Exploiting a corpus of business letters from a phraseological, functional perspective*

LYNNE FLOWERDEW

Hong Kong University of Science and Technology, Clear Water Bay, Kowloon,
Hong Kong, SAR
(email: lclynne@ust.hk)

Abstract

This paper illustrates how a freely available online corpus has been exploited in a module on teaching business letters covering the following four speech acts (functions) commonly found in business letters: invitations, requests, complaints and refusals. It is proposed that different strategies are required for teaching potentially non-face-threatening (invitations, requests) and face-threatening (complaints, refusals) speech acts.

The hands-on pedagogic activities follow the 'guided inductive approach' advocated by Johansson (2009) and draw on practices and strategies covered in the literature on using corpora in language learning and teaching, *viz.* the need for 'pedagogic mediation', and the 'noticing' hypothesis from second language acquisition studies.

Keywords: inductive, pedagogic mediation, noticing, pattern-hunting, speech acts, functions, phraseology

1 Introduction

A theory of phraseology is now well established (cf. Cowie, 1998; Sinclair, 1991; Hunston & Francis, 2000) in which the lexical item is seen to have primacy (Sinclair, 2004). Corpus searches have shown that lexemes display preferred collocational and colligational patterning (the lexical and grammatical company that words keep), and also have preferred semantic preferences and semantic prosodies, i.e., lexical items tend to co-occur with certain semantic sets and items are imbued with either a negative or positive connotation. For example, CAUSE tends to occur with the semantic set of 'diseases' and usually has a negative semantic prosody (Stubbs, 1996). This linear, syntagmatic approach to language reveals that meaning does not reside in individual lexemes, leading Sinclair to argue for the existence of 'extended units

---

* This is a revised and expanded paper of part of a plenary presentation delivered on 11th October 2010 at the International Symposium on Corpus Linguistics and Language Learning at the National Taipei College of Business, Taiwan and also a presentation given on 2nd July 2010 at the 9th Teaching and Language Learning Conference held at Masaryk University, Brno, Czech Republic.

of meaning'. It is to be noted, however, that Cowie's approach to collocation is somewhat different from that of Sinclair's in that a 'textual' over a 'statistical' identification is preferred on the grounds that individual restricted collocations may recur to only a limited extent within a given text or across several texts on the same topic. As I am working with a small, specialized corpus of one million words, the Business Letter Corpus (BLC), I take a 'textual' approach to identification of collocational patterns.

Moreover, Stubbs (2006: 26) has signalled a relationship between Sinclair's 'extended units of meaning' and speech act theory: ''Although they are based on very different kinds of data, both speech acts and extended units are functional and build the speaking agent into units of language structure and use''. Speech acts can be of a non-threatening or face-threatening kind. As Brown and Levinson (1987) point out, certain speech acts are likely to damage a person's 'face', a concept first proposed by Goffman (1967) to signify one's reputation or good name. The hearer's positive face can be damaged by the speaker expressing disapproval of the hearer's action; the hearer's negative face has the potential to be damaged if the speaker gives an order which impinges on their freedom of action. Similarly, the speaker's own positive or negative face can be damaged if they are pushed into admitting some error or an imposition is made upon them. Syllabus designers drew on speech acts to provide the theoretical underpinning for the functional approach to language learning, which was at the heart of communicative language teaching in the 1970s (Wilkins, 1976). While speech act theory and a functional approach to language have mainly been discussed in relation to speaking, they are also of relevance for writing.

The purpose of this paper is to illustrate how a freely available online corpus has been exploited in a module on teaching business letters from a phraseological, functional perspective covering the following four speech acts (functions) commonly found in business letters: invitations, requests, complaints and refusals. It is proposed that different strategies are required for teaching more neutral, non-face-threatening (inviting and requesting) and face-threatening (complaining and refusing) speech acts.

In the first part of this article I first review briefly how a functional approach to language learning has been addressed in corpus-based materials. Most of these applications, however, focus on English for Academic Purposes (EAP) or spoken material specifically directed toward English Language Teaching (ELT). I then review the 'Noticing Hypothesis' which underlies much corpus-based pedagogy, although only in a few accounts is this made explicit. In the second part of the paper I illustrate how corpus consultation focusing on functions with particular emphasis on their phraseologies has been addressed in the teaching of a less explored genre, i.e., that of business letters. I also illustrate how the 'Noticing Hypothesis' underpins much of the corpus consultation, but is an aspect which has been little commented on in the literature.

2 Treatment of functions in corpus-based instruction

The teaching of the lexico-grammar of functions has been addressed in a variety of ways in corpus-based pedagogy. (Here, I use the term 'corpus-based' in a general sense to cover both hands-on and pen-and-paper activities derived from concordance

output.) One early key endeavour is that by Thurstun and Candlin (1998a & 1998b) for teaching the functions associated with general academic English, for example, stating the topic, reporting the research of others, starting from a key lexical item. Within each broad function, each key word (e.g., *claim*, *identify*) is examined using concordance output within the following chain of activities (Thurstun & Candlin, 1998b: 272):

- LOOK at concordances for the key term and words surrounding it, thinking of meaning.
- FAMILIARIZE yourself with the patterns of language surrounding the key term by referring to the concordances as you complete the tasks.
- PRACTISE key terms without referring to the concordances.
- CREATE your own piece of writing using the terms studied to fulfil a particular function of academic writing.

Corpus-based instructional material for English for general academic purposes has also been produced by Charles (2007, 2011), Thompson and Tribble (2001) and Bloch (2009, 2010). Like Thurstun and Candlin, Charles (2007) also targets key rhetorical functions, using a corpus of PhD theses written by native speakers, in this case the combinatorial function of *defending your work against criticism*, a two-part pattern: 'anticipated criticism→defence and its realization using signals of apparent concession, contrast and justification' (*op. cit.*, 296). Another feature of Charles' materials is that she approaches these functions by first using a top-down approach, providing students with a suite of worksheet activities to sensitize them to the extended discourse properties of this rhetorical function. She then supplements these with a more bottom-up approach by having students search the corpus of theses to identify typical lexico-grammatical patterns realizing these functions.

