Corpora for materials' development in language teacher education: underlying principles and useful data

Introduction

Student teachers and teacher educators, when thinking about the kinds of materials they ideally want to use during their pre-service or in-service programmes, will immediately think of words like: relevant, based on real contexts, local or localised, interesting and motivating, easily accessible, time-proof, effective, up-to-date, and many other such adjectives. Those involved in the creation of language teacher education (LTE) materials, most of whom are teacher educators, aim to achieve most of this when they set about design and development, whether for their own programmes or for commercial publication and distribution. Based on our experience, this is often a very demanding task, and while it may be possible to effectively satisfy some of the criteria that are set, it is usually difficult to satisfy them all. For example, consider the question of locally relevant materials. Understandably, commercially produced materials, due to market constraints, cannot afford to be localised to such an extent that they become unviable in terms of sales. In the case of English Language Teaching, this has meant that the major geographical locations (Britain, America, Australia, parts of Asia etc) tend to be represented more in terms of teaching contexts and also language variety. If we consider materials produced locally for other LTE contexts, their relevance is much higher and they are arguably more motivating for student teachers who plan to work in the same or a similar location after they have achieved the necessary qualification. The difficulty here, however, is to be found in

the fact that many such materials, despite a huge amount of effort during their collation, tend not to be used to their full potential, or by any substantial number of users. This happens for many reasons: lack of confidence in making them available for other students/teacher educators, lack of time and resources to fully exploit them, potential copyright issues, lack of awareness of the most appropriate means for integration, and many other such causes. In this chapter, after outlining what we consider to be the five key principles for the development of LTE materials, we discuss the potential that corpus-based approaches have to facilitate and integrate these principles into effective and motivating activities. Although creative and appropriate use of corpora would suggest that they might be applied in numerous ways in many facets of teacher education programmes, we focus on three here: corpora as language resources, corpora of classroom discourse, and corpora for reflective practice and professional development, as these generally tend to be core components of education programmes for teachers of languages. Our focus is primarily on identifying materials and suggesting what insights they might offer, rather than on providing a detailed account of writing or evaluating corpus-based materials. While we limit our illustrations in this chapter to examples taken from the English and French languages, mainly because these are the fields in which we have most experience and expertise, we believe that our general discussions and suggestions can be applied to the education of teachers in other languages also.

Underlying principles for the development of LTE materials

In this section, in identifying what we consider to be the five key underlying principles for LTE materials, we draw on a framework proposed by Tomlinson (1998) for the development of language learning materials based on previous research from the field of second language acquisition research. Here, we take the liberty to select, adopt and adapt, combine, and add to Tomlinson's fundamentals to take account of the different learning environment and target group within an LTE context. Nonetheless, much of what he originally suggests for language learning materials equally applies.

1. Materials should achieve impact

In general, those investing time, money and effort in a programme as specific as language teacher education tend to be relatively highly motivated, usually in an instrumental way. LTE programmes often operate at postgraduate level, which also affords a certain amount of maturity to students enrolled on them, and in the case of English Language Teaching (ELT), the students are very likely to already have had some previous practical teaching experience before embarking on this route. For all of these reasons, the type of learner in LTE can be quite focussed, and materials will often achieve impact most saliently through students' perception of their relevance and usefulness. Impact will be enhanced in other ways, by such things as novelty, variety, and attractive presentation, as outlined by Tomlinson (1998, p.7).

2. Materials should help learners to feel at ease and to develop confidence

It is now generally accepted that one's emotional state and readiness plays a major role in the learning process. As a profession, we have been giving affect the

importance that it deserves in language education environments and activities for a number of years (see, for example, Arnold 1999). However, much of what we know about how anxiety and confidence operate tends to be based on traditional classroom modes of learning. Much less is known about how these variables function when the learning is taking place using computer-based learning either in on-line mode or using local software and materials. One notable exception to this is work done by Hauck on on-line language learning courses (see for example, Hauck and Hurd 2005), the findings of which suggests that the relationship between affective factors and learning through the use of a computer is quite a complex one. Depending on different learner-related variables, styles, and strategies (Oxford 1990, McLoughlin 1999) the development and use of corpus-based materials for LTE may be a greater or lesser challenge than non-computer based materials in terms of lowering anxiety and increasing confidence and motivation among student teachers.

3. Materials should require and facilitate critical cognitive and emotional development

We are truly beyond conceptualising and practising education in a purely transmission-based mode where the learner was seen as an empty vessel to be filled with knowledge emanating from educators and books. LTE now operates on a principle of learner self-investment, where student teachers move from a position of peripheral participation (Lave and Wenger 1991) to being fully fledged members of the teaching profession through a process of drawing on their own practical knowledge (Golombek 1998) and combining this with theoretical content through processes of reflection and further practice. This reflective practice process and cycle in LTE, which often integrates simple action research activities, demands that student

teachers think in critical and questioning ways, with an ability to find solutions in a context dependent and flexible way. For these reasons, LTE materials should encourage critical intellectual and emotional involvement to prepare students for the ever changing world in which they will develop their language teaching careers.

4. Materials should allow for attention to be drawn to specific features

In investigations of language use or classroom practice, student teachers should ideally be able to extract, quantify, and reshape the output of data in ways which will allow them to answer specific hypotheses. To do this they will often need a concentration of examples, or occurrences, which is rarely present in full-length spoken or written texts. In essence, the automatic extraction and processing of language as it is afforded by corpus linguistic techniques (see for example McEnery et al. 2006) provides an efficient mechanism by which this can be done, while also maintaining the means by which context can be examined and scrutinised in a more qualitative way, depending on the design and interface associated with the corpus in question. And at a time when it is acknowledged by many (see for example, McDonough and Shaw 2003, p.41) that teachers need to be able to focus on form and function, taking the best of the traditional and communicative approaches to teaching language, corpus techniques provide a means by which forms and patterns can be quickly extracted and presented. All the while, a focus on function and communication can equally be attended to through a complementary reading of the full texts, many of which now provide adequate information on those who produced the language for this purpose.

