

# USING A CORPUS TO ENHANCE PRAGMATIC AWARENESS

**Brian Clancy and Anne O’Keeffe**  
Mary Immaculate College, University of Limerick

## **1. Introduction**

Over the past two decades, corpus linguistics has had an increasing influence on the field of language teaching pedagogy. This chapter will demonstrate, in particular, the benefits that the fortuitous blend of corpus linguistics and pragmatics offers language teachers and learners. One of the primary benefits is that the language represented in corpora is authentic and naturally-occurring. Corpus linguistics is “the study of language based on examples of real life language use” (McEnery and Wilson, 2001: 1). It involves the assembly of a number of spoken and/or written texts that are collected according to a principled set of design criteria. These texts are then stored electronically as a whole, or a ‘corpus’, and analysed by specifically designed computer software. The electronic nature of a corpus affords the researcher access to the results of quantitative analysis, represented as, amongst other things, frequency-based information. These results allow researcher intuition about language to be measured against a naturally-occurring language sample, something that was not possible prior to the advent of corpora. The development of modern spoken corpora and the speaker information they contain, such as number of participants, speaker relationship, conversation topic and channel of communication, has also facilitated detailed qualitative analysis. This allows for the “explanation, exemplification and interpretation of the patterns found in quantitative analyses” (Biber et al., 1998: 5). In relation to language teaching and learning, corpus-based studies have highlighted a frequent mismatch between authentic language use and the language that is presented in language textbooks (see, for example, Holmes, 1988; McCarthy and Carter, 1995; Römer, 2004). Corpora

provide teachers and learners with quantifiable evidence to test their language intuitions and hypotheses against or a resource that can impart an answer to a question about language. O’Keeffe, McCarthy and Carter (2007) point out that corpora have also highlighted features about language that had previously eluded intuition.

According to Crystal (1985: 240), “pragmatics is the study of language from the point of view of users, especially of the choices they make, the constraints they encounter in using language in social interaction and the effects their language has on other participants in the act of communication”. Pragmatics might, therefore, be interpreted as the study of how speakers and writers successfully accomplish their goals, i.e. to get things done, while at the same time attend to the relationship they have with others (Leech, 1983; Crystal, 1985; Kasper, 1997). The notion of context is crucial to this. Context is complex and multi-faceted in nature, and is comprised of a number of cultural, social and discursal factors such as interpersonal shared knowledge. The importance of the interpersonal to the study of pragmatics and, indeed, to language use in general, cannot be overstated. It is the interpersonal that has traditionally separated the study of pragmatics from that of syntax or semantics. Yule (1996: 4) maintains that pragmatics explores “the relationship between linguistic forms and the user(s) of these forms.” Contextual factors such as interpersonal shared knowledge impact on the particular pragmatic choice made by speakers and writers. Corpus linguistics has allowed for the comparison of this pragmatic choice at a number of levels. These include language variety (e.g. between Irish English and British English), medium (e.g. spoken language versus written language) and, of great import to the study of pragmatics, recent studies in corpus linguistics have highlighted the fact that specific, local-level discourse domains (e.g. radio phone-in or family discourse) use language in pragmatically specialised ways (see O’Keeffe et al., 2011). This enables the researcher/teacher to examine in detail the contextual factors associated with a particular language choice at a

particular level. In relation to language learning, the ability of learners to make the linguistic choice appropriate to the context within which they find themselves is referred to as their ‘pragmatic competence’. Therefore, it might be argued that a corpus provides language teachers and learners with a wide range of authentic linguistic choices as made by spoken and written language participants, and the contextual information surrounding these choices, thus allowing the examination of what might be called ‘real’ pragmatic competence.

Edwards and Csizér (2004: 17) describe pragmatic competence as an “organic part of the learners’ communicative competence”, commensurate to, not contained within, grammatical competence. Traditionally, pragmatic competence has been subdivided into two components – *pragmalinguistics* and *sociopragmatics* (Leech, 1983). Pragmalinguistics refers to “the resources for conveying communicative acts and relational or interpersonal meaning” (Kasper, 1997: 2). These resources include a range of speech acts and politeness strategies and the devices available for intensifying or softening these acts. Sociopragmatics refers to the knowledge of how to make an appropriate pragmatic choice based on a particular goal in a particular setting. It is necessary to consider addressing these components in the language classroom for a number of reasons. Pragmatic misunderstandings can lead to a negative evaluation of a non-native speaker by a native speaker, due to the fact that grammatical errors are expected of non-native speakers and acknowledged as part of the language learning process, whereas pragmatic errors are “often interpreted on a social or personal level” (Bardovi-Harlig and Mahan-Taylor, 2003: 38). It also appears that pragmatic competence does not develop in tandem with a learner’s grammatical competence (Bardovi-Harlig and Dörnyei, 1998). In addition, it seems that exposure to a second language alone without instruction is insufficient for the acquisition of pragmatic competence (Schmidt, 1993; Kasper and Rose, 2002a, 2000b; Rose, 2005). Finally, as will be highlighted

here, it has been shown that the pragmatic content in much of the commonly used ELT material is restricted or inconsistent.

## **2. Corpus linguistics, pragmatics and language learning and teaching**

Corpus linguistic research of relevance to the teaching and learning of pragmatic competence is characterised by the fact that authentic language use yields a variety of patterns that are not readily evident in ELT materials or traditional grammars. One particularly fertile area for corpus-based studies is the comparison of the distribution, meaning and context of use of modal particles in corpus data and in ELT textbooks. Möllering and Nunan (1995: 41) characterise modals as important “indicators of pragmatic competence”. Modals often function pragmatically as politeness markers in the form of hedges (Brown and Levinson, 1987), allowing speakers and writers to downtone the force of an utterance or an argument respectively. Corpus-based studies have illustrated that the presentation of many modal particles in textbooks differs markedly from their use in the authentic, everyday speech of native speakers. Broadly speaking, modal devices have two different types of meaning: ‘deontic’ refers to modal meaning associated with permission, obligation or volition, whereas ‘epistemic’ modality refers to the indication of likelihood, e.g. possibility or prediction. Holmes (1988) uses four different corpora, the Brown corpus of written American English, the parallel Lancaster-Oslo-Bergen corpus of British English, semi-formal and informal sections from the London Lund Corpus of spoken English, and a 50,000 word ‘base corpus’ of spoken and written English across a variety of contexts, to determine the frequency of lexical epistemic devices in spoken and written English. She then compared these results to occurrences of epistemic devices in four textbooks. She claims that textbooks devote an “unjustifiably large amount of attention” to modal auxiliary verbs while neglecting other devices that express doubt or uncertainty such as lexical verbs (e.g. ‘appear’,

‘hope’, ‘think’) or adverbials (e.g. ‘at first sight’, ‘maybe’, ‘obviously’) (Holmes, 1988: 40). She provides a wealth of frequency counts to support this finding and argues that these counts could provide textbook writers with “a more principled basis on which to select the forms to be included in their ESL materials” (Holmes, 1988: 40).

