Hong Kong’s New Senior Secondary (NSS) English Language Curriculum: Perspectives from Corpus Linguistics

May L.-Y. Wong
Department of Linguistics
University of Hong Kong

Abstract

This study explores the Hong Kong’s New Senior Secondary (NSS) English language curriculum and discusses potential implications for using corpus linguistics (CL) in a pedagogical context. While spoken and written corpora have mainly been used to facilitate language description, they can also have a place in the language-teaching context by providing learners with typical patterns of language use some of which are not open to intuition. Specifically, this paper discusses how corpora might aid the teaching of formulaic sequences and workplace English in the classroom, which is of prime concern in the task-based, learner-centred teaching approach upheld in the new curriculum for English.

Introduction: The Hong Kong Context

English has always held an official and very important position in Hong Kong. While Cantonese is spoken as a first language by the majority (over 95 per cent) of the population (Census & Statistics Department, 2006) and has long been viewed as the language of solidarity and community ties (Cheung, 1985; Lai, 2009), English is seen as the language of success leading to higher education and better career prospects (Evans, 2009). It is also valued as a global language, and thus if Hong Kong has to gain a firm foothold in the international economy, good English skills among its workforce are considered to be essential. Given these circumstances, the Hong Kong government continues to stress the importance of English in its language policy for education after the return of Hong Kong to China, as evidenced in the compulsory benchmarking of all English language teachers (Qian, 2008). This is also reflected in the recent reform of the senior secondary English curriculum launched in 2009. The reform is geared towards achieving greater efficiency in language teaching so that learners are able to communicate well in English and perform better in higher education and in the workplace.

This paper focusses on the secondary education system in Hong Kong, with particular reference to the integration of corpora and corpus evidence into the new curriculum for English. Corpora are structured collections of texts from spoken and written sources (and of English in particular) in the electronic form. As will be outlined in the section on corpus linguistics (CL) and English language teaching (ELT), teaching based on corpora entails the use of information technology (IT) in the language classroom by teachers and/or students, which is generally in line with the government directions for education. Since the late 1990s, the government has called for ‘at least 25 per cent of the curriculum to be supported through IT’, as stated in Hong Kong’s first Chief Executive’s first policy address (Tung, 1997). A substantial amount of resources has been allocated to the development of IT facilities and support in Hong Kong schools, including multimedia laboratories, computers, technical staff,
and IT training programmes. However, reactions to IT applications in language teaching are mixed. The corpus linguistic approach to teaching English described in this paper therefore tries to examine how IT can contribute to the teaching and learning of English in the secondary school context. Teachers are encouraged to evaluate the possibilities that CL offers and arrive at their own decisions about the value and potential of corpus-based language pedagogy.

**New Senior Secondary (NSS) English Language Education: An Overview**

The New Senior Secondary (NSS) curriculum, which has come into effect since September 2009, was designed to provide greater flexibility for Hong Kong schools to cater for individual learners’ varied interests, needs and capabilities. The NSS framework was a product of concerted efforts by educational academics, secondary head teachers, practising teaching professionals and officers from the Education Bureau. It is based on the key recommendations made in the recent curriculum changes documented in four government publications, namely, *Learning for Life, Learning through Life* (Education Commission, 2000), *Learning to Learn—the Way Forward in Curriculum Development* (Curriculum Development Council [CDC], 2001), *Basic Education Curriculum Guide—Building on Strengths* (CDC, 2002), and *Senior Secondary Curriculum Guide* (CDC, 2007). The new curriculum is composed of nine Key Learning Areas (KLAs), of which English language education is the prime interest of this paper.

The English language education component of the NSS curriculum is based on a constructivist view of learning (CDC & Hong Kong Examinations and Assessment Authority [HKEAA], 2007). Its main intentions are summarised in Appendix 1. These important intentions or government’s goals will be dealt with thoroughly in the subsequent sections where a corpus-linguistic approach to teaching English is introduced with a view to realising these goals. However, before we proceed to these sections, it should be helpful to explain the potential link between the current curriculum reform and the corpus-based teaching approach. In the following two subsections, a brief description of two of the curriculum objectives—i.e. (i) facilitating self-study and lifelong learning among secondary school students, and (ii) promoting learner autonomy (see CDC & HKEAA, 2007, pp. 4 & 56) — is outlined and the feasibility of corpus-linguistic teaching in achieving these objectives is discussed.

**Corpus-related Curriculum Objectives**

CL becomes relevant to ELT when it comes to analysing authentic texts with computer software and developing learner autonomy through learning from online language resources as an out-of-class activity. Text analysis programs (e.g. concordancers; see the next section for a description) are cited in the curriculum and assessment guide as one type of computer-based resources for self-access work; they are exploited ‘for studying how words are used in authentic texts’ (CDC & HKEAA, 2007, p. 94). As far as the effective use of IT for teaching purposes is concerned, the essence of self-access learning, as described in the guide, is that it should ‘provide opportunities for learners to take charge of their own learning through selective use of online resources’ (p. 97). One kind of online resources that can be used by English
teachers are online corpora, as will be shown in the section on corpus applications in the teaching of formulaic expressions. Online corpora can then be used by students with adequate training.

Another aspect of the curriculum stated in the guide which is related to CL is the creation of a learning community where ‘teachers and learners become partners/joint investigators in the process of developing knowledge’ (p. 100). As will be demonstrated in the sections on corpus applications in teaching formulaic expressions and workplace English, corpus-based activities usually call for learners’ active participation in a knowledge-constructing process. A corpus linguistic approach to teaching suggests that teachers are not the sole source of knowledge or expertise (Hunston, 2002). All participants in the learning community have a joint purpose, which they pursue and negotiate together. In this approach, knowledge is co-constructed by the participants who have something to contribute and something to learn. In other words, a student can be at various points both a teacher and a learner. Since corpus-based learning aims to develop collaboration between teachers and students in constructing knowledge (see the notion of ‘distributed knowledge or expertise’ in Graves [2008, p. 167]), this is largely embodied in the idea of a learning community enshrined in the new curriculum.