5. Materials should permit a silent period and take into account that the positive effects of instruction are usually delayed

Reflection, as one of the fundamentals in LTE, by its very nature requires a processing period. Student teachers need to be given and to take the time to think and to activate the appropriate cognitive processes which will allow them to integrate their practical and theoretical knowledge in order to come to plausible conclusions about the nature of language, teaching and learning. Only when the necessary processing has taken place are the teachers in a position to be able to share and develop their thoughts in the practical context of a socio-constructivist and socio-cultural learning environment (Wertsch 1998), which will facilitate learning. And indeed, this may still be quite far removed from the effects that this increased learning might have in implementation terms in their own teaching practices. It is a reality that little is known about the actual effect of LTE or reflection on practice (Akbari 2007), for precisely that reason; any positive effects of instruction are usually delayed and therefore not easily captured in empirical terms.

Using corpora as language resources

Corpora are already present in many language classrooms in the form of ready-made resources such as dictionaries, grammars and, less frequently, course books. Teachers and learners, however, may not be aware that they are already familiar with corpus data. While these resources give learners access to attested rather than invented examples, they cannot provide the large number of occurrences which promote noticing (Schmidt 1993), nor can the examples be selected from a context specifically

chosen to be relevant to the learners' needs, as is possible when using a corpus as the source. Concordances provide a novel way of providing a variety of relevant examples, thus conforming to the first principle listed above by producing a strong impact on the learners. They learn to read in a different way, concentrating on the various patterns which can be observed when using the word or expression they have searched for. In addition, as Johns (2002) has shown, concordance lines provide the basis for inductive learning or, as he terms it, data-driven learning, thus facilitating the cognitive development of the learners. LTE programmes can therefore fulfil an important role in preparing teachers for the use of corpora in more active and inductive ways than simply by consulting published corpus-based materials. This section suggests ways in which student teachers can be prepared for the use of corpora in their future careers, concentrating mostly on the initial steps of such training, namely the use of three types of corpora: freely available web-based access to vast corpora using simple built-in concordancers, web-based pedagogic corpora, namely corpora created specifically to be used by teachers (Willis 1998, p.46), and corpora of specialised language use in one particular genre. As we shall see, even these simple uses of corpus resources illustrate the five principles listed above. The decision to avoid exploiting the potential of corpora to the full in this section is motivated by the fact that, in empirical studies of corpus consultation by learners, the generally positive responses are very often accompanied by a small but significant number of negative comments to the effect that working with corpus data is time-consuming (Yoon and Hirvela 2004, p.274; O'Sullivan and Chambers 2006, p.xx), laborious and tedious (Cheng *et al.* 2003, p. 183; Chambers 2005, p.120). It is thus important that teachers are familiar with simple applications which they can use to introduce the use of corpora without giving rise to negative reactions from their learners.

As a number of large reference corpora have samplers which can be easily accessed via the Internet, teachers with very basic web skills can easily produce concordances to illustrate aspects of the language which they are teaching, or to provide additional examples of one occurrence in a text. The Bank of English (Cobuild Concordance and Collocations Sampler) and the British National Corpus, for example, provide 40 and 50 concordance lines respectively as the result of a simple search, and WebCorp (Renouf *et al.* 2007) allows the entire web to be searched. Thus, without any prior training a teacher could produce the following examples of adjectives qualifying *majority*.

```
 - Mr Walter Sisulu was elected deputy President by a substantial
majority.
 - The main problem is that the vast majority of motorists over-rate
their own driving ability.
- The Italian government won a confidence vote in parliament by a
comfortable majority averting a political crisis
 - the great majority will confidently tell you that they are the
smartest kid in their class.
 - The overwhelming majority of Black actors on prime-time television
are employed in comedic roles or as criminals;
 - the US House of Representatives passed a civil rights bill this
week by a large majority,
 - the National Party is on course for a big majority in parliament.
- The euphoria of 1966 had given Harold Wilson a decisive majority in
March
 - The declaration of sovereignty was passed by a massive majority of
three hundred and fifty-five votes to four
```

Figure 1: Concordances of *majority* from the Cobuild Concordance and Collocations Sampler

A simple concordance such as this will have immediate impact on learners intending to produce a written or spoken text where the word *majority* is likely to occur. It could lead them, for example, to run searches for expressions such as *vast majority* and *great majority* to help them to find out which is more suitable for their needs. It can also increase the confidence both of learners and of a non-native speaker teacher by providing attested examples. In keeping with the fourth principle, it draws attention to a specific feature, in this case the choice of adjective, in a way which the study of a single text cannot do. The few examples listed above do not enable us to focus on patterns, but even in 40 or 50 examples certain patterns become clear, such as the use of *vast majority* in a general context such as that in the second concordance line above. Occurrences of many problematic aspects of English, such as phrasal verbs, can easily be produced, thus allowing teachers not only to provide their learners with multiple examples, but also to check their own native speaker intuition or non-native speaker command of the language. A limitation of these resources is that access to the full context is not available, as it would be if the teacher were using the corpus in question with a concordancer such as Wordsmith Tools (Scott 2004) or MonoConc (Barlow 2000). For some teachers, the samplers might provide a quick solution to the problem of finding attested examples, while others might decide to progress to the more advanced possibilities provided by a full concordancer.

While the samplers and the reference corpora provide vast corpus resources for teachers, pedagogic corpora specially created for teachers can provide teaching material more closely targeted at the learners' specific needs. The web-based

SACODEYL corpora of teen talk, for example, include video recorded interviews and transcripts of approximately ten minutes with 20-25 teenagers in each of the following languages: English, French, German, Italian, Lithuanian, Romanian, and Spanish. Teenage learners can thus watch the video-recorded interviews, and then focus on specific aspects of the language use through concordances. Interestingly, the French teenagers, when giving their age, often use forms different from the classic *J'ai douze ans*, mostly using the present tense to focus on the forthcoming birthday, a pattern which is less frequently included in course books.

- Tu as des frères et des sœurs? Oui, j'ai une sœur qui va avoir 13 ans
- J'ai bientôt douze ans
- Doris, quel âge as-tu ? J'ai 14 ans, bientôt 15.
- Je m'appelle Claire. Quel âge as-tu ? 15 ans et j'aurai 16 ans au mois de novembre. - J'ai 15 ans et demi, bientôt 16 ans.
- Tu as des frères et des sœurs ? Alors, j'ai une sœur qui a bientôt 16 ans
- tu as des frères et des sœurs ? J'ai une grande sœur de 21 ans
une petite sœur, qui aura son anniversaire dans deux jours, le 22, elle aura dix ans
- Ma sœur a vingt-deux ans et j'ai un frère qui a bientôt vingt- sept ans, c'est un demi -frère en fait, qui est plus avec nous.