The hands-on activities by Thompson and Tribble (2001) and Bloch (2009, 2010) target a specific function, that of citations. Thompson and Tribble's tasks focus on having students categorize the citations identified in a dedicated corpus according to their range, purpose and forms. Bloch (2009) describes a user-friendly program for teaching the use of reporting verbs in academic writing in corpora compiled in-house to meet the needs of specific learners. The interface presents users with only a limited number of hits for each query and a limited number of criteria for querying the database, namely integral/non-integral; indicative/informative; writer/author; attitude towards claim; strength of claim, categories devized from Bloch's (2010) research on the use of reporting verbs from a rhetorical perspective.

The teaching of functions has also been the focus of corpus-based materials of spoken communication in general ELT (Ackerley & Coccetta, 2007; Coccetta, 2011). Coccetta's materials are based on multimodal approaches to corpus analysis, which take a systemic-functional orientation to determine how different semiotic resources (language, gaze, gesture, etc.) interact to create meaning (cf. Baldry & Thibault, 2006). Coccetta explains how an online multimodal concordancer incorporating a search engine allows students to find and isolate sequences in a corpus sharing the same characteristics by means of a functionally tagged corpus. For example, to see if the function of 'declining an offer' occurs in a subcorpus relating to *requests, invitations* and *offers* and to find the linguistic forms realizing this function, students

Table 1 *Concordance results for the 'declining an offer' function (Coccetta, 2011: 131)*

1.	No thanks.
2.	no thanks.
3.	No thanks. I mean, that – that water's been there for ages.
4.	No.
5.	No thanks. I'm not – I'm not hungry.
6.	Uh, no thanks.
7.	I've already had one thanks.

choose a parameter from a drop-down menu to retrieve the relevant concordance lines exemplified in Table 1. Each concordance line has access to the film clip which provides non-linguistic information drawing on semiotic resources such as gesture, posture, gaze and facial expressions.

The account in this article of corpus-based pedagogy aimed at writing business letters addresses the use of corpora for English for Occupational Purposes (EOP), thus contributing to the existing literature on using corpora for teaching functions in English for Academic Purposes (EAP) and ELT settings.

3 Noticing Hypothesis and corpus consultation

The 'Noticing Hypothesis' discussed in second language acquisition (SLA) studies underpins many corpus studies. The principle underlying this cognitive concept is that learners' acquisition of linguistic input is more likely to increase if their attention is drawn to salient linguistic features. Schmidt (1990, 2001), one of the main proponents of this hypothesis, maintains that it precedes understanding and is a condition which is necessary for converting input into intake. Moreover, as Boulton (2011) points out, 'noticing' overlaps with other features such as focus on form, consciousness-raising, and language awareness. Swain (1998: 66) links 'noticing' to frequency counts of form, remarking that there are several levels of noticing, one of which is that: "Learners may simply notice a form in the target language due to the frequency or salience of the features themselves". In spite of its detractors, most notably Truscott (1998) and Robinson (1997) whose empirical research on implicit and explicit second language learning under four conditions (implicit, incidental, rule-search, instructed) found learning to be fundamentally similar across all four conditions, the concept does, in general, hold currency in corpus-based pedagogy on various accounts. One reason is that as Key Word in Context (KWIC) concordance lines highlight *recurrent* phrases, scrutiny of corpus data would seem to be an ideal means of enhancing learners' input with attention paid to frequency counts, a level indicated by Swain earlier (1998). Another reason, as signalled by Boulton (2011), is that inductive approaches, the mainstay of data-driven learning (DDL), are entirely dependent on noticing (although the concept itself is only explicitly referred to in a few corpus-based endeavours (cf. Johns, Lee & Wang, 2008; Flowerdew, 2008, in press, 2012b).

A purely inductive approach to corpus consultation implies spontaneous noticing by the learners. However, as Johansson (2009) points out, in reality, the pedagogic

use of corpora combines inductive and deductive approaches, involving some kind of teacher intervention.

> Is the use of corpora to be grouped with the explicit or implicit method? The term 'data-driven' learning suggests that it is an inductive approach and therefore comparable with the implicit method, though the emphasis is on gaining insight rather than establishing habits, and in this sense it is mentalistic. I believe that the dichotomy explicit-implicit is far too simple. In the case of corpora in language teaching, I would favour a guided inductive approach or a combination of an inductive and deductive approach where the elements of explanation and corpus use are tailored to the needs of the student.
>
> (Johansson, 2009: 41–42)

Several studies providing a framework for corpus consultation mediate this inductive/deductive continuum. They involve some type of 'pedagogic mediation', a term first introduced by Johns (1997) and subsequently taken up by McCarthy (1998) and Widdowson (2000), for teacher-directed noticing activities. For example, Chujo, Anthony and Oghigian (2005: 1) propose the following four-step DDL approach to incorporate cognitive processes such as noticing and hypothesis formation: (1) hypothesis formation through inductive corpus-based exercises; (2) explicit explanations from the teacher to confirm or correct these hypotheses; (3) hypothesis testing through follow-up exercises; and (4) learner production. Meanwhile, Flowerdew (2009: 407) proposes modifying Carter and McCarthy's (1995) '3 Is' strategy: Illustration (looking at data); Interaction (discussion and sharing observations and opinions); Induction (making one's own rule for a particular feature) to accommodate the concept of noticing through adding 'Intervention' as an optional stage between Interaction and Induction.

