



Literatura y Lingüística

Universidad Católica Cardenal Raúl Silva Henríquez

jdela Fuente@ucsh.cl

ISSN (Versión impresa): 0716-5811

ISSN (Versión en línea): 0717 - 621X

CHILE

2006

Leonardo Juliano Recski

CORPUS LINGUISTICS AT THE SERVICE OF ENGLISH TEACHERS

Literatura y Lingüística, número 017

Universidad Católica Cardenal Raúl Silva Henríquez

Santiago, Chile

pp. 303-324

Red de Revistas Científicas de América Latina y el Caribe, España y Portugal



Universidad Autónoma del Estado de México

<http://redalyc.uaemex.mx>

Corpus linguistics at the service of English teachers

Leonardo Juliano Recski
brasileiro
UFSC / NUPdiscurso / Faculdades Barddal
recski@gmail.com

Abstract:

This paper contemplates how corpus evidence might be used to address teachers' questions about English grammar and suggests that corpus linguistics has an important role to play in raising their awareness to linguistic features and patterns. The article surveys a range of grammatical questions posted by EFL/ESL teachers on four Orkut communities devoted to the teaching and learning of English. It concentrates on three specific types of questions: synonymous lexical items which function differently and are reported to be difficult to teach and explain; linguistic evidence that contradicts the prescriptive grammar rules that teachers have been taught during their education; and collocations that teachers attempt to explicate. In attempting to address the teachers' questions, corpus evidence is used to offer possible explanations. It is suggested that the use of corpus data in addressing these questions is not only convincing but also leads to discoveries of patterns and meanings which might not be found in other reference materials such as grammars and dictionaries.

Keywords: corpus linguistics; EFL/ESL teachers' questions; linguistic description; language awareness

Resumen:

Este artículo contempla cómo la evidencia del corpus lingüístico podría ser usado para responder preguntas de los profesores de inglés respecto de la gramática inglesa, y sugiere que el corpus lingüístico tiene un rol importante que jugar para incrementar la advertencia respecto de los patrones y aspectos lingüísticos. El artículo describe una selección de preguntas en torno a gramática inglesa enviadas por profesores de inglés como idioma extranjero y como segunda lengua en cuatro comunidades Orkut dedicadas a la enseñanza y aprendizaje del idioma inglés. El artículo se concentra en tres tipos específicos de preguntas: ítems léxicos sinónimos que funcionan de manera diferente y se reportan como difíciles para enseñar y explicar; evidencia lingüística que contradice las reglas de la gramática prescriptiva, la cual se ha enseñado a los profesores de Inglés por siempre en su formación, y expresiones de uso frecuente que los profesores intentan explicar. Al intentar responder las consultas de los profesores, se usó evidencia recogida de un corpus lingüístico para ofrecer posibles explicaciones. Se sugiere, con este estudio, que el uso de corpus lingüístico para fundamentar las explicaciones, no solamente es convincente, sino que además conlleva a descubrimientos de patrones y significados que podrían no hallarse en otros materiales de referencia, tales como libros de gramática y diccionarios.

Palabras claves: corpus lingüístico – profesores de inglés como lengua extranjera y como segunda lengua – descripción lingüística – conciencia lingüística

1. Corpora and linguistic description

Before the existence of corpora, linguistic description relied very much on native-speaker intuition and introspection. Native-speakers normally describe what they *know about* language, or what they *perceive* language to be, rather than how language *is* used. The easy accessibility of huge bodies of naturally occurring texts on the computer has made it possible for us to test the robustness of linguistic descriptions which were based on introspection and elicitation, and to gain new insights into language structure and use. It has helped us to gain a better understanding of how language is actually used rather than how language is perceived to be used. Examining specific instances of language use gives us insights into how language works which would never have been obtained by simply introspecting about the language system. Such insights result in our construing the linguistic system in a different way. So far, corpus-based studies have focused on four main types of description and analysis: *lexical collocation* by examining the frequency and context of occurrence of linguistic items (see for example, Sinclair 1991; see also Kjellmer's (1994) dictionary of collocations based on the Brown Corpus), *syntactic patterning* based on co-occurrence of grammatical word-class tags, *genre analysis* based on the co-occurrence of groups of linguistic items and processes (see for example, Biber, 1988), and *discourse structure* and cohesion in spoken and written English (see for example, Carter and McCarthy's Spoken English Corpus at the University of Nottingham - e.g. Carter & McCarthy, 1997) (see Kennedy, 1998, for a sound summary of corpus-based studies). The findings of the above studies, particularly word-based studies, have important implications for second or foreign language teaching, as we shall see in the next section.

