

The integration of corpus-based data into grammar instruction: Using *advise, recommend, and suggest* as an example

Wen-Shuenn Wu

Instructor, Department of Foreign Languages and Literature
Chung Hua University

Abstract

It seems that explicit focus on form in the classroom still arouses considerable interest among EFL teachers in many countries. At the same time, the rapid development of computer technology has made it possible for language learners to access large and principled databases of naturally occurring language. This paper summarizes the integration of corpus-based data and lexico-grammar instruction. The paper first gives an overview of corpora in terms of accessibility. Next, the relations between corpus-based linguistics and lexico-grammar instruction are introduced. The applications of corpus-based findings on language learning and teaching are also summarized. Then, pedagogical issues of the use of corpora in a classroom setting are fully discussed. Finally, by using *advise, recommend, and suggest* as sample lesson material, the linguistic and pedagogical challenges that English practitioners face are brought up. These challenges include: (1) technical and statistical challenges (2) theoretical foundation (3) intuition vs. competence (4) absence or rare instances of target examples (5) real language.

Keywords: corpus, grammar instruction, DDL, inductive approach

語料庫資料和文法教學的整合：以 **advise, recommend,** 和 **suggest** 為例

吳文舜

中華大學 外國語文學系 講師

摘要

許多國家的英文為外國語言(EFL)的教師似乎對於在教室裡以「著重形式」(focus on form)為主的教學法仍然有很大的興趣。在此同時，電腦科技的快速發展使得語言學習者可以讀取大量且有條理的自然發生的語言資料。此篇論文在探討語料庫資料和文法教學的整合。這篇論文首先就語料庫的可得性(accessibility)做一個摘要性的介紹，其次，介紹語料庫資料和文法教學的關係。接著，分析語料庫的發現，可以應用在語言學習和教學的部分也會稍作說明。最後，用 **advise**、**recommend**、和 **suggest** 三個意思類似的英文單字為範本教材，來說明英文教師將語料庫應用在教學上所可能遇到的語言和教學的挑戰為何？這些挑戰包括：(1) 技術和統計上的挑戰；(2) 理論基礎的挑戰；(3) 語言直覺對應語言能力的挑戰；(4) 語料庫例句稀少或付之闕如的挑戰；(5) 語料庫的資料是否可視為實際真實語言的挑戰。

關鍵詞：語料庫、文法教學、資料導向學習、誘導式教學法

Introduction

Corpora, defined as a collection of naturally occurring examples of written texts or spoken data for linguistic study, have been mainly used in lexicography, linguistic research (e.g., lexical cohesion, word frequency, language variation and genre study, etc.), material design, and language pedagogy. Corpora first captured the attention of most English language teachers in 1987 with the publication of *Collins COBUILD English Language Dictionary*, the first corpus-based learner's dictionary (Gabrielatos, 2005). Later, *Longman Dictionary of Contemporary English*, *Cambridge International Dictionary of English* and *Macmillan English Dictionary* were also published with the help of Longman Corpus Network, Cambridge International Corpus, and World English Corpus respectively. At the same time, much corpus-related research on different linguistic and pedagogical fields has been conducted in the past 10 years, such as collocation (Claveau & L'Homme, 2006; Deignan, 1999), word frequency (Biber & Reppen, 2002; Coniam, 1997), genre study (Flowerdew, 2005; Lee, 2001), pedagogical application (Conrad, 2000; Frankenberg-Garcia, 2005; Leech, 1997; Tribble, 1997), grammar instruction (Hunston, & Francis, 1998) and teacher education (Renouf, 1997), L2 writing feedback (Gaskell & Cobb, 2004) to name just a few. Recently, the use of corpus has also been applied to EFL teaching materials. For instance, *Touchstone*, a series of EFL textbook compiled by Michael McCarthy, Jeanne McCarten, and Helen Sandiford, draws on the *Cambridge International Corpus* to build a syllabus based on how people actually use English.

With the rapid development of information technology and World Wide Web, many researchers advocated that even with moderate computer literacy and minimal computer resources, language teachers could and in fact should integrate corpora into their teaching (e.g., Gabrielatos, 2005; Stevens, 1995; Tribble, 1997). However, are there a lot of large and principled web-based corpora which are readily accessible to teachers and students? Are they user-friendly and properly compiled? Are the structural markups or annotations transparent to the ordinary users? A general overview of corpora will help us decide how we should integrate corpora into our teaching.