**Authentic Materials and Autonomy in Language Teaching and Learning**

Authenticity is vital to language teaching as it is always regarded as a motivating force for learners. Various studies have suggested that authentic materials are intrinsically more interesting than examples of contrived, isolated texts (see, for example, Little, Devitt & Singleton, 1989; Little & Singleton, 1991; Swaffar, 1985). While the former is intended to communicate a real message, the latter highlights a specific form of the target language that shows no direct relationship with a genuine communicative need and renders itself far less motivating for learners.

In the NSS English language curriculum, authentic materials are regarded as ‘cultural artifacts like books, newspapers, magazines, radio and TV broadcasts, web sites, advertising, music and so on’ (CDC & HKEAA, 2007, p. 107). This kind of discourse has been included in an elective component containing four language arts modules (dramas, short stories, poems and songs, and popular culture). This corresponds to some studies (e.g. Benson & Voller, 1997; Morrow, 1977; Nunan, 1988, 1998; Swaffar, 1985) which categorise texts with a real communicative objective as authentic. As Morrow (1977, p. 13) puts it, ‘[a]n authentic text is a stretch of real language, produced by a real speaker or writer for a real audience and designed to convey a real message of some sort’. Corpus linguists define authenticity in the same way as Morrow and they have built both corpora of native speaker’s English (e.g. the British National Corpus) and those of learner English (e.g. the International Corpus of Learner English). In their view, authentic texts are not the privilege of the native speakers. Rather, they can be produced by any speaker/writer (native or non-native user of a particular language) in a real communicative context.

In addition to authenticity, CL focusses on autonomy, which lends itself well to teaching languages. Autonomy is defined as ‘the ability to take charge of one’s own learning’ (Holec, 1981, p. 3), which involves planning, selection of materials and resources, monitoring learning progress and self-assessment (p. 4). Autonomy is also
a matter of degree. Littlewood (1999, p. 75) distinguishes between ‘reactive’ autonomy and ‘proactive’ autonomy. The former deals with some degree of learners’ individuality under explicit instructions and sets up directions of learning which learners have partially created, whereas the latter shows a high degree of autonomy with which learners are able to organise their resources autonomously in order to reach a self-initiated goal.

Existing educational norms in Hong Kong, however, seem to be at odds with the above two levels of learner autonomy (Carless, 1999). Lee and Ng’s (2009) study reveals that over the past two decades teachers have dominated talk in the classroom, while secondary school students are never involved in planning and selecting teaching materials, seldom seek clarification or pose questions, and respond to teacher questions minimally. Interestingly, findings of some studies into Hong Kong students’ responses to the idea of autonomy have indicated that local students do value freedom and the opportunity to direct their own learning (Braine, 2003; Chan, Spratt & Humphreys, 2002; Littlewood, 1999, 2000). Learner autonomy is beneficial to language acquisition because not only does it recognise learners’ individuality and willingness to engage in language learning processes, but it can also enhance motivation (Benson, 2006; Ryan & Deci, 2002).

Corpus Linguistics and English Language Teaching

CL is an approach which generally involves three key stages: extraction of data from large banks of naturally occurring texts, processing the output (i.e. sorting out data according to specific needs), and interpretation of output. A concordance is the output of using a piece of corpus analysis software to search for all occurrences of one word/phrase in a corpus. It takes a keyword-in-context (KWIC) format in which all the occurrences are presented with the node word/phrase (i.e. the search term/phrase) in the centre of the line, with a predefined number of words in the context presented at either side of the node word/phrase. The first and second stages are usually assisted by using appropriate software. Corpus software such as WordSmith Tools (Scott, 2009) can afford a suite of functions, including producing wordlists, counting occurrences of individual search terms, (re-)sorting data for identifying meaningful patterns, and producing collocation or cluster lists. The software can also provide a keyword tool which allows for a comparison of lexico-grammatical items between corpora to identify words with significantly higher or lower frequencies.

A useful distinction between a ‘corpus-based’ approach and ‘corpus-driven’ approach has been developed in CL and language teaching. While a corpus-based approach uses the corpus to test theoretical frameworks which are typically formulated from introspection, or analyses the corpus data on the basis of these theoretical frameworks (McCarthy, 1998), a corpus-driven approach does not presuppose any frameworks or taxonomies and emphasises that theoretical statements are a product of corpus evidence (Tognini-Bonelli, 2002). The latter approach is favoured by those working in a purely inductive data-driven learning tradition. According to Johns (1991), ‘data-driven learning’ (DDL) means that the language learner is at the same time a language researcher, being able to have available abundant examples of authentic language samples, analyse them and construct knowledge in an environment which is stimulating, interesting, playful and interpersonal. Such an approach echoes the changing roles of the teacher and the
student in the new syllabus for English as outlined in the previous section. The teacher becomes a facilitator of language study instead of being seen merely as the knowledge transmitter or language expert, and the student is imbued with a new role as a language investigator in addition to that of a language learner. In this study, the corpus-linguistic teaching under consideration is taken to include both corpus-based and corpus-driven approaches in order to suggest as many potential areas as possible in which CL would be useful for language teachers and learners. Those areas to which CL has contributed greatly are to be explored in the following section.

Pedagogical Applications of Corpora

The impact of corpus linguistic research on language teaching has been immense (McCarthy, 2001; Murison-Bowie, 1996) since the past decade when researchers in ELT (e.g. Barlow, 2000; Davies, 2000, 2004; Kirk, 2002; Osborne, 2000; Partington, 2001; Peters, Picchi & Biaagini, 2000; Römer, 2004, 2005; Wang, 2000) strongly criticised traditional approaches based on out-of-context, fabricated sentences supplied by teachers from intuition and/or found in most textbook exercises, and consequently called for the use of corpora in language teaching. It has led to improved dictionaries (e.g. Gillard & Gadsby, 1998) and grammar reference materials (e.g. Carter & McCarthy, 2006) as well as pedagogically oriented books (e.g. Aston, 2000; Burnard & McEnery, 2000; Sinclair, 2004; Wichman, Fligelstone, McEnery, & Knowles, 1997). Many online corpus resources, activities and discussion groups have also sprung up in recent years (see David Lee’s website 2 for a selection of ‘teaching and miscellaneous links’).