Sixteen years later, this situation appeared not to have been addressed. Römer (2004) compares the distribution of modals in the 10-million-word spoken component of the British National Corpus (BNC) with six ELT textbooks and a reference grammar. She found, for example, that in textbooks 78.3% occurrences of ‘could’ are used to express ability, in comparison to the BNC where the corresponding figure is 34%. Also, ‘may’ is used to express permission in 41.5% of the occurrences in textbooks, but only in 13% of the occurrences in spoken English; ‘may’ is predominantly employed to express uncertainty in real language use (83%). Römer, like Holmes, suggests that these findings are used to improve teaching materials. She suggests, for one, a changing of the order in which modals are introduced from ‘can’ → ‘must’ → ‘may’ → ‘could’ → ‘would’ → ‘should’ → ‘will’ → ‘shall’ → ‘ought to’ → ‘might’ to ‘will’ → ‘would’ → ‘can’ → ‘could’ → ‘should’ → ‘might’ → ‘must’ → ‘may’ → ‘shall’ → ‘ought to’, to reflect their frequency of occurrence in the BNC, thereby equating frequency of occurrence with degree of importance in communication. She also suggests that more focus be placed on the relationship between past tense modals and politeness, which, she maintains, is “an important concept which is still very much neglected in the EFL context” (Römer, 2004: 197). Although not a study in the ELT context, Farr and O’Keeffe’s (2002) study on the variational distribution of the hedging devices ‘I would say’ and ‘I’d say’, highlights the importance of investigating past tense modals, politeness and context of use. They examined the frequency of occurrence of these hedges across three one-million-word corpus samples: the Limerick Corpus of Irish English (LCIE), the Cambridge and Nottingham Corpus of Discourse in English

(CANCODE) the Cambridge International Corpus (CIC), representing Irish, British and American English respectively. They found that these devices are used most frequently in an Irish context, the Irish being twice as ‘hedgy’ as their American counterparts. For language teachers and learners, this may point toward the need for ELT to take account of language variation when designing teaching materials (see also Conrad, 2004).

Corpus linguistics has played an especially prominent role in exploring the relationship between other pragmatic devices utilised in spoken language and their contexts of use. Corpus analysis has shown that discourse markers are among the most frequent forms employed in spoken discourse (see, for example, Biber et al., 1999; Carter and McCarthy, 2006). One of the primary functions of discourse markers is as pragmatic devices (see, for example, Blakemore, 1987; Schiffrin, 1987; Fraser, 1996; Aijmer, 2002) in that they “indicate an interactive relationship between speaker, hearer and message” (Fung and Carter, 2007: 411). Discourse markers (DMs) are pedagogically relevant for learners at an advanced level because, through their, they can progress towards a C2 level in the *Common European Framework of Reference for Languages* (Mukherjee and Rohrbach, 2006). Müller (2005) explored the pragmatic functions of the discourse markers ‘well’, ‘like’, ‘you know’ and ‘so’ in the Giessen Long Beach Chaplin Corpus, which consists of recordings of English and German-speaking university students. She assigned ‘you know’ five different functions (labelled ‘imagine the scene’, ‘see the implication’, ‘reference to shared knowledge’, ‘appeal for understanding’ and ‘acknowledge that the speaker is right’) and found that there are two of these functions (‘see the implication’ and ‘appeal for understanding’) for which there is no significant difference between German students speaking English and native speakers of English. There was, however, a considerable difference in the use of the other three pragmatic functions of ‘you know’.

Fung and Carter (2007) explored the use of DMs in two pedagogical corpora – a 14,157-word corpus from group discussions of intermediate-advanced learners of English in a secondary school in Hong Kong, and the pedagogic sub-corpus from CANCODE (460,055 words, native speakers of English). They found that DMs, while present in the student corpus, are generally less frequent than in British English. However, their frequency counts demonstrated that those DMs that function interpersonally, for example, ‘you know’, ‘well’, ‘sort of’, ‘yeah’, have only limited occurrences in the student data. The exception to this was ‘I think’, which was found to be used markedly more frequently in the student corpus. Previous research into this marker has suggested that it is used primarily as a marker of politeness and uncertainty, rather than as a verb of cognition (Brown and Levinson, 1987; Holmes, 1990). Fung and Carter, however, indicate that ‘I think’ is used in their student corpus to mark speaker thoughts to an extent which, they claim, indicates pragmatic fossilisation (see also Romero Trillo, 2002). They also found evidence of pragmatic fossilisation in the students’ frequent use of ‘but’ and ‘because’. In terms of pedagogical implications, Fung and Carter suggest that the students’ range of DMs, and the frequency with which they use them, reflects the unnatural and restricted input they receive in the ELT context. They cite the example of ‘well’, whose adverb, adjective and noun meanings receive attention in ELT, but whose pragmatic meanings are largely ignored (although Mukherjee and Rohrback (2006) suggest that ‘well’ is not as underrepresented as other discourse markers in modern materials). They advocate awareness-raising approaches such as the Illustration-Interaction-Induction model (McCarthy and Carter, 1995) in order that students become competent and intelligible interactionally.

Vague language, which, as Carter (1997) observes, is almost always highly significant pragmatically, is another prevalent feature of everyday spoken language. Drave (2002) studied the use of vague language in intercultural conversations between native speakers of English

(NSE) and native speakers of Cantonese (NSC), recorded in Hong Kong. Using two corpora, a 98,310-word corpus of NSE and an 84,208-word NSC corpus, he generated frequency counts and found that NSE are ‘vaguer’ than NSC. However, he noted that the range of vague expressions used by the two groups was broadly similar. Functionally, he maintains that in intercultural conversation, vague language is used for promoting politeness and intersubjectivity and for managing asymmetries of knowledge. Through an analysis of the vague language marker ‘stuff’, he discovered that NSE use ‘stuff’ for a range of pragmatic functions which suggest that speakers share knowledge and assumptions. The NSC, however, does not use ‘stuff’ in this way. He offers the tentative explanation that “perhaps prevailing pedagogical methods do not allow for sufficient exposure to native language models which contain vague language, such as informal conversation” (Drave, 2002: 38).

Corpus linguistics has shown spoken grammar to be rich, flexible, emergent and intensely interpersonal. O’Keeffe et al. (2007) analyse occurrences of *which*- and *if*-clauses and *wh*-cleft clauses in spoken language across a range of corpora. They note that all three patterns have important interpersonal functions – *which*-clauses serve to evaluate and encode attitude and stance, *if*-clauses are used in hierarchical speech situations to help create a non-threatening context, and *wh*-clefts can also be used to encode attitude and stance. They maintain that these three language patterns depart from canonical grammatical ‘rules’ which lends support to the theory of emergent grammar (Hopper, 1998), “where structure is not seen as a pre-ordained system through which discourse realises its communicative intent, but rather, the opposite: grammar is always ‘deferred’, temporally negotiable, and is always emergent from the exigencies of discourse, moment by moment” (O’Keeffe et al., 2007: 136). This has one very important implication for what corpora can tell us about the teaching of pragmatics. O’Keeffe et al’s analysis of *if*-clauses in particular highlights the relationship between grammatical form and



context. This grammatical form appears to be sensitive to context and has emerged as a result of particular participant ‘needs’ in a specific context – in this case, feedback sessions between English language teacher trainers and trainees (see also Farr and McCarthy, 2002). Therefore, it is logical to expect that different contexts have, due to their differing demands, a range of distinct grammatical forms which do not always behave as previously prescribed in grammar references. O’Keeffe et al. (2007: 138) maintain that “learners wishing to focus on their special needs should be given the opportunity to work with the typical grammatical patterns which characterise the special contexts in which they work or study”, and corpora afford these learners the opportunity to do so should they desire.