Figure 2: Concordances of *ans* from SACODEYL

Similarly, when explaining how long they have been doing something, the teenagers are more likely to choose the informal expression *ça fait x ans que ...* than the standard *je joue au rugby depuis 8 ans*.

- Oui, je fais du rugby. Ça fait bientôt huit ans que j'en fais.
- Oui, c'est de la batterie et ça fait presque cinq ans que j'en fais. Ça fait cinq ans d'ailleurs.
- ça fait deux ans maintenant que je suis dans une école, une petite école de musique sympathique
- Depuis combien de temps tu as un blog ? Ah ça fait longtemps. Je l'ai depuis quatre ans presque , je l'ai créé en 2003 donc oui, ça fait quatre ans à peu près
- Ça fait trois ou quatre ans que je suis dans ce club -là.
- Et tu joues au football depuis combien de temps ? Alors, ça fait six ans que je joue au football.

Figure 3: Concordances of *ça fait x ans que* from SACODEYL

The SACODEYL corpora thus combine web-based video-recorded material, intended to be of interest to teenage language learners, with corpus-based resources which make it easy to produce materials based on the interviews. The learning of grammar can thus be integrated into the content of the class, emanating directly from the language use of the learners' French counterparts. Observation of concordances such as these encourages learners to participate actively in the learning process, and forces both learners and teachers to reflect on the differences between standard written language, the expressions which are included in their course book, and the concordance examples of what French teenagers actually say.

While teen talk is a common form of everyday language, one could consider these corpora as specialised in that they focus on the language use of one specific group. Corpora of more specialised registers are also used to provide advanced learners with attested examples of the genre which they wish to master. Hyland's publications, for example, are illustrated with references to a large corpus of academic writing in English (see, for example, 1998). To give just one illustration of how a corpus of academic writing in French could be used by a teacher or learner, the following examples, based on a search of the expression *dans un premier temps* show the variety of verbs used to describe the research process in a one-million corpus of academic writing in French (Chambers and Le Baron 2007). While this corpus is freely available via the web, it does not have a built-in concordancer, as in the case of the examples cited earlier, so access to a concordancer and training in its use is necessary. The concordancer used here is Wordsmith Tools (Scott 2004).

Dans un premier temps, nous aborderons brièvement la notion d'égalité
Dans un premier temps, nous allons tenter de mettre en contexte

Dans un premier temps, nous préciserons notre cadre conceptuel en soulignant
Dans un premier temps, nous rendrons compte des pratiques telles qu'elles sont
Dans un premier temps nous présenterons l'adventisme dans l'hexagone.
Dans un premier temps, nous tenterons une étude comparée de deux dessins

Figure 4: Concordances of *dans un premier temps, nous* from the Chambers and Le Baron Corpus of Research Articles in French

In addition to revealing the prevalence of *dans un premier temps* as an alternative to *d'abord* to introduce the content of the first section when giving the plan of the article, a concordance of several occurrences of this expression will provide the writer of an article or essay with a wide variety of verbs used to describe the writer's activity, providing an overview, situating the topic in context, describing the conceptual framework, doing a comparative study etc. This will increase not only the vocabulary but also the confidence of the writer by providing attested examples, and also encourage cognitive learning and reflection on the writing process, as the learner selects the most appropriate expression from the concordance. Lee and Swales (2006) take this use of corpora a stage further by providing training to researchers in the creation of a corpus to meet the needs of their own disciplinary area. There is thus a wide variety of possibilities open to teachers, from quick consultation of a sampler, to the use of pedagogic and specialised corpora, to the creation of a corpus to meet a specific need. Wynne (2005) provides guidance on the process of creating such a corpus. In a significant number of universities learners are given direct access to the corpora and training in the use of concordancing software (see Chambers 2007 for an overview of quantitative and qualitative empirical studies).

LTE programmes of study can thus make an important contribution to the integration of corpora in language learning and teaching by giving the student teachers knowledge about the corpora which exist and the skills necessary to make use of the

resources. In this way, the student teachers can improve their own linguistic abilities and awareness, and develop the critical skills needed for them to create their own materials for the pedagogic career paths ahead of them. Returning to our key principles, because such materials are based on real contexts, they have the potential to have a strong impact. This is increased by the fact that they can be local or localised, which makes them interesting and motivating. As we have seen, they have the potential to inspire confidence, facilitate cognitive development and reflection, and to draw attention to specific features. Because these corpora are available on-line and, in some cases, have built-in concordancers which require no prior training, they are well within the familiar realm of our digital savvy students and also form an easy introduction to corpora for technologically shy students.

Corpora of classroom discourse

So far in this chapter, we have made the case for the inclusion of corpora as an extra resource for teachers and learners for the purposes of increasing language awareness and teaching aspects of language systems. In fact, as recently highlighted, the use of corpus data as a complementary resource supporting a multifaceted approach to initial and continuing language teacher education is a necessary condition for the integration of corpora in the repertoire of teachers' skills (McCarthy 2008, Breyer 2009). For instance, as a complement to an in situ classroom observation, an increasing number of classroom discourse corpora, based on transcribed text from recorded classes, are used as resource material to provide language teacher educators with examples of teacher and learner talk. These corpora are used to enhance pedagogic (Walsh 2006, Chambers and O'Riordan 2007) as well as linguistic awareness (Amador Moreno et

al. 2006, O'Riordan 2009). After exploring the availability of classroom discourse corpora, this section will discuss the potential benefits of using corpus-based approaches for the investigation of language classroom discourse within LTE, and return again to the ways in which these meet the criteria for the development of effective LTE materials.

