Corpus Linguistics, the World Wide Web, and English Language Teaching

Charles F. Meyer

University of Massachusetts Boston

meyer@cs.umb.edu

Abstract

This article explores how the Web can serve as a resource for teaching modality in English to prospective teachers of ESL, EFL and ESP. The first section demonstrates how texts taken from two different registers on the Web can be used to teach epistemic modality (e.g. how modal verbs such as *may* and *could* express notions of probability and certainty). The second section explores how the Michigan Corpus of Academic Spoken English (MICASE) can be searched to study deontic modality, specifically how the expressions *should* and *might want to* are used by teachers and advisors to give advice to students. The discussion demonstrates that teaching grammar is most effective when it is based on real data rather than <u>contrived</u> and decontextualized examples in grammar books.

Key words: corpus linguistics, the World Wide Web, data-driven learning, modality

Resumen

Lingüística del corpus, la Web y la enseñanza de la lengua inglesa

El presente artículo explora cómo la Web puede servir de recurso para el aprendizaje de la modalidad en inglés por futuros profesores de inglés para segunda lengua, inglés como lengua extranjera e inglés para fines específicos. La primera parte del artículo explica cómo textos procedentes de dos registros distintos y obtenidos de la Web pueden aplicarse a la enseñanza de la llamada modalidad epistémica (en concreto, los verbos modales como *may* y *could* para la expresión de probabilidad y certeza). La segunda parte del artículo se centra en el estudio de la modalidad deóntica utilizando el Michigan Corpus of Academic Spoken English (MICASE) y lo ilustra analizando las expresiones *should* y *might want to*, utilizadas éstas por profesores e instructores para dar consejos a los estudiantes. Finalmente, el apartado de discusión valora la mayor efectividad de la enseñanza de la gramática basando ésta en ejemplos de uso real del idioma, a diferencia de los ejemplos inventados o descontextualizados que ofrecen las gramáticas clásicas.

Palabras clave: lingüística de corpus, Internet, aprendizaje basado en corpus, modalidad

Until recently, corpus analyses had to be based on one or more of the many corpora available, such as the Brown Corpus (containing edited written American English) or

the British National Corpus (comprised of spoken and written British English). However, the emergence of the World Wide Web and the vast amount of texts available on it has prompted many linguists and language teachers to conduct their corpus analyses using data from the Web. As de Schryver (2002) observes, the Web can be used as or for a corpus. To exploit the Web as a corpus, one can use WebCorp, an online concordancing program that can be used to search for words and phrases on the Web (Morley, 2006). To exploit the Web for a corpus, one can use KWiKFinder, a program that allows for the collection of texts off the Web for inclusion in a custom-made corpus of the user's design (Fletcher, 2005).[1] The web has additionally proven useful as a gateway to existent corpora: web-based interfaces are now available for searching such corpora as the British National Corpus (BNC) or the Michigan Corpus of Academic Spoken English (MICASE).

One of the distinct advantages of all of these web-based technologies is that they enable quick and easy access to corpus data. In the past, one had to purchase a corpus and software tools before conducting any analyses. Now, it is possible to use, for instance, the freely available VIEW (Variation in English Words and Phrases), a textual analysis program that is available online,[2] and that can be used to conduct sophisticated searches of the British National Corpus. Although web-based technologies have many applications, I will focus in my discussion on how I have used these technologies to educate future teachers whose goals are to teach English for specific purposes or English as a foreign/second language.