Kennedy and Miceli's (2010) corpus work also favours a guided-inductive approach. They note that corpus work of a purely inductive nature without any kind of pedagogic mediation or guidance would make high demands in terms of language proficiency, observation and inductive reasoning such as 'the learner-as-researcher' model, proposed by Bernardini for her advanced translation students (Bernardini, 2002, 2004). As their students are intermediate level Italian and not advanced like Bernardini's, they propose two modes of apprenticeship training, 'pattern-hunting' and 'pattern-defining', using a 500,000-word corpus of contemporary Italian, to aid intermediate-level Italian students with personal writing on everyday topics. For example, when writing about their sense of personal space for an autobiography, students were first prompted to come up with some key words for pattern-hunting. Many students suggested the common term *spazio*, which not only turned up ideas and expressions, for example, *rubare spazio* (take space) but also triggered further searches on words encountered in the concordance lines, for example, *percorso* (path). Other pattern-hunting techniques included browsing through whole texts on the basis of the title and text-type, and scrutinizing frequency lists for common word combinations. The pattern-defining function was used when students did have a specific target pattern in mind to check. For example, one student wanted to establish if the pattern "so" <adjective> "that" could be rendered in Italian with *così* <adjective> *che* and if the subjunctive mood was required after *che*. Both types

of apprenticeship training involve noticing, 'pattern-hunting' instigated by the teacher and 'pattern-defining' by the student.

The following section first provides a brief overview of the business letters module and the approach taken to the corpus consultation.

4 Business letters module: background and approach

At the tertiary institution in Hong Kong where I work a 15-hour module on writing business letters is offered to final-year undergraduate science students to prepare them for the professional workplace. There is a comprehensive set of in-house textbook material, covering four key genre sets of business letters, namely invitations and thanks, requests and replies (refusals), complaints and adjustment letters, and sales letters. A one-million-word freely available business letters corpus comprising American and British business letters was used to supplement the existing course materials and textbook activities (see www.someya-net.com/concordancer for further details of this corpus).

This paper describes how I integrated corpus consultation into classroom activities based on the teaching of the module to six different groups of students, with approximately eighteen students in each class. My account differs from that of other initiatives in the literature in several aspects. First, there was only one computer in the classroom, which necessarily constrains the corpus-based activities. Another difference is that the students did not undertake a training session in how to formulate queries and search the corpus beforehand. Leading practitioners have emphasized the importance of incorporating strategy training into corpus consultation (cf. Chambers, 2005, 2007; Lee & Swales, 2006; O'Sullivan & Chambers, 2006; O'Sullivan, 2007; Kennedy & Miceli, 2010), which has been identified as one of the possible reasons for lack of uptake (cf. Frankenberg-Garcia, 2011). While not denying the importance of strategy training, my attempts at using a set of in-house-produced materials for corpus training with two classes proved somewhat unsuccessful, the main reason being that the materials targeted searches in corpora for academic writing. It was found that building strategy training into the corpus consultation was a more effective and efficient mode of instruction in this particular teaching scenario where time was limited. Also, the freely-available business letters corpus proved to be ideal for teaching purposes on account of its restricted size and user-friendly search facilities (see Braun, 2005). Moreover, Boulton's (2009) empirical study has provided evidence indicating that even lower-level learners can cope with corpus data with no prior training, with Boulton remarking that "We are perhaps beginning to see something of a retreat on this strong insistence on training" (*op. cit.*: 40).

The third way in which the pedagogic activities differ from other accounts in the literature which describe how corpus consultation has been incorporated *systematically* (cf. Flowerdew, 2008; Kennedy & Miceli, 2010) is that the business letters corpus was used on an *ad hoc* basis. By this I mean that corpus searches were primarily either used at the initial stage of the unit in hypothesis-type activities or conducted whenever students were faced with problematic lexico-grammatical aspects arising from the in-house-produced textbook materials. A description of the corpus-based activities for the four functions is provided in the following section. I relate this to the phraseological approach and noticing hypothesis described earlier in the paper.

5 Corpus consultation for speech acts/functions

5.1 Invitations

Stubbs (1987) has noted that the formal written language of business correspondence is a context which produces a large number of explicit performatives, for example, *May I wish you a successful and interesting conference, I emphasise that…* (*op. cit.*: 10). This was the starting point for the corpus task on writing letters of invitation. Before looking at the sample letter in the textbook from an undergraduate student to the executive director of a well-established engineering company in Hong Kong inviting them to be a guest speaker at the quarterly dinner of the student alumni association, I asked students to write the opening sentence of this letter. After eliciting a variety of responses, most of which began with the phrase 'I am writing to invite you…' or 'We would like to invite you…', I asked students to search the corpus for various verb forms for invitations, e.g. *invite, inviting, invited*, i.e., the performative verb for the speech act of 'inviting'. The main purpose behind this task was to discourage students from adopting a 'phrase book' mentality and to expose them to a wide variety of exponents. This was achieved by encouraging students to read the corpus both paradigmatically (from top to bottom) and syntagmatically (from left to right); O'Keeffe and Farr (2003) and Flowerdew (2009) have indicated that students need training in 'reading' concordance lines in this way. For example, reading the corpus paradigmatically familiarizes students with a variety of phrases, e.g., *we would like to invite you…; we are pleased to invite you…; you are cordially invited to attend…*. Students then compared these openings with the ones they had written, noting that the phrase *you are cordially invited to attend…* from the corpus might not be quite appropriate for their context of writing.

An additional task in the textbook required students to circle an appropriate adjective from a choice of three or four to collocate with a particular noun, a language point which it would be difficult to find the answer to from grammars or dictionaries. This task was exploited to introduce students to reading the corpus syntagmatically through hypothesis testing. In the sentence below from the textbook, students were first asked to circle which adjective they thought most appropriate. When checking their responses I found that most students thought either *cordial* or *kind*was best but none of them could explain why.

> We extend a (*friendly/sincere/cordial/kind*) invitation to you to join the … Young Scientists Society and to participate in our exciting educational programmes.

