The marriage between corpus-based linguistics and lexico-grammar instruction: Using advise, recommend, and suggest as an example

Wen-Shuenn Wu*

*Department of Foreign Languages and Literatures, Chung Hua University, 707, Sec. 2, Wu-Fu Rd., Hsinchu 300, Taiwan, Republic of China
wswu@chu.edu.tw

This paper talks about 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. 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.

1. 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.

1.1 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.

1.2 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. ICAME, COLT, and CSPA are some examples of paid corpora.

1.3 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. Cobuild Concordance and Collocations Sampler (http://www.collins.co.uk/Corpus/CorpusSearch.aspx)
3. JustTheWord (http://193.133.140.102/JustTheWord/)
4. MICASE (Michigan Corpus of Academic Spoken English) (http://quod.lib.umich.edu/m/micase/)

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.

2. Corpus-based linguistics and lexico-grammar instruction
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 reply 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 only once or twice in a textbook, 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 lexi-co-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.

3. 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 student 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.

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. However, 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.

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’s time-consuming for both students and the teacher (if the 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.

4. Sample lesson: use advise, recommend, and suggest as an example

A sample lesson of using corpora was designed to see whether students can discern patterns and regularities in naturally occurring input. An inductive pattern-discovering exercise was given. Before that a couple of online dictionaries and corpus websites had been introduced.

<table>
<thead>
<tr>
<th>Instruction: Use the online dictionary and corpus-based websites that I recommend in the introduction to find out whether the following sentence patterns are grammatically acceptable to these three words (advise, suggest, and recommend).</th>
</tr>
</thead>
<tbody>
<tr>
<td>Example:</td>
</tr>
<tr>
<td>V + to N + that-clause:</td>
</tr>
<tr>
<td>advise: No such usage</td>
</tr>
<tr>
<td>suggest: I suggest to Miss Johnson that she sit down on the chair and wait.</td>
</tr>
<tr>
<td>recommend: No such usage</td>
</tr>
<tr>
<td>1. V + N + to-V or V + N + not to-V</td>
</tr>
<tr>
<td>2. V + wh-clause/phrase</td>
</tr>
<tr>
<td>3. V + that-clause</td>
</tr>
<tr>
<td>4. V + -ing</td>
</tr>
</tbody>
</table>
5. Challenges of the integration of corpora and grammar instruction in language pedagogy

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.

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.

   It’s a challenge to non-native speaker English teachers to give satisfactory answer to students if they also find these examples from the corpus.

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).

   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.

6. 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.

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.

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.

References