Corpus-linguistic research has successfully revolutionised ELT in that it has led to the discovery of patterning that differs from the prescriptive models of the English language and thus highlights the drawbacks of relying solely on the intuitive models of language in use usually found in school textbooks. Indeed, numerous studies have demonstrated the mismatch between actual language use and the language presented in textbooks. Holmes (1988), for example, compares epistemic modality in ESL textbooks with that in corpus data and states that modal verbs are given too much attention in the textbooks at the expense of alternative linguistic strategies for expressing possibility used in real-life data. Kettermann (1995) also makes a case for such a mismatch in which the backshift rule for tenses in reported speech constructions is not always followed in authentic language use, in contrast to the prescription in pedagogical grammars. Boxer and Pickering (1995) look at speech acts in textbook dialogues as compared with real, spontaneous encounters found in a corpus. Carter (1998) conducted a similar study based on a different corpus (the Cambridge and Nottingham Corpus of Discourse in English [CANCODE]) and notes that dialogues from textbooks are completely void of core spoken language features such as ellipsis, hedges, discourse markers and vague language. Based on the same corpus, McCarthy (1998) points out that the past perfect verb forms have a broader and more complex function in spoken discourse than previously described. Recently, Römer (2006) compares the use of progressive forms (e.g. ‘re looking) in huge collections of spoken British English with that in a small corpus of ‘spoken-type’ texts from German EFL textbooks and reports on a lack of naturalness or authenticity in the latter. Cheng and Warren (2007) contrast the interactional strategies of checking understanding in school textbooks with real life communications found in the Hong Kong Corpus of Spoken English (HKCSE). Their findings suggest that
textbook writers should have more awareness of the realities of language use. These and many other findings from corpus research as well as work done on how to exploit and evaluate corpus use in the language classroom (e.g. Chambers, 2005; Tribble, 2000) consistently point to the fact that corpora and corpus evidence have made a profound contribution to the field of ELT, by deepening our understanding of lexico-grammatical units, collocation, colligation and language patterning in general.3

However, almost every approach to language teaching and learning has its ups and downs, and CL is no exception. While McCarthy (2001, p. 128) states that ‘[t]he language of the corpus is, above all, real, and what it is that all language learners want, other than “real” contact with the target language’, many practitioners and applied linguists have cast doubt on the usefulness of adopting corpus-based materials in the language classroom (see, for example, Cook, 1998; Owen, 1996; Prodromou, 1997a, 1997b; Seidlhofer, 1999; Widdowson, 2000). Their arguments centre on the type of data that corpus texts represent and how far ELT syllabuses and materials should be ‘corpus-driven’ in order to better reflect linguistic reality. As Widdowson (2000, p. 7) puts it, using ‘real’ corpus data does not necessarily mean more ‘reality’ in the classroom. He further argues that corpus findings represent third- rather than first-person reality and provide just a ‘partial description’ of ‘decontextualised language’.

Meanwhile, the pedagogical value of language corpora should not be undermined in the light of the benefits they bring to English language description as noted above. Equally importantly, the role of the teacher as a facilitator of language learning should not be ignored:

[I]n the end it is teachers who will engage in the process of recontextualising corpora and any useful findings from corpus-based description. It is teachers who will mediate between corpus-based content and the needs of the learners in their individual classroom contexts. (O’Keeffe & Farr, 2003, p. 391)

If acquiring communicative competence is of paramount importance as communicative language teaching has traditionally held, the focus should be shifted away from reality:

[T]he question is not whether corpora represent reality, but rather whether their use can create conditions that will enable learners to engage in real discourse, authenticating it on their terms—and whether this engagement can lead to language learning. (Gavioli & Aston, 2001, p. 240)

Language corpora can indeed be a useful resource for teachers and learners. Farr (2008) outlines the findings from survey results which examine 25 student teachers’ perspectives on their experiences of using corpora; there is generally a positive predisposition towards the use of corpora. In Cheng, Warren and Xu’s (2003) study, undergraduate English language majors who were trained to become language researchers and engaged in corpus-driven mini-research projects found both project experiences and outcomes interesting and useful. While we have already seen some success in using corpora at the university level, no attempt has been made to adopt a corpus-linguistic teaching approach in the secondary school context. The next two
sections therefore aim to provide some examples of how corpus data can be integrated into the NSS curriculum, focusing on formulaic expressions and workplace English, both of which are two key features in the new curriculum.

**Corpus Applications in Teaching Formulaic Expressions**

Formulaic language is a key element in the communicative competence of native speakers (Foster, 2001) and, therefore, has important pedagogical implications. R. Ellis (2005) stresses that the development of L2 proficiency depends on both formulaic expressions for fluency and knowledge of specific grammatical rules for accuracy (see also Skehan, 1998; Wray, 2000, 2002). N. Ellis (1997, p. 129) argues that ‘speaking natively is speaking idiomatically using frequent and familiar collocations, and the job of the language learners is to learn these familiar word sequences’. In the Hong Kong’s NSS English curriculum, 21 out of a total of 26 ‘language items’ which ‘learners need to develop as they perform the communicative functions’ are concerned with formulaic expressions (CDC & HKEAA, 2007, p. 1321; see also Appendix 2 for two of these major functions illustrated in the curriculum guide), explicitly indicating that these prefabricated patterns and routines are regarded as a fundamentally important part of learners’ communicative repertoire. Despite the importance of these recurring combinations of words, textbooks rarely deal with them in a systematic way. The scarce attention given to formulaic chunks in the classroom might be explained by the fact that teachers find it difficult to select the right formulaic sequences to present to learners.