It should also be remembered that the notion of pragmatic competence does not apply exclusively to the spoken environment. Epistemic modality is also an essential discursual tool for presenting an argument in the context of academic writing. According to Hyland (1994: 241), “in persuasive writing, hedges are an important means of both supporting the writer’s position and building writer-reader relationships”. He maintains that the problems faced by L2 university students in employing modality in their writing results in an important area of pragmatic failure. Hyland compiled a corpus of 22 ELT textbooks, representative of those used around the world for the teaching of academic writing skills that covered a broad time scale, a range of writing materials and levels of proficiency. He found that, generally, the importance of hedging and hedging devices is under-represented in textbooks making the information contained therein both inadequate and misleading. For example, none of the textbooks give much attention to the use of epistemic adjectives (e.g. ‘apparent’, ‘evident’ or ‘possible’), adverbs (e.g. ‘essentially’, ‘probably’ or ‘undoubtedly’) and nouns (e.g. ‘assumption’, ‘claim’ or ‘evidence’), despite their widespread use in academic writing. Holmes (1988) suggests that these grammatical classes comprise 27% of devices used to express epistemic modality in written discourse. In common

with Holmes (1988), he calls for materials developers to employ the authentic and empirical data provided by corpora in order to raise learner awareness of socio-pragmatic variation in specialist registers.

Pronoun usage is an area that is generally under-represented in pragmatic research (Levinson, 2004). Our use of pronouns is connected to our deictic system, a system that facilitates contextual orientation. For example, in the utterance “I’ll meet you here at five o’clock”, the meanings of ‘I’, ‘you’, ‘here’ and ‘five o’clock’ are all determined by reference to the context – ‘I’ and ‘you’ are the speakers, ‘here’ is where the conversation took place and ‘five o’clock’ is determined by the moment of utterance. Hyland (2002a) examined the presence of author pronouns in two corpora – an ‘expert’ corpus of 240 published journal articles, 30 from each of eight disciplines, and a ‘novice’ corpus of 40 project reports across six fields written by final-year undergraduates in Hong Kong (see also Hyland, 2002b). He found that there were 12 author pronouns per text in the novice corpus compared to 22 in the expert corpus. He also noted that in the expert corpus there is considerable disciplinary variation with 75% of author pronouns occurring in the social sciences and humanities, while the sciences and engineering accounted for only 25%. However, he found that this variation was largely absent in the novice corpus, in fact, when the results from the two corpora were normalised, expert writers were three times more likely to use author pronouns in their texts, and this applied to both the hard and soft disciplines. He maintains that academic writing is commonly portrayed as impersonal and faceless in textbooks and style guides. However, to portray academic writing as such is to ignore the degree of subject-specific variability present therein. He claims that “by avoiding the use of author pronouns, and failing to stand behind their interpretations, these emerging writers run the risk of not establishing an effective authorial identity, and of failing to create a successful academic argument” (Hyland,

2002a: 354). Hyland also advocates an awareness-raising approach where students' critically evaluate the use of 'I' in their own writing.

### 3. Searching a corpus for pragmatic features

This section presents an overview of the basic corpus analysis techniques that are possible using various corpus software programmes. The tables and figures presented in this section have been generated using *WordSmith Tools™, Version 5.0* (Scott, 2009). There are, however, a large variety of other software programmes available for use with a corpus.

#### Frequency

Frequency lists are often identified as a good starting point for the analysis of a corpus. For the language teacher and learner, raw frequency lists, which simply rank the order an item appears in a corpus based on the number of times it occurs, can prove a useful teaching tool for the illustration of the relationship between the frequency of occurrence of linguistic items, and the variety, genre or context in which they appear. This enables the identification of pragmatic items that may be characteristic of a particular variety, genre or context. When applied to the study of pragmatics and the development of students' pragmatic competence, corpus frequency lists can be quite beneficial, especially when presented in direct contrast with one another. Table 1 features the first 25 words from three different corpora:

Table 1. Comparison of word frequencies for the 25 most frequent words across three corpora.

<b>Rank</b>	<b>LCIE</b>	<b>BNC</b>	<b>LIBEL</b>
1	the	the	the
2	I	I	and
3	and	you	of

4	you	and	you
5	to	it	to
6	it	that	that
7	a	a	a
8	that	's	in
9	of	to	it
10	yeah	of	is
11	in	n't	I
12	was	in	's
13	is	we	so
14	like	is	what
15	know	do	we
16	he	they	this
17	on	er	they
18	they	was	on
19	have	yeah	there
20	there	have	have
21	no	what	for
22	but	he	okay
23	for	to	amm
24	be	but	ahh
25	what	for	are

- The Limerick Corpus of Irish English (LCIE): a one-million-word corpus of spoken English collected in Southern Ireland;
- The British National Corpus (BNC): a ten-million-word corpus of spoken British English;
- The Limerick and Belfast Corpus of Academic Spoken English (LIBEL CASE, hereafter LIBEL): a one-million-word corpus of academic English collected on the island of Ireland.

Table 1 demonstrates that a considerable amount of pragmatic information can be gleaned simply by working with the most frequent 25 words in each corpus. In relation to pragmatics and variety, when spoken Irish English (LCIE) is compared to spoken British English (BNC), it can be seen that, for example, ‘like’ and ‘know’, positions 14 and 15 respectively on the LCIE frequency list, do not appear in the top 25 words in the BNC. ‘Like’ and ‘know’ (and its associated chunks such as ‘you know’, see below) have been shown to function frequently as pragmatic markers of uncertainty in Irish English (see, for example, Clancy, 2011a, 2011b; Schweinberger, forthcoming). Indeed, corpus-based research into Irish English in general has shown that Irish English speakers may feel the need to mark uncertainty where other English speaking cultures may not (see also Farr and O’Keeffe, 2002; Vaughan and Clancy, 2011). Table 1 also demonstrates that the response token ‘yeah’ is present on both the LCIE (position 10) and BNC (position 19) frequency lists. This finding points toward the importance of the use of ‘yeah’ as a response token in informal, spoken English, both in Ireland and Britain. McCarthy (2002: 70) maintains that cross-corpora, inter-varietal studies offer “a powerful tool for an overall understanding of the common ground that typically exists alongside differences between one variety and another”. He also claims that these studies could contribute in some way toward an ‘average’ list for English as an international language, something which has important implications for English language pedagogy.

In relation to pragmatics and context, personal pronouns feature prominently in corpus frequency lists, especially spoken ones. Personal pronouns are strongly associated with deictic reference; in particular, personal pronouns facilitate the identification of conversational participants. In both the LCIE and BNC lists, ‘I’ and ‘you’ appear in the top four most frequent items. This is indicative of the high level of interactivity between participants in casual conversation. However, in marked comparison to both the LCIE and BNC lists, in the LIBEL

corpus frequency list, ‘you’ (position 4) is more frequent than ‘I’ (position 11). This is, perhaps, illustrative of the different interactive nature of the context – the information presented in the academic sphere could be interpreted as largely mono-directional from lecturer/teacher/tutor to student where ‘you’ may be used to refer to ‘*you* the audience’. Of added interest is that the BNC frequency list features two plural personal pronouns ‘we’ (position 13) and ‘they’ (position 16), neither of which are present in the 25 most frequent items in the LCIE wordlist. ‘We’ and ‘they’, however, do feature in the spoken academic corpus LIBEL, in positions 15 and 17 respectively. It seems that ‘we’ is used in academic spoken discourse to invoke a professional academic community that both lecturer and student are part of. For the language teacher and learner, distinguishing between what ‘we’ refers to in different contexts appears to be essential to successful communicative competence within these domains (see, for example, McCarthy and Handford, 2004; Harwood, 2005; Vaughan, 2007). The LIBEL list is also characterised by the high frequency of spoken discourse markers such as ‘so’ and ‘okay’. These discourse markers are necessary in the academic spoken context, as they are used by speakers to ‘manage’ the discourse allowing them to mark phases such as openings and closings, new topics or rhetorical shifts for the students (Carter and McCarthy, 2006: 214-216).

