) estimates that 1 hour of talk can take 12 to 15 hours to transcribe, and that process becomes more complex depending on the number of speakers (see Chapter 3, this volume). Despite its

labour-intensive nature, the transcription of spoken data by the researcher conducting the study brings serious advantages. For one thing, it augments familiarity with the data the corpus contains, and, as transcription proceeds, particular language features which will be interesting to look into in greater detail can be identified for preliminary analysis. While written text sources, especially those derived via web scraping, may not require the same type of processing, the same type of attention is required. If, for example, a corpus of student essays or other written work is not submitted electronically, these will also need to be transliterated, and the same sort of familiarity and insight proceeds as with spoken transcription. Transcription itself should be sensitive, consistent and principled (see Jenks 2011).

Two final points should be made on the building of corpora which are implicated in the third of the issues outlined earlier, the position of the researcher and the broader micro-ethics of this type of research. When teachers create their own spoken corpora, they are often 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-analyst can be extraordinarily positive, but it is also important to acknowledge the potential biases this dual role may bring. In setting out to record authentic, spoken interaction, an oft-mentioned Catch-22–like situation occurs; inevitably, a degree of artificiality is introduced into a previously authentic situation, which is physically present as the recording device (no matter how discreet it is). Any negative impact that recording has on the comfort and behaviour of participants can be mitigated in the early stages of the study by ensuring that participants are clear about the purpose of the study and why it is useful to do. As previously mentioned, informed consent is a fundamental consideration, and participants should be confident that their contributions will be treated ethically, for example, by protecting the anonymity of participants thoroughly, changing names, place names and any other features in the transcript that could identify the precise location or participants. A short time into recording, most participants forget that the microphone is there and so the "microphone effect" is by and large mitigated. It would be responsible practice to ensure that the participants have access to the corpus, or at least the transcribed texts, for review, approval, use in instruction or their own research. This should be discussed and agreed in advance and is not without its own caveats. While institutions and organisations have their own ethical guidelines, for teachers doing research with corpora outside of research institutions, the British Association for Applied Linguistics (BAAL) *Recommendations on Good Practice in Applied Linguistics Research* (2016) is a good resource. De Costa (2015) is an insightful guide on the micro-ethical issues that arise in conducting educational research as a professional insider/institutional outsider doing research.

4 What can be done with corpora inside and outside the classroom?

Ultimately, the only possible answer to the question "how can teachers use a corpus for their own research?" is "it depends on the question". We might ask instead how corpora have been deployed in investigating the *doing*, *knowing* and *being* of language teaching and considering what sort of insights a corpus approach have already been disseminated. Discussions of the relevance of corpus studies for educational domains tend to distinguish between different points of impact, both indirect and direct (Leech 1997; Johansson 2009; Römer 2009, 2011; Flowerdew 2012; Timmis 2015). Indirect applications are seen as

reference works, materials and the design and validation of language tests, for example; applications construed as direct include students using corpora in the classroom, by directly accessing corpora guided by teachers conversant in their pedagogical application. Some areas, such as compiling dictionaries or the production of grammar reference books (e.g. Biber *et al.* 1999; 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 arguably less widespread (see Timmis 2015). Where testing and assessment are concerned, corpora have informed major reviews and evaluation of language tests as well as informing current practices in developing and validating tests (Barker 2014; Park 2014 and Chapter 45, this volume). A major project, English Profile, which has a core aim of developing and substantiating the reference level descriptions of L2 users of English based on the Common European Framework of Reference (CEFR), uses a subset of the *Cambridge Learner Corpus* containing learner language produced under examination conditions (see Saville and Hawkey 2010 for an overview) (see Chapter 22, this volume).