A General Overview of Corpora

A corpus is always designed for a particular purpose, and the type of corpus will depend on its purpose. According to the functions and purposes given by Hunston (2002), commonly used corpus types include: specialized corpus, general corpus, comparable corpora, parallel corpora, learner corpus, pedagogic corpus, historical or diachronic corpus, and monitor corpus (p. 14-17). However, in terms of corpus accessibility, three types of web-based corpus can be categorized: *closed corpora*, *paid corpora*, and *free online*

corpora.

1. Closed corpora

Longman Corpus Network, *Cambridge International Corpus* (CIC), and *World English Corpus* are three typical closed corpora. Only a few researchers, editors, lexicographers and textbook writers affiliated with *Longman*, *Cambridge*, and *Macmillan* publishers have access to these corpora. Rest of us can only use corpus-based printed products of these publishers.

2. Paid corpora

Many well-known corpora can be bought in a CD-ROM format; price varies from one corpus to another. Most creators of corpus also offer special discount to non-commercial users or educational institutes. Table 1 shows the price list of some major corpora.

Table 1. Price List of Major Corpora

Corpus abbr.	Corpus full name	Personal license	Institutional license
BNC (100 million words)	British National Corpus	BNX XML: GBP£ 75 (single) GBP£ 500 (10 packs)	BNC XML: GBP£ 500
ANC	American National Corpus		US\$75
LDC	Linguistic Data Consortium		Standard: US\$2,400
ICAME*	International Computer Archive of Modern and Medieval English	NOK\$3,500 (single user)	NOK\$8,000 (10 users)
COLT (500,000 words)	Bergen Corpus of London Teenage English		NOK\$1,250
CSPAЕ (about 2 million words)	Corpus of Spoken Professional American-English	US\$49	US\$179
ICE (A variety of million-word corpora)	International Corpus of English	GBP£ 346.96**	single copy: GBP£ 413.04** multiple copy: GBP£ 652.17**

*ICAME corpus collection includes about 20 different corpora such as Brown, LOB, Helsinki, etc.

**The price of 'The British Component of the International Corpus of English' bought outside EU.

3. Free online corpora

With the growing popularity and outstanding performance of wired computers, corpus linguists have created more and more online corpora which are open to the public. Some free corpora are listed below:

1. BNC Sampler (<http://www.natcorp.ox.ac.uk/>)
2. BNCWeb (CQP-Edition)
(<http://bncweb.lancs.ac.uk/bncwebSignup/user/login.php>)
3. Cobuild Concordance and Collocations Sampler
(<http://www.collins.co.uk/Corpus/CorpusSearch.aspx>)
4. JustTheWord (<http://193.133.140.102/JustTheWord/>)
5. Corpus of Contemporary American English (<http://www.americancorpus.org/>)
6. MICASE (Michigan Corpus of Academic Spoken English)
(<http://quod.lib.umich.edu/m/micase/>)
7. Virtual Language Centre (<http://www.edict.com.hk/default.htm>)
8. Compleat Lexical Tutor (<http://www.lextutor.ca/>)
9. WebCorp (<http://www.webcorp.org.uk/>)
10. *IWiLL* collocation explorer (<http://www.iwillnow.org/iwill/default.aspx>)

If language teachers would like to integrate corpora into their English teaching but won't use free online corpora, they can only buy corpus data (usually in CD-ROM format) and sometimes also concordancing software. However, to English teachers who are not computer literate, to design and maintain a fast, reliable, large, powerful and quality corpus without the support of a team of engineers and programmers is a formidable, if not impossible, task. Therefore, it is more practical to use free online corpora that are readily available to teachers as well as students.

Corpus-based Linguistics and Lexico-grammar Instruction

Tim Johns, who developed data-driven learning (DDL) and exploited the computer corpus for the teaching of international students at the University of Birmingham, is a pioneering university teacher who combined corpus with language teaching (Hunston, 2002; Leech, 1997). The essence of DDL is the student's inductive acquisition of grammar

rules and regular patterns through exploring corpus materials. DDL, based on the theory that students can act as ‘language detectives’, in fact does not ‘teach’ language features, but presents learners with authentic examples and asks them to interpret clues and draw conclusions from context (Huston, 2002, p.184).