Frequency information can also be used to analyse how words systematically ‘cluster’. Rather than asking the computer to generate single word frequencies, it can instead be asked to generate frequencies for recurrent strings of words, often referred to in corpus literature as ‘lexical bundles’ or ‘chunks’ (Greaves and Warren, 2010). Corpus software can provide this information in the form of 2-, 3-, 4-, 5- or 6-word bundles. In addition to providing a wealth of information about the vocabulary of a language, lexical bundles can also provide illuminating insights into the importance of the pragmatic system of a given language. Table 2 features the most frequent 2-, 3- and 4-word chunks in LCIE:

Table 2. Ten most frequent 2-word, 3-word and 4-word units in LCIE results per million words.

<b>Rank</b>	<b>2-word units</b>	<b>3-word units</b>	<b>4-word units</b>
1	you know	I don't know	you know what I
2	in the	do you know	know what I mean
3	of the	a lot of	do you know what
4	do you	you know what	I don't know what
5	I don't	do you want	do you want to
6	I think	I don't think	are you going to
7	It was	you know the	you know the way
8	I was	you have to	I don't know I
9	going to	going to be	thank you very much
10	on the	yeah yeah yeah	the end of the

Table 1 has already demonstrated the presence of the items ‘know’ and ‘like’ in the top 25 words of the LCIE frequency list, and noted that these are not present on either the BNC or LIBEL lists. Table 2 reinforces the importance of these items to the pragmatic system of Irish English in that it contains expressions such as ‘you know’, ‘I think’, ‘I don’t think’, etc., which are commonly associated with relational language, language that is used to establish or sustain our relationships with others. The presence of these items in the frequency lists perhaps indicates the importance of this relational work in Irish culture and society and, therefore, has implications for teachers and students working with this language variety.

In terms of language learning, when looking at word unit frequency lists, it is also useful to distinguish between units which seem complete (I don’t know) compared with those that seem incomplete or fragmented (know what I mean). O’Keeffe et al. (2011) point out that there is a temptation to dismiss the latter as fragmented units, but that these are in fact operating as frames for different structures, for example:

**You** know what I mean?

**If** you know what I mean?

**Do you** know what I mean?

**Does he** know what I mean?

**Does she** know what I mean?

**Do you** know what I mean **by X**?

**Does he** know what I mean **by X**?

**Does she** know what I mean **by X**? ... etc. (O’Keeffe et al., 2011: 10-11)

Mukherjee (2009) maintains that it is worthwhile to input items such as ‘you know’ as part of larger chunks that are frequently used in casual conversation at a formal level in the classroom. This results in these chunks becoming ‘automatised’ and, therefore, easily accessible in a variety of different spoken contexts.

### **Keyness**

Similar to frequency lists, keyword lists provide language teachers and learners with a tool for the development of pragmatic competence. Key words are words that occur with unusual frequency or ‘keyness’ in a target corpus relative to a norm. Using corpus software it is possible to identify keywords whose frequency is unusually high (positive keywords) or low (negative keywords) in comparison to a reference corpus. A reference corpus is, usually, a larger corpus such as the BNC which acts as a ‘baseline’ for comparison. In order to achieve this measure of keyness, the wordlist from the target corpus is compared to a wordlist in the reference corpus, and the statistical significance of difference is calculated using chi-square or log-likelihood tests. This distinguishes between frequencies that are a matter of chance, and those that “are likely to be motivated by some characteristic of the communicative event” (Anderson and Corbett, 2009: 37).



Scott and Tribble (2006) maintain that it is impractical for teachers to employ detailed statistical analysis of corpora in the classroom in order to demonstrate differences between, for example, speech and writing. Instead, they recommend the use of keyword analysis which, they maintain, gives teachers and learners ‘an accessible means’ for exploring difference.

Due to the statistical nature of the list, keywords are, according to Baker (2006), indicative of saliency as opposed to only providing frequency. Scott and Tribble (2006: 56) maintain that keywords “reflect what the text is really about, avoiding trivia and insignificant detail. What the text ‘boils down to’ is its keyness, once we have steamed off all the verbiage, the adornment, the blah, blah, blah”. Therefore, the keywords in a corpus are often attributed to its ‘aboutness’ (see Scott, 2010; Cheng, 2009; Scott and Tribble, 2006). In relation to keyword analysis, Scott (2010: 165) cautions that researchers “don’t compare apples with phone boxes!”, meaning that the selection of the reference corpus to be used is of importance, for example, if the researcher wishes to generate a keyword list for a selection of newspaper articles and uses a spoken corpus, then the characteristics of spoken versus written language may affect the keyword list. A suitable comparison in this case would be with a larger written corpus or, ideally, a larger corpus of newspaper articles (although Scott and Tribble (2006) note that when the BNC was used as a reference corpus for studying a Shakespeare play, useful items for follow-up still emerged).

Keyword analysis can be used by the language teacher and learner according to their needs to explore different conventions of different contexts. Spoken keyword lists generally contain three types of words: proper nouns, ‘aboutness’ words and high frequency grammatical words. It is often these ‘aboutness’ words that are the focus of corpus studies, however, for the study of pragmatics, both the proper nouns and the high frequency grammatical words also offer a rich vein of material that can be exploited in the language classroom. For example, Table 4 contains

the top ten keywords, minus proper nouns and ‘aboutness’ words, of C-MELT, a corpus of meetings of English language teachers,<sup>1</sup> generated with LCIE as the reference corpus.

Table 4. Top 10 keywords, minus content items, of C-MELT with LCIE as reference corpus.

<b>Rank</b>	<b>Keywords</b>
1	we
2	think
3	they
4	okay
5	so
6	kind
7	mean
8	maybe
9	if
10	could

This keyword list offers a wealth of information into the pragmatic system of a specific discourse domain. The list contains a number of epistemic markers that may indicate uncertainty such as ‘think’, ‘maybe’ and ‘could’. It also contains ‘kind’, which is connected to the vague language item ‘kind of’. Finally, ‘we’ is shown to be the most significant grammatical item in the corpus. The work of Vaughan (2007, 2010) and McCarthy and Handford (2004) has illustrated the particular pragmatic practices employed in professional context-types such as business meetings. For example, ‘we’ has been shown to be a device that can be employed as an unthreatening means of proffering identity in meetings, whereas the epistemic markers serve to soften the presentation of suggestions or directives.

---

<sup>1</sup> C-MELT is a 40,000-word corpus of English language teacher meetings. Thanks to our colleague Elaine Vaughan, University of Limerick, for allowing us to use her corpus.

This is not to say that the aboutness words should be discarded in the teaching of pragmatics.

Table 5 features the keyword list for lexical verbs used in a corpus of first year undergraduate philosophy essays with the written component of ICE-Ireland as the reference corpus.<sup>2</sup>

Table 5. Lexical verb keyword list for first year undergraduate philosophy essays with ICE-Ireland (written) as reference corpus.

<b>Rank</b>	<b>Lexical verb</b>
1	die
2	believed
3	believe
4	prove
5	commit
6	believing
7	believes
8	claims
9	says
10	defend
11	speaks
12	brought
13	convince
14	committed
15	accused
16	shows
17	states
18	makes
19	argue
20	knew

---

<sup>2</sup> The first year undergraduate corpus is a 50,000-word corpus of philosophy essays collected in Mary Immaculate College, University of Limerick, Ireland. Thanks to our colleague James Binchy for allowing us the use of this corpus.