2. Corpus-based studies and ESL / EFL teaching

In EFL and ESL situations, learners do not have the same amount of exposure to the target language as they do in L1 situations. Therefore, it is safe to presume that in most cases they are unlikely to acquire the language efficiently without systematic guidance on linguistic forms. By focusing on words which have a high frequency of occurrence and by concentrating on the usual rather than the exceptional, teachers can help learners acquire the language more efficiently, especially at elementary and intermediate levels. The findings of corpus analysis can be used as a basis for selecting and sequencing linguistic content, as well as for determining relative emphases. A number of studies have observed discrepancies between corpus findings and the selection of and emphasis given to linguistic content in ESL and EFL textbooks and curriculum. As early as the sixties, George (1963, cited in Kennedy 1998: 283) studied a corpus of English that was based on written texts and found that the highest frequency of occurrence of the simple present is not to indicate habitual or iterative actions, such as "I go to school by bus every day" (5.5%), but rather the actual present, such as "I agree with you" (57.7%) or neutral time, such as "My name is Mary" (33.5%). His findings converge with a more recent grammar of English compiled by Mindt (2000) based on corpora totaling 240 million words of spoken and written English. Mindt found that the three prototypes which make up the majority of all cases of the present forms of verbs are the extended present, the actual present and the timeless present. This is contrary to the emphasis given to the habitual present in most ESL and EFL textbooks as the major function of the simple present.

Holmes (1988) compared a corpus analysis and a textbook analysis of epistemic modality and found that, like most textbooks, important epistemic uses of modal verbs are under-taught and that lexical verbs expressing modality, such as *appear*, *believe*, *doubt*, and *suppose*, nouns such as *possibility*, *tendency*, and *likelihood*, and adverbials, such as *perhaps*, *of course*, and *probably*, tend to be given little pedagogical attention. Ljung (1991) compared the EFL textbooks at upper secondary level in Sweden with the Cobuild corpus and found that 20% of the most frequent one thousand words in the learners' texts did not occur in the most frequent one thousand words in Cobuild. Biber, Conrad and Reppen (1994) examined the structural options for postnominal modification and the attention given to these options in popular ESL and EFL textbooks. They found that typically more pedagogical attention was paid to finite and non-finite relative clauses than prepositional phrases as noun modifiers, in contrast with their analysis of the Lan-

caster Oslo/Bergen (LOB) corpus, which shows prepositional phrases as noun modifiers occurring far more frequently than relative clauses (see also Quirk et al., 1985: 1274). Kennedy (1998) observes that similar incompatibility can be found in the pedagogical focus on grammatical quantifiers such as *all* and *every* in many textbooks to indicate the concept of totality when in both written and spoken corpora totality is much more commonly lexically marked, such as *entirely*, *completely*, *whole*, *throughout*.

The above very brief summary of some comparative studies of corpora and ESL and EFL textbooks show the relevance of corpus studies to ESL and EFL teaching and learning. One of their major contributions is to provide objective quantitative evidence of the distribution of linguistic items on which the goals and content of the curriculum can be based.

3. Corpus analyses and teachers' language awareness

One area that is under-explored is the relevance of corpus linguistics to teacher education, particularly in the area of teachers' language awareness (see also Berry, 1994; Hunston, 1995). In the last decade or so, more attention has been paid to the importance of raising teachers' language awareness (see for example the collected papers in Bygate et al., 1994; Hawkins, 1999). This paper hopes to emphasize that teachers' language awareness is one area in which corpus linguistics has an important contribution to make. It examines grammatical questions that English teachers of EFL/ESL posted on Orkut communities devoted to the teaching and learning of English to seek advice, and attempts to demonstrate how empirical linguistic data which show the context and frequency of occurrence of the linguistic items in question can be a powerful tool to raise teachers' linguistic sensitivity, to help teachers question long-standing assumptions, and to gain new insights into language structure and use.

4. The Orkut¹ communities of EFL/ESL Teachers

Currently, there are innumerable Orkut communities of EFL/ESL teachers on the web. Most of these communities were created in 2004 and some of them (e.g. *English as a Second Language*) have close to 20.000 members. The criterion I have adopted for choosing which communities to scrutinize was solely based on the number of members, since communities with a large number of participants were more likely to yield more questions. A total of four communities were investigated in search of grammatical questions posed by teachers:

- a) *English as a Second Language* (<http://www.orkut.com/Community.aspx?cmm=90629>): **Description:** "A community for students and teachers of ESL discussing general questions about English, the best ways of teaching and learning, where to find resources for teaching, and diverse cultural aspects". **Founded:** June 12, 2004. **Members:** 18.884. **Language:** English. **Location:** US
- b) *English Language Teachers* (<http://www.orkut.com/Community.aspx?cmm=67884>): **Description:** "For English Language Teachers from around the world who want to share thoughts and learn. Use our forum to get info, exchange ideas, make friends or simply practice your English". **Founded:** May 22, 2004. **Members:** 7.106. **Language:** English. **Location:** Brazil
- c) *English Teachers in Brazil* (<http://www.orkut.com/Community.aspx?cmm=65989>). **Description:** "A community for English teachers living (or not) in Brazil. It's a place for us to share lessons, ideas and teaching experience". **Founded:** May 21, 2004. **Members:** 6.490. **Language:** Portuguese. **Location:** Brazil
- d) *English Language Teaching* (<http://www.orkut.com/Community.aspx?cmm=99548>). **Description:** "A place to share your teaching

1 Orkut is defined as "an online community which connects people through a web of reliable friends". To make sure that abusive content (e.g. pornography, racism, pedophilia, etc...) does not circulate among the different communities the idealizers have created standards which are the shared values of the orkut.com community. According to the idealizers "the standards are a living document and will change based upon the needs of the broad community and the available tools. The Community Standards will be upheld through a combination of human and automated moderation. As you may have noticed, an automated abuse-detection system is already at work. The system temporarily suspends the accounts of individuals who are abusing the orkut.com community." (quoted from www.orkut.com)

ideas and experiences". **Founded:** June 17, 2004. **Members:** 2.238.
Language: English. **Location:** Brazil.

In the rest of this paper, I shall examine the grammatical questions sent by teachers over a period of two years, discussing some of the ways in which corpus data could be used to help them tackle the sorts of questions they are faced with in their everyday teaching.

5. Teachers' grammatical questions and corpus evidence

An analysis of the grammatical questions posted on the Orkut communities of EFL/ESL teachers over the period of two years shows that they largely fall into one of the following six types. The first type has to do with synonymous lexical items. Some lexical items are largely synonymous but have different usage. Teachers are aware of the difference in usage but have problems explaining the difference to students, for example, *tall* and *high*. Some lexical items appear to be synonymous but teachers are not sure if they are "absolute synonyms" (Partington, 1998), for example, *day by day* and *day after day*. The second type relates to linguistic evidence that contradicts the prescriptive grammar rules that teachers had been taught when they were learners, the most frequently asked being subject-verb agreement and the use of the definite article. The third type concerns lexical collocations which teachers try to rationalize but sometimes cannot. The fourth type consists of lexical items which teachers take to be absolute synonymous but have been asked by students to explain whether there is any difference in meaning, for example, *big* and *large*, *lastly* and *finally*. The fifth type are prescriptive stylistic rules which seem to have been passed on from generation to generation but are queried by teachers, for example, the rule that one should not begin a sentence with *because*, *and* and *but*. Finally, the sixth type concerns lexical items which students find confusing because their translation into the students' first language is either identical or very similar, for example *find* and *look for* which will have very similar translations in Portuguese. Because of the limit of space, I shall focus on the first three types of questions. Teachers' questions will be cited, followed by the use of corpus evidence in hypothetically addressing their questions. I shall also discuss how in the course of addressing the teachers' questions, analyses of corpus data may lead to insights about linguistic patterns and meanings which apparently have not been given much attention in reference grammars, dictionaries, or in common practice.

5.1 Synonymous lexical items

One of the most frequently asked questions is whether there is any difference between words that are commonly taken as synonymous. There are cases in which teachers were not aware of any difference in meaning and usage, such as *big* and *large*, *lastly* and *finally*. But there are some in which they were aware of a difference in usage but could not quite articulate what the difference was, for example, *tall* and *high*.

Tall vs. High

The following is a message sent by a teacher who said that she knew that *tall* and *high* have different usage, but she could not quite explain the difference to her students:

The words “tall” and “high” have similar meaning but different usage. I have no problems in using the words myself. However, I find it difficult to explain the difference between these two words to my students. Is there any suggestions of teaching these two words?

The difference in usage between *tall* and *high* is particularly difficult for Brazilian learners to grasp because there is no such distinction in Portuguese. Both words will be translated as the same word in written Portuguese. Therefore explaining the words in Portuguese does not really help.

The *Longman Dictionary of Contemporary English* states that *high* is used for measurement of most things but not people, especially when we are thinking only of distance above the ground, such as *a high shelf*, *a high building* and *a high mountain*, whereas *tall* is used for people and ships, for example, *a tall man* and *a tall ship*. It further adds that *tall* is used for things that are high and narrow (e.g. *a tall/high building* and *a tall/high tree*).

In response to the teacher’s question, I have searched the Freiburg-Brown Corpus of American English (FROWN)² with roughly one million words, and examined the nouns that were modified by *high* and *tall*. What emerged from the search was that there was a tendency for *high* to be used in a metaphorical sense with more abstract nouns whereas *tall* tended to be used more frequently with concrete nouns such as *people*, *trees* and *buildings*. The following concordance lines elicit this difference in use.