Background

The approach I take to presenting grammar to my students is grounded heavily in the methodology of data-driven learning (DDL). DDL, as described by Tim Johns, is intended "to confront the learner as directly as possible with the data, and to make the learner a linguistic researcher […] [someone who is able] to recognize and draw conclusions from clues in the data […]" (Johns, 2002: 108). In other words, instead of having students read grammar books and then, perhaps, complete exercises based on made-up sentences, instructors present a point of grammar by having students study, say, conditional clauses in a corpus and then draw their own conclusions based on the examples they find. Strictly speaking, my students are teachers, not learners of English. But I find that many of them, particularly the Americans, have a poor understanding of English grammar. In the American educational system, grammar is often not taught in English classes for native speakers, or it is presented very traditionally and prescriptively, with discussions based exclusively on notional definitions of grammatical categories ("A

noun is a person, place, thing, or idea"). Many of the non-native speakers of English in my classes know grammar well, but need exposure to the use of grammatical constructions in real rather than contrived linguistic contexts. I have therefore found that DDL is a very useful methodology for teaching all aspects of grammar to my students. In fact, although the literature on DDL promotes this methodology for use with all levels of second language learners of English, it works best for advanced learners, who speak and write English adequately but are interested in exploring some of the more advanced points of English grammar. To teach grammar in this manner, I have devised a series of grammatical exercises that use the Web as a medium of instruction. In the sections below, I discuss how I use DDL to teach modality in English. Throughout the discussion I cite sources that I either refer to in class or have my students read prior to class.

Modality in English

Although theories of modality abound, most discussions recognize the difference between *deontic* (or *root*) modality and *epistemic* modality. While deontic modality "involve[s] some kind of intrinsic control over events," epistemic modality is more concerned with "human judgment of what is or is not likely to happen" (Quirk et al., 1985: 219). If I say to a student of mine "You should study hard for tomorrow's exam," I am using the modal verb *should* deontically: I am attempting to regulate my student's behavior by telling her what she needs to do. If, on the other hand, I say to the student "Study hard and you should do well on the exam," I am using *should* epistemically: I am making a judgment that her success on the exam is predicated upon her studying hard. If I had used *will* instead of *should*, my certainty of her success would have been even higher.

Students can typically grasp the general distinction between epistemic and deontic modality. However, when they have to apply the distinction to each of the individual modal verbs, modality can become a difficult notion to grasp, largely because there are subtle semantic and pragmatic differences between the choice of the various modal verbs. Moreover, grammatical treatments of modals assign a range of different meanings for the modals: Quirk et al. (1985: 221) posit seven meanings for the modals; Mindt (1995: 164) seventeen meanings. My goal in presenting modality is to not overwhelm my students with all of the potential meanings of modal verbs. Instead, I have developed two modules for investigating epistemic and deontic modality: one based on the choice of modals (and certain adverbs such as *perhaps*) to express varying degrees of probability in texts taken from two different registers, and a second based on the use of two directives-*you might want to* and *you should*-in two registers of MICASE. My goal in both modules is not only to demonstrate that the meanings of modals are

interrelated, but to show, as Biber et al. (1999: 489) demonstrate, that modal usage varies considerably by register: that meaning is very context dependent and closely associated with the goals of the speaker/writer in a particular communicative situation.

Epistemic Modality

My discussion of epistemic modality is based on an analysis of two texts taken off the Web from two different registers: scientific writing and press reportage. To represent scientific writing, I selected a semi-technical report on the topic of global warming ("Climate Change Impacts on the United States" [Cambanis, 2000]). To represent press reportage, I used a newspaper article describing Shi'ite resistance in Iraq (Cambanis, 2005). After downloading the texts from the Web, I used a shareware program, "Htmasc,"[3] that removes all html markup from a document, and produces a markup-free text file suitable for use in any text analysis program. Since the two texts were of different length, I excerpted approximately 1000 words from each so that valid statistical comparisons between them could be made. Each text was loaded individually into a freeware concordancing program called "Simple Concordance Program."[4] I use this concordancing program rather than others (such as *WordSmith Tools*) because it is suitable for basic concordancing tasks and freely available for students to download and use on their own. In class, I have students suggest modal verbs and adverbs expressing varying degrees of probability that we can search for in each of the texts. After retrieving examples, we construct tables and discuss concordance views.

Table 1 lists the linguistic items examined and their raw frequencies in each of the texts.