If verbs such as ‘die’, ‘commit’, ‘defend’ and ‘accused’ are omitted (these are connected to specific essay titles assigned to the students in this particular philosophy module), a number of reporting verbs associated with citation in academic writing such as ‘believe’, ‘claim’, ‘show’, ‘state’ or ‘argue’ emerge as key. Hyland (1999) illustrates that philosophy has the highest citation rate in comparison to seven other academic disciplines – sociology, applied linguistics, marketing, biology, electronic engineering, mechanical engineering and physics. Based on corpus analysis, he also identifies the seven most frequent reporting verbs used by expert philosophical writers – ‘say’, ‘suggest’, ‘argue’, ‘claim’, ‘point out’, ‘propose’ and ‘think’. Table 5 demonstrates that although these students are at an early stage of their academic careers, they have already identified and utilised some of the key reporting verbs, ‘say’, ‘claim’ and ‘argue’, in their discipline. This has obvious teaching applications. For example, the keyword list in Table 5, coupled with Hyland’s (1999) findings, offer a starting point for awareness-raising activities in English for Academic Purposes, where the difference in reporting verb use between novice and experts writers is highlighted and discussed. This has important pragmatic outcomes for EAP students. Pragmatics, as discussed, is the study of how speakers and writers accomplish their communicative goals while maintaining interpersonal relationships. Hyland (1999: 359) maintains that “research in any field has significance only in relation to existing literature, and citation helps demonstrate accommodation to this community knowledge”. However, Hyland has demonstrated that expectations of how this accommodation to community knowledge is presented are different across different disciplines. Therefore, part of writers successfully accomplishing their goals and maintaining interpersonal relationships with their reader is to choose the reporting verbs that demonstrate appropriate accommodation to community expectations.

## Concordance

Concordances are perhaps the most familiar manifestation of a corpus in the language classroom. As demonstrated, frequency lists often identify items that merit further investigation. Concordances provide the language teacher and learner with detailed contextual information about these specified items. Concordance lines are often presented with the search, or ‘node’, word in the centre surrounded by a number of words on either side. Figure 1 shows a randomised selection of 20 concordance lines generated from LIBEL with ‘okay’ as the node word:

Figure 1. Random sample of 20 concordance lines for *okay* in LIBEL.

N	Concordance
1	. To vote out of office okay . Okay yeah? . <b>Okay</b> so right to vote and freedom of speech .
2	in at the end you know is absolutely fine . <b>Okay?</b> It's it's again kind of something that
3	going to come up with that standard of writing <b>okay?</b> In the terms of the language that
4	present it in a bad way and you lose marks . <b>Okay</b> . Now that point that is unfortunate but
5	either the essay questions or in the exam . <b>Okay?</b> Is the book related to the course? If it
6	sentence . Now everybody has a voice . <b>Okay?</b> If you understand what you say . I
7	get the gist and find out what was going on . <b>Okay?</b> So ye learned success . What was
8	two thousand c cs and Bob's your uncle . <b>Okay?</b> If you've any questions about that just
9	articles that were recommended by Stevens <b>okay?</b> That case . We'll have a look at that
10	people who are up to that great challenge . <b>Okay</b> . Graphic design then is all we need
11	Are there any problems with reading it? No? <b>Okay</b> . If you just go back to the applied
12	in . Like I said less employment available . <b>Okay</b> so the the rockites ahh under the guise
13	is we want a function that will do this for us <b>okay?</b> So I want a function that will look here
14	in bold but that's kind of it's in ordinary font <b>okay?</b> So what that means is that means
15	anywhere you want . So you might say <b>okay</b> I might maybe prefer down at the bottom
16	or comments or anything before we go? No . <b>Okay</b> well I'll leave it at that then . Amm if
17	clear why everyone would want a solution . <b>Okay</b> and then the next block . And that's
18	you ahh generally buy something you pay vat <b>okay?</b> Now that's an indirect tax right?
19	the company auditors have a statutory right <b>okay?</b> ten second pause . And lastly
20	power and took a look around and she said <b>okay</b> I've a problem here . First and foremost

Concordances are often read vertically or from the centre moving to the left or the right and, therefore, it may be necessary to introduce students to how to read concordances. Sinclair (2003)

recommends a seven-step process for reading concordances: Initiate → Interpret → Consolidate → Report → Recycle → Result → Repeat (this process is explored in detail with examples in Tribble (2010)). Many corpus software programmes allow the re-sorting of the concordance to the left or right of the node word in order to examine patterns that occur before or after the node. Corpus software also facilitates a quick switch from an individual concordance line to the original text in which it appears. Concordances are important for the study of pragmatics because they allow the researcher to determine whether or not an item has, in fact, a pragmatic function. For example, the LCIE frequency list in Table 1 highlights the potential that ‘like’ is pragmatically significant; however, concordances allow instances of the lexical verb ‘like’ to be separated from occurrences of ‘like’ as a pragmatic marker.

Although concordance lines can be an interesting and motivating classroom resource, it is necessary to treat them with some caution. Johns (1997: 114) warns that the selection and editing of concordance lines is “time-consuming and requires fine linguistic and pedagogic judgement”. Spoken language features such as ellipsis, repetitions and false starts often mean that, visually at least, spoken discourse can appear ‘disorganised’. In addition, because concordances only present approximately eight or nine words to the left or right of the node word, then students are frequently dealing with truncated sentences. Johns notes that short, complete sentences are easier for the student to deal with than longer, incomplete ones. Some corpus transcription conventions are also quite dense and, therefore, difficult to read for the uninitiated. Accordingly, the language teacher needs to carefully select examples and, in some cases, may have to ‘clean them up’ by eliminating unnecessary and confusing aspects of spoken language and/or its transcription. Mukherjee (2009) recommends using concordance lines based on language generated by the learners themselves. This ‘individualisation’ of corpus analysis is, he claims, more psychologically rewarding for, cognitively accessible to, and typical of, learners at a certain level.

In the language classroom, concordances might function to allow student hypotheses about a variety, genre or context to be proven or disproven. Figure 2 shows 20 sample concordance lines generated from LCIE with the item ‘now’ as node word:

Figure 2. Sample of 20 concordance lines for *now* in LCIE.

N	Concordance
1	they're in that . So you can rename them <b>now</b> yourself I'm not going to do it at all for
2	you go . Its up there . Yeah . There you are <b>now</b> the only thing is its pulling all you can
3	when it flashes on charge you put it like that <b>now</b> . Yeah . There tis now . No you see I
4	to kind of you know you can see it there <b>now</b> . Yes . The guy who sings has dyed red
5	song was nice . The German song was . <b>Now</b> how do you listen to those kind of
6	. Now . There's the fire . There's the fire . <b>Now</b> is that a good shot or is that not a
7	the religious artefacts to this fella ! fellow . <b>Now</b> . There's the fire . There's the fire . Now
8	onto your system? Do shure stick them on . <b>Now</b> am where is my pictures do you know?
9	Syl Adley's place look . Syl . He has his <b>Now</b> do you want those things will I put them
10	good I'm going to print that off . That's it <b>now</b> Dermot Lynch said he'd send me all
11	two three laughing It's sent . Is it? It's gone <b>now</b> . background talking He'll go that Derek
12	and she goes I was telling you about it <b>now</b> in Paris and I said what's this and she
13	like yeah sure there's Ah Eva's off to Canada <b>now</b> in+ she's off in two weeks . +she's going
14	\$1> How are you Gerry? Not too bad Eileen <b>now</b> . Ah yeah with the exams . ah you'd get
15	I don't know about being mechanically minded <b>now</b> . If you'd do well Derek listen have a go
16	. That's her that's the mother . The other one <b>now</b> is just Yeah . The other thing is on
17	you want they're not holding at all Tommy <b>now</b> ? Which no . The bananas . Are they
18	at all Terry I only bought them yesterday <b>now</b> and the fruit there+ Yeah . +and they go
19	laughing I know well that's another story <b>now</b> . you know what I mean shure give John
20	take care of yourself alright . Alright bye <b>now</b> . See ya

From the concordances, students, supported by the teacher, should be able to draw conclusions on the use of ‘now’ in spoken Irish English:

- ‘now’ can be used in the initial (e.g. line 5), medial (e.g. line 1) and final (e.g. line 3) positions in an utterance and also occurs as a stand-alone item (e.g. line 7);

- ‘now’ has a variety of functions – it is used as a temporal adverb (e.g. line 14), and as a discourse marker which indicates a change in topic (e.g. line 6);
- ‘now’ can be used with a variety of time references – present time reference (e.g. line 14), past (e.g. line 18) and future (e.g. line 13);
- ‘now’ time references can occur either to the right (e.g. line 13) or to the left (e.g. line 18);
- ‘now’ may be used in phrases connected to greeting (e.g. line 14) and leave-taking (e.g. line 20).