To test their hypotheses, students were then asked to look up the illocutionary noun *invitation*, a concealed performative, in the BLC which yielded the following concordance lines:
  However, students were only able to interpret the corpus data through teacher mediation to encourage 'noticing' beyond the collocational level. When prompted to read the lines syntagmatically in the spirit of Sinclair's 'extended units of meaning', i.e., to examine the subject + verb + adj. + noun, students were then able to work out that when *cordial* was used with *invitation* it was most commonly used when the sender was offering the invitation, e.g. *Please accept our cordial invitation to visit…* On the other

Thanking you for your	kind invitation	to address the audience, I rem
Thank you very much for your	kind invitation	to your Christmas party.
Many thanks for your	kind invitation	to join in your celebrations on
I thank you for your	kind invitation	on the occasion of your openin

. May we extend to you a	cordial invitation	to call in at White's and make a
I would like to extend my	cordial invitation	to you to visit our Tokyo office
Province of <name> extends a	cordial invitation	to <name> attend a reception on
This is a	cordial invitation	to become one of Myer's charge

Figure 1. Collocations for 'invitation'.

hand, when *invitation* was preceded by *kind*, it tended to be used for thanking by the receiver, for example, *Thank you for your kind invitation to attend*… The final mini-task involved drawing students' attention to frequency data, which elicited some surprise on the part of the students to discover that the noun form *invitation* was more common in this corpus. These examples thus serve as a means of acquainting students with the probabilistic nature of language and that language does not always consist of rule-governed behaviour.

In addition to noting form-function correlations, students also have to be made aware that business language is highly context-sensitive. Widdowson (1998) has pointed out that a corpus is transposed from its original context, which obscures the communicative intent and socio-cultural purpose. However, as Gavioli and Aston (2001: 240) have argued, it is not so much a question of whether corpora are divorced from their original setting, but rather "whether their use can create conditions that will enable learners to engage in real discourse, authenticating it on their terms" (see also Mishan, 2004). Sometimes the co-text can provide enough clues to the context, enabling students to 'authenticate' the corpus output for their own learning situation. For example, in Figure 1 the co-text of the collocation 'kind invitation' provides some help with interpretation at the pragmalinguistic level (the collocation of certain linguistic features in a certain register). A student would be able glean from the co-text of *kind invitation* the pragmalinguistic knowledge that 'I thank you for your kind invitation on the occasion of…' belongs to a more formal register than 'Many thanks for your kind invitation to join in…'.

However, sociopragmatic appropriacy, which is influenced by social, cultural and personal preferences and the dynamics of the unfolding interaction (Kasper, 2001), is more difficult to discern in corpus data. Here, the use or non-use of certain direct or indirect speech acts can pose problems for interpretation. In Figure 1 for 'cordial invitation' it could well be that the example 'Province of <name> extends a cordial invitation to <name> to attend…' is a polite directive disguised as an invitation. But we have no way of knowing this unless we are familiar with the situation and social roles (although admittedly, as business letters are somewhat conventionalized these may be fairly obvious in some cases). But in cases such as these, the lack of situational context can serve as a consciousness-raising activity. Students can be asked to supply

what they consider to be possible scenarios for this exponent, thus sensitising them to the necessity of using corpus data judiciously and avoiding a cut-and-paste mentality.

5.2 Requests

One input task in the student textbook required students to formulate a polite request from one member in a company to another of the same status, asking them to get in contact about a fairly routine matter. This yielded student writing such as in the following two examples:

> * *I would appreciate if you can contact me regarding*….
> * *I would be very appreciated if you could contact with me*…

However, general unfamiliarity with various patterns containing *appreciat** made it necessary to adopt Kennedy and Miceli's 'pattern-hunting' strategy outlined in the first part of this article. Through scrutinizing the concordance lines students were able to see the correct pattern for 'appreciate' and note the obligatory object 'it', as illustrated by the concordance lines in Figure 2. The corpus data also revealed the prevalence of modals with illocutionary lexical verbs, i.e., hedged performatives, which Stubbs (1987) has noted as the commonest surface form of verbs in his small corpus of business correspondence.

Likewise, students were able to discern the correct lexico-grammatical patterning with phrases for 'appreciated', and to work out that the string '…appreciated if…' requires the dummy subject 'it'. The concordance lines also revealed that the collocation 'very appreciated' as found in the students' suggestions is a non-harmonic one, with possible ones being 'very much', 'greatly' and 'highly', as noted by students from their paradigmatic reading of the concordance lines (see Figure 3). Again, these tasks are an attempt to apply Sinclair's concept of 'extended units of meaning' to pedagogy.

However, while the BLC is very useful for revealing lexico-grammatical patterning, this is only half the story. Students have to carry out further analysis to decide which

```
[BLC2:32:03428] We would          appreciate   it if you could select a suitable library, prefer
[BLC2:31:02117] I would greatly    appreciate   it if you could send any notes or minutes of the
              1] First of all, I would really   appreciate   it if you could send me two or three sample issue
thin two weeks from today, we?d    appreciate   it if you could send them by Air Express.
ng of May, so I Would therefore    appreciate   it if you could send them to arrive here by May 1
[BLC2:33:00315] We would          appreciate   it if you could send us a brief outline or pointe
[BLC2:35:00613] Therefore, we shall appreciate   it if you could send us a copy of the said document
days to remit payment, we would    appreciate   it if you could supply us with some references.
f shipment ourselves, and would    appreciate   it if you could treat the matter as urgent.
[BLC2:28:01047] I would really     appreciate   it if you could understand our position in this m
[BLC2:02:01186] I would greatly    appreciate   it if you could.
[BLC2:35:00354] As such, we shall  appreciate   it if you will advise us by telex whether you can
[BLC2:35:00162] If so, I shall     appreciate   it if you will direct any additional information
[BLC2:35:01177] We shall          appreciate   it if you will initial the correction and return
is acceptable to you, and shall    appreciate   it if you will kindly send the additional informa
[BLC2:34:03134] We would          appreciate   it if you will reply immediately on the following
town Refinery Project, we shall    appreciate   it if you will send them to attention Z Kirino, M
[BLC2:32:01398] We would          appreciate   it if you would  act as a liaison with your hometo
[BLC2:33:00906] We would greatly   appreciate   it if you would allow us an extension of three mo
[BLC2:40:01776] We should         appreciate   it if you would allow us monthly account terms.
[BLC2:31:00670] We               appreciate   it if you would answer the following questions;
[BLC2:32:01131] We would          appreciate   it if you would assist us in making this position
recently moved office, and would   appreciate   it if you would change all listings in your direc
[BLC2:05:02441] I would          appreciate   it if you would check your records and let me kno
[BLC2:32:01718] We would          appreciate   it if you would come again, this time to see me w
busy time for you, but we would    appreciate   it if you would consider appearing at our Center.
```

