Corpus-Informed Pedagogical Grammar of English: Pros and Cons

Denis Samburskiy*
State University of New York at Albany, 1400 Washington Ave, Albany, NY 12222, USA

Abstract

In this paper I discuss some advantages and disadvantages of incorporating corpus-data instruction into language classrooms, using my experience of teaching a graduate course at State University of New York at Albany titled Corpus-Informed Pedagogical Grammar of English. Some relevant examples are provided, with a discussion of potential benefits of using corpus methods to enrich language pedagogy. In addition, I address some common challenges with utilizing corpora.

© 2014 The Authors. Published by Elsevier Ltd. This is an open access article under the CC BY-NC-ND license (http://creativecommons.org/licenses/by-nc-nd/3.0/).

Peer-review under responsibility of National Research Tomsk State University.

Keywords: Corpus linguistics; language teaching and learning; COCA

1. Introduction

In recent years second language acquisition has seen a steady rise in interest regarding corpus linguistic studies of multi-word units language production. How is corpus linguistic analysis different from others? Firstly, it offers a rigorous inductive approach to language inquiry, which allows for quantification of authentic language patterns. Moreover, corpus linguistics revealed an intricate interplay between form and meaning, thus promoting a fresh stance on language pedagogy labeled lexicogrammar. Lexicogrammar considers lexicon and grammar to be generally intertwined, forming a single entity with no distinction between form and meaning. This view challenges the traditional “wisdom of postulating separate domains of lexis and syntax” (Sinclair, 1991, p. 104) and offers language learners and teachers a window into real native language usage. In addition, corpus linguistics provides user-friendly tools that could be easily incorporated into classroom activities or used independently at home (e.g. Corpus of Contemporary American English www.americancorpus.org)

* Corresponding author. Tel.: 5184425020; fax: 5184425008.
E-mail address: denisalbany@gmail.com
2. What are the advantages of incorporating corpus-linguistics into language classrooms?

Corpus-based research has consistently demonstrated a connection between lexical patterns and registers. Native speakers do not utilize the same language patterns in all communicative situations, but rather opt for certain idiosyncratic linguistic units. For instance, academic prose has been reported to contain more elaborate means of developing, supporting or countering arguments, and leading the readership through logical steps to a conclusion (e.g., therefore, as a result, thus, etc) (Conrad, 2001). In contrast, newspaper writing tends to be organized around chronological sequence. This knowledge might foster learners’ understanding of the purposes of various written genres and help them acquire appropriate linguistic inventory for each type of writing.

Apart from register variation, corpus linguistics demonstrates a complex relationship between lexicon and grammar. This relationship is reflected in frequency distributions of linguistic units in various contexts of usage. For example, many grammar textbooks report that certain verbs can have either to-complements (e.g., want, attempt) or that-complements (e.g., show) or both (e.g., think, say). What those textbooks do not indicate is that think, say and know are much more commonly used with that- than to-complements. Such frequency distribution data could be of great importance to language learners, who need to know not just what is grammatical usage but also what is most common usage. According to corpus analyses, that-clauses are most common in conversation, and their primary purpose is to express opinions and ideas or report the opinions and ideas of others (Conrad, 2001; Biber & Conrad, 2004).

Additionally, corpus-driven analysis of language patterns allows learners to investigate the frequency of any formulaic lexical bundle in any register. Grant (2007) found a number of frequently used idiomatic expressions in spoken academic English, such as bottom-line, come into play, down the line, flip a coin, etc. However, certain idiomatic expressions are favored more in conversation than writing, e.g. come up with (Liu, 2003). Language learners may definitely find such frequency information useful to determine which collocations are more appropriate in a given context.

The affordances that corpus linguistics provides for language educators motivated me to design a TESOL teacher-training course that would incorporate both pedagogical grammar theory and corpus investigations. The course, titled Corpus-Informed Pedagogical Grammar of English, was already offered 3 times to a class of 25-30 pre-service ESL educators at the State University of New York at Albany, NY. The students were approximately 85-90% native speakers of American English, and 15-10% were speakers of other languages.