Corpus-based evidence can offer new insights into how words are distributed in a language and allow learners to explore more deeply into the nature and extent of the use of formulaic sequences. A prominent example is Schauer and Adolphs’s (2006) study on expressions of gratitude (e.g. thanks, cheers, thank you, and thank you very much) and pedagogical implications for using corpus evidence in ELT. Their work is now discussed with greater detail as it forms the basis for the sample corpus-based tasks for teaching formulaic sequences.

Expressions of gratitude are one of the discourse functions that learners are likely to encounter in a variety of contexts (Coulmas, 1981, as cited in Schauer & Adolphs, 2006). Although these expressions appear to comprise a closed set, language learners are often unable to produce correct formulaic replies when expressing their gratitude, highlighting yet another failure of textbook English. According to Schauer and Adolphs, learners are not aware of the single expressions cheers or ta as casual synonyms for thanks; nor do they realise that the routinised utterances of expressing gratitude can take several various forms. Their corpus data taken from CANCODE have revealed three commonly used patterns by native speakers: (i) thanking + complimenting interlocutor/positive evaluation of previous speaker’s utterance, e.g. Thank you. That’s very good of you; (ii) thanking + stating reason, e.g. Thanks for coming; (iii) thanking + refusing, e.g. Ah, no, it’s okay thanks (pp. 127–130). The last category is the most common in the data and suggests that learners may need to possess the ability to express gratitude and at the same time to refuse a proposition as one of the main communication skills.

To connect theory with practice, and to address one set of communicative functions of formulaic expressions stated in the government’s NSS curriculum guide
i.e. to request, offer, accept and decline help (see Appendix 2), I have devised two tasks that incorporate the use of corpus evidence for raising students’ awareness of effective thanking strategies. The tasks centre on the stem thank (Figure 1) and cheers (Figure 2). The sample of concordance lines for these two tasks was produced with the British National Corpus (BNC) using the BNCweb (n.d.) interface. The BNC is a 100-million-word principled collection of naturally occurring text representative of the spoken and written English used in Britain at the end of the 20th century. It includes 90 million words of written English from eight genres (80% informative prose; 20% imaginative prose) and 10 million words of spoken English both contextually-governed (e.g. lectures, news broadcasts, parliamentary proceedings, and radio chat shows) and demographically sampled from four social class groupings and 38 geographical regions in the UK. Because of its size and grammatical tagging, the BNC is widely considered as one of the most significant research tools currently available for corpus-based and corpus-driven studies of British English. To illuminate classroom application of corpus findings, Figure 1 displays a task that focusses students on formulaic sequences, i.e. thanks and thank you, commonly used for expressing gratitude, and Figure 2 presents examples with cheers and aims to develop a sense of questioning about what textbooks say about language use in context.

Discover-for-yourself Task
How do people say thank you?

Below are concordance lines for the expressions of gratitude thanks and thank you.
(a) Identify words or word sequences that come before/after thanks and thank you which are closely related to the expression of gratitude.
(b) What is the preposition that is often used with thanks and thank you?
(c) Find examples for the five major categories of expressions of gratitude:
   i) Thanking + Ø
   ii) Thanking + adverbs e.g. very much
   iii) Thanking + another thanking
   iv) Thanking + praising the other speaker
   v) Thanking + giving reason

<table>
<thead>
<tr>
<th></th>
<th>reminiscences from people in the audience. Thank you very much indeed.</th>
</tr>
</thead>
<tbody>
<tr>
<td>2</td>
<td>I am here taking notes of people's comments thank you. Thank you. Okay. Can</td>
</tr>
<tr>
<td>3</td>
<td>see that standard maintained. Well thank you for that's a very good start to the evening Thank you. Jan &lt;name&gt; erm I</td>
</tr>
<tr>
<td>4</td>
<td>good start to the evening. &lt;laugh&gt;. thank you Jan &lt;name&gt; erm I</td>
</tr>
<tr>
<td>5</td>
<td>reward to people who live in the town. Thank you. I think there's another way of</td>
</tr>
<tr>
<td>6</td>
<td>spoken to. Yeah. Thank you. Right has the board consider giving this place</td>
</tr>
<tr>
<td>7</td>
<td>say please fill in your short questionnaire thank you for your attendance and I can assure you as Chairman</td>
</tr>
<tr>
<td>8</td>
<td>please ladies. &lt;unclear&gt;. Thank you all very much. Thank you.</td>
</tr>
<tr>
<td>9</td>
<td>Filling up a blank week. Thank you Pauline, very, that er very good report on</td>
</tr>
<tr>
<td>10</td>
<td>that been interesting for you. Thank you. Very nice. &lt;clapping&gt;.</td>
</tr>
<tr>
<td>11</td>
<td>&lt;unclear&gt; you know keep it. Well thank you. There's, there's another thing I</td>
</tr>
<tr>
<td>12</td>
<td>Yeah, your off the hook. Thank you very much, thank you Chris. &lt;clapping&gt;.</td>
</tr>
<tr>
<td>13</td>
<td>&lt;unclear&gt;. They &lt;unclear&gt;. Thanks a lot. Well now, now you know</td>
</tr>
<tr>
<td>14</td>
<td>the scenes. That's right. thanks a lot Chris.Okay &lt;unclear&gt;. That's</td>
</tr>
</tbody>
</table>
Okay <unclear>. That's great. thanks very much. Right, were now on your pen and I'll write that down. Thanks . Are Red Cross still taking furniture <unclear> Yeah. Thanks ever so much. Because er <pause> Caroline <pause> to have it. No I won't thank you . Well there's one at erm <pause> <unclear> can always divide up. <unclear> . See you then. Thanks Katherine. Your welcome. I'll see you all on Wednesday <unclear> thanks for coming. <pause> Erm <pause> Bye Liz Oh.okay. Thanks . okay. Thanks. Thank you very much. They say Wednesday erm activity at Latton Bush, thanks for sending that information, I've sent off the would be interesting <unclear dur=3 > Thank you very much you fascinate, battling against others.