Figure 2. Sample concordance output from the BLC for 'appreciate'.

Mr Eric Sweeny, for I have	appreciated	his sterling qualities for many years.
on 16ᵗʰ January, it would be	appreciated	if we could have your reply by 13ᵗʰ January at
B2:02940] It would be greatly	appreciated	if we could receive your statement by the end of
81 and it would be very much	appreciated	if you could deliver a lecture on any subject al
2:40:00152] It would also be	appreciated	if you could give us some idea of the time requ
sequently, it would be highly	appreciated	if you could kindly arrange your visit in the
future it would be greatly	appreciated	if you notified the office ? in advance ? if
intime, it would be very much	appreciated	if you would kindly arrange to meet with him eit
00494] It would be very much	appreciated	if you would kindly arrange to meet with me eitl
[BLC2:32:00105] It would be	appreciated	if you would pass on sincere expressions of grat
understanding would be much	appreciated	in this matter.
rpreting, would be very much	appreciated	it will also, we hope, lead to the publication
I wanted to let you know we	appreciated	Ms. Maskoff?s efforts.
BLC2:41:01360] We have always	appreciated	our friendly relations with you firm and shall
You to know how much I have	appreciated	our pleasant association during the past year.
:01336] To begin with, it is	appreciated	that you have recommended three excellen, 1981
to write to say how much I	appreciated	the favour you did for me last week.
to know that my wife and I	appreciated	the generous bonus you gave this year.
You to know how much I have	appreciated	the hard work you have put into welding your dep
[BLC2:04:02752] I also	appreciated	the information about Levi Strauss, and I especi
[BLC2:25:05073] We very much	appreciated	the opportunity of serving you again in the men
Or your time this week, as I	appreciated	the opportunity to become personally acquainted
032] Again, Mr. Feyerherm, as I	appreciated	the opportunity to meet with you
[BLC2:09:03293] We certainly	appreciated	the opportunity to see your set-up firsthand and
2:15:0505] In particular, I	appreciated	the perseverance you showed in the fact of many
[BLC2:32:02224] The group	appreciated	the pleasure of visiting Governor Naqato and we

Figure 3. Sample concordance output from the BLC for 'appreciated'.

pattern would be suitable for their context of writing. Again, this is another instance in which the co-text and also frequency counts can provide some help. Students' attention was drawn to the fact that while there were 105 hits of patterns with 'appreciate', only nine instances of the pattern '*It…appreciated if…*' were recorded, thus suggesting this is a marked form. That this pattern usually occurred with a high degree of modalisation and adverbials was an indication that it was used for making requests which placed onus on the recipient, or was used as a deferential marker for making requests to someone of a rank higher than the writer. Applying the concept of 'noticing' of salient linguistic features and frequencies helped students to decide on appropriate patterns for their own context of writing.

5.3 Complaints

In some situations, it may be more sociopragmatically appropriate to use an indirect speech act, especially for those functions which are regarded as somewhat face-threatening or make impositions on the hearer or reader. A case in point is the speech act of disagreeing which has the potential to damage the hearer's positive face. McCarthy (1998) notes that in the 5-million-word Cambridge Nottingham Corpus of Discourse in English (CANCODE) there are only eight occasions when someone says, *I disagree*, and, interestingly, no examples followed by *with you*. McCarthy further notes that all eight occurrences of *I disagree* are modified in some way with a mitigating device, e.g. *I'd er, I'd disagree*. Speech acts may not involve performative verbs, but may unfold indirectly and in negotiation "with due sensitivity to interlocutors' personal face" (McCarthy, 1998: 19). However, such types of indirect speech act present the dilemma of how to search for these in a corpus. An indirect speech act common to business letters, namely complaining, is discussed below.

One of the input tasks on style and tone in the textbook reads thus: Maria Wong is thinking about changing the wording of parts of her letter. Circle any phrases you consider inappropriate and rewrite them:

I am writing to complain about a contamination incident which has recently occurred on campus…

Students suggested a wide range of lexico-grammatical patterns to replace the direct speech act 'complain', the most common of which are given below. This procedure is somewhat similar to Kennedy and Miceli's (2010) 'pattern-defining' technique as they had specific target patterns in mind.

I am writing to lodge a complaint about…
I am writing to comment on…
I am writing to inform you…
I am writing to express my opinion on…
I am writing to express my concern about…

Students' responses indicated that they had a grasp of the sociopragmatics but lacked pragmalinguistic knowledge. While the task did not specify the kind of contamination incident that had occurred, students were able to indicate that the phrase 'I am writing to lodge a complaint about…' might be used for a serious incident, whereas 'I am writing to express my concern about …' would commonly be used for something of less severity. A search in the BLC confirmed students' intuition on the use of 'lodge a complaint' with the following phrase found:

The <NAME> has lodged a formal complaint with me…

Sometimes corpus data triggered other queries, one such query being when the noun vs. verb pattern was used (possibly motivated by the search queries for the function of 'inviting' discussed earlier). A search in the corpus for *complain** showed that of a total of 118 tokens, 82 were nouns, e.g., 'We have received your complaint regarding…', but invariably used for acknowledging a complaint. When the verb form was used, it was found to occur in two scenarios: as a follow-up to a previous complaint or in a reporting statement (similar to the function of *disagree* in the CANCODE corpus), e.g.:

We sent an e-mail complaining of the late shipment last week
…back to the old standard that brought about my original complaint.
Several secretaries have complained of major and frequent breakdowns…

The BLC was used to verify students' other suggestions for making a complaint, realized by implied performatives such as 'express my concern'. However, as noted by several students, the corpus data do not exactly correspond to the student's suggestion as no examples of 'my concern' were found in the one-million-word BLC. This observation alerted students to the principle expressed in Carl Sagan's well known aphorism 'absence of evidence is not evidence of absence'. A Google search revealed that this combination was indeed possible.