Marker	Climate Change	Shi'ite Resistance
Modal verb		
can(not)	5	1
Could	4	0
may	1	0
might	0	1
will	9	6*
would	1	2
Adverbial/modal combinations		
will probably	1	0
could possibly	2	0
will (very) likely	5	0
	Total: 28	Total: 10

*4 instances in quoted speech; 1 instance in reported speech

Table 1. Markers of Modality in Two Texts.

Even though the two texts are relatively short, Table 1 reveals some notable differences in the occurrence of markers of modality in each. The modal verbs in Table 1 express differing degrees of probability: *can, could, may,* and *might* mark relatively low degrees of probability, and *will* and *would* relatively high degrees. All of the adverbials are members

of a class of content disjuncts that, according to Quirk et al. (1985: 620), express "some degree of doubt": *probably, possibly,* and *(very) likely*. In general, as Biber et al. (1999: 489) report, modals of possibility (with the exception of *could*) are more common in academic writing than news; in contrast, modals of high probability (which they categorize as expressing "prediction") are more common in news.

That modals of low probability occurred more commonly in the Climate Change article is not surprising: as many of the examples in the concordance views in Figure 1 illustrate, what is known about the causes of climate change is still being determined.

```
may
    century. Such "surprises" may stem from unforeseen changes in
can
    long-lived infrastructure can be designed taking projected clim
    egy.//Adaptation measures can, in many cases, reduce the magnit
    nt. Adaptations, however, can involve trade-offs, and do involv
    uctivity elsewhere, which can shift international patterns of f
could
    ems due to climate change could possibly have negative conseque
    oduction in fertile areas could cause prices to fall, benefitin
    that some US food exports could increase, depending on impacts
will
    ore widespread ecosystems will also be vulnerable to climate ch
    st. Reduction in snowpack will very likely alter the timing and
    r example, sea-level rise will very likely cause further loss o
    recreation such as skiing will very likely be reduced, and air
    nd air conditioning usage will very likely increase.//Highly ma
    ting of sea ice in Alaska will further damage forests, building
    ry likely.//These changes will, at minimum, increase discomfort
    agriculture, many farmers will probably be able to alter croppi
    d our potential responses will likely depend in part on impacts
```

Figure 1. Concordance Views (Global Warming).

The modals in each of the examples below express varying degrees of certainty: "unforeseen changes" may be the result of "unanticipated" events; increased "crop production" might affect the cost of food; it is possible that "changes in agricultural productivity" will affect global distributions of food.

There are also very likely to be unanticipated impacts of climate change during the next century. Such "surprises" **may** stem from unforeseen changes in the physical climate system […]

It is also possible that a rise in crop production in fertile areas **could** cause prices to fall, benefiting consumers.

Effects of climate variability and change on US agriculture will depend critically on changes in agricultural productivity elsewhere, which **can** shift international patterns of food supply and demand.

Somewhat unexpected is the frequency of the modal *will*, which expresses a higher degree of certainty than *may* or *could*. But note that many of these instances are tempered by the use of adverbials expressing doubt:

> Reduction in snowpack **will very likely** alter the timing and amount of water […]

> Many farmers **will probably** be able to alter cropping and management […]

The infrequency of low probability markers of modality in the Shi'ite Resistance article is a consequence of the nature of reportage: reporters write about the facts of a particular news story. As illustrated in Figure 2, when a low probability marker is used, reporters may simply be interpreting the facts as they see them:

> Many clerics and politicians consider themselves Sadrists even though they **might** not take orders from Moqtada, a junior cleric with much less stature than his father, who was the top Shi'ite cleric in Iraq when he was killed in 1999.

Many of the instances of *will* also come not from the reportage itself but from quotes or reported speech:

> "We **will** be the judges in the parliament," Sheikh declared. "If we see that a decision has been influenced by the Americans, we will denounce it and boycott it."