In fact, language learners do not browse through sentences from a corpus. Instead, they use a concordancing program (which is also called ‘concordancer’) to search a corpus for a selected word or phrase. Depending on the corpus one uses, the node word that one searches will be usually highlighted and displayed in a whole sentence or a truncated sentence. Sometimes, a concordancer may also offer two different kinds of view: the traditional sentence view and KWIC (key word in context) view in which the target words appear in the center of the screen. A corpus KWIC display of *interested* from BNCweb is shown in Figure 1.

Figure 1
A Corpus KWIC Display of *Interested* from BNCweb

You are the actor on show at your audition — no-one is	<u>interested</u>	in auditioning your coach. Valuable coaching ca
newspapers, and do n't assume that the whole world is as	<u>interested</u>	in acting as you are. 5 Do n't neglect your healt
VQ the people who sit in the audience are usually either professionally	<u>interested</u>	, or are fellow students and friends. Any progra
financially justified. Agents do n't want deadwood and if they are	<u>interested</u>	in you it is because they feel they can sell you ir
, casting director, producer or director. If an agent is	<u>interested</u>	in you he may well have useful comments on th
. Remember this will be a professional opinion from someone who is	<u>interested</u>	in the qualities you now have to offer, so bear s
P.R. I'm sure it did. The NT is obviously	<u>interested</u>	in new talent but also in staying power — track
the peak. I think everyone needs the director to be vitally	<u>interested</u>	in the work well beyond the first week of actua
partially a façade, but one which it was felt by both	<u>interested</u>	parties had to be maintained, probably so as ne
Still uncertain, he wrote. Have been growing less and less	<u>interested</u>	in titles that are other than purely descriptive. T
to give up. Rather, a wonder that I was ever	<u>interested</u>	in starting. Card from Goldberg to say overwh
need for a fish restaurant in the surrounding area. However,	<u>interested</u>	parties have to fulfil the following prerequisites
etween. The report clearly stresses that the catering industry is more	<u>interested</u>	in quality than price. The number of establishm
nly catering industry representative on this working party, I would be	<u>interested</u>	to hear from any companies which have any ex
(Caterer, 22-28 August), I thought you might be	<u>interested</u>	to hear the other side of the coin. As an individ
staff member who regularly uses the computer equipment and sufficiently	<u>interested</u>	to study the manuals, that person will, given en

A concordance is not a list of dictionary entry and can't replace the function of ESL learner's dictionary. Nevertheless, if an English learner with intermediate English proficiency is not sure what he/she has to use after ‘*be interested*’ and is presented with data in Figure 1. He/she is expected to interpret the data and make a conclusion like this:

‘*Be interested*’ is commonly followed by a preposition *in* and noun/pronoun or gerund (V-ing) such as ‘is interested in you’, ‘is interested in new talent’, ‘be interested in the work’, or ‘is interested in auditioning your coach’, ‘was interested in starting’. However, it can also be followed by infinitive (to+V) such as ‘would be interested to hear’ ‘might be interested to hear’.

Using corpora in a language class is not beneficial to students only; it also benefits language instructors. English teachers of non-native speaker now can rely not on intuition of native speakers but on principled corpora to solve some grammatical and usage problems. In addition, because a particular grammatical feature may occur in a textbook only once or twice, additional corpus material may be useful to expose the learner to a recurrent pattern rather than a single occurrence.

Viewing the incorporation of corpus research into language teaching from a register-specific and lexico-grammatical perspective, Conrad (2000) further argued that three major influences of corpus-based linguistics on the teaching of grammar can be seen in the 21st century. These three significant impacts she mentioned are (p. 549):

1. Detailed descriptions of grammar rules will be replaced by register-specific descriptions.
2. The grammar instruction will be closely integrated with the teaching of vocabulary. Lexico-grammar patterns will be central to language description and language learning.
3. Emphasis on structural accuracy will be shifted to the appropriate use of alternative grammatical constructions.

Finally, in an article responding to Widdowson's question about corpus analysis, Stubbs (2001) uses syntagmatic and paradigmatic relations to describe how a concordance can effectively help language learners access a large amount of corpus data and hopefully can unearth regular patterns of collocational co-occurrence:

In an individual text, neither repeated syntagmatic relations, nor any paradigmatic relations at all, are observable. However, a concordance makes it possible to observe repeated events: it makes visible, at the same time, what frequently co-occurs syntagmatically, and how much constraint there is on the paradigmatic choices (p.152).