One could make a pragmatic distinction between phrases with and without a possessive prefacing 'concern', but whether it is worth covering such a fine distinction

with intermediate-level students and in light of the debate on English as a *lingua franca* is open to debate (see Seidlhofer, 2011). One student query triggered by the focus of discussion on verbs collocating with 'concern', was whether 'voice concern' would be appropriate. As a search in the BLC did not yield any patterns of this kind, another Google search was conducted. Of interest is that this pairing most often occurred as reported speech such as that associated with news reporting and writing of minutes, thus sensitizing students to the genre-specific nature of collocations. Other search engines such as WebCorp, which would provide a more 'linguistic' presentation of results from the web, or a large 'principled' corpus such as the BNC would arguably have been more useful. However, given the context of my teaching situation, a 15-hour module on business letter writing to final-year undergraduate science students, my approach tended towards expediency and efficiency.

5.4 Refusals

Refusals, like complaints, are another potentially face-threatening speech act. Similar to the previous task in the textbook, students were asked to identify the problem in the following sentence and rewrite it in an appropriate tone for the scenario provided:

I refuse your request to return the items that you ordered from our company.

Students were asked to provide patterns to replace 'refuse', which were then discussed in class and checked against the corpus data. One student suggested the phrase 'May we remind you that items ordered are not returnable', saying that it had appeared on a notice in a shop and printed on the receipt she received. While corpus data may be decontextualized as Widdowson has pointed out, having students reflect on genres and contexts when defining patterns helps to overcome this obstacle to some extent. The most common phrases proposed by students for the function of 'refusal' were as follows:

I am sorry that…
I am afraid that…
I regret that…
*I apologise that…[1]

These patterns were subsequently searched in the BLC to verify students' suggestions. One student follow-up query concerned whether it was necessary to use 'that' to introduce the reporting clause. After examination of the concordance data, students concluded that although 'that' was optional, it was preferred in formal writing. Of interest is that students discovered that the string '…sorry that…' was found to be multifunctional (Moon 1997), as illustrated in Figure 4. When 'sorry' occurred with a verb in the past or present perfect in the reporting clause, it had the function of apologizing, for example:

I am very sorry that this happened to you.
We are very sorry that we have not replied earlier.

---

[1] The * is used to indicate that this expression is somewhat unnatural in the way students are using it, e.g., I apologise that items cannot be returned.

```
       [BLC2:05:01862] I am       sorry    that the instruction manual for operating
       [BLC2:17:00430] We are      sorry    that the model 88b handsaws you purchased
       [BLC2:18:03208] I am very   sorry    that the outcome proved us wrong
       [BLC2:15:03606] I am        sorry    that the person you spoke with spoke to yo
       [BLC2:15:05692] I am        sorry    that the timming worked out like this, but
       [BLC2:32:03403] We are      sorry    that there is not a great selection of lar
       [BLC2:23:02332] We are      sorry    that these increases have been made necess
             friends will doubtless be      sorry    that they will see him much less frequentl
      r several years, and I am    sorry    that this change of policy affects you, to
       [BLC2:18:06031] I am very   sorry    that this happened to you.
       [BLC2:32:02049] I am        sorry    that this message comes to you so late, bu
       [BLC2:31:01823] We are      sorry    that two of the products were not in worki
       [BLC2:16:01012] We are      sorry    that we are not able to be more helpful.
       [BLC2:18:03197] I am very   sorry    that we are unable to complete < name of
       LC2:29:00208] We are very    sorry    that we are unable to honor your request,
       LC2:24:00529] We are very    sorry    that we can not allow you the special disc
       [BLC2:41:01971] We are      sorry    that we can not be more helpful in this ca
       [BLC2:27:01330] We are      sorry    that we can not be of more assistance to
             your application and are    sorry    that we can not extend your credit at this
       [BLC2:22:01575] I am        sorry    that we can not help you further, but the
       LC2:22:00590] We are very    sorry    that we can not help you in this case, but
       LC2:41:00492] We are very    sorry    that we can not send you a large dummy  but
             r thinking of us, and are    sorry    that we can'nt give you a more favorable
       LC2:31:02327] We are very    sorry    that we could not accept your offer for th
       [BLC2:32:00819] I am very   sorry    that we could not facilitate such a reward
       [BLC2:13:02580] I am        sorry    that we couldn?t meet your needs on this
```

Figure 4. Sample concordance output for 'sorry'.

In most cases, though, it was found to function as a polite refusal, which, it could be argued, also serves as an implicit performative for apologizing, for example:

> *We are sorry that we can not allow you the special discount.*

The elicitation of students' patterns for polite refusals revealed confusable speech acts. Students were therefore instructed to scrutinize the concordance output in the 'that' clause, again echoing Sinclair's concept of 'extended units of meaning'. Follow-up class discussion of the data revealed that students had noticed the overlap in speech acts, i.e., while 'I am sorry / I regret / I apologize for…' can all be used to realize the speech act of apologizing, only 'I am sorry / I regret that' also function as polite refusals. It is also important to draw students' attention to overused, stereotypical phrases such as 'we regret to inform you that the train will be delayed' (Henry Tye, personal communication). In other words, consideration should be given to the perlocutionary effect of the speech act, i.e., the effect of the apology on the addressee and whether they regard it as genuine or merely a standard, impersonal response.