The main goal of this web-enhanced course was to inform future ESL educators about essential pedagogical grammar concepts from a traditional prescriptive position, and elaborate on those concepts by applying descriptive data from authentic written and spoken texts in American English corpora (www.americancorpus.org). Students were expected to investigate information from the course book, learn how to analyze data gathered from online corpora, and use both to design authentic classroom activities for intermediate and advanced English learners. In addition, we addressed challenges that learners might generally have mastering American English grammar, as well as some common errors made by native speakers. The final project in the course was developing a lesson plan with corpus-informed activities and exercises; hence, the course had a hand-on perspective as well as a theoretical component. The course contained an online section via a popular teaching platform Edmodo (www.edmodo.com), which helped my students to share their expertise, accrue a database of quality grammatical activities, and discuss relevant ESL issues.

Since my student population was never homogeneous, I had to modify my lesson content to appeal to both native and non-native speakers. Interestingly, non-native speakers tended to display a better understanding of and overall grasp on grammatical theory; therefore, it was sometimes harder to keep them motivated, especially, when we discussed some ‘elementary’ grammatical aspects. As I came to realize, what appeared to be elementary to English learners was not necessarily easy to native speakers of English. Thus, I strove to use the corpus to illustrate some basic principles of English syntax and semantics. In addition, I asked non-native speaker students to help natives as often as possible, and American students to help English learners with their unmistakable feel for grammatical accuracy. Thus, my goal was to create a collaborative environment, in which the strengths of both native speakers and learners could foster a mutually beneficial learning setting.

Apart from general readings on various grammatical topics, each week my students were to complete a ‘language investigation’ by means of an online corpus. I designed each task so that they increased in complexity: the students
were required to analyze more complex syntactic combinations and utilize more sophisticated searches. My objective was to demonstrate a connection between lexicon and grammar in language in use, so I focused at first on collocational and colligational patterns in English. In the lexicogrammatical view of language, “a grammatical structure may be lexically restricted” (Francis, 1993, p. 142) and, conversely, lexical items are often grammatical in nature, because the use of a lexical item often has grammatical implications (Conrad, 2001; Hunston & Francis, 2000). Different words often tend to demonstrate unique patterns of co-usage with other lexical items (e.g., in the meaning of caretaking/educating a child/children, *raise/rear a child/children* typically collocate but *lift/elevate a child/children* do not, example from Liu (2011)). Along with collocational restrictions, which affect word combinations on the semantic level, there are also colligational restrictions, which affect combinations on the syntactic level (e.g., of the synonyms astonishing/shocking/surprising, *only surprising* often appears in the negative structure “It’s not surprising”, example from Liu (2011)). Thus, Liu (2011) argues that ‘lexicogrammar covers not only traditional grammatical topics such as syntax and morphology (i.e., sentence and word formation rules) but also important lexical usage issues not considered within the purview of grammar traditionally’ (p. 357).

In my class, the students were required to discover semantic motivation behind certain collocations of English. For example, in a home assignment about prepositions, the students received the following corpus task:

a. What is the most common preposition following ‘in case’? What are the most common nouns following ‘in case’ with that preposition? What can you say about the connotation of such expressions? Positive or negative?

b. Search for prepositions that occur after all forms of the verb ‘smile’. How many of the most common preposition ‘at’s’ do you find? Then, click on the preposition ‘to’ and look at the examples. What words do typically occur after ‘smile to’?

This task required looking into the collocational pattern of the preposition *in case of* and its lexical combinations with noticeable negative connotations, such as *in case of emergency* (91), *fire* (37), *war* (36), *attack* (21), etc (the numbers in parenthesis indicate the frequency of the collocation in the corpus). Then, a search for *smile* (as a verb) with a prepositions revealed that *at* was the most frequent preposition associated with this verb. However, *smiled to* (in the past tense) was sixth on the list. Further analysis of the context in which this collocation was used showed that it was always followed by a reflexive pronoun, for example, *I smiled to myself when I reached the paper-packed shelves, In the kitchen Malini smiled to herself, Walking down the side of the narrow road in the darkness, Melody smiled to herself*, etc. Such tasks were aimed at raising my students’ awareness of the combinatorial nature of language in use and a powerful aspect of collocational selectiveness.