Pedagogical Issues of the Use of Corpora in a Classroom Setting

One practical issue about using corpora in the classroom is whether teachers should use 'raw', unedited corpus search or specially-compiled, teacher-edited presentation of corpus data (Hunston, 2002, chap. 7). The former one is feasible if the school can provide teachers with enough computer facilities so that students and the teacher in a class have their own computer to look at the corpus together. The advantage of this kind of study is to maximum student motivation: the student has a question which needs to be dealt with (e.g., to complete a piece of written work) and is therefore highly motivated to discover the answer from the corpus data consulted. Of course, the disadvantage is that teachers have

little control over what happens. For example, if the corpus is consulted and no answer is apparent to students or teacher, or some unacceptable answers are found, the teacher has to depend on his/her language intuition (for native speakers) or linguistic competence (for non-native speakers) to solve the problem. Take the last sentence of KWIC display of ‘*interested*’ from Figure 1 for example. Once you click the key word ‘*interested*’, the target sentence is shown in the following paragraph:

Realistically, most staff never will want to plough through the manuals that come with most software. If the independent business has one staff member who regularly uses the computer equipment and sufficiently **interested** to study the manuals, that person will, given encouragement, build up expertise which can usefully be passed on.

It seems that this target sentence, which misses a *BE* verb, is not universally acceptable, compared to the other two target sentences in Figure 1:

As the only catering industry representative on this working party, I would be **interested** to hear from any companies which have any experience, or knowledge of any examples, which make them feel that this is an area that needs to be examined quickly and subjected to public debate.

and

FURTHER to your article TECs slammed over funding red tape (Caterer, 22 - 28 August), I thought you might be **interested** to hear the other side of the coin.

If a student brings up this kind of unexpected grammar issue, the teacher must be confident and competent enough to give an answer in a professional manner. This instance also shows that not all the concordance-derived data are grammatically correct. On the contrary, the latter pedagogy – specially-compiled, teacher-edited presentation of corpus data – is realistic when a school is not equipped with enough computer facilities for the students. Materials then can be printed on to paper to be used with a whole class. At the same time, teacher has more control over the prepared data. Nonetheless, the disadvantage is that, as the teacher selects the topic for study, the students will probably be less motivated to look for or remember the target information.

The way that a learner achieves the understanding of a grammatical rule through studying examples is called inductive (rule-discovery) approach. Deductive (rule-driven) approach, as opposed to inductive one, starts with the presentation of a rule and is followed by examples in which the rule is applied. Research findings about the benefits of deductive and inductive learning have been inconclusive (Thornbury, 1999). Corpus-based,

pattern-recognition, self-directed learning is an inductive approach, which has many advantages. However, applying the corpus-based inductive approach to language pedagogy, we should pay more attention to its potential disadvantages so that they can be minimized by all means. There are at least four disadvantages of inductive approach:

1. It is time-consuming for both students and teachers if a teacher would like to check the corpus before she/he asks students to do so. The time taken to work out a rule may be optimally used to integrate the rule directly into some productive activities.
2. Students may reach a wrong conclusion about some grammatical features, or their interpretation of these rules is either too broad or too narrow.
3. Some students just do not like this kind of learning style and some kinds of language items are better 'given' than 'discovered'. Personal learning preferences will definitely influence one's learning results.
4. Using corpus-based data search in a classroom setting, a teacher has to make a couple of digressions from the main topics of a lesson, which definitely will interrupt the flow of the lecture or discussion. Whether this kind of digression will distract students' attention or has negative impact on learning results should be further investigated.

Sample Lesson: Use *advise*, *recommend*, and *suggest* as an Example

Biber, Conrad, and Reppen (1998) pointed that one of the major functions of corpora is to investigate the lexico-grammatical associations and pattern differences between nearly synonymous words (p. 84). A sample lesson of using corpora was designed to see whether students can discern patterns and regularities in naturally occurring input.¹

Step 1: Lesson instruction

A handout describing the detail of the lesson was given.

We usually look up a word in a dictionary to find out its definition and usage to help us improve our grammar and probably also writing. However, the definition, example sentences, and usage notes of a dictionary is sometimes not enough to solve our English learning problems. Take *advise*, *suggest*, and *recommend* for example. The *Longman Dictionary of Contemporary English* and *Longman Language Activator* define *advise*, *suggest*, and *recommend* as follows:

advise: to tell someone what you think they should do, especially when you have more knowledge or experience than they have

suggest: to tell someone your ideas about what they should do, where they should go etc

recommend: to advise someone to do something, especially when you have special knowledge of a particular subject or situation

Of course, as an ESL learner's dictionary, *Longman Dictionary of Contemporary English* also gives users examples, sentence patterns, and usage notes. However, it is still not enough to solve our writing problems when we would like to use these three words. For example, the following three sentences are correct:

1. I **suggested** a new procedure to the committee.
2. A broker will **advise** you how to invest your money.
3. I would always **recommend** buying a good quality bicycle rather than a cheap one.