In addition to the concordancer, another search facility in the BLC which proved useful was the *Bigrams Plus* function. It was found that students might be familiar with two or three key words in a pattern, but unsure how to 'string' them together. Figure 5 below shows the output for the bigram 'sorry……attend'. However, students need to select and modify the corpus data to suit their own context of writing and be trained to carry out further searches (in this case, using Google) on more unusual patterns. For example, a Google search revealed the one occurrence of 'sorry to say…' in the BLC was usually restricted to more informal contexts.

6 Concluding remarks

This paper has illustrated how a freely available online corpus has been exploited in a module on teaching business letters. With reference to concepts derived from speech

---

Bigram Plus search result:

Word 1: sorry
No. of intervening words: 8
Word 2: attend
Search Corpus: 01. Business Letter Corpus (BLC2000)

The combination "sorry+[W=<s]+attend" was found 7 time(s):

1   sorry to say I will be unable to attend
1   sorry that you were unable to attend
1   sorry that I can not attend
1   sorry that you could not attend
1   sorry I was out of town and unable to attend
1   sorry that I was unable to attend
1   sorry I will not able to attend

Figure 5. Using the *Bigram Plus* function to search for patterns.

act theory and the 'Noticing Hypothesis' I hope to have shown how a small, specialized corpus can profitably be used to supplement existing course materials. Many of the student queries centred on phraseological aspects of language which are not always covered in dictionaries and grammars. Secondly, hands-on concordancing activities have advantages over using purely text-based material in which only one or two sample letters are provided, as recurrent phraseologies alert students to the 'preferred ways of saying things' (Sinclair, 1991).

It is hoped that the strategies discussed in this paper will contribute to the growing body of literature on teaching functions from a corpus-based perspective and to corpus-based pedagogy in general. There remain several limitations to this DDL initiative, however. Only observational data gleaned from teaching the business letters module across six classes are reported. No systematic experimental data were collected on students' evaluation of using corpora or to what extent their writing profited from corpus consultation. In any case, it is only very recently that studies have been conducted on students' performance based on DDL activities (see Flowerdew 2012a for a review of these studies). Boulton (2011: 39) raises the issue that "the real advantages of DDL lie in longer-term benefits, cognitive/constructivist as well as purely linguistic; in addition to 'incidental' learning and greater learner autonomy, these include language awareness and noticing ability". What are now needed are more studies to determine the long-term benefits of DDL. This article has merely touched on heightening language awareness through teacher-directed 'noticing' activities related to the phraseologies of speech act functions (Austin 1962) – How to do things with words, with corpora.

Acknowledgements

References

Ackerley, K. and Coccetta, F. (2007) Enriching language learning through a multimedia corpus. *ReCALL*, **19**(3): 351–370.

Austin, J. L. (1962) *How to do things with words*. Oxford: Oxford University Press.

Baldry, A. and Thibault, P. (2006) Multimodal corpus linguistics. In: Thompson, G. and Hunston, S. (eds.), *System and Corpus. Exporing Connections*. London: Equinox, 164–183.

Bernardini, S. (2002) Exploring new directions for discovery learning. In: Kettemann, B. and Marko, G. (eds.), *Teaching and Learning by Doing Corpus Analysis*. Amsterdam: Rodopi, 165–182.

Bernardini, S. (2004) Corpora in the classroom. An overview and some reflections on future developments. In: Sinclair, J. (ed.), *How to Use Corpora in Language Teaching*. Amsterdam: John Benjamins, 15–36.

Bloch, J. (2009) The design of an online concordancing program for teaching about reporting verbs. *Language Learning and Technology*, **13**(1): 59–78.

Bloch, J. (2010) A concordance-based study of the use of reporting verbs as rhetorical devices in academic papers. *Journal of Writing Research*, **2**(2): 219–244.

BNC (British National Corpus). Available at: www.natcorp.ox.ac.uk/. Retrieved 6 February 2012.

Boulton, A. (2009) Testing the limits of data-driven learning: language proficiency and training. *ReCALL*, **21**(1): 37–54.

Boulton, A. (2011) Language awareness and medium-term benefits of corpus consultation. *New Trends in CALL– Working Together*. Madrid: Macmillan ELT.

Braun, S. (2005) From pedagogically relevant corpora to authentic language learning contents. *ReCALL*, **17**(1): 47–64.

Brown, P. and Levinson, S. (1987) *Politeness: Some universals in language usage*. Cambridge: Cambridge University Press.

Business Letter Corpus (BLC). Available at: http://www.someya-net.com/concordancer/. Retrieved 6 February 2012.

Carter, M. and McCarthy, M. (1995) Grammar and the Spoken Language. *Applied Linguistics*, **16**(2): 141–158.

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

Chambers, A. (2007) Editorial. Integrating corpora in language learning and teaching. *ReCALL*, **19**(3): 249–251.

Charles, M. (2007) Reconciling top-down and bottom-up approaches to graduate writing. Using a corpus to teach rhetorical functions. *Journal of English for Academic Purposes*, **6**(4): 289–302.

Charles, M. (2011) Using a corpus to teach rhetorical functions: students' evaluation of a hands-on concordancing approach. In: Frankenberg-Garcia, A., Flowerdew, L. and Aston, G. (eds.), *New Trends in Corpora and Language Learning*. London: Continuum, 26–43.

Chujo, K., Anthony, L., and Oghigian, K. (2005) DDL for the EFL classroom. http://www.ucrel.lancs.ac.uk/publications/cl2009/48_FullPaper.doc

Coccetta, F. (2011) Multimodal functional-notional concordancing. In: Frankenberg-Garcia, A., Flowerdew, L. and Aston, G. (eds.), *New Trends in Corpora and Language Learning*. London: Continuum, 121–138.