Classroom discourse has long been a focus of investigation for teacher educators (Sinclair and Coulthard 1975, Cazden 1988) and not only in language teacher education, but also in science (Wells and Mejia Arauz 2006) and in primary education (Nassaji and Wells 2000). Transcripts from classroom discourse feature in most textbooks for student teachers as such data is useful to illustrate particular features of teacher-learner interactions. In the late 1980s, however, Brumfit and Mitchell (1989, p.12) deplored the fact that '[w]e actually know remarkably little about typical practice in language learning, and there is a great need for additional comparative studies'. This call was heard by researchers such as Seedhouse (1995, 2004) and Walsh (2002, 2006) for instance, who, undoubtedly influenced also by Sinclair and Coulthard's (1975) seminal work on turn-taking in classroom discourse, focused their research on the interactional architecture of the language classroom (Seedhouse 2004). Alongside this research, others such as Lyster (2002) concentrated on corrective feedback in the language classroom, putting forward the provision of teacher recasts as the most prominent manifestation of content-focused corrective feedback in the communicative classroom. All these areas of classroom discourse investigation aim at raising pedagogic awareness amongst language teachers.

While such studies were extremely useful in raising awareness on some key aspects of classroom interactions, to our knowledge, none of these studies were corpus-based in the corpus linguists' sense of the term, in that they did not use corpus linguistics tools, such as concordancers or frequency lists, to query their data. These studies were in that sense text-based. While this is of course justified in that some studies focus mainly on the qualitative assessment of the data, there is nonetheless also scope for more quantitative research to take place, which can be achieved more easily with corpus-based methods. Such study of classroom discourse for the purpose of investigating, in both linguistic and pedagogic terms, the ways in which teachers and learners talk in the foreign language classroom, is a research area which can make an important contribution to the integration of corpora in language learning and teaching.

To illustrate the potential contribution of the study of quantitative linguistic data about classroom language, Figure 5 presents a list generated by Wordsmith Tools (Scott 2004) of the most frequently used words during a 40-minute French class taught by a native French speaker at Leaving Certificate Level[1] in Ireland in 2008.

---

[1] This corresponds to the British A-level year i.e. the last year of secondary school in Ireland.

N	Word	Freq.
1	EST	242
2	C	170
3	EUH	146
4	OUI	136
5	QUE	101
6	LE	92
7	TU	66
8	IL	60
9	JE	60
10	LA	58
11	UNE	48
12	D	44
13	MAIS	44
14	PAS	44
15	A	41
16	CE	41
17	DE	39
18	SUBJONCTIF	39
19	UN	38
20	ACCORD	36
21	ÇA	33
22	ON	32
23	VOUS	32
24	ET	28
25	L	27

Figure 5: Frequency list from a Leaving Certificate French Class

At a glance, one can see that all the most frequently used words are in French, suggesting a high target language use which can be an issue in secondary level education (Turnbull 2006). After taking a closer look, one notices that, unsurprisingly, the most frequently used words in this class are small grammatical words such as the French equivalent of *is* (*est* -1st rank)[2] or *has* (*a* – 15th), articles such as *the* (*le/la*- 6th – 10th), *a/an* (*une/un* 11th – 19th), but more interestingly, the hesitation marker *euh*, a typical feature of spoken French, comes 3rd and the content word *subjonctif* (ranked 18th), conveying metalinguistic information, shows that this aspect of French conjugation appears prominently within the class's discussion.[3] All these findings

---

[2] The second word on the list is the elided form of *ce* in front of *est* which becomes *c'est* in French. A concordance of the two words *c + est* with Wordsmith Tools reveals that, indeed, *c'est* occurs 170 times in this particular class.

[3] The word *accord,* which could at first sight look like a content word, is actually the second part of the expression *d'accord,* which is mostly used as a rhetorical question by the teacher to check pupils' comprehension.

could trigger interesting discussions with student teachers about classroom language such as the use of metalinguistic terminology in the target language. However, the most prominent linguistic feature illustrated by the frequency list is the use of personal pronouns which rank 7th (*tu*), 8th (*il*), 9th (*je*), 22nd (*on*) and 23rd (*vous*) within the twenty-five most frequent words uttered in this particular class. Leaving the pronoun *il* aside as it is often used in impersonal expressions such as *il y a* (*there is/there are*) the remaining four pronouns seem to reveal a symmetry between *je* and *tu* on the one side and *on* and *vous* on the other. In fact, further concordancing queries inform us that there is some but no systematic correlation between these pairs of pronouns. This teaches us to remain cautious when interpreting quantitative data such as those emanating from frequency lists. Interpretations should always be corroborated by further KWIC (Key Word in Context) queries through the subsequent study of concordances. Nevertheless, were it not for this specific frequency list, one might not have identified, from intuition, the issue of the use of the pronoun *on* in French classroom discourse as a relevant item for discussion with student teachers as part of a classroom language awareness activity. Indeed, it could be puzzling to a student teacher that *on* appears 32 times while *nous* appears only 12 times (ranking 70th) in this class. Although it has been obvious to French linguists for quite some time now that *on* is commonly used instead of *nous* in French conversation (Grévisse 1975, p.787), it is relevant to investigate what the uses of these pronouns may be in a classroom context. Concordances of *on* as shown in Figure 6 and of *nous* reveal that, in the transcribed class, the teacher uses *nous* only for the purpose of practising the conjugation of verbs in the first person plural whereas *on* is used in a variety of ways but especially to signpost to students what is going on in the class as in *Allez on continue* (*Come on let's go on*), *on fait la correction* (*Let's do the correction*) and to

announce the types of activities to come in the future as in *on va voir celui-là maintenant* (*We're going to see this one now),* or *après on arrêtera un petit peu* (*Afterwards we'll stop a little bit*). These are interesting examples as they show that, when addressing the class, the teacher seems to use *on*, which is translated as *we*, to be at one with the learner group and the fact that she does not use the more formal sounding *nous* in these instances softens the directions she gives through her signposting strategies, which are a key pedagogic skill that student teachers need to develop.

Figure 6: Concordances of *on* from a French Leaving Certificate class

Learning how to plan, structure and reflect on one's discourse in a teaching setting whether at primary, secondary or tertiary level is arguably a core component of teacher education. It is thus logical to call for the integration of classroom discourse corpora as teaching and learning materials for teacher educators as well as student teachers. The potential of such resources is exemplified by the Michigan Corpus of Academic Spoken English (MICASE), whose website offers activities aimed at EFL/ESL learners for the purpose of self study on commonly used multi-word expressions in academic talk (see Simpson et al. 2007), as well as at EFL/ESL teachers for the purpose of exploring topics such as giving instructions in the context of Content and Language Integrated Learning (CLIL) (see Mendis 2002). A separate word search interface also enables users to query the use of specific words or word clusters within a sample of 152 transcripts available on-line and free of charge. Undeniably, the MICASE project paved the way for more spoken academic discourse corpora to be created, in other varieties of English such as British English and Irish English as illustrated by BASE (British Academic Spoken English) and LIBEL CASE (Limerick-Belfast Corpus of Academic Spoken English). These corpora all contain tertiary level classroom discourse. However, a recent large-scale move towards making primary and secondary level language classroom interaction data available was initiated by the SCoRE project (Singapore Corpus of Research in Education) which is composed of EFL and Mandarin Chinese classroom interactions. Such corpus-based classroom discourse research acts as a flagpole for similar smaller scale corpora to develop and they also provide relevant data for primary and secondary level student teachers.