**Figure 1.** Sample material based on the British National Corpus for *thanks* and *thank you*.

With respect to category (c)(ii) above, it is worth noting that in the context of expressing gratitude in spontaneous conversations, ‘gratitude clusters’ usually occur. These clusters are referred to as ‘sequences and lexical items of gratitude linked and often repeated in a single turn, as well as across turns’ (Schauer & Adolphs, 2006, p. 126). This is because the next speaker often employs *thank you* as a response to the previous speaker’s expression of gratitude. Additionally, the use of *cheers* merits special attention from language teaching professionals because in most teaching materials it is usually not considered as an expression of gratitude employed by native speakers of English (Schauer & Adolphs, 2006, p. 125). Rather, it is taught as an expression to mark the raising of glasses during dinner parties and other festivities for a toast. In the corpus data, *cheers* is used both as part of a formulaic sequence for expressing gratitude and as a responder to an expression of gratitude as well as functions as a discourse marker that signals the end of an encounter or discourse episode.

**Discover-for-yourself Task**

**When do people say *cheers***?

Below are concordance lines for the word *cheers*.

(a) How often does *cheers* occur with words or word sequences that express gratitude, and how often is it said at leave-taking?

(b) Identify its different uses from these examples.

(c) Do you notice any differences between textbooks and corpus data in explaining the meaning of *cheers*?

<table>
<thead>
<tr>
<th>1</th>
<th>often the situation is one that</th>
<th>Cheers</th>
<th>! Ha! Thank you!</th>
</tr>
</thead>
<tbody>
<tr>
<td>2</td>
<td>so you could Thanks very much.</td>
<td>Cheers</td>
<td>Dad. Put away your er luggage. &lt;pause&gt;</td>
</tr>
<tr>
<td>3</td>
<td>there. Yeah Right,</td>
<td>cheers</td>
<td>.See you next week or before then.</td>
</tr>
<tr>
<td>5</td>
<td>Okay &lt;unclear&gt;, &lt;unclear&gt;</td>
<td>Cheers</td>
<td>. Bye. Could be interesting</td>
</tr>
<tr>
<td>6</td>
<td>that's grand, thanks, Melvin,</td>
<td>cheers</td>
<td>.yes, yes, okay</td>
</tr>
<tr>
<td>7</td>
<td>. Yeah good. Okay</td>
<td>cheers</td>
<td>thanks a lot for your help. Okay</td>
</tr>
<tr>
<td>8</td>
<td>knock when you've finished here.</td>
<td>cheers</td>
<td>.Bye, &lt;unclear&gt; &lt;unclear&gt; Oh what er</td>
</tr>
<tr>
<td>9</td>
<td>cards. Thank you.</td>
<td>cheers</td>
<td>. Keep well. Yes</td>
</tr>
<tr>
<td>10</td>
<td>. Keep well. Yes</td>
<td>cheers</td>
<td>.He's got See you</td>
</tr>
<tr>
<td>11</td>
<td>.Two pound two change thanks very much</td>
<td>cheers</td>
<td>now. Six size three please.I</td>
</tr>
</tbody>
</table>
er, about two o’clock? Magic! Cheers

now. Bye <“telephone conversation ends”>! It’s enough? Yeah. Okay. Cheers!

! Yeah. Bye <“telephone conversation ends”>! Okay. Cheers . thanks. Aye, we were gonna

you later. Right. Cheers . Bye!


the thick stuff out of the bottle. Cheers . Oh that’s ver-- very nice.

Right now. Cheers ! Cheers! Mm! Thank you

now. Cheers! Cheers Mn! Thank you so much for <unclear>.

Yeah see you again. Cheers then. See you again.

Figure 2. Sample material based on the British National Corpus for cheers

Corpus Applications in Teaching Workplace English

Workplace communication seems to be introduced for the first time by the Hong Kong government as an elective module to raise students’ awareness of the practical needs of learning English. This corresponds to the work of a number of researchers who have investigated the use of English as a lingua franca (ELF) in international business contexts, with an increasing concern with those strategies that are closely associated with effective communication in business. As Nickerson (2002) so eloquently describes,

there has been an increasing acknowledgement in their [other researchers’] work that there needs to be a shift from a focus on language proficiency, i.e., from a concern with how we can ensure that non-native speakers of English sound more like native speakers of English, to a concern with teaching the rhetorical strategies, and their linguistic realisations, that have been identified as effective strategies through the investigation of various business genres. (p. 374)

Not surprisingly, research on business discourse has often focussed on the strategic nature of the discourse, by Bilbow (2002) and Rogerson-Revell (1999) for instance, for meetings in Hong Kong. Bilbow’s study looks at commissive speech acts such as promises and statements of commitment, discussing when they are used and how they are realised linguistically, by native (Western) speakers and non-native (Chinese) speakers of English, and across different types of meetings in a large multinational airline corporation. Rogerson-Revell, based on the findings from an international airline corporation, discusses interactive strategies such as procedure-related strategies (for managing a talk) and message-related strategies (for indicating how the content is to be understood).

The change in emphasis in the literature, from proficiency to strategic communication, might perhaps be a bit problematic for the Hong Kong’s new English syllabus. The learning objectives of the newly-introduced module are largely proficiency-oriented rather than strategy-oriented as they tend to enhance English proficiency without explicitly stating ways of doing so: ‘to familiarise learners with the different types of workplace correspondence’, and ‘to develop learners’ understanding of the vocabulary, language, formats, styles and conventions used in spoken and written communication in the workplace’ (CDC & HKEAA, 2007, p. 47).