Early work by Mindt (1996) identified a mismatch between corpus evidence based on natural contexts of language use and the content of English language textbooks; Meunier and Gouverneur (2009) provide an overview of studies of this type, and Chapter 26, this volume, looks at the applications of contemporary corpus research in designing and developing coursebooks. Researchers are interested in investigating learner language and compiling large learner corpora to do so in the consideration of the relationship between materials used to teach and what could be prioritised in language teaching (Meunier and Reppen 2015; Paquot 2018). Learner corpora are used to investigate the practical and theoretical bases of foreign language teaching, and learner corpus research has been one of the more rapidly developing independent fields of research at the intersection of corpus linguistics and language education (see Granger *et al.* 2015; Brzezina and Flowerdew 2017). From a language education perspective, edited volumes such as Sinclair (2004) and Aijmer (2009) present case studies of how corpora can be used directly and indirectly in the classroom by informing pedagogical practices and priorities, but also, as with language testing, to add to theory development in, for example, second language acquisition research (see also Seedhouse 2019 and Chapter 23, this volume). Chambers (2019) focuses on closing the research–practice gap between studies that report on the direct or mediated use of corpora in the classroom, studies that use corpora to interrogate the content, theoretical bases and practices of language teaching and the criticality of language educators who are not corpus linguists actively engaging with corpora data in the classroom (O'Keeffe and Mark 2017; O'Keeffe 2020; see also Chapter 29, this volume).

Where teacher professional language is concerned, the aspects of teacher language that have been prioritised in the existing (not exclusively corpus-based) research is generally centred on understanding teacher language in connection to how classrooms work and how the profession considers its practices within them reflexively. Walsh (2006, 2011, 2013) shows that teacher language in the L2 context is typified by 1) teachers controlling patterns of communication, 2) question and answer routines, 3) "repair" or correction of learner errors being the prerogative of the teacher and 4) modification of teacher speech to accommodate learners. He posits a framework (*Self-Evaluation of Teacher Talk*, or *SETT*) to facilitate the description of language in the classroom and as a conduit for understanding the complex interactional processes that occur within it (*ibid*: 62–92; see also Sert 2019). While SETT developed as an applied conversation analytic framework, it provides an excellent discourse-level structure for teachers using a corpus to research classroom discourse. Elements of the framework and the potential linguistic proxies associated with them can be validated and

explored. For example, Yang (2014) blends SETT and corpus linguistics to investigate teacher's use of discourse markers in the classroom.

Findings from the field of LTE have illuminated the concerns of language teachers in training. For example, (corpus) analyses of post-observation interaction in the LTE context have led to unique insights into its characteristics, the discursive construction of supervisory interventions and how they are responded to, and this has enhanced pedagogical practices developed around them (Vásquez 2004; Vásquez and Reppen 2007; Copland 2008; Farr 2011; Farr *et al.* 2019). This specific language event in the life of the trainee – feedback meetings on trainee's observed classes – is inherently face-threatening and necessitates deft interpersonal and linguistic negotiation. Corpus-based work on discourse produced in contexts of initial teacher education, such as teaching practice, post-observation feedback and reflective blogs produced during third-level courses also provide important insights into specific elements of the trajectories of professional development with participants that are in the process of *becoming* teachers (see Farr *et al.* 2019 for a comprehensive overview of these LTE contexts and Chapter 32, this volume). The final aspect of how teachers have been using corpora to carry out research such as in the broad categories outlined earlier, 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.

5 What can a corpus tell us about a professional language teaching context? The case of C-MELT

This study we discuss here aimed to investigate teacher talk in the workplace and required recourse to a corpus which contains samples of that language in use. As this very specific corpus did not exist, it was necessary to build one. The *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 were recorded (c. 40,000 words). It therefore represents a sample of the situated language use and practices of two local communities of teachers who form part of a hypothesised larger, global community (Vaughan 2007). Its underlying purpose was to put practising 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. In 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. To place C-MELT in the larger context of teacher professional interaction, some of the contexts of language use *inside* and *outside* the classroom are outlined in Figure 33.1. This should come with the caveat that, in the first instance, professional language use happens in multiple modes and within each will be overlaps and blurred boundaries; in addition, encompassing all of the suggested contexts could be digital environments, as well as within each element of the schematic. Another way of characterising the contexts is to consider whether they occur in the professional *frontstage* (classroom) or *backstage* (outside the classroom) (Goffman 1971; this is a blunt division, and the practical rationale for making it in this way can be found in Vaughan 2007). 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". This was the crux of the larger study which emerged from close analysis of C-MELT.