As argued, quantitative corpus data can be of assistance to teachers and student teachers in a number of ways, particularly in relation to reflecting on their own language use in the classroom, and also on their pedagogic practice. Indeed, when asked about what words and phrases French teachers use the most in their classes, student teachers often quote directives such as *ouvrez-vos livres* (*open your books*) or positive evaluative remarks such as *très bien* (*very good*). Provided with a frequency list from a corpus of fifteen secondary level French classes, student teachers can discover that words like *alors* and *donc*, which have a strong discourse marking function, tend to feature much more prominently in teacher talk. As part of a corpus-based tutorial on classroom discourse in French, within a Graduate Diploma in Education group at an Irish third level institution during the academic year 2008-2009, two student teachers (out of a cohort of 8) chose to focus on the different uses of *alors* for an assessed task requiring them to identify an aspect of classroom discourse which they considered as relevant for preparing to teach through the target language. Although these two students had never done a concordance-based linguistic analysis on their own before, both were able to demonstrate the polysemic nature of *alors*. For instance, in addition to being used as a discourse marker especially to introduce new activities (see Amador Moreno et al. 2006 for more functions of *alors* as a discourse marker), one student highlighted the fact that *alors* is also used to express consequence as in the following extract: *Mais si c'est une question ou si la phrase est négative alors tu utilises le présent du subjonctif* (*But if it is a question or if the sentence is negative then you use the present subjunctive*). She also noted that it could convey emotions such as impatience through the tone of voice taken by the teacher: *Et alors, où est votre livre?* (*And where is your book exactly/then?*). This

study led the student to draw the following conclusion in her essay showing the possible impact of this type of study on the student teacher's pedagogic awareness:

'C'est très important de sensibiliser nos étudiants aux marqueurs du discours pour les aider à suivre le discours de la salle de classe et aussi pour promouvoir une compétence pragmatique dans la langue.' (*It is very important to make our pupils aware of discourse markers to help them follow the discourse of the classroom and also promote pragmatic competence in the language*).

In short, although the results of this pedagogic application of a corpus of classroom discourse to help non-native speaker student teachers to prepare to teach are still being processed, there is clear evidence that a corpus of classroom discourse can be a very valuable resource to answer specific linguistic questions about this genre thus enhancing student teachers' critical cognitive development, as outlined in the third key principle earlier. Moreover, the example on the pervasive use of *on* compared to the lesser use of *nous* by one native speaker teacher highlights, through a quick and easy frequency list analysis, how the corpus-based study of classroom language can draw attention to specific features (fourth key principle). The added value of such materials is that they have been collected from local Irish contexts, making them highly relevant, interesting and motivating for the student teachers, who will find themselves in the same or similar situations during and after their LTE programmes.

Corpora for continuing professional development

Arguably, much, or most, of what teachers need to know is learned 'on the job' when they begin or continue on their career paths. For this reason, LTE programmes now place a strong emphasis on continuing professional development. As corpora are ever evolving and growing, they have the potential to be a really good resource to support on-going professional development activities, with the added advantage that they can be utilised in independent study mode by individuals with a computer and internet access. These are factors which are likely to be more significant when teachers are no longer in a position to rely on the institutional resources and supports which they enjoyed during their LTE programmes. One variable which is constant in teaching is that new students with different and specific needs and wants will find their way into our classrooms very frequently. This is the case within schools and will be accentuated in the case of teachers moving to different schools, often in different countries and continents. In the case of ELT, the question of language variety often becomes an issue. With the high mobility of teachers, it is quite likely that they will find themselves working in a geographical location where the variety of English being taught, learned or used may be different from the variety which the teacher grew up using, or studied formally. Therefore, as well as the local context being important, the need to consider internationalisation is also a concern. In our experience, there is an interesting side to such scenarios in the case of Irish teachers of English, who will be quite familiar with the standards of British or American English which they may have to use in their teaching, but they may not be at all familiar with those aspects of their own language usage which are specific to Irish English. For these reasons, corpora which aim to represent a specific variety of language use may be quite useful at certain junctures of one's career.

