THE XXV ANNUAL INTERNATIONAL ACADEMIC CONFERENCE, LANGUAGE AND CULTURE, 20-22 October 2014

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.

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 form 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 prescriptivist 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