> He [Jabouri] said he **will** work from outside the political system, coordinating closely with Sheikh, to block any constitution that's not sufficiently Islamic or gives too much autonomy to Kurds.

```
can
    s of our new constitution cannot ignore this book," Amri said
might
    Sadrists even though they might not take orders from Moqtada,
will
    s American stooges.//''We will be watching in the National As
    nced by the Americans, we will denounce it and boycott it."//
    American interests.//''We will be the judges in the parliamen
    d by Islam.//''Everything will be conducted according to the
    ed Iraqi Alliance leaders will have to answer clerics who acc
    ng," he said.//He said he will work from outside the politica
would
    , and Arab nationalism he would bring to the body.//Yesterday
    il of Shi'ite clerics who would review -- and have the author
```

Figure 2. Concordance Views (Shi'ite Resistance).

Deontic Modality

I use a similar strategy in presenting more pragmatically conditioned uses of modals, their so-called deontic meanings. To illustrate this use of modals, I have students examine the difference in usage of the constructions *you might want to* and *you should* using examples taken from MICASE (the Michigan Corpus of Academic Spoken English).[5] This is a very large corpus (currently 1,848,364 words in length) consisting of spoken American English occurring in various academic contexts. It has a search interface available for free online, and can easily be used in a technology enhanced class containing a computer with Internet access that is connected to an LCD projector.

Within Searle's (1976) taxonomy of speech acts, constructions containing *you might want to* and *you should* having deontic meanings are classified as 'directives': speech acts that attempt to get someone to do something. In the examples below, taken from advising sessions, both *you should* and *you might wanna* are directing individuals to do something: inquire about personal characteristics in the first example, and think about taking a particular course in the second example.

> i think **you should** think about, your strengths and interests (ADV700JU047)

> yeah. **you might wanna** take Polish or you might say to yourself well gee, i want, uh you know i want biology or, or biomedical engineering (ADV700JU023)

Although both of these examples are directives, they are different kinds of directives than those found in imperative sentences such as *Close the door*. Both *you should* and *you might want to* are not necessarily commanding someone to do something. Instead, they are offering advice. In addition, the force and directness of the two forms differ markedly. *You might want to* would be classified as a mitigated directive (or command). Within the context of an advising session, such a directive is pragmatically quite desirable. As Mitchell (2003: 143) notes,

> A speaker who backs up his directives with a considered judgment of what is in the performer's interest makes them more palatable to the hearer than directives based on the apparently arbitrary exercise of the speaker's authority or volition.

In other words, a student coming to an advising session does not expect (at least in the American context) to be told what to do. Instead, he/she wants to explore options with an advisor. Thus, a form such as *you might want to* serves this communicative goal quite well. However, even though this form is well suited to this communicative context, *You should*, the more direct form, occurs as well. To explain the usage of this form, it is necessary to investigate further variables: gender, the power relationships between speakers, and differences in the force of directives used in an advising session vs. office hours.

Research has shown that gender is a variable in the use of directives. For instance, West (1998) investigated the types of directives used by male and female doctors in conversations with their patients, a communicative context somewhat analogous to college professors issuing directives to students during office hours. West (1998: 341) found that mitigated directives, such as the example below, were much more commonly used by female doctors than male doctors:

> PHYSICIAN: We:ll, you could *try*: taking .h two: ev'ry four *hou::rs* if yuh needed to:…you could take that many an' *see:* an' that's- .h it's a *ver:y strong* medicine fer arthri [:tus.]

Overall, as Table 2 demonstrates, females used *you might want to* much more frequently than males.[6] Even though females predominated in both of the genres, it is nevertheless striking that only one male used this form.[7]

Form	Advising and Office Hours	
	Male	Female
You might want to/wanna	1	51
You should	15	59
Total	16	110

Table 2. Modal Breakdown by Gender.

Gender differences, however, only partially explain the usage of the two forms. Table 2 reveals that females do not necessarily use *You might want to* instead of *You should*. In fact, in most instances (80%), they use *You should*. This usage requires the examination of additional variables: the power relationships between speakers and considerations of politeness and tact.

Table 3 lists the usage of the two forms by academic role.