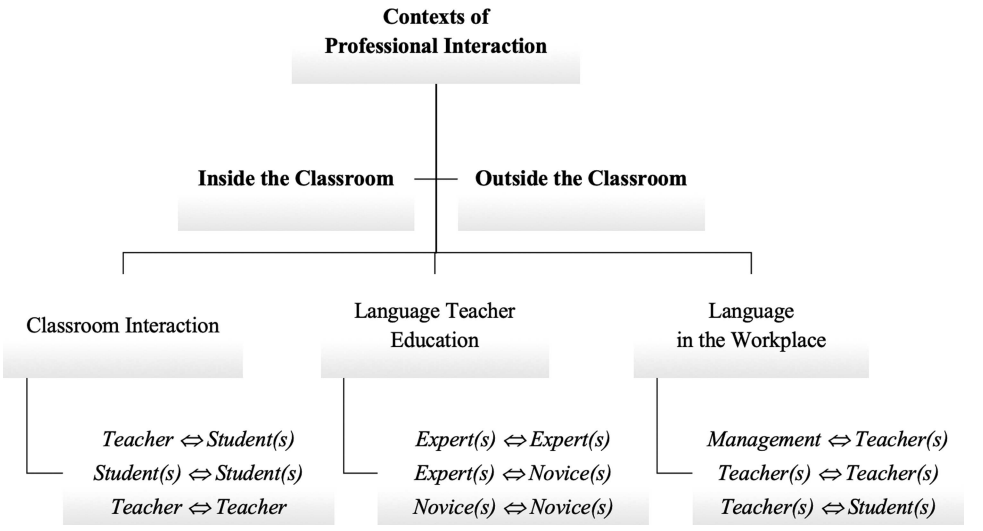

Figure 33.1 Some contexts of professional interaction

The output of corpus analysis tools such as wordlists, clusters, keywords and concordancing (see Chapters 9, 10 and 14, this volume) were allied with the aim of investigating what linguistic markers of community and identity could be extrapolated from them. Lists of words and clusters on their own compared with larger corpora meant that sometimes "small" items carrying multiple meanings were working in specialised ways in the interaction, such as highly contextualised referential aspects of pronouns such as *you*, *we* and *they* (see Vaughan and Clancy 2013). As the corpus was small, it was possible to explore specific aspects of the teachers' shared linguistic and professional repertoire (see Vaughan 2007: 179) and go in-depth where particular practices associated with language teaching came to the fore. Particular shorthand for talking about student ability and negotiating which classes students enrolling in the school should be placed in is one example of the types of practices that were particularly salient; extract 1 next, taken from a meeting in an Irish language school, shows an example of how this practice is negotiated in the interaction.

(1)
SIOBHÁN: He's not strong.
SALLY: Now he's he's weak in it you know.
SIOBHÁN: Hm.
SALLY: The others would be all stronger than him.
NIALL: Ali? I had him on Friday.
SALLY: Yeah did how did you find him he'd be weak now in that class.
NIALL: Yeah I would then I'd suggest maybe.
AOIFE: Switch.
NIALL: Swapping the two of them.

[C-MELT04: Sub-corpus 2: Ireland]

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 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 is a quite typical example of how the teachers in C-MELT use humour (note that *ye* is used in Irish English to denote *you* plural and <$E>…<\$E> is used to mark extralinguistic information, such as laughter, or laughing as part of a turn, in this case).

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

[C-MELT05: Sub-corpus 2: Ireland]

These are just two aspects of community that corpus-based analysis prompted and revealed. It would be interesting to see corpus research that focuses on, for example, more informal staffroom talk, or any context which pertains to the backstage practices of being a teacher (for a conversation analytic perspective, see Richards 2006). The interaction taking place on public fora online is a rich resource for the sorts of stances new and experienced teachers take up with relation to their profession, some of which is undeniably gloves off (e.g. a 2013 post on one forum stating 'work in a language school and it's death by a 1000 cuts to your soul'). This would, to some extent, extend our understanding 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. O'Keeffe *et al.* (2007: 246) have underlined the need for reciprocity in the relationship between language researchers and language teachers and have claimed 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'. Teachers defining the scope of the agenda for research based on their profession and, above all, using corpora for their own research will be the decisive factor for this in the future.

Further reading

O'Keeffe, A., McCarthy, M. J. and Carter, R. A. (2007) *From Corpus to Classroom: Language Use and Language Teaching*, Cambridge: Cambridge University Press. (This practical introduction to the discipline of corpus linguistics remains one of the most popular for good reason. It provides an extensive overview of the relationship between corpus-based research and language teaching, using numerous practical examples of the types of linguistic information a corpus can provide.)