How about the following sentences? Are they grammatically acceptable sentences?

1. Her uncle **suggested** her to get a job in a bank.
2. He **advised** to leave as quickly as possible.
3. Can you **suggest** something for us to do this weekend?
4. If you go to Paris, we **recommend** you visiting Louvre Museum.
5. They **advised** him that the tour would proceed.

Step 2: Exercise

An inductive pattern-discovering exercise was given. Before that a couple of online dictionaries and corpus websites (see the free online corpora section of this article) had been introduced.

Instruction: Use the online dictionary and corpus-based websites that I recommended in the introduction to find out whether the following sentence patterns are grammatically acceptable to these three words (*advise*, *suggest*, and *recommend*).

Example:

V + to N + that-clause:

advise: No such usage

suggest: I **suggest** to Miss Johnson **that** she sit down on the chair and wait.

recommend: No such usage

1. **V + N + to-V or V + N + not to-V**
2. **V + wh-clause/phrase**
3. **V + that-clause**
4. **V + -ing**
5. **V + for/to someone + to-V**
6. What is the commonest modal (e.g. would, should, might, can, could, etc.) that is used in the that-clause of the third sentence pattern (i.e., V +that-clause)? Which adverbs are usually used with these three words?

modal:

adverbs:

Challenges of the Integration of Corpora and Grammar Instruction in Language Pedagogy

Because both the speed of computer and the size of hard disk have been greatly improved, the use of corpus-based, inductive approach has been much easier. Not only can we access more free, large, and representative corpora, but also the pattern-seeking can be carried out more reliably and much faster. However, it seems that corpora are widely acknowledged to be a valuable resource in language description, but there is a lack of consensus about the use of corpus-based material in language classrooms (Hunston, 2002). Since the use of corpus-based, inductive, DDL approach requires highly motivated and autonomous learners with the ability to analyze and interpret linguistic data, this discovery learning is more suitable for advanced students. Even if a teacher has advanced and active students, she/he still has to give careful consideration to the following issues.

1. Technical and Statistical Challenges

Applying corpora to the language instruction, a teacher has to be familiar with some basic terms of corpus linguistics such as *node*, *token*, *lemma*, *tag*, *parse*, *annotate*, and *n-gram*, to name just a few. If a teacher would like to give a more detailed explanation to students about the frequency and distribution of corpus data, she/he had better also know something about mutual information (MI) score, *t*-score, and *z*-score, which is pretty difficult to most English teachers.

2. Theoretical Foundation

Whether grammar should be taught implicitly or explicitly is still debatable (Ellis, 2006). Similarly, whether inductive approach is more effective than deductive one or the other way around is not conclusive. A string of words and a series of examples from corpora may show language learners some frequently occurring lexical and syntactic phenomena, but corpus linguistics needs more solid theoretical foundation to support its pedagogical effectiveness.

3. Intuition vs. Competence

To native speaker English teachers, consulting corpus data may help them confirm their language intuition and explanation of grammatical features can be based on statistical evidence rather than just anecdotal evidence. However, it can also happen that native speakers find collocations or grammatical features contrary to their intuition. A famous example is from Owen (1996). An invented figure Ah Peng from Owen's article in 1996 wrote the following sentence:

Many more experiment studies *require to be done* before we can say that...

This sentence sounded unacceptable to Owen's native-speaker intuition. Nevertheless, Owen also noticed that COBUILD corpus had several examples of *require* followed by a passive to-infinitive clause. For instance (p. 222-223):

...decided that a large number of laws would *required to be passed* by a two-thirds majority

Yes, your cordon pears do *require to be pruned* in summer

Owen gave these examples to imply that reference to a corpus did not necessarily solve the discrepancy between descriptive grammar and prescriptive grammar; intuitive prescription was still essential to language teaching and learning (p.224). Hunston (2002), after checking COBUILD corpus thoroughly, argued that Owen's question about the usage of *require* can be resolved if the corpus data are examined through phraseological rather than grammatical perspective. She concluded:

Although *REQUIRE to be* is found in the Bank of English, and fairly frequently, the past participle that follows is usually that of a verb with a specific meaning, not a general verb such as *do*. There are plenty of examples of the type *These roses require to be pruned each spring* but very few indeed of *require to be done* (only 3 out of 302). Thus Owen's intuitions are backed up by evidence of the corpus, but on phraseological rather than grammatical grounds (p.21-22).