Other potential problems are related to the availability of appropriate teaching materials. Firstly, most of the published materials tailored for the Hong Kong’s
business world developed by local researchers/practitioners are geared to the English language needs of business students at the tertiary level (e.g. Bhatia & Candlin, 2001; Chew, 2005; Jackson, 2004), which might not be well suited for the secondary school context. Secondly, those pedagogic materials outside Hong Kong for the teaching of English for Specific Business Purposes (ESBP) reflect a global concern that they regretfully lack a link with empirical research in business discourse reported in scholarly publications. A brief survey by Nickerson (2005, pp. 375–376) of recent ELT business English books reveals that none of these books refer to the findings of empirical research in their descriptions, with only two exceptions which are derived from corpus-based research, namely Collins Cobuild (2004) and Mascull (2004). Collins COBUILD’s Business Vocabulary in Practice (2004) provides comprehensive coverage of business vocabulary illustrated with thousands of examples of real English from the Bank of English, to help students write, speak, and understand English better, whereas Mascull’s (2004) Advanced Business Vocabulary in Use has been carefully researched using the Cambridge International Corpus to ensure that the 2,000 new words and expressions in the book represent the English ones that native speakers actually use. However, both of these books are intended for advanced learners of English who need vocabulary for business and professional purposes. Given that currently available resources may not suit the specific needs of Hong Kong secondary schools, teachers may consider developing their own materials with the help of a corpus for the teaching of business language.

Perhaps a free useful online resource that secondary school teachers can turn to is the Business Letter Corpus (BLC) Online KWIC Concordance. The corpus can be used to search for terminology that is commonly used in business correspondence, for example, the word grateful, which is frequently employed in business writing as a politeness strategy, typically followed by an if-conditional for making a request as in I would be grateful if you could ... (cf. Kong, 2006). Given corpus examples of how native speakers use this word in real-life business contexts, teachers can devise tasks to explore these examples with their students, focussing on different lexicogrammatical aspects of language use in context: for example, the prepositions for and to that can be used with grateful; the following if-clause, that-clause and non-finite clause introduced by to-infinitive; the adverbs that modify grateful, e.g. very, most, extremely, so, and deeply (see Appendix 3 for illustrative examples in an extract of concordance lines of grateful taken from the BLC via the web-based concordancer).

Discussion

In the above sections, we have seen two illustrations of how the teaching of formulaic expressions and workplace English can benefit from the support of CL. There are at least three advantages of employing a corpus linguistic teaching approach. Firstly, the corpus exploration work enables learners to foster an enquiring and critical outlook towards the description of English that they encounter in the classroom, textbooks and reference materials which do not always correspond to English in use. Secondly, sharing discoveries from corpus data provide learners with motivation for participating in interaction because they have information to contribute and a real experience to talk about. This is a primary goal of task-based language teaching (TBLT). Pairs of learners can investigate a concordance together and discuss their findings. Different learners can examine different concordance lines selected at random. By reporting and comparing their analyses, they will be engaged actively in
using the target language to negotiate their findings. This is in line with what TBLT preaches: it dedicates more attention to designing tasks where learners can use the language for genuine communicative purposes than to drilling the learners in grammatical rules described in textbooks. Discussing corpus data therefore helps to engage learners in both processes of communication and of learning. Thirdly, the task of exploring the corpus can also provide the basis for self-directed learning. As well as being given tasks which direct them at specified items, learners can explore and find out about formulaic sequences on their own. The teacher can assist this process by suggesting and illustrating possible activities, asking questions that can stimulate probing investigation and reasoned interpretation, and giving constructive feedback during/after the task. These strategies accord completely with the quality interaction approaches recommended by the government in its new curriculum guide, i.e. ‘scaffolding’, ‘questioning’ and ‘feedback’ (CDC & HKEAA, 2007, pp. 99–100). In fact, this interactive orientation appears to be mirrored in the education literature, since learners who have both high challenge and high support are believed to be able to learn most effectively in both Bruner’s (1983) model of ‘scaffolding’ and Vygotsky’s (1978) notion of zone of proximal development (ZPD). In the ZPD model, learning only takes place when the challenge of a task is just beyond the learner’s level of competence and it can only be achieved with support. Some researchers tend to define scaffolding more broadly, as operating at two levels (Hammond & Gibbons, 2005; Swain, Brooks & Tocalli-Beller, 2002). For example, Hammond and Gibbons (2005) state that students can work within their ZPD with carefully designed planning, selection and sequencing of teaching materials (i.e. at the macro ‘designed-in’ level); at the micro ‘interactional’ level, teachers and learners jointly construct knowledge and skills in the classroom. In both narrow and broad sense, scaffolding means that learning is a joint activity and a constructive interaction between the teacher and learners, and the corpus linguistic teaching approach, by encouraging learners to be researchers and teachers to be facilitators, works exactly in this spirit.

Corpora will become a good resource for learners to use directly, inside and outside the classroom. With respect to ELT, there has been increasing recognition that learners develop skills of learning how to learn (Gardner & Miller, 1997). For this reason, the Hong Kong government has, in its new English language curriculum for senior secondary schools, promoted independent and lifelong learning as one of its key curriculum development initiatives. Corpora provide a good opportunity for learners to work independently on language problems outside schools so that they are able to cope effectively with the challenges of learning English later in their lives.

To ensure effective use of corpora by learners, Gavioli and Aston (2001) have insightfully pointed out three main requisites. Firstly, learners need to be able to access a wide collection of corpora that reflect different language use conditioned by different types of text, topic and genre. Large corpora such as the Bank of English and the British National Corpus may prove useful. Secondly, learners should have access to user-friendly concordancers such as MicroConcord (Scott & Johns, 1993) and AntConc (Anthony, 2008) to interrogate corpora. This requirement might not be easily met in the context of language teaching at the secondary school level in Hong Kong, where anecdotal evidence has shown that IT support for everyday classroom use is still rather limited. However, each school can consider its multi-media computer room which English teachers can use for corpus-based activities. Thirdly,
corpus-based tasks need to be carefully designed in terms of their selection and grading. To this end, various degrees of mediation are involved. Learners can begin with corpus texts and concordances selected and pre-edited by teachers (Tribble & Jones, 1990). As Gavioli and Aston (2001) remark, ‘[t]his may help learners pass gradually from “easy” concordances where recurrent patterns are evident, to those which call for more complex categorisation and interpretation’ (p. 244). In addition, corpora are graded before they are used in the classroom. Learners move from small collections of texts of a similar type probably created by teachers to larger and more heterogeneous ones which require more complex analysis (see Ghadessy, Henry & Roseberry, 2001, for illustrative case studies).