Timmis, I. (2015) *Corpus Linguistics for ELT*, London: Routledge. (This book surveys the relationship between research in corpus linguistics and practice in ELT critically and in depth. It advocates for greater practitioner awareness and engagement vis-a-vis corpus research and asserts this a a critical factor for the reflective professional.)

Vaughan, E. and McCarthy, M. J. (2017) 'Research in Corpora in Second Language Teaching and Learning', in E. Hinkel (ed.) *Handbook of Research in Second Language Teaching and Learning*, Vol.*III*. New York: Routledge, pp. 173–85 (Provides a broad-ranging, critical survey of the impact corpora and corpus research have had on second language teaching and learning, including corpora and reference/pedgagogical materials for language learners, corpora and L2 pragmatics, SLA and learner corpora and the importance of questions of classroom models and world Englishes.)

References

Aijmer, K. (ed.) (2009) *Corpora and Language Teaching*, Amsterdam: John Benjamins.

Barker, F. (2014) 'Using Corpora to Design Assessment', in A. J. Kunnan (ed.) *The Companion to Language Assessment*, NJ: Wiley-Blackwell, pp. 1013–28.

Biber, D. (1990) 'Methodological Issues Regarding Corpus-Based Analyses of Linguistic Variation', *Literary and Linguistic Computing* 5(4): 257–69.

Biber, D. (1993) 'Representativeness in Corpus Design', *Literary and Linguistic Computing* 8(4): 243–57.

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

Borg, S. (2013) *Teacher Research in Language Teaching*, Cambridge: Cambridge University Press.

Breyer, Y. (2009) 'Learning and Teaching with Corpora: Reflections by Student Teachers', *Computer Assisted Language Learning* 22(2): 153–72.

British Association for Applied Linguistics (BAAL) (2016) *Recommendations on Good Practice in Applied Linguistics*, Available https://baalweb.files.wordpress.com/2016/10/goodpractice_full_2016.pdf.

Brzezina, V. and Flowerdew, L. (eds) (2017) *Learner Corpus Research*, London: Bloomsbury.

Callies, M. (2019) 'Integrating Corpus Literacy into Language Teacher Education,' in S. Götz and J. Mukherjee (eds) *Learner Corpora and Language Teaching*, Amsterdam: John Benjamins, pp. 245–63.

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

Chambers, A. (2019) 'Towards the Corpus Revolution? Bridging the Research–Practice Gap. *Language Teaching* 52(4): 460–75.