However, this kind of subtle differences sometimes cannot be easily spotted by native speaker linguists, not to mention a non-native speaker who does not have English intuition. To non-native speaker English teachers, we can only rely on our linguistic competence to decide the acceptability of language input and output. Hence, the recurrent target examples can bolster our linguistic competence. However, what really challenges non-native speaker English teachers is the absence or rare instances of the target examples.

4. Absence or Rare Instances of Target Examples

The data obtained through concordancing is closely connected to the corpus examined. The absence of target examples or specific collocations in a certain corpus does not mean they do not exist or are grammatically unacceptable. Thorough search of other

reference materials such as dictionaries, grammars or other corpora is needed. Alternative solution is the language intuition of an educated native speaker of English. The sample lesson of *advise*, *suggest*, and *recommend* is a prime example to illustrate the judgment dilemma of the absence or rare instances of target information. Pattern 1 in the following was one of the patterns that students were asked to check while Pattern 2 was not included in the sample lesson.

Pattern 1: V + N + to-V

To my knowledge, *advise* and *recommend* can be used in this kind of sentence pattern, but *suggest* is not acceptable in this pattern. For instance:

- I *advise* you *to consult* a doctor, Mrs. Johnson.
- Although they have eight children, they do not *recommend* other couples *to have* family of this size.
- *Her uncle *suggested* her *to get* a job in a bank.

However, one example from *Collins Cobuild Advanced Learner's English Dictionary* (4th edition, 2003, p.1449) seems to 'break' this rule:

Could you *suggest* someone *to advise* me how to do this?

Furthermore, examples of this pattern are also found from BNCweb, although the frequency is pretty low: only 8 out of 8778 matches have this kind of pattern. They are:

- I would *suggest* you *to ask* that question in fact you'll ask Gordon afterwards. (spoken)
- It may be advisable to speak to your local pharmacist who is professionally qualified to offer advice on how to cope with the symptoms of flu. The pharmacist can also *suggest* medicines *to help* reduce your symptoms. (written)
- Can you *suggest* ways *to help* my weight reduction? (written)
- My daughter is really knocked out by it and very depressed. Can you *suggest* anything *to hasten* their recovery? (written)
- But Wednesday's meeting could *suggest* ways *to stop* things getting worse. (spoken)
- Patricia Beachy and Jane Deacon, writing in the American Journal of Gynecologic and Neonatal Nursing, *suggest* strategies *to increase* the safety of neonates while they are in hospital. (written)
- Firstly, the basic hypotheses to be tested must be quite clear and these will immediately *suggest* the topics *to be covered*. (written)
- A vet, may be able to *suggest* some treatment *to help* ease tight muscles causing the pain. (written)

It's a challenge to non-native speaker English teachers to give satisfactory answer to students if students also find these examples from the corpus. Using corpus examples in classroom setting, English teachers, both native speakers and non-native speakers, should view grammar teaching in a broader perspective. The question of grammar rules is usually not about the possibility but about the probability of usage (Tsui, 2004).

Pattern 2: V + NP + V-ing

In *Longman Grammar of Spoken and Written English* (LGSWE), both *suggest* and *recommend* are classified as 'communication/speech act verbs'. They can be followed by a noun phrase (NP) and then -ing clause, which is attested in the *Longman Spoken and Written English Corpus*, but no examples are given (p.742). *Longman Dictionary of Contemporary English* (updated edition CD-ROM), on the contrary, gives an example which 'breaks' the rule:

- If you go to London, we *recommend visiting* (NOT *recommend you visiting*) Tate Modern.

Only rare instances of target pattern can be traced from BNCweb: to be exact, 4 out of 8778 matches for *suggest* and 2 out of 1784 matches for *recommend*.