This engagement with concordances should be followed up by the teacher with the opportunity to use and practice the item selected for analysis (Mukherjee, 2009).

#### **4. Using a corpus in the language classroom to enhance pragmatic awareness**

The first challenge that is faced by a language teacher who wishes to use a corpus is the question of which corpus. Here we will suggest some basic templates from which materials can be created using easily accessible corpora, some of which are freely available and others which are commercially available at a reasonable price. Another issue which faces the teacher is how to present the material. Should the students be given hands-on work with corpora using PCs? Should the teacher mediate and carefully select the material on hand-outs? (For a detailed discussion of these dilemmas, see Gilquin and Granger (2010) and Sripicharn (2010)). The consensus seems to be that students need training before they can cope with working with ‘live’ concordance searches. This training would help them to become accustomed to how to read concordance lines vertically and from the node word out so as to see the patterns that words make. The best way to build up students’ awareness of how to use corpora then is to carefully select sentences from a corpus to illustrate a teaching point and to guide the learners through the initial tasks. In the initial tasks, it is wise to use complete or near complete sentences where possible.



We are also mindful that corpus data does not have to be presented to learners in the form of concordances. A corpus is at its most basic a collection of naturally-occurring texts in spoken or written form and these texts can simply be drawn on for classroom use to raise pragmatic awareness. What follows therefore is a selection of ideas for corpus-based materials.

**Sample 1: Exploring the modal verb *may* in the context of presenting scientific facts**

**Corpus:** Michigan Corpus of Spoken Academic English (MICASE)

**Details:** 152 transcripts of interactions in academic settings at Michigan University (1,848,364 words)

**Availability:** Freely available online

(<http://quod.lib.umich.edu/cgi/c/corpus/corpus?page=home;c=micase;cc=micase>)

**Rationale:** The focus of this task is the use of the modal verb *may* to present factual information in a hedged manner, in a science lecture. The modal verb *may* expresses epistemic modality and it is often the case that in the presentation of facts in English that their certainty is hedged by the use of this, and other, modal verbs.

**Below are extracts from a Physical Sciences and Engineering lecture on ‘the Dynamic Earth’, recorded at Michigan University, USA. Notice how the lecturer uses the modal *may* very often when presenting scientific information. Look closely at the examples and try to identify the function of *may* in these examples (the longer extracts are presented below the table):**

<b>Modal</b>	<b>Function</b>	<b>Example (see below for fuller context)</b>
May (extract 1)		<i>we <b>may</b> be able to observe at hundred kilometers away, units one two three four five and then six</i>
May (extract 2)		<i>And in fact you <b>may</b> find that on top of that, there are other sediments deposited</i>
May (extract 3)		<i>Uh you <b>may</b> find, that, uh immediately below an unconformity like so, uh there are series of intrusions we call dikes.</i>

May (extract 4)		<i>well we think that one <b>may</b> have happened sixty-six million years ago, when the dinosaurs go extinct.</i>
--------------------	--	--

**Extracts:**

1 ...we're going to start talking about stratigraphic correlations, we **may** be able to observe at hundred kilometers away, units one two three four five and then six. That's how we infer, that a unit five ever existed.

2 okay so sedimentary layers would look like this. This **would** be say unit one two, three and four. And in fact you **may** find that on top of that, there are other sediments deposited.

3 Say this is a line, along which there's been some erosion, uh take place. Uh you **may** find, that, uh immediately below an unconformity like so, uh there are series of intrusions we call dikes. And let's have them represent several different uh, uh generations, okay so let's call this one dike A, and this one is going to be dike B, uh and to make it interesting let's put in, (a) third one. okay so this is cross sectional view, uh we're looking at a road cut, and we see sedimentary layers, that are horizontal like so <PAUSE:04> okay and these sedimentary units have been cut, uh by these intrusions that I'm calling, uh dikes. okay so sedimentary layers would look like this.

4 I would like you to remember those three terms, Paleozoic Mesozoic and Cenozoic the reason being that uh later on in our discussions we will need to know what they are I'll simply tell you life-forms of the Mesozoic and you should know what Mesozoic means. If you, if you want to you can remember, uh exactly when each one of these ends but it's not absolutely necessary, uh the Paleozoic begins five hundred and thirty million years ago ends two hundred and forty-five million years ago, .... well we think that one **may** have happened sixty-six million years ago, when the dinosaurs go extinct. In other words reptiles, uh many reptiles go extinct at this particular time, and mammals become the dominant life-form, uh subsequently.

**Sample 2: Looking at numbers and vague language**

**Corpus 1:** Corpus of Contemporary American English (COCA)

**Details:** 425 million words, spoken, fiction, magazine, newspaper and academic genres

**Availability:** Freely available online (<http://corpus.byu.edu/coca/>)

**Corpus 2:** British National Corpus (BNC), Spoken Component

**Details:** 10 million words of spoken language

**Availability:** Commercially available at a reasonable price

**Rationale:** Here we focus on the use of vague language when talking about numbers so as not to make them sound too direct (see O’Keeffe et al., 2011). This is prevalent in spoken language but not as much so in written language. If you choose two corpora, one spoken and one written, then use a number as your search word, you should be able to select good examples to compare the use of vague language in speaking with the more on-recordness of written language. Note that you may need to write out the number in alphabetical form for your search, depending on how each corpus transcribed numbers.

**1) Below is an extract from an American magazine, *American Scholar*, taken from the COCA corpus. Read through it and notice how numbers are presented.**

According to the Joint Center for Political and Economic Studies, black men represented 7.9 percent of the 18- to 24-year-olds in the U.S. population in 2000, but they constituted just 2.8 percent of undergraduate enrolments in 50 of the best public universities in the nation in 2004. In each of the 30 flagship universities, fewer than 500 black male undergraduates were enrolled that year. # Even after being enrolled, less than half of all black male students who start college at a four-year institution graduate in six years or less, a rate more than 20 percentage points lower than the white graduation rate. That is not good news: it is the lowest college completion rate among all racial groups for both sexes. Perhaps most striking about these discouraging figures is that many black male students at some of the best institutions would likely not be enrolled at all if they were not athletes. The same Joint Center study reveals that more than one out of every five black men at 21 flagship public institutions was a student athlete in 2004. At 42 of these universities...

(Source: CHACE, W. M. (2011): “Affirmative Inaction”, *American Scholar*, 80(1): 20-31)

**2) Below is an extract of a recording from the British National Corpus. Speakers A and B are talking about their school days. Notice here how numbers and times are talked about with more vague language.**

- a) What words are used by the speakers to make the numbers more vague?
- b) Why do you think speakers do this?
- c) Look back at the first example, why is the use of vague language not so common when talking about numbers in writing?