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

Coniam, D. (1997) 'A Practical Introduction to Corpora in a Teacher Training Language Awareness Programme', *Language Awareness* 6(4): 199–207.
Copland, F. (2008) 'Deconstructing the Discourse: Understanding the Feedback Event', in S. Garton and K. Richards (eds) *Professional Encounters in TESOL: Discourses of Teachers in Training*, Basingstoke: Palgrave, pp. 5–23.
Davies, M. (2010) 'The Corpus of Contemporary American English as the First Reliable Monitor Corpus of English', *Literary and Linguistic Computing* 25(4): 447–64.
De Costa, P. (2015) 'Ethics and Applied Linguistics Research', in B. Paltridge and A. Phatiki (eds) *Research Methods in Applied Linguistics: A Practical Resource*, London: Bloomsbury, pp. 245–57.
Farr, F. (2011) *The Discourse of Teaching Practice Feedback: An Investigation of Spoken and Written Modes*, London: Routledge.
Farr, F., Farrell, A. and Riordan, E. (2019) *Social Interaction in Language Teacher Education: A Corpus and Discourse Perspective*, Edinburgh: Edinburgh University Press.
Fligelstone, S. (1993) 'Some Reflections on the Question of Teaching, from a Corpus Perspective', *ICAME Journal* 17: 97–109.
Flowerdew, L. (2012) *Corpora and Language Education*, London: Routledge.
Frankenberg-Garcia, A. (2012) 'Raising Teachers' Awareness of Corpora', *Language Teaching* 45(4): 475–89.
Frankenberg-Garcia, A. (2016) 'Corpora in ELT', in G. Hall (ed.) *The Routledge Handbook of English Language Teaching*, London: Routledge, pp. 383–98.
Goffman, E. (1971) *The Presentation of Self in Everyday Life*, Harmondsworth: Penguin.
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., Gilquin, G. and Meunier, F. (eds) (2015) *The Cambridge Handbook of Learner Corpus Research*, Cambridge: Cambridge University Press.
Healy, M. and Onderdonk, K. (2013) 'Looking at Language in Hotel Management Education', in F. Farr and M. Moriarty (eds) *Language, Learning and Teaching: Irish Research Perspectives*, Bern: Peter Lang, pp. 141–65.
Jenks, C. (2011) *Transcribing Talk and Interaction*, Amsterdam: John Benjamins.
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.
Kubanyiova, M. (2008) 'Rethinking Research Ethics in Contemporary Applied Linguistics: The Tension between Macroethical and Microethical Perspectives in Situated Research', *The Modern Language Journal* 92(4): 503–18.
Leech, G. (1997) 'Teaching and Language Corpora: A Convergence', in A. Wichmann, S. Fligelstone, T. McEnery and G. Knowles (eds) *Teaching and Language Corpora*, London: Longman, pp. 1–23.
McCarthy, M. J. (2008) 'Accessing and Interpreting Corpus Information in the Teacher Education Context', *Language Teaching* 41(4): 563–74.
McCarthy, M. J. and O'Keeffe, A. (2010) 'Historical Perspective: What Are Corpora and How Have They Evolved?', in A. O'Keeffe and M. J. McCarthy (eds) *The Routledge Handbook of Corpus Linguistics*, London: Routledge, pp. 3–13.
McEnery, T. and Hardie, A. (2012) *Corpus Linguistics: Method, Theory and Practice*, Cambridge: Cambridge University Press.
McEnery, T. and Xiao, R. (2011) 'What Can Corpora Offer in Language Teaching and Learning', in E. Hinkel (ed.) *Handbook of Research in Second Language Teaching and Learning*, VolII, New York: Routledge, pp. 364–80.
McKinley, J. (2019) 'Evolving the TESOL Teaching-Research Nexus', *TESOL Quarterly* 53(3): 875–84.
Mann, S. (2005) 'The Language Teacher's Development', *Language Teaching* 38(3): 103–18.
Medgyes, P. (2017) 'The (Ir)relevance of Academic Research for the Language Teacher', *ELT Journal* 71(4): 491–8.
Meunier, F. and Gouverneur, C. (2009). 'New Types of Corpora for New Educational Challenges: Collecting, Annotating and Exploiting a Corpus of Textbook Material', in K. Aijmer (ed.) *Corpora and Language Teaching*, Amsterdam: Benjamins, pp. 179–01.