Cowie, A. P. (ed.) (1998) *Phraseology. Theory, Analysis and Applications*. Oxford: Clarendon Press.

Flowerdew, L. (2008) Corpus linguistics for academic literacies mediated through discussion activities. In: Belcher, D. and Hirvela, A. (eds.), *The Oral-Literate Connection. Perspectives on L2 Speaking, Writing and Other Media Interactions*. Ann Arbor, MI: University of Michigan Press, 268–287.

Flowerdew, L. (2009) Applying corpus linguistics to pedagogy: A critical evaluation. *International Journal of Corpus Linguistics*, **4**(3): 393–417.

Flowerdew, L. (in press, 2012a) *Corpora and Language Education*. Basingstoke, UK: Palgrave Macmillan.

Flowerdew, L. (in press, 2012b) Corpora and the classroom: An applied linguistic perspective. In: Hyland, K., Chau, M.H. and Handford, M. (eds.), *Corpora in Applied Linguistics: Current Approaches and Future Directions*. London: Continuum.

Frankenberg-Garcia, A. (2011) Raising teachers' awareness of corpora. *Language Teaching* doi:10.1017/S026144481000080, Cambridge University Press. http://journals.cambridge.org/action/displayAbstract?fromPage=online&aid=7931797

Gavioli, L. and Aston, G. (2001) Enriching reality: language corpora in language pedagogy. *ELT Journal*, **55**(3): 238–246.

Goffman, E. (1967) *Interaction ritual: essays on face-to-face behavior*. Garden City: New York.

Hunston, S. and Francis, G. (2000) *Pattern Grammar*. Amsterdam: John Benjamins.

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

Johns, T. (1997) Contexts: The background, development and trialling of a concordance-based CALL program. In: Wichmann, A., Fligelstone, S., McEnery, T. and Knowles, G. (eds.), *Teaching and Learning Corpora*. London: Longman, 100–115.

Johns, T., Lee, H- C. and Wang, L. (2008) Integrating corpus-based CALL programs in teaching English through children's literature. *Computer Assisted Language Learning*, **21**(5): 483–506.

Kasper, G. (2001) Classroom research on interlanguage pragmatics. In: Rose, K. and Kasper, G. (eds.), *Pragmatics in Language Teaching*. Cambridge: Cambridge University Press, 33–60.

Kennedy, C. and Miceli, T. (2010) Corpus-assisted creative writing: introducing intermediate Italian learners to a corpus as a reference resource. *Language Learning and Technology*, **14**(1): 28–44.

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

McCarthy, M. (1998) *Spoken Language and Applied Linguistics*. Cambridge: Cambridge University Press.

Mishan, F. (2004) Authenticating corpora for language learning. *ELT Journal*, **58**(3): 219–227.

Moon, R. (1997) Vocabulary connections: multi-word items in English. In: Schmitt, N. and McCarthy, M. (eds.), *Vocabulary: Description, Acquisition and Pedagogy*. Cambridge: Cambridge University Press, 40–63.

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

O'Sullivan, I. (2007) Enhancing a process-oriented approach to literacy and language learning: The role of corpus consultation literacy. *ReCALL*, **19**(3): 269–286.

O'Sullivan, I. and Chambers, A. (2006) Learners' writing skills in French: corpus consultation and learner evaluation. *Journal of Second Language Writing*, **15**(1): 49–68.

Robinson, P. (1997) Individual differences and the fundamental similarity of implicit and explicit adult second language learning. *Language Learning*, **47**(1): 45–99.

Schmidt, R. (1990) The role of consciousness in second language learning. *Applied Linguistics*, **11**(2): 129–158.

Schmidt, R. (2001) "Attention". In: Robinson, P. (ed.), *Cognition and Second Language Instruction*. Cambridge: Cambridge University Press, 3–32.

Seidlhofer, B. (2011) *Understanding English as a Lingua Franca*. Oxford: Oxford University Press.

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

Sinclair, J.McH. (2004) The lexical item. In: Weigand, E. (ed.), *Contrastive Lexical Semantics*. Amsterdam: John Benjamins, 1–24.

Stubbs, M. (1987) 'A matter of prolonged field work': Notes towards a modal grammar of English. *Applied Linguistics*, **7**(1): 1–25.

Stubbs, M. (1996) *Text and Corpus Analysis*. London: Blackwell.

Stubbs, M. (2006) Corpus analysis: the state of the art and three types of unanswered questions. In: Thompson, G. and Hunston, S. (eds.), *System and Corpus. Exploring Connections.* London: Equinox, 15–36.

Swain, M. (1998) Focus on form through conscious reflection. In: Doughty, C. and Williams, J. (eds.), *Focus on Form in Second Language Acquisition*. Cambridge: Cambridge University Press, 64–81.

Thompson, P. and Tribble, C. (2001) Looking at citations: using corpora in English for Academic Purposes. *Language Learning and Technology*, **5**(3): 91–105.

Thurstun, J. and Candlin, C. (1998a) *Exploring Academic English: A workbook for student essay writing*. Macquarie University: NCELTR.

Thurstun, J. and Candlin, C. (1998b) Concordancing and the teaching of vocabulary of academic English. *English for Specific Purposes*, **17**(3): 267–280.

Truscott, J. (1998) Noticing in second language acquisition: a critical review. *Second Language Research*, **14**(2): 103–135.

WebCorp. Available at: www.webcorp.org.uk/. Retrieved 6 February 2012.

Widdowson, H. G. (1998) Context, community and authentic language. *TESOL Quarterly*, **32**(4): 705–716.

Widdowson, H. G. (2000) On the limitations of linguistics applied. *Applied Linguistics*, **21**(1): 3–25.

Wilkins, D. (1976) *Notional Syllabuses*. Oxford: Oxford University Press.