The following example illustrates varietal differences in the usage of verb phrases with the combined auxiliary structures *do* and *be* in Irish English. Many might consider this to be infrequent in contemporary Irish English and confined to very isolated rural areas or speakers from less well-educated and agricultural backgrounds, but the Limerick Corpus of Irish English (LCIE) (Farr et al. 2004), which is a one-million word corpus of spoken Irish English, the majority of whose speakers fall within the educated 18-30 year old age group living in both urban and rural areas, suggests that this is not the case. In LCIE *does be* is found 12 times and *do be* is found 14 times, marking it as a feature of contemporary Irish English. And as is suggested by the following concordance lines, varietal features of a language tend to cluster providing further data for consideration. We see this in Line 2 with the use of the collocations *ould ones* meaning old women, and *the whole time* meaning frequently. Line 3 gives a good example of the double negative in Irish English, with *don't tell her nothing* occurring twice. There are also three examples of subject verb disagreement in the use of *I does be* in Lines 4, 6, and 7. Line 5 illustrates the use of the utterance final *like*, and *perished* with the meaning of very cold can be seen in Line 9. And finally in Line 11 we find the colloquial phrasal verb *pissed off* meaning very annoyed, and a couple of swearwords illustrate another aspect of Irish English.
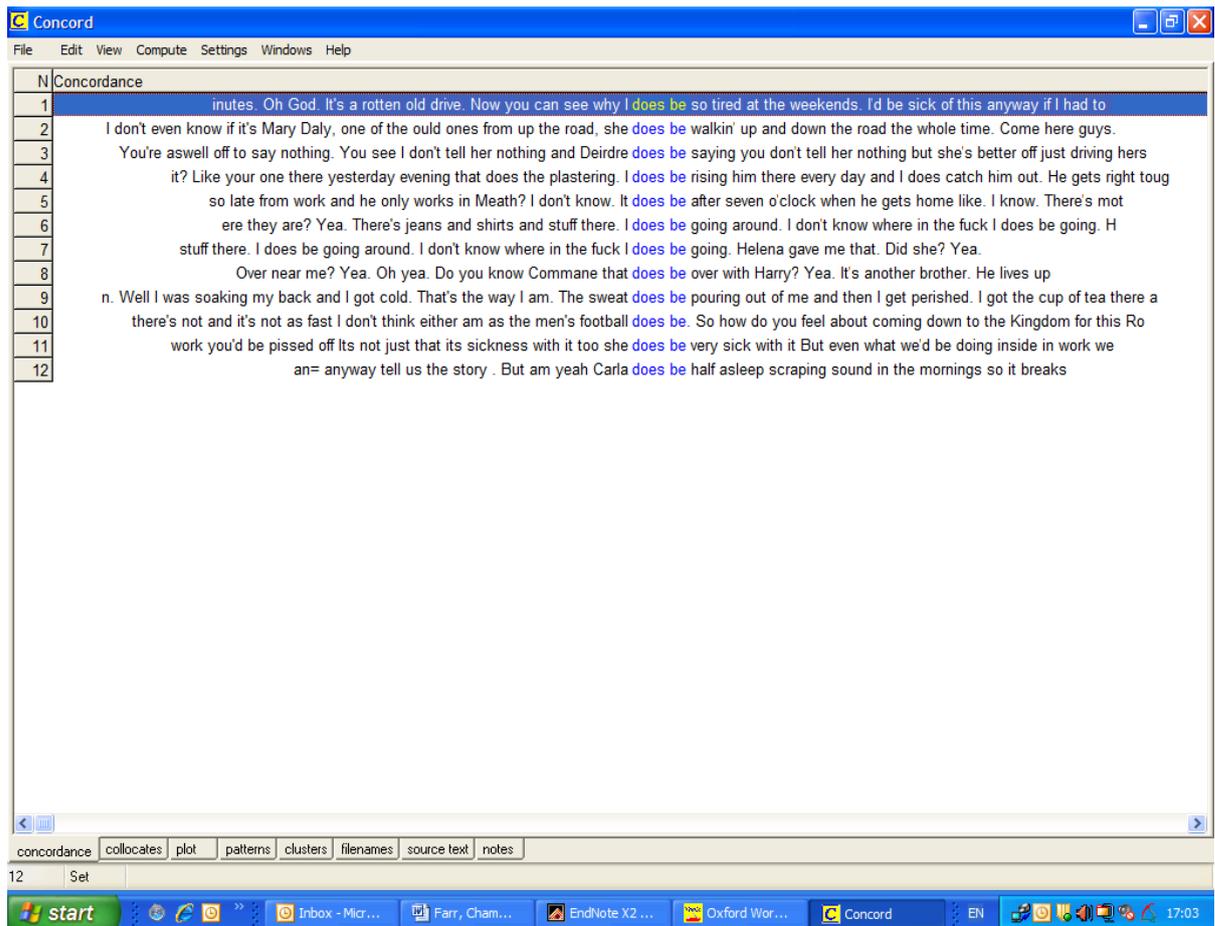
Figure 7: Concordances of *does be* from the Limerick Corpus of Irish English

The co-occurrence of *do* and *be,* coming from a similar structure in Gaelic, has been cited as one of the traditional characteristics of this variety of English and, according to Kallen (1985), its most common usage is to express frequency. Such corpus-based investigations provide a concentration of examples and features which allow a teacher to characterise a language variety of English, either for their own awareness, or if they want to share something of the richness of the English language with their students. While some of this may have been part of initial LTE programmes, it is unlikely that it will be a customised fit for all the future teaching contexts of teachers, and with the growing reach of corpus collections, there is an increasing representation of varieties and languages for continued use in the future. Such national, regional, and even local

representations of language in electronic format will make them highly relevant and should allow teachers to feel more confident in presenting attested examples of language use from real speakers of the variety under consideration.

In addition to variety specific representations of language use, there are many corpora based on specific context based language use. Already mentioned earlier was the case of academic language, and there are others including business language, medical language, workplace language, and many others. In addition, a very interesting development has also been taking place at the University of Vienna in the study of English as a Lingua Franca (ELF), and has led to the recent release of the VOICE corpus (Seidlhofer 2005). This one-million word corpus of spoken English represents language use from a number of contexts including interviews, press conferences, service encounters, seminar discussions, working group discussions, workshop discussions, meetings, panels, question-answer-sessions and casual conversations. This freely available resource provides a rich bed of language for language for specific purposes, and also to examine the linguistic features of speakers from different language backgrounds as well as the characteristics of English as a Lingua Franca, which has growing significance in sociolinguistic circles.

Continuous professional development includes a strong element of reflective practice, and, as has been argued elsewhere (Farr 2005b), this can be most effective if it is evidence based. In other words, when there is some real material evidencing how we behave and what we do in our professional lives, then we can use this to critically assess practices and possibly change future behaviours. In the previous section we spoke about learning teaching through the use of corpora of classroom discourse, but

there are also the many out of classroom contexts in which teachers operate as professionals all the time. These contexts, which view teaching as a much more holistic activity, have been the subject of research conducted by Vaughan (2007) in which she collects and analyses a type of backstage teacher discourse in the form of teacher meetings. She analyses and reflects on a total of six teacher-teacher meetings in both Ireland and Mexico and identifies the type of professional discourse that is produced in these contexts. For example, she finds that on a descending scale of frequency, the following lexical items are used to describe language students' ability: *good, weak, strong, high, better, low, bad, best,* and *worst*. She also examines affective variables such as humour, and demonstrates how this is used to consolidate professional friendships, to attend to the interpersonal realm of professional discourse, but also on a subversive level 'by a member of the group *against* the group itself' (Vaughan, 2007, p.184). This type of evidence in the form of a corpus, whatever size (Vaughan's corpus was 3.5 hours of data), can offer very valuable insights into how teachers work as professionals and develop communities and sub-communities of practice. Such insights can be invaluable on teacher education programmes where the student teachers are still very much engaging in peripheral participation with the professional community and are learning through apprenticeship (Hawkins 2004). Professional context corpora provide the means by which students can fully participate in an observational way in various scenarios without the potential threat and fear that being present in person might impose. Teacher educators, with their vast professional networks and cohorts of graduates should easily be in a position to collect data and create some such corpora as they perceive relevant for use on their programmes.