- "Well, would you like me to approach your parents and *suggest Jennifer coming* to Rosemount?" (written)
- Perhaps I could *suggest her getting* someone in to look after him and she could come for half time. (written)
- As a whole, though, they do *suggest an author looking* back over his own work and trying to reduce it to order. (written)
- It is probably best to take this piece seated, to *suggest a character looking* back reflectively at the events recalled. (written)
- Zelda West-Meads of the marriage guidance counselling service, Relate, believes it's perfectly acceptable to change your partner in the way Madeline and Judith have done, although she would *recommend you using* more tact than Judith did. (written)
- "True psychologists do not *recommend siblings competing* directly because they could become jealous and envious, and might even come to hate each other," says Polgar. (written)

It seems that *V + NP + V-ing* pattern is mainly used in written form, and in fact is not a frequent occurring pattern. Frequency information should not be the sole factor used to decide what to teach, but it definitely can offer valuable information about what kinds of pattern are more popular or acceptable.

5. Real Language

Whether corpus data are real language is the last but not the least challenge to the use of corpus in the classroom. Some researchers express weak reservations about the real language issue of corpus data (e.g., Owen, 1996; Gavioli, 1997) while others have strong reservations about it (e.g., Widdowson, 2000). Widdowson's (2001) major concern about the use of corpora in language teaching is that through "the pronouncements of linguistics and the promotion of publishers", many teachers believe that "the only language worth teaching is that which has been certified as authentic or real by the findings of corpus analysis" (p.531). In Widdowson's term, a corpus comprises traces of texts, not discourse:

For obviously enough, the computer can only cope with the material products of what people do when they use language. It can only analyse the textual traces of the processes whereby meaning is achieved: it cannot account for the complex interplay of linguistic and contextual factors whereby discourse is enacted...., we can say that corpus analysis deals with the textually attested, but not with the encoded possible, nor the contextually appropriate (2000, p.6-7)

The issue whether language in a corpus is 'de-contextualized' is controversial. On the one hand, language in a corpus, in a sense, is real because it is naturally occurring language recorded in written or spoken form. On the other hand, the 'de-contextualized' language from a corpus is often sentence-based data and the corpus analysis usually does not take account of what makes a sequence of clauses a text. For example, coherence and cohesion can not be easily analyzed from a few sentences, let alone one complete sentence with one truncated sentence added before and after it. In the classroom setting, EFL students are unlikely to be motivated by corpus-based data if the instances of language use that they are studying are taken from contexts which are detached from their interests and concerns (Tribble, 1997).

Conclusion

The quantitative analysis of text by computer reveals facts about actual language use

that are not obvious to our language intuition, which shows tremendous possibilities for language learning and teaching. However, concordance lines just present information; they do not interpret it. The interpretation of corpus data requires the insight and competence of the observer.

Corpus-based language teaching is not necessarily suitable for every advanced English learner. The three momentous impacts that Conrad (2000) argued may never happen if most corpora available right now do not have register or genre annotation and the background sources of corpora are not transparent to the public users.

In the classroom setting, teachers should not accept corpus evidence uncritically (Hunston, 2002). Using corpus-based materials in language teaching, non-native speaker English teachers, who do not have language intuition, face more pedagogical challenges than native speaker English teachers. If non-native speaker English teachers find corpus evidence which is beyond their comprehension or contrary to their expectation, or the concordance findings are too rare to be statistically significant, they should be more cautious about their interpretation and explanation.

Dr. Ellis played it safe when he discussed the ‘best’ approach of the grammar instruction:

It is certainly true that I do not believe (and do not think the research demonstrates) that there is just one preferred approach to teaching grammar. The acquisition of the grammatical system of an L2 is a complex process and almost certainly can be assisted best by a variety of approaches. (Ellis, 2006, p.103)

However, it is pretty sure that the consensus about the best approach of grammar teaching will never be reached. Corpus-based grammar instruction is a potential option, but language teachers should use it with care. The marriage between corpus-based linguistics and grammar instruction can have a happy ending if competent English teachers are fully prepared and find adequate methods to integrate corpus-based materials into relevant grammar learning.

Notes

1. Because of the limit of length, the results of students’ exercise are not discussed here.

References

Biber, D., Johansson, S., Leech, G., Gonrad, S., & Finegan, E. (1999). *Longman grammar of spoken and written English*. London: Longman.