Role	*You might want to/wanna*	*You should*	Total
Graduate Student	31 (54%)	22 (46%)	53
Staff	19 (37%)	31 (63%)	50
Faculty	2 (5%)	12 (95%)	14

Table 3. Modal Breakdown by Academic Role.[8]

There are two ways of looking at power relationships in the table. First of all, in all contexts, students will be lower on the power hierarchy in conversations they have with graduate students, staff, and faculty. However, there is an additional power hierarchy holding between the three roles listed in Table 3, with graduate students lowest on the hierarchy and faculty highest. Thus, in a conversation with a student, a graduate student might feel less of a superordinate than a faculty member would and thus, as the figures suggest in Table 3, use *You might want to* more frequently than *You should*. However, the figures for graduate students in Table 3 need to be interpreted somewhat cautiously, since 26 of the 31 instances were uttered by a single speaker. For faculty members, however, their status on the power hierarchy neutralizes any gender differences in the use of the two forms: while two instances of *You might want to* can be attributed to female faculty members, the 12 instances of *You should* were split among five male and seven female faculty members.

Related to academic role is the register in which the two forms were used. All instances of the two forms attributed to staff occurred only in the register of advising. For graduate students and faculty, the forms occurred only in the register of office hours.

While directives are given in both of these registers, the type of advice given in a directive varies considerably in the two genres, and is subject to one maxim of Leech's (1983: 132) politeness principle: the tact maxim, which states that one should "minimize cost to *other*, maximize benefit to *other*." And different kinds of directives, Leech (1983: 107) notes, will violate this maxim in varying degrees. Because *Peel these potatoes* does not maximize benefit to other, it is relatively low on the scale of politeness. In contrast, *Have another sandwich* does maximize benefit and is thus high on the scale of politeness. An examination of the various uses of *You might want to* and *You should* reveals that the two forms interact with the tact maxim in very different ways.

In an advising session, students meet with staff members—university employees who do not teach—to discuss their academic schedules. On one level, staff members are providing a service to students and thus are engaged in a type of customer/client discourse. On another level, however, advisors have knowledge that is vital to a student's academic career—knowledge of degree requirements, course prerequisites, and so forth. Therefore, this knowledge places the advisor in a position of power over the student. Given this kind of relationship between student and advisor, it is not surprising to see a mixture of the two forms in this register: the more direct form *You should* as well as the mitigated *You might want to*. Although *You should* is relatively low on the politeness hierarchy of Leech's Tact Maxim, the examples below would hardly be characterized as impolite:

> so maybe **you should** take the one-eighty-five one-eighty-six you realize that you're going to lose the A-P credit, if you take one-eighty-five and one-eighty-six. **[S3: mhm]** but, four credits, aren't going to make you or break you. (ADV700JU023)

> **you should** do your senior audit next fall (ADV700JU047)

> **you should** take Intro Comp next semester. (ADV700JU047)

The main difference between these examples and those with *You might want to* is that *You might want to* presents an option as one choice among many. In the first two examples below, the choices are implicit; in the third example, they are explicitly given:

> well **you might wanna** major in English (ADV700JU047)

> okay so that sounds like **you might want to** take a mathematics class next semester, do you remember what math you placed into? (ADV700JU047)

> but **you might wanna** sign up for Astronomy one-eleven or you know Intro to Astronomy. (ADV700JU047)

In meeting with an instructor during office hours, in contrast, students are more clearly in a subordinate position relative to their instructors. For this reason, if the 26 examples of *You might want to* attributed to one graduate student are removed, *You should* predominated in this register. In addition, the role of directives is different in this register. As the examples below illustrate, instructors are not presenting options to their students; they're directly telling them what they need to do to write better papers or have a clearer understanding of a concept:

> in general **you should** do this throughout the paper too. you need to go through and ask yourself what the point of each paragraph is, right? um and make sure it has a point, and make sure it says what its point is, okay? (OFC115SU060)