A: ... And after that you'd go to your own classes and you had a, a set < pause > er what did they call it?

B: Set er programme where you had er maybe an hour's arithmetic.

A: Addition, subtraction, multiplication, and so on, do so many sums a day, each lesson.

B: And after that you'd maybe have er well be playtime then.

A: And about ten minutes playtime running round the yard, come back and maybe have a history lesson.

B: Er maybe half an hour history and half an hour geography and er science er < pause > but it were all very elementary stuff.

A: Er < laugh > wasn't nothing technical, you know.

(Source: BNC spoken Oral History: talking about their school days)

### **Sample 3: Exploring politeness in business correspondence**

**Corpus:** Corpus of Business Correspondence

**Details:** Corpus of business letters (c.5,900 words), business memos (c.10,000 words), business reports (c.15,000 words)

**Availability:** Freely available online

([http://langbank.engl.polyu.edu.hk/corpus/business\\_correspondence.html](http://langbank.engl.polyu.edu.hk/corpus/business_correspondence.html))

**Rationale:** The focus of this task is the formal use of language in business letters so as to attend to negative politeness. It focuses on speech acts of apologies, requests and offers, as well as the

polite use of the adjective 'near' in the phrase 'in the near future'. This task could easily be limited to one speech act.

**Look closely at the extracts from the business letters below and answers to the following questions:**

- a) Find a polite more polite way of saying "soon".
- b) Find a very polite way of saying "I'm sorry" that is used in formal letters but not usually in speaking.
- c) Find a polite way of saying "Get in touch if you need any help".
- c) Find a very polite way of requesting that someone sign a contract.

**Extract 1**

Dear Fiona,

Firstly, please accept my sincere apologies for not responding sooner with regard to consignment stock; I am now able to detail below our proposal, which I trust you will find acceptable...

**Extract 2**

I would be grateful if you would indicate your acceptance by signing below and returning the original to us. In the meantime, I assure you of our closest attention at all times and remain,

Yours sincerely

**Extract 3**

I am currently looking into the cost issues relating to the Select feed Plus service and will contact you in the near future. Yours sincerely

#### **Extract 4**

I have sent a copy of your letter to City and District Health Authority and United Healthcare NHST with a request that they reply directly to your request for information. I hope you find the information contained in this letter helpful. If I can be of any other assistance to you, please do not hesitate to contact me.

#### **5. Conclusion**

In this chapter, we have attempted to make the case for two areas that are often under-represented in relation to language teaching, namely the raising of pragmatic awareness and the potential of language corpora. The dearth of awareness in relation to both of these areas can ultimately be traced to their absence from many teacher education courses. While pragmatics has gained much more attention in recent years, particularly within the study of English language and philology programmes, it is generally absent from contemporary grammars or vocabularies of English even though pragmatic meaning in relation to grammar and vocabulary is crucial. Equally, much insight has been gained over the years from the empirical study of language through the use of corpora but so many of these insights remain frozen on the pages of academic journals. There may be many reasons for this, not least of all lack of training in the use of corpora and their software. For the areas of pragmatics and corpus linguistics to become more integral to language teaching, more publications such as the present volume will play an important role.

To this end, we hope to have shown in this chapter that there is a wealth of research findings from pragmatics research using corpus linguistics that is relevant to the language classroom and course materials. We hope to have given an insight into the basic functions of corpus software (namely the generation and use of word frequency lists, concordances and

keyword lists), along with some examples of the types of information about language use that can be gleaned from these analyses. We also hope to have enthused our readers to try out corpora even if it is just for the purpose of exploring some authentic language use. We do accept that using corpora is not without its challenges. They are often messy; the language often needs more context than invented examples, and so on. However, any user of a corpus will become enthralled at the evidence-base for language use that it offers.

## References

AIJMER, K. (2002): *English Discourse Particles: Evidence from a Corpus*. Amsterdam: John Benjamins.

ANDERSON, W. and J. CORBETT (2009): *Exploring English with Online Corpora*. Basingstoke: Palgrave.

BAKER, P. (2006): *Using Corpora in Discourse Analysis*. London: Continuum.

BARDOVI-HARLIG, K. and R. MAHAN-TAYLOR (2003): "Introduction to teaching pragmatics", *English Teaching Forum*, July, 37-39.

BARDOVI-HARLIG, K. and Z. DÖRNYEI (1998): "Do language learners recognise pragmatic violations?: Pragmatic versus grammatical awareness in instructed L2 learning", *TESOL Quarterly*, 32, 233-259.

BIBER, D., S. CONRAD and R. REPPEN (1998): *Corpus Linguistics: Investigating Language Structure and Use*. Cambridge: Cambridge University Press.

BIBER, D., S. JOHANSSON, G. LEECH, S. CONRAD and E. FINEGAN (1999): *Longman Grammar of Spoken and Written English*. Harlow: Pearson Longman.

- BLAKEMORE, D. (1987): *Semantic Constraints on Relevance*. Oxford: Blackwell.
- BROWN, P. and S. LEVINSON (1987): *Politeness: Some Universals in Language Usage*. Cambridge: Cambridge University Press.
- CARTER, R. (1997): "Orders of reality: CANCODE, communication, and culture", *ELT Journal*, 52, 43-56.
- CARTER, R. and M. MCCARTHY (2006): *Cambridge Grammar of English*. Cambridge: Cambridge University Press.
- CHENG, W. (2009): "Income/interest/net: using internal criteria to determine the aboutness of a text". In K. AIJMER (ed.): *Corpora and Language Teaching*. Amsterdam: John Benjamins, 157-178.
- CLANCY, B. (2011a): "Complementary perspectives on hedging behaviour in family discourse: the analytical synergy of corpus linguistics and variational pragmatics", *International Journal of Corpus Linguistics*, 16, 372-391.
- CLANCY, B. (2011b): "Do you want to do it yourself like?: hedging in Irish Traveller and settled family discourse". In B. DAVIES, M. HAUGH and A. MERRISON (eds.): *Situated Politeness*. London: Continuum, 129-146.
- CONRAD, S. (2004): "Corpus linguistics, language variation and language teaching". In SINCLAIR, J. (ed.): *How to use Corpora in Language Teaching*. Amsterdam: John Benjamins, 67-85.
- CRYSTAL, D. (1985): *A Dictionary of Linguistics and Phonetics* (2nd edition). Oxford: Blackwell.



DRAVE, N. (2002): "Vaguely speaking: a corpus approach to vague language in intercultural conversations". In PETERS, P., P. COLLINS and A. SMITH (eds.): *New Frontiers of Corpus Research. Papers from the Twenty First International Conference on English Language Research on Computerized Corpora. Sydney, 2000*. Amsterdam: Rodopi, 25-40.

EDWARDS, M. and K. CSIZÉR (2004): "Developing pragmatic competence in the EFL classroom", *English Teaching Forum*, July, 16-21.

FARR, F. and M. MCCARTHY (2002): "Expressing hypothetical meaning in context: theory versus practice in spoken interaction". Paper delivered at *The Teaching and Language Corpora (TALC) Annual Conference*. Bertinoro, Italy, July 2002.

FARR, F. and A. O'KEEFFE (2002): "*Would* as a hedging device in an Irish context: an intra-varietal comparison of institutionalised spoken interaction". In REPPEN, R., S. FITZMAURICE and D. BIBER (eds.): *Using Corpora to Explore Linguistic Variation*. Amsterdam: John Benjamins, 25-48.

FRASER, B. (1996): "Pragmatic markers", *Pragmatics*, 6, 167-190.