Meunier, F. and Reppen, R. (2015) 'Corpus versus Non-Corpus-Informed Pedagogical Materials: Grammar as the Focus', in D. Biber and R. Reppen (eds) *The Cambridge Handbook of English Corpus Linguistics*, Cambridge: Cambridge University Press, pp. 498–514.
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. 232–47.
Mukherjee, J. (2006) 'Corpus Linguistics and Language Pedagogy: The State of the Art of and Beyond', in S. Braun, K. Kohn and J. Mukherjee (eds) *Corpus Technology and Language Pedagogy: New Resources, New Tools, New Methods*, Frankfurt am Main: Peter Lang, pp. 5–23.
O'Keeffe, A. (2020) 'Data-Driven Learning – A Call for a Broader Research Gaze', *Language Teaching* 1–14. 10.1017/S0261444820000245
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.
O'Keeffe, A. and Mark, G. (2017) 'The English Grammar Profile of Learner Competence: Methodology and Key Findings', *International Journal of Corpus Linguistics* 22(4): 457–89.
O'Keeffe, A. and Farr, F. (2019) 'Using Corpus Approaches in English Language Teacher Education', in S. Walsh and S. Mann (eds) *The Routledge Handbook of English Language Teacher Education*, London: Routledge, pp. 268–82.
O'Keeffe, A., McCarthy, M. J. and Carter, R. A. (2007) *From Corpus to Classroom: Language Use and Language Teaching*, Cambridge: Cambridge University Press.
Paquot, M. (2018) 'Corpus Research for Language Learning and Teaching', in A. Phakiti, P. de Costa, L. Plonsky and S. Starfield (eds) *The Palgrave Handbook of Applied Linguistics Research Methodology*, London: Palgrave Macmillan, pp. 359–74.
Paran, A. (2017) '"Only Connect": Researchers and Teachers in Dialogue', *ELT Journal* 71(4): 499–508.
Park, K. (2014) 'Corpora and Language Assessment: The State of the Art', *Language Assessment Quarterly* 11(1): 27–44.
Richards, K. (2006) *Language and Professional Identity. Aspects of Collaborative Interaction*, Basingstoke: Palgrave Macmillan.
Römer, U. (2009) 'Corpus Research and Practice', in K. Aijmer (ed.) *Corpora and Language Teaching*, Amsterdam: John Benjamins, pp. 83–98.
Römer, U. (2011) 'Corpus Research Applications in Second Language Teaching', *Annual Review of Applied Linguistics*, 31: 205–25.
Sato, M. and Loewen, S. (2019) 'Do Teachers Care about Research? The Research-Pedagogy Dialogue', *ELT Journal* 71(4): 1–10.
Saville, N. and Hawkey, R. (2010) 'The English Profile Programme – The First Three Years', *English Profile Journal* 1(1). 10.1017/S2041536210000061
Seedhouse P. (2019) 'The Dual Personality of 'Topic' in the IELTS Speaking Test', *English Language Teaching Journal* 73(3): 247–56.
Sert, O. (2019) 'Classroom Interaction and Language Teacher Education', in S. Walsh and S. Mann (eds) *The Routledge Handbook of English Language Teacher Education*, London: Routledge, pp. 216–38.
Sinclair, J. M. (ed.) (2004) *How to Use Corpora in Language Teaching*, Amsterdam: John Benjamins.
Swales, J. M. (1996) 'Occluded Genres in the Academy: The Case of the Submission Letter', in E. Ventola and A. Mauranen (eds) *Academic Writing: Intercultural and Textual Issues*, Amsterdam: John Benjamins, pp. 45–58.
Timmis, I. (2015) *Corpus Linguistics for ELT*, London: Routledge.
Tsui, A. B. M. (2005) 'ESL Teachers' Questions and Corpus Evidence'. *International Journal of Corpus Linguistics* 10(3): 335–56.
Vásquez, C. (2004) '"Very carefully managed": Advice and Suggestions in Post-Observation Meetings', *Linguistics and Education* 15(1–2): 33–58.
Vásquez, C. and Reppen, R. (2007) 'Transforming Practice: Changing Patterns of Interaction in Post-Observation Meetings', *Language Awareness* 16(3): 153–72.
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): 173–89.

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.
Vaughan, E. and Clancy, B. (2013) 'Small Corpora and Pragmatics' in J. Romero-Trillo (ed.) *Yearbook of Corpus Linguistics and Pragmatics*, Vol. I, Dordrecht: Springer, pp. 53–73.
Vaughan, E. and O'Keeffe, A. (2015) 'Corpus Analysis', in K. Tracey (ed.), *The International Encyclopedia of Language and Social Interaction*, Hoboken, NJ: John Wiley and Sons, pp. 252–68.
Vaughan, E. and McCarthy, M. J. (2017) 'Research in Corpora in Language Teaching and Learning', in E. Hinkel (ed.) *Handbook of Research in Second Language Teaching and Learning*, Vol. III, New York: Routledge, pp. 173–85.
Walsh, S. (2006) *Investigating Classroom Discourse*, London: Routledge.
Walsh, S. (2011) *Exploring Classroom Discourse: Language in Action*, London: Routledge.
Walsh, S. (2013) *Classroom Discourse and Teacher Development*, Edinburgh: Edinburgh University Press.
Walsh, S., Morton, T. and O'Keeffe, A. (2011) 'Analysing University Spoken Interaction', *International Journal of Corpus Linguistics* 16(3): 326–45.
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. 210–29.
Xerri, D. (2017) 'Split Personality/Unified Identity,' *ELT Journal* 71(1): 96–8.
Yang, S. (2014) 'Interaction and Codability: A Multi-Layered Approach to Discourse Markers in Teachers' Spoken Discourse', in J. Roméro-Trillo (ed.) *Yearbook of Corpus Lingustics and Pragmatics 2014*, New York: Springer International Publishing, pp. 291–313.