**Conclusion**

The decision of integrating CL into the classroom lies with all stakeholders—policy makers, curriculum planners, materials designers, teacher educators, teachers and learners. Top-down support from higher administration is essential in curriculum innovation; in the case of Hong Kong, the government has started to recognise the pedagogic value of CL in language teaching, as delineated in the second section of this paper. However, the new curriculum needs to be adapted to or modified by the realities of the local classroom. Naturally teachers play a decisive role in the process. As O’Keeffe and Farr (2003) note,

> [t]o do this [to effectively integrate CL into the new curriculum], teachers will need to be able to make informed decisions, and, not least of all, they will need to be able to access the validity of the arguments that are made in relation to corpus findings and corpus use. (p. 391)

This statement clearly necessitates corpus integration into teacher education courses (see also Chapelle, 2001; Conrad, 2000). However, few published accounts of this kind of integration have been documented to date. Hunston (1995) illustrates some of the ways in which teachers of mother-tongue English on grammar awareness courses learn how to use corpora. Renouf (1997) reports on the corpus component of a Language Teacher Education (LTE) course for postgraduate students at the University of Birmingham. Similarly, in Hong Kong, Coniam (1997) discusses his teacher training programme which contains corpus-based tasks for language investigation. O’Keeffe and Farr (2003) provide a practical account of how the applications of corpus-based techniques in LTE increase student teachers’ understanding of word classes, register-related grammatical choices and socioculturally conditioned grammatical patterns, and most recently, Farr (2008) evaluates the use of corpus-based activities on teacher education programmes. Hence, if the Hong Kong government is to incorporate corpus-based instruction into its new English language curriculum for senior secondary schools, it has to implement necessary changes in teacher education.

It is hoped that corpus linguistic teaching can find its way into the ELT classroom in the secondary school context. Meanwhile, more research needs to be carried out, for example, to investigate teachers’ and students’ willingness to use corpora in the teaching and learning of English, which would be an interesting topic for a follow-up article.
The nine KLASs are Chinese language education, English language education, mathematics education, personal, social and humanities education, science education, technology education, arts education, physical education and liberal studies.

Palmer (1933, p. i) defines a collocation as ‘a succession of two or more words that must be learned as an integral whole and not pieced together from its component parts’, whereas Firth’s (1957, p. 195) collocational meaning, one of the most widely cited definitions in the literature, is: ‘[y]ou shall know a word by the company it keeps’. Hence, collocation refers to the fact that certain words are more likely to occur in combination with other words in certain contexts (e.g. lexical words such as waste, devote, spend, and spare are more likely to occur with time than other words). A form of collocation which involves relationships at the grammatical level rather than the lexical level is colligation. For example, nouns tend to colligate with adjectives whereas verbs tend to colligate with adverbs. CL has generally supported the concept of a lexical grammar in that it is difficult to make a clear distinction between lexicon and grammar. A lexical item must be studied in terms of their grammatical context. In this view, the basic unit of analysis is a lexico-grammatical unit which is characterised in terms of its distributions in grammatical patterns. See Hunston and Francis (1999).

Answers to the questions in Figure 1 are provided in the following:

(a) words/sequences before thanks/thank you: well, that’s right, that’s great; words/sequences after thanks/thank you: very much indeed, Pauline, very much, a lot, ever so much

(b) the preposition for

(c) an example for each of the five thanking categories:

(i) thanking + Ø: Thank you. (line 5)
(ii) thanking + adverbs e.g. very much: Thank you very much indeed. (line 1)
(iii) thanking + another thanking: Thank you all very much. Thank you. (line 9)
(iv) thanking + praising the other speaker: Well thank you for that. That’s a very good start to the evening. (line 4)
(v) thanking + giving reason: Thank you for your attendance. (line 8)

Answers to the questions in Figure 2 are given below:

(a) In the data, cheers as a gratitude expression occurs as often as cheers as a leave-taking signal (exact figures need not be given).
(b) When followed or preceded by thanks or thank you, cheers is used as an expression of gratitude (see line 1); when followed or preceded by bye or see you, it is used as a leave-taking signal (lines 3 & 5).
(c) As noted in the paper, it is taught as an expression to mark the raising of glasses during dinner parties and other festivities for a toast but in the corpus data it is actually used to thank someone or to indicate that the speaker is leave-taking.


The BLC is a freely available corpus at http://www.someyanet.com/concordancer/index.html (accessed 6 January 2010). It comprises one million words of business letters. The creator of the corpus, Yasumasa Someya, made a considerable effort to ensure that a large majority of works that were included were produced by a native speaker of English. Most business letters in the corpus were taken from handbooks and reference guides published in the UK and US in the 1990s. A handful of business letters were based on books published in Japan but they were all proofread by native English speakers. I am very grateful to Yasumasa Someya for sending me a full list of the sources of the business letters.
The idea of incorporating a corpus-based approach into ESBP research and materials development is not entirely new. In their exemplary account of cultural differences in application letters written in English, Upton and Connor (2001) have used WordSmith Tools to analyse the linguistic features that realise politeness strategies in letters. Rogerson-Revell (1999) proposes a series of activities to develop students’ language skills, including case studies, presentations, task-based language activities in a simulated classroom setting, and collection of a corpus of samples of ‘real workplace interactions’.

Part of the Bank of English (56 million words of contemporary spoken and written texts) can be accessed through Corpus Concordance Sampler. See [http://www.collins.co.uk/corpus/CorpusSearch.aspx](http://www.collins.co.uk/corpus/CorpusSearch.aspx) for details.