> so probably actually what **you should** do is go back and actually just, um, read this one more time, go back and read the Crawford one more time, and see if there're any sort of other sort of arguments that help you out. (OFC115SU060)

> but if you're thinking about it at all **you should** think of it as a system, of of signs and pointers to your reader, that that allows that person to see your logic and to anticipate your logic. (OFC300JU149)

> don't interpret your confidence interval level, with just one interval. **you should** interpret it as being looking at many intervals. (OFC575MU046)

> **you should** know the poem well enough so that you can make a point about it and make an argument about it. (OFC300JU149)

Mitigated forms do occur, but they can raise unintended consequences, such as a student misinterpreting the directive as presenting options when in fact fairly explicit advice is being given. For instance, in the examples below, are instructors merely suggesting possible changes in student papers, or insisting that they be made?

> it doesn't change your argument necessarily, but **you might wanna** qualify it in that kinda way (OFC115SU060)

> o you **you might wanna** say that, in order to understand um, the, programs and sort of missions of, these two organizations, **[S5:** mhm **]** we have to understand them within the context of, the, you know political and economic situations in these two cities. right? (OFC115SU060)

> that's one of the factors that **you might want to** consider for noisy-or, and, another one is the fact that you don't know tha- that these four causes are not complete, (OFC270MG048)

> **you might wanna** focus on those. and then just think about the coloration variation and the variation of weaves, and see where that leads you. (OFC320SU153)

Many non-native speakers of English have told me that they find mitigated directives such as these quite difficult to interpret, especially if they are used by a teacher—someone from whom they expect explicit and direct advice.

Conclusions

With the increased emphasis in language instruction on the use of authentic material, the Web has become an increasingly important resource for teachers to draw upon. To facilitate data-driven learning, the Web provides instructors with a wide variety of texts that they can use to illustrate points of grammar, software (such as concordancing programs) that can display linguistic information from the texts being analyzed, and web-based interfaces to corpora such as MICASE or the British National Corpus.

While I have described in detail how I have used the Web to discuss modality in English, I have used data taken from the Web for other purposes, such as studying collocations and usage trends. For instance, shortly following the first Bush/Kerry debate in the autumn of 2004, I downloaded a transcript of the debate, and created two files: one containing Bush's responses to questions and the other Kerry's responses. Using kfNgram (a program for studying collocations),[9] I had my class analyze "lexical bundles" in the debate: "multi-word sequences," as Biber et al. (2004) define them, that occur repeatedly in texts. Although Biber et al. (2004) restricted their discussion to four-word lexical bundles, such as *what does that mean*, because of the relatively short length of the debate responses, we investigated lexical bundles of three-words or more that occurred more than three times. We discovered, for instance, that Kerry used numerous "stance bundles," constructions that "express attitudes or assessments of certainty that frame some other proposition" (Biber et al., 2004: 384): *I'm going to* (11 occurrences), *we need to* (9), *We have to* (7), *I believe that* (6), *I have a* (6), and *you have to* (6). As a challenger for president, we noted, Kerry used these bundles to discuss precisely what he would do if elected president.

The Web is also a good resource for tracking usage trends. During a discussion of the movement in English towards gender-neutral language, I asked one of my classes to supply examples of words in English ending in *-man* (e.g. *chairman*) that have equivalent forms ending in *-person* (e.g. *chairperson*). Someone wondered whether the word *fisherman* had undergone a similar shift and become *fisherperson*. I commented that for me this was an awkward expression. But a quick check with Google revealed numerous instances of *fisherperson*, such as the example below:

> A fishing rod supporting device allows a fisherperson to wear it at waist-level for hands-free support of a fishing rod. ("Portable Handsfree Fishing Rod Supporting Device,")

As this and other examples I have discussed in this article illustrate, the Web provides data for analysis that are not readily available from either textbooks or the repository of knowledge that instructors bring with them to class; it is an interactive medium allowing for engagement with the material being presented; discussions can be tailored to the specific point of grammar being covered; and because the Web contains current and often interesting material, it can overcome the sometimes monotonous and decontextualized discussions of grammar so prevalent in textbooks. For these reasons, all language teachers should consider the Web a resource that is as indispensable as a textbook.