FUNG, L. and R. CARTER (2007): "Discourse markers and spoken English: native and learner use in pedagogic settings", *Applied Linguistics*, 28, 410-439.

GILQUIN, G. and GRANGER, S. (2010): "How can data-driven learning be used in language teaching?". A. O'KEEFFE and M. J. MCCARTHY (eds) *The Routledge Handbook of Corpus Linguistics*. London: Routledge, 359-370.

GREAVES, C. and M. WARREN (2010): "What can a corpus tell us about multi-word units?". In O'KEEFFE, A. and M. MCCARTHY (eds.): *The Routledge Handbook of Corpus Linguistics*. London: Routledge, 212-226.

HARWOOD, N. (2005): "Nowhere has anyone attempted...In this article I aim to do just that: a corpus-based study of self-promotional I and we in academic writing across four disciplines", *Journal of Pragmatics*, 37, 1207-1231.

HOLMES, J. (1988): "Doubt and certainty in ESL textbooks", *Applied Linguistics*, 9, 21-44.

HOLMES, J. (1990): "Hedges and boosters in women's and men's speech", *Language and Communication*, 10, 185-205.

HOPPER, P. (1998): "Emergent grammar". In M. TOMASELLO (ed.): *The New Psychology of Language*. Hillsdale, NJ: Lawrence Erlbaum Associates, 155-175.

HYLAND, K. (1994): "Hedging in academic writing and EAP textbooks", *English for Specific Purposes*, 13, 239-256.

HYLAND, K. (1999): "Academic attribution: citation and the construction of disciplinary knowledge", *Applied Linguistics*, 20, 341-367.

HYLAND, K. (2002a): "Options of identity in academic writing", *ELT Journal*, 56, 351-358.

HYLAND, K. (2002b): "Authority and invisibility: authorial identity in academic writing", *Journal of Pragmatics*, 34, 1091-1112.

JOHNS, T. (1997): "Contexts: the background, development and trialling of a concordance-based CALL program". In WICHMANN, A., S. FLIGELSTONE, T. MCENERY and G. KNOWLES (eds.): *Teaching and Language Corpora*. London: Longman, 100-115.

KASPER, G. (1997): "Can pragmatic competence be taught?" (NetWork #6). URL: [www.nflrc.hawaii.edu/NetWorks/NW06/](http://www.nflrc.hawaii.edu/NetWorks/NW06/) [Retrieved 03 May 2011]. Honolulu: Second Language Teaching and Curriculum Centre, University of Hawai'i.

KASPER, G. and K. ROSE (2002a): *Pragmatic Development in a Second Language*. Oxford: Blackwell.

KASPER, G. and K. ROSE (2002b): “The role of instruction in learning second language pragmatics”, *Language Learning*, 52, 237-273.

LEECH, G. (1983): *Principles of Pragmatics*. London: Longman.

LEVINSON, S. (2004): “Deixis”. In HORN, L. and G. WARD (eds.): *The Handbook of Pragmatics*. Oxford: Blackwell, 97-121.

MCCARTHY, M. (2002): “Good listenership made plain: British and American non-minimal response tokens in everyday conversation”. In REPPEN, R., S. FITZMAURICE and D. BIBER (eds.): *Using Corpora to Explore Linguistic Variation*. Amsterdam: John Benjamins, 49-72.

MCCARTHY, M. and R. CARTER (1995): “Spoken grammar: What is it and how do we teach it?”, *ELT Journal*, 49, 207-218.

MCCARTHY, M. and M. HANDFORD (2004): “‘Invisible to us’: A preliminary corpus-based study of spoken business English”. In CONNOR, U. and T. UPTON (eds.): *Discourse in the Professions: Perspectives from Corpus Linguistics*. Amsterdam: John Benjamins, 167-201.

MCENERY, T. and A. WILSON (2001): *Corpus Linguistics: An Introduction*. Edinburgh: Edinburgh University Press.

MÖLLERING, M. and D. NUNAN (1995): “Pragmatics in interlanguage: German modal particles”, *Applied Language Learning*, 6, 41-64.

MUKHERJEE, J. (2009): “The grammar of conversation in advanced spoken learner English”. In K. AIJMER (ed.): *Corpora and Language Teaching*. Amsterdam: John Benjamins, 203-230.

MUKHERJEE, J. and J-M ROHRBACH (2006): "Rethinking applied corpus linguistics from a language-pedagogical perspective: new departures in learner corpus research". In KETTEMANN B. and G. MARKO (eds.): *Planning, Gluing and Painting Corpora: Inside the Applied Linguist's Workshop*. Frankfurt am Main: Peter Lang, 205-232.

MÜLLER, S. (2005): *Discourse Markers in Native and Non-native English Discourse*. Amsterdam: John Benjamins.

O'KEEFFE, A., B. CLANCY and S. ADOLPHS (2011): *Introducing Pragmatics in Use*. London: Routledge.

O'KEEFFE, A., M. MCCARTHY and R. CARTER (2007): *From Corpus to Classroom: Language Use and Language Teaching*. Cambridge: Cambridge University Press.

RÖMER, U. (2004): "A corpus-driven approach to modal auxiliaries and their didactics". In SINCLAIR, J. (ed.): *How to use Corpora in Language Teaching*. Amsterdam: John Benjamins, 185-199.

ROMERO TRILLO, J. (2002): "The pragmatic fossilisation of discourse markers in non-native speakers of English", *Journal of Pragmatics*, 28, 769-784.

ROSE, K. (2005): "On the effects on instruction in second language pragmatics", *System*, 33, 385-399.

SCHIFFRIN, D. (1987): *Discourse Markers*. Cambridge: Cambridge University Press.

SCHMIDT, R. (1993): "Consciousness, learning and interlanguage pragmatics". In G. KASPER and S. BLUM-KULKA (eds.): *Interlanguage Pragmatics*. Oxford: Oxford University Press, 21-42.

SCHWEINBERGER, M. (forthcoming): "Analysis of the sociolinguistic distribution of LIKE in Irish English". In A. KOLL-STOBBE and M. BIESWANGER (eds.): *Proceedings from the 10 Norddeutsches Linguistisches Kolloquium, March 2009*. Bern: Peter Lang.

SCOTT, M. (2009): *WordSmith Tool Version 5.0*. Liverpool: Lexical Analysis Software.

SCOTT, M. (2010): *WordSmith Tool Version 5.0: User's Manual*. Liverpool: Lexical Analysis Software.

SCOTT, M. and C. TRIBBLE (2006): *Textual Patterns: Key Words and Corpus Analysis in Teacher Education*. Amsterdam: John Benjamins.

SINCLAIR, J. (2003): *Reading Concordances*. Harlow: Pearson Longman.

SRIPICHARN, P. (2010). "How can we prepare learners for using language corpora?". In O'KEEFFE, A. and M. MCCARTHY (eds.): *The Routledge Handbook of Corpus Linguistics*. London: Routledge, 371-384.

TRIBBLE, C. (2010): "What are concordances and how are they used?". In O'KEEFFE, A. and M. MCCARTHY (eds.): *The Routledge Handbook of Corpus Linguistics*. London: Routledge, 167-183.

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

VAUGHAN, E. (2010): "How can teachers use a corpus for their own research?". In O'KEEFFE, A. and M. MCCARTHY (eds.): *The Routledge Handbook of Corpus Linguistics*. London: Routledge, 471-484.

VAUGHAN, E. and B. CLANCY (2011): "The pragmatics of Irish English", *English Today*, 27, 47-52.

YULE, G. (1996): *Pragmatics*. Oxford: Oxford University Press.