- Biber, D., & Reppen, R. (2002). What does frequency have to do with grammar teaching? *Studies in Second Language Acquisition*, 24, 199-208.
- Claveau, V., & L'Homme, M.-C. (2006). Discovering and organizing noun-verb collocations in specialized corpora using inductive logic programming. *International Journal of Corpus Linguistics*, 11, 209-243.
- Coniam, D. (1997). A preliminary inquiry into using corpus word frequency data in the automatic generation of English language cloze tests. *CALICO*, 14(2), 15-33.
- Conrad, S. M. (2000). Will corpus linguistics revolutionize grammar teaching in the 21st century? *TESOL Quarterly*, 34, 548-559.
- Deignan, A. (1999). Linguistic metaphors and collocation in nonliterary corpus data. *Metaphor and Symbol*, 14, 19-36.
- Ellis, R. (2006). Current issues in the teaching of grammar: An SLA perspective. *TESOL Quarterly*, 40, 83-107.
- Flowerdew, L. (2005). An integration of corpus-based and genre-based approaches to text analysis in EAP/ESP: Countering criticisms against corpus-based methodologies. *English for Specific Purposes*, 24, 321-332.
- Frankenberg-Garcia, A. (2005). Pedagogical uses of monolingual and parallel concordances. *ELT Journal*, 59, 189-198.
- Caskell, D., & Cobb, T. (2004). Can learners use concordance feedback for writing errors? *System*, 32, 301-319.
- Gabrielatos, C. (2005). TESL-EJ: Corpora and language teaching: Just a fling or wedding bells? *TESL-EJ*, 8(4). Retrieved from <http://writing.berkeley.edu/TESL-EJ/ej32/a1.html>
- Gavioli, L. (1997). Exploring texts through the concordancer: Guiding the learner. In A. Wichmann, S. Fligelstone, T. McEnery, & G. Knowles (Eds.), *Teaching and language corpora* (pp. 83-99). London: Longman.
- Hunston, S. (2002). *Corpora in applied linguistics*. Cambridge: Cambridge University Press.
- Hunston, S., & Francis, G. (1998). Verbs observed: A corpus-driven pedagogic grammar. *Applied Linguistics*, 19, 45-72.
- Johns, T. (1997). Contexts: The background, development and trialling of a concordance-based CALL program. In A. Wichmann, S. Fligelstone, T. McEnery, & G. Knowles (Eds.), *Teaching and language corpora* (pp. 100-115). London: Longman.
- Lee, D. (2001). Genres, registers, text types, domains, and styles: Clarifying the concepts and navigating a path through the BNC jungle. *Language Learning & Technology* 5(3), 37-72. Retrieved from <http://llt.msu.edu/vol5num3/pdf/lee.pdf>
- Leech, G. (1997). Teaching and language corpora: A convergence. In A. Wichmann, S. Fligelstone, T. McEnery, & G. Knowles (Eds.), *Teaching and language corpora* (pp.

- 1-24). London: Longman.
- Owen, C. (1996). Do concordances require to be consulted? *ELT Journal*, 50, 219-224.
- Partington, A. (1998). *Patterns and meanings: Using corpora for English language research and teaching*. Amsterdam: John Benjamins.
- Renouf, A. (1997). Teaching corpus linguistics to teachers of English. In A. Wichmann, S. Fligelstone, T. McEnery, & G. Knowles (Eds.), *Teaching and language corpora* (pp. 255-266). London: Longman.
- Stevens, V. (1995). Concordancing with language learners: Why? when? what? *CAELL Journal*, 6(2), 2-10.
- Stubbs, M. (2001). Text, corpora, and problems of interpretation: A response to Widdowson. *Applied Linguistics*, 22, 149-172.
- Thornbury, S. (1999). *How to teach grammar*. Essex, England: Pearson Education Limited.
- Tribble, C. (1997). Improvising corpora for ELT: Quick-and-dirty ways of developing corpora for language teaching. In J. Melia. & B. Lewandowska-Tomaszczyk (Eds.), *PALC '97 Proceedings, Practical Applications in Language Corpora* (pp. 106-117). Lodz: Lodz University Press.
- Tsui, A. B. M. (2004). What teachers have always wanted to know – and how corpora can help. In J. M. Sinclair (Ed.), *How to use corpora in language teaching* (pp. 39-61). Amsterdam: John Benjamins.
- Widdowson, H. G. (2000). On the limitations of linguistics applied. *Applied Linguistics*, 21, 3-25.
- Widdowson, H. G. (2001). Interpretations and correlations: A reply to Stubbs. *Applied Linguistics*, 22, 531-538.