Concordances of up to 50 lines from the British National Corpus can be obtained free of charge from the BNC web site. See [http://www.natcorp.ox.ac.uk/](http://www.natcorp.ox.ac.uk/) for details.

References


Bhatia, V., & Candlin, C. (Eds.). (2001). *Teaching English to meet the needs of business education in Hong Kong*. Hong Kong: City University of Hong Kong.


Gardner, D., & Miller, L. (1997). *A study of tertiary level self-access facilities in Hong Kong*. Hong Kong: ESEP, City University of Hong Kong.


Appendix 1  A Summary of Main Intentions of the English Language Education Component of the NSS Curriculum

- the development of specific learning targets to provide a clear direction for learning;
- the use of learning tasks to promote ‘learning by doing’ and to involve students in ‘three interrelated strands which define the general purposes of learning English’ (CDC & HKEAA, 2007, p. 11): INTERPERSONAL STRAND (for interpersonal communication), KNOWLEDGE STRAND (for developing and applying knowledge), EXPERIENCE STRAND (for responding and giving expression to real and imaginative experience);
- catering for individual learner differences so as to adapt teaching and learning to different student abilities and learning styles;
- the promotion of learner independence and lifelong learning so that students can become more actively involved in constructing knowledge and skills in classroom activities and in their own time;
- task-based learning as an integral part of teaching, learning and assessment;
- a greater emphasis on school-based assessment rather than one-off assessment based on public exams;
- the first-ever incorporation of an elective part (25%) of modules to allow for more flexibility for both schools and students to choose their desired topics (e.g. poems and songs, popular culture, sports communication, social issues) with which to learn English more creatively.

Appendix 2  Two Major Communicative Functions of Formulaic Expressions with Selected Illustrative Examples Listed in Hong Kong’s NSS English Language Curriculum

<table>
<thead>
<tr>
<th>Communicative functions of formulaic expressions</th>
<th>Example(s)</th>
</tr>
</thead>
</table>
| To request, offer, accept and decline help | Can you give me a hand?  
What can I do for you?  
Thank you, that’s very kind of you.  
No, thank you. I can finish this project on my own. |
| To express gratitude and regret | May I take this opportunity to express our deepest gratitude to all our teachers for their guidance, patience and support?  
It’s a pity that the concert has to be cancelled because of the typhoon. |

Appendix 3  Extract of Concordance Lines for the Word Grateful Sorted 1R (First Word to the Right) Generated with the Business Letter Corpus Online KWIC Concordancer (n.d.)

3 on as possible so we would be grateful for a prompt reply.  
4 ement and this school is most grateful for all that you are doing.  
13 C2:41:02057] We should be grateful for an early settlement.  
14 LC2:02:00469] I would be very grateful for answers to these question  
15 C2:41:02360] I should be most grateful for any assistance or advice  
29 25:04161] I, for one, am very grateful for it.  
31 BLC2:41:01270] We shall be grateful for prompt delivery.  
36 5759] Don't get me wrong: I'm grateful for the inquiries.  
37 BLC2:23:02379] I am very grateful for the interest you have sho  
51 he workshop and they are also grateful for your contribution.
I would be very grateful for your cooperation in this
ending period, and I am deeply grateful for your dedication.

I would be most grateful for your help.

We are grateful for your generous assistance.

I would be most grateful for your help.

financing, so I would be very grateful for your prompt response.

us to serve you and should be grateful for your reply.

duled deadline, and I am very grateful for your willingness to work

So I will simply say how grateful I am and express the hope that

I’d like you to know how grateful I am for your valuable assist

I can not tell you how grateful I am that you took the time to

We would be grateful if you could confirm whether

We would be grateful if you could make enquiries on

Iready been sent, I should be grateful if you could now do so to ena

We would be grateful if you could provide us with

We would be grateful if you could provide us with

e circumstances, we should be grateful if you could send us your che

I’d therefore be most grateful if you would arrange for him

We should be very grateful if you would arrange for Mr B

As such, we would be grateful if you would be kind enough to

We would be grateful if you would check into what

We would therefore be grateful if you could confirm whether

We should be grateful if you could make enquiries on

We would be grateful if you could make enquiries on

We would therefore be grateful if you could confirm whether

We would be grateful if you could confirm in writing

I should therefore be grateful if you would consider my requi

ring August, and we should be grateful if you would forward us detail

We would be grateful if you would furnish us with

We should therefore be grateful if you would grant us a 3 mon

We would be grateful if you would grant us an exte

I should be most grateful if you would let me know if t ment of Account and should be grateful if you would let us have your

e meantime, we should be most grateful if you would let us know what

do carrier bags and would be grateful if you would send us samples

of knitting Machines and would be grateful if you would send us your cat

I will be most grateful if you would seriously consid

We should be grateful if you would supply us with a

We should be very grateful if you would give me a call to le

s in the village send you our grateful thanks and sincere congratula

My grateful thanks for writing me such a

Do please extend my grateful thanks to everyone on your pr

May I extend our grateful thanks to you, also on behalf

Please accept our grateful thanks.

I am very grateful that our organization has bee

We are fully aware and grateful that such a speedy and favora

We are grateful that we could offer the people

I am very grateful that you made the extra effor

In fact, we would be so grateful to have you fill it out that

these situations we should be grateful to hear from you, as we are m

We should be grateful to receive your cheque at you

We are all very grateful to you for accepting this ass

e Executive Committee is very grateful to you for all your efforts d

I’m also grateful to you for helping to arrange

this oversight and that I’m grateful to you for pointing the error

I am very grateful to you for raising my salary

We are most grateful to you for settling this matt

ice, for which I am extremely grateful.

me, for which I am extremely grateful.

show, Chris, and I am deeply grateful.

convenience, I would be very grateful.

additional costs, we would be grateful.

months on arrival, I would be grateful.

r participation, and I’m very grateful.

ditionnaire, we would be very grateful.

deferment, we should be most grateful.

stitutions, we would be very grateful.