Another example of a collocational task was the following assignment, which was given after we discussed the topic Modal Verbs:

a. What are the most common combinations of |pronoun + had better + verb| and |pronoun + would rather + verb|? (a couple are enough)

b. Can would rather or had better be followed by infinitives with a particle ‘to’?

This assignment required the students to explore some colligational behavior of two semi-modal verbs: *had better* and *would rather*. We found out that *It had better be, You had better be, and We had better get* were the most frequent combinations with *had better*, whereas *I would rather have, I would rather be, and He would rather be* were most frequent for *would rather*. This finding points to the tendency of *would rather* to be express personal volition, while illustrating the typical sense of irritability, threat, or advisability toward others of *had better*. Although native speaker’s intuition undoubted helped them to answer the second question, my requirement was not provide hard evidence to support their proposition. The corpus had no instances of any infinitives with particle to after either of the semi-modals except a couple of ungrammatical patterns in character dialogues from fiction.

3. What are the disadvantages of incorporating corpus-linguistics into language classrooms?

Despite the obvious advantages of employing corpus-driven methods in English teaching, researchers have identified a number of drawbacks in this methodology. Liu (2011) reported that participants in his study found it hard to use the corpus; so conducting even some basic searches was time-consuming. I could share a similar observation in my course, since some students persistently struggled with corpus searches despite my step-by-step description of each assignment. One the one hand, the corpus like Corpus of Contemporary American English
requires its users to learn special commands (called syntax) to perform any searches that go beyond specific words or phrases. If a student is interested, for example, in finding most common adjectives modifying the noun car, he or she would have to use the corpus command \([j^*]\) to indicate any adjective before car and the entire search command would look like this:

\([j^*]\) car

After clicking search, the student would see the results: new car (1422) is the most common combination in American English. More complex searches require a good understanding of the corpus syntax and other specifics. However, this hindrance is only temporary and will become less significant as users hone the skill of corpus searching.

A more critical skill, in my opinion, is having some understanding of grammatical structure of the English language.

Knowing what restrictions are placed on sentence segments and what combinations are grammatically feasible in English. For example, one assignment required the students to find some examples of before, after, and since as both subordinators and prepositions. Despite some examples provided, the students had difficulty understanding the distinction between these two functions. Subordinators, by definition, act as logical or temporal links between dependent and independent clauses; therefore, in a dependent clause, a subordinator must be followed by a subject (denoted by a noun phrase). For instance, Before I arrived, I called my parents. As a preposition, before may be followed by a noun phrase or a gerund, but neither will have a function of subjects, as in Before arriving, I called my parents. Such functional distinctions pose a challenge for students because they require a high degree of grammatical knowledge and planning before embarking on a corpus search. Simply searching for a noun phrase after before, after and since will yield many examples. However, they will contain sentences that have before, after, and since as both subordinators and prepositions. Consequently, the student would have to manually select relevant examples for each function of the words. It is this pre-search planning that baffles many students and may discourage them from turning to the corpus for fear of not being able to obtain trustworthy findings.

One final note about corpus disadvantages is that language educators are often reluctant to incorporate corpus work into their classes because they do not understand how to do it properly. A concur with Conrad’s (2001) suggestions for corpus use in classrooms:

- Corpus analysis should not be utilized as the only means of language instruction;
- Corpus tasks may accompany regular language tasks to develop learners’ linguistic intuition, sensitivity to register variation, good command of collocational and colligational patterns;
- The ability to use corpora empowers learners, especially those who do not have a chance to interact with native speakers much, because they can use a database of cumulative knowledge of native speakers at any time of day and night for free.
- Lastly, frequency findings from the corpus should not mean that a rare word combination or grammatical structure should be neglected. Rather, it should inform learners of a changeable nature of language in use and teach to be vigilant about chronological variations.

4. Conclusion

In conclusion, I can suggest that language educators familiarize themselves with this new trend in language pedagogy and take advantage of what it has to offer. This technological age provides us with a plethora of tools but corpus linguistics is one method that has proven to be extremely beneficial for both instructors and language learners alike. Despite struggling with some initial difficulties, everyone will soon see advantages of this tool and improve their teaching practices or language learning outcomes.

References