Finally, in this section, we would like to move to the question of professional development for teacher educators. The same principles as those discussed in this and the previous section apply, but in the case of the educators examining their own practices, the methodology fits much closer with an action research approach with the corpus compilation constituting the data collection phase. The corpus ultimately becomes the material on which professional reflection and development can be based. However, we would strongly suggest that most benefit can be derived in this context from combining data sources in order to reach a more holistic interpretation of the situation. Taking dyadic teaching practice feedback between student teachers and tutors as a case in point, Farr (2005a) demonstrates how she uses the perceptions of the participants in the interactions to classify a number of her 80,000 word POTTI (post-observation trainer trainee interaction) sessions as being either positive or negative in tone. On this basis it is possible to isolate the ways in which the positive and negative corpora differ in terms of the language used. The following emerge as general overall trends:

- levels of participation and interactivity show very little deviation between the positive and negative corpora,

- both corpora share a core of 41 of the top 50 most frequent words,

- from the top 50 frequency list, the following words are more frequent the positive corpus: *that, of, a, were, at, it's, be,*

- from the top 50 frequency list, the following words are more frequent in the negative corpus: *they, so, this, because,*

- *know, think* and *would* are significantly more frequent in the positive corpus

- the negative corpus contains almost five times as many non-minimal response tokens (for example, *right, okay, good*) than the positive corpus

- modal verb usage is different qualitatively, with *would* and *could* featuring more prominently in the positive corpus and *should* as a preferred option in the negative corpus.

More detailed follow-up analysis of these specific items, using concordance and cluster searches can identify patterns of usage associated with each context, and while it is very difficult to conclude that any or all of these are the determining cause for the positive of negative perceptions of the participants, the heightened awareness resulting from such analyses may lead to more careful attention to the ways in which we, as teacher educators, conduct this particular activity. Vásquez and Reppen (2007) created a similar corpus of teaching practice feedback in the first semester of their teacher education programme, used it as a tool for professional development with the tutors, and showed a significant improvement in interaction patterns in the second semester as a result of this type of activity. And these are just some of the potentials offered by the collection of corpora of our professional activities to use as evidence in our reflections and developments as teachers and teacher educators. Returning to the five criteria outlined earlier in this chapter, corpora illustrating local and international varieties have great potential to achieve impact through relevance for the specific contexts of student teachers when used in LTE programmes. Peripheral participation through the use of data emanating from a range of professional contexts allows users to experience a variety of experiences vicariously before they are immersed in full participation. This helps to develop confidence and reduce feelings of anxiety associated with fear of the unknown. Cognitive and

emotional development is equally facilitated through these types of corpora but especially through the use of POTTI type data by teacher educators who can easily focus on specific features and incrementally implement changes to their practices where appropriate.

Conclusion

At one end of the scale, using existing corpora in the development of LTE materials is perhaps the less difficult route, and can indeed be perfectly adequate for many types of activities as demonstrated in this chapter. And these often comply well with the key principles of developing confidence, drawing attention to specific features, requiring cognitive development and allowing a silent period. However, the question of impact, which is arguably one of the most important of the key principles we outline, based on relevance and usefulness, may not be fully satisfied in this way. Therefore, the teacher educator must consider two courses of action. The first involves adaptation. McDonough and Shaw (2003, p.85) provide a framework for the adaptation of materials which involves three stages. The initial stage involves identifying a need to localise, personalise, individualise etc. The second step entails deciding on the techniques by which this can be done, and they suggest adding, deleting, modifying, simplifying, and reordering. And thirdly, these tailored materials can then be applied to the relevant content areas, for example, teaching practice, grammatical awareness etc. Corpus materials, as they already exist in electronic mode and can be automatically reordered effectively, make adaptation a possible option. The second course of action available to teacher educators who want to develop corpus materials which achieve maximum impact, is to collect locally relevant corpora on which materials can be based. This is really the best course of action, and corpus resources

tend to go a long way in terms of their potential for exploitation on LTE programmes. Arguably, the larger and available corpora can be reasonably suitable for language awareness and acquisition purposes (with the exception of varietal features), but locally compiled corpora are most definitely preferable for pedagogic awareness and induction to the community of teaching professionals in which student teachers will find themselves in their future careers.

References

Akbari, R. (2007) 'Reflections on reflection: a critical appraisal of reflective practices in L2 teacher education', System 35(2), 192-207.

Amador Moreno, C. P., Chambers, A. and O'Riordan, S. (2006) 'Integrating a corpus of classroom discourse in language teacher education: the case of discourse markers', ReCALL 18(1), 83-104.

Arnold, J. (1999) Affect in Language Learning, Cambridge: Cambridge University Press.

Barlow, M. (2000) MonoConc Pro, Houston: Athelstan.

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

Brumfit, C. and Mitchell, R. (1989) 'The language classroom as a focus for research' in Brumfit, C. and Mitchell, R. (eds.) Research in the language classroom, London: Modern English Publications and the British Council, 3-15.

Cazden, C. B. (1988) Classroom Discourse: The Language of Teaching and Learning, Portsmouth, NH: Heinemann.

Chambers, A. (2005) 'Integrating corpus consultation in language studies', Language Learning and Technology 9(2), 111-125.

Chambers, A. (2007) 'Popularising corpus consultation by language learners and teachers' in Hidalgo, E., Quereda, L. and Santana, J. (eds.) Corpora in the Foreign Language Classroom, Amsterdam: Rodopi, 3-16.

Chambers, A. and Le Baron, F. (2007) The Chambers-Le Baron Corpus of Research Articles in French/Le Corpus Chambers-Le Baron d'articles de recherche en français, Oxford: Oxford Text Archive.