(Paper received April 2006)

REFERENCES

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

Biber, D., S. Conrad & V. Cortes (2004). "*If you look at*….Lexical Bundles in University Teaching and Texbooks." *Applied Linguistics* 25: 371-405.

Cambanis, T. (2000). "Climate Change Impacts on the United States: The Potential Consequences of Climate Variability and Change." The National Assessment Synthesis Team, US Global Change Research Program. [accessed on March 16, 2006].

Cambanis, T. (2005). "Iraq's hard-line Shi'ites vow to resist US agenda." *Boston Globe* (Feb. 12).

De Schryver, G.-M. (2002). "Web for/as corpus: a perspective for the African languages." *Nordic Journal of African Studies* 11,2: 266-282 (also available at: http://tshwanedje.com/publications/webtocorpus.pdf)

Fletcher, W. H. (2005). "Concordancing the Web: Promise and Problems, Tools and Techniques" [accessed on April 5, 2006].

Johns, T. (2002). "Data-Driven Learning: The Perceptual Challenge" in B. Kettemann & G. Marko (eds.), *Teaching and Learning by Doing Corpus Analysis*, 107-117. Amsterdam: Rodopi.

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

Mindt, D. (1995). *An Empirical Grammar of the English Verb*. Berlin: Cornelsen.

Mitchell, K. (2003). "*Had better* and *might as well*: On the margins of modality?" in R. Facchinetti, M. Krug & F. Palmer (eds.), *Modality in Contemporary English*, 129-149. Berlin: Mouton de Gruyter.

Morley, B. (2006). "WebCorp: A tool for online linguistic information retrieval and analysis" in A. Renouf & A. Kehoe (eds.), *The Changing Face of Corpus Linguistics*, 283-296. Amsterdam: Rodopi.

Quirk, R., S. Greenbaum, G. Leech & J. Svartvik (1985). *A Grammar of Contemporary English*. London: Longman.

Searle, J. (1976). "A Classification of Illocutionary Acts." *Language in Society* 5: 1-23.

West, C. (1998). "Not Just "Doctor's Orders": Directive-Response Sequences in Patients Visits to Women and Men Physicians" in J. Coates (ed.), *Language and Gender: A Reader*. Oxford and Malden, MA: Blackwell [originally published in 1990 in *Discourse and Society* 1: 85-112].

Charles F. Meyer is Professor of Applied Linguistics at the University of Massachusetts Boston, where he teaches students studying to be teachers of ESL and EFL. He is very interested in how linguistic corpora can be used to gain insights into the English language and is author of *English Corpus Linguistics: An Introduction* (Cambridge University Press, 2002).

NOTES

[1] Fletcher (2005) also provides an overview of the many ways that linguists are using the Web to conduct corpus-based research.

[2] Developed by Mark Davies and available at: http://view.byu.edu/.

[3] Available from Bite'n'Byte http://www.bitenbyte.com/.

[4] Developed by Alan Reed and available from Textworld.com http://www.textworld.com/.

[5] http://micase.umdl.umich.edu/m/micase/.

[6] For ease of reference, I will refer to this form as *you might want to*, even though the MICASE transcription system distinguishes between both *want to* and *wanna*.

[7] According to demographic information taken from the MICASE website (http://www.lsa.umich.edu/eli/micase/MICASEStats.htm), in the register of advising, 56.5% of speakers were females and 14.2% males; in the register of office hours, the breakdown was 67.9% females and 26.9% males.

[8] Excluded from this table were a small number of instances of *You should* used by undergraduate students (5) and Researchers (4). Although undergraduates figure prominently in MICASE, in the registers of advising sessions and office hours, they are not in a position to issue directives to staff or faculty.

[9] Developed by William Fletcher and available for download at: http://www.kwicfinder.com/kfNgram/kfNgramHelp.html.