Chambers, A. and O'Riordan, S. (2007) 'Learning to teach through French from a corpus of classroom discourse: giving corrective feedback' in Conacher, J. E. and Kelly-Holmes, H. (eds.) New Learning Environments for Language Learning, Frankfurt: Peter Lang, 87-101.

Cheng, W., Warren, M. and Xun-feng, X. (2003) 'The language learner as language researcher: putting corpus linguistics on the timetable', System 31(2), 173-186.

Farr, F. (2005a) Reflecting on reflections: a corpus-based analysis of spoken post teaching practice interactions in an English language teaching academic environment, unpublished thesis (PhD), University of Limerick.

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

Farr, F., Murphy, B. and O'Keeffe, A. (2004) 'The Limerick Corpus of Irish English: design, description and application' in Farr, F. and O'Keeffe, A. (eds.) Corpora, Varieties and the Language Classroom. Special Edition of Teanga 21, Dublin: IRAAL, 5-29.

Golombek, P. R. (1998) 'A study of language teachers' personal practical knowledge', TESOL Quarterly 32(3), 447-464.

Grévisse, M. (1975) Le Bon Usage, Gembloux: Duculot.

Hauck, M. and Hurd, S. (2005) 'Exploring the link between language anxiety and learner self-management in open language learning contexts', European Journal of Open, Distance and E-Learning II, on-line.

Hawkins, M. R. (2004) 'Social apprenticeships through mediated learning in language teacher education' in Hawkins, M. R. (ed.) Language Learning and Teacher Education. A Sociocultural Approach, New York: Multilingual Matters, 89-110.

Hyland, K. (1998) Hedging in Scientific Research Articles, Amsterdam: John Benjamins.

Johns, T. (2002) 'Data-driven learning: The perpetual challenge' in Kettemann, B. and Marko, G. (eds.) Teaching and Learning by Doing Corpus Analysis, Amsterdam: Rodopi, 107-117.

Kallen, J. (1985) 'The co-occurrence of *do* and *be* in Hiberno- English ' in Harris, J., Little, D. and Singleton, D. (eds.) Perspectives on the English Language in

Ireland, Dublin: Centre for Language and Communication Studies, Trinity College, Dublin, 133-148.

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

Lee, D. and Swales, J. (2006) 'A corpus-based EAP course for NNS doctoral students: Moving from available specialized corpora to self-compiled corpora', English for Specific Purposes 25(1), 56-75.

Lyster, R. (2002) 'Negotiation in immersion teacher-student interaction', International Journal of Educational Research 37, 237-253.

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

McDonough, J. and Shaw, C. (2003) Materials and Methods in ELT, Oxford: Blackwell.

McEnery, T., Xiao, R. and Tono, Y. (2006) Corpus-Based Language Studies. An Advanced Resource Book, London and New York: Routledge.

McLoughlin, C. (1999) 'The implications of the research literature on learning styles for the design of instructional materials', Australian Journal of Educational Technology 15(3), 222-241.

Mendis, D. (2002) 'Giving instructions about an assignment.', [online], available: http://lw.lsa.umich.edu/eli/micase/Worksheet_GivingInstructions.pdf [accessed December 2002].

Nassaji, H. and Wells, G. (2000) 'What's the use of 'triadic dialogue'?: An investigation of teacher-student interaction', Applied Linguistics 21, 376-406.

O'Riordan, S. (2009) 'Etude de la métalangue grammaticale employée dans deux cours de FLE en Irlande: Analyse des recours à la première langue des apprenants', Mélanges 31, 57-78.

O'Sullivan, I. and Chambers, A. (2006) 'Learners' writing skills in French: Corpus consultation and learner evaluation.', Journal of Second Language Writing 15, 49-68.

Oxford, R. (1990) Language Learning Strategies: What Every Teacher Should Know, New York: Newbury House.

Renouf, A., Kehoe, A. and Banerjee, J. (2007) 'WebCorp: an integrated system for web text search' in Hundt, M., Nesselhauf, N. and Biewer, C. (eds.) Corpus Linguistics and the Web, Amsterdam: Rodopi, 47-68.

Schmidt, R. (1993) 'Awareness and second language acquisition', Annual Review of Applied Linguistics 13, 206-226.

Scott, M. (2004) Wordsmith Tools 4, Oxford: Oxford University Press.

Seedhouse, P. (1995) 'L2 Classroom Transcripts: Data in Search of a Methodology?', TESL-EL 1(4), 2-37.

Seedhouse, P. (2004) The Interactional Architecture of the Language Classroom: A Conversation Analysis Perspective, Oxford: Blackwell.

Seidlhofer, B. (2005) 'English as a Lingua Franca', English Language Teaching Journal 59(4), 339-341.

Simpson, R., Leicher, S. and Chien, Y.-H. (2007) 'What are 'spoken academic formulas'?', [online], available: http://legacyweb.lsa.umich.edu/eli/micase/ESL/FormulaicExpression/Definition.htm [accessed January, 2007].

Sinclair, J. M. and Coulthard, R. M. (1975) Towards an Analysis of Discourse. The English used by Teachers and Pupils, Oxford: Oxford University Press.

Tomlinson, B. (1998) 'Introduction' in Tomlinson, B. (ed.) Materials Development in Language Teaching, Cambridge: Cambridge University Press, 1-24.

Turnbull, M. (2006) 'Employons le français en français de base!', The Canadian Modern Language Review 62(4), 611-628.

Vásquez, C. and Reppen, R. (2007) 'Transforming practice: changing patterns of participation 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 within the context of community of practice', Language Awareness 16(3), 173-188.

Walsh, S. (2006) Investigating Classroom Discourse, London and New York: Routledge.

Wells, G. and Mejia Arauz, R. (2006) 'Dialogue in the classroom', Journal of the Learning Sciences 15(3), 379 - 428.

Wertsch, J. V. (1998) Mind as Action, Oxford: Oxford University Press.

Willis, J. (1998) 'Concordances in the classroom without a computer: assembling and exploiting concordances of common words' in Tomlinson, B. (ed.) Materials Development in Language Teaching, , Cambridge: Cambridge University Press, 44-66.

Wynne, M., ed. (2005) Developing Linguistic Corpora: a Guide to Good Practice, Oxford: Oxbow Books.

Yoon, H. and Hirvela, A. (2004) 'ESL student attitudes towards corpus use in L2 writing', Journal of Second Language Writing 13, 257-283.