2 For further information access <http://khnt.hit.uib.no/icame/manuals/frown/INDEX.HTM>.

cers of the bank or financial institution; the prestige of **high bureaucratic position** means that any lesser offi

ce are generally considered to be those students with **high ability**, particularly, high mathematical ability (D
ected decline in fishing boat production, coupled with **high costs** in the change-over to totally enclosed lifebe
ve revealed that individuals with the most potential for **high academic achievement** in mathematics and scie
rooted in the English, American and French history of **high comedy**, and it's the tradition that I want to work
conomic policies - rigid European exchange rates and **high interest rates** - adopted to cope with German re
ated return of 13 rupiahs for each rupiah spent. These **high economic returns** often justify a higher priority fo
ice but to lower interest rates again given the fact that **high federal budget deficits** have eliminated the possi
100 of the viable seeds in the first tray and 100 with a **high incidence of genetic load** in the other, and also i
e boxing and baseball. Those who held their health in **high esteem** played baseball, so like everyone else i
aximum output level is reduced to conserve power. A **high output** is needed to maximize signal-to-noise rati
red to be those students with high ability, particularly, **high mathematical ability** (Davis, 1965; Green, 1989;
urreal, spontaneous, mixing off-hours pop culture with **high political meanings**, public behavior with private c
rty as an institution could afford to waffle <quote>on a **high moral principle**.</quote> The Bush campaign's r
eir graphics accelerator hardware to get reasonable or **high performance**, however. </p> <p> The X Consorti
factors that contribute to addiction should be given a **high priority**, as this should result in crucial advances
ortunate, given that such expenditures often have very **high rates of return**. For instance, the expected retur
has been battered by layoffs, corporate cutbacks and **high unemployment rates**, and it would signal the en
ndary education and on preventive health usually have **high returns** and are central to increasing the producti
araderie and competition of an all-women's race. The **high spirits** generated by women running with women

severe biological ramifications in the Great Lakes too. A **tall aquatic weed**, purple loosestrife, is edging out native
ented six feet deep in places. The gargantuan funnel, a **tall**

four-story building, collapsed and spewed secti blankets in the middle of sagebrush country, toward the **tall brown of snowy mountains**. The city had almost disa York, with its cold stone base, its pagan portico with six **tall columns**, its central doorway with a <quote>squared re exists an undifferentiated land mass of red barns and **tall corn** or golden wheat growing in flat, featureless land airy and joyous, she proves that, contrary to tradition, a **tall dancer** can give a lot to Bournonville. </p> <p> Henri ader combines last year, running them in wheat, barley, **tall fescue and blue grass**. <quote>We were impressed ld manage. </p> <p> Another night, we went to lie in the **tall grass** behind the practice field; again I fled. My serming to someone else, someone with a red shirt. He was **tall man** with wavy brown hair. He was lighting her cigar scattered holes and trenches visible from the sidewalk. **Tall poles**, topped with identifying letters or the logical o Dr. Mega, I'm glad I ran into you!</quote> </p> <p> **The tall red-headed guy** with a Starfleet patch on his jacket I small boys, down the gallery, wafting them on like some **tall sailing ship** - a sort of covey of noble English life. ... r evening tightened around them, and the windows of the **tall sitting room** with its fine provincial furniture gave b ere piled in leaning towers all over the room. There were **tall stacks of newspapers and magazines** as well, and b g board and a sandbox and a long swing hanging from a **tall tree**. She imagined the lift she could get out of that s aight across the street to the house on the other side, a **tall yellow house** with dark green shutters and an overgr es in the world, and they traveled the globe. </p> <p> **A tall, dark, handsome man** also journeyed to New York to d feel vulnerable passing by places such as alcoves and **tall, dense shrubs** that afford potential offenders refuge (

These concordance lines are a useful start to get the teachers to think about a word not in isolation but in terms of its “semantic preference” (Sinclair, 1991). A further analysis of the FROWN corpus revealed that there are 605 instances of *high* whereas there are only 54 instances of *tall*. Except for 9 instances in which *tall* is used idiomatically, such as *a tall order*, *walk tall*, the rest are used in the context of talking about the height of people, buildings, and vegetation (see the above concordance lines for *tall*), with the highest frequency of *tall* co-occurring with people (about 50%) followed by buildings and structures (about 35%). In other words, the semantic preference of *tall* is quite restricted. By contrast, the contexts in which *high* is found is much more wideranging, including

amount, intensity, quality and relative quantity. Taking 5 as the cut-off point for frequency yielded the following nouns that co-occur with *high* in the corpus (see Table 1).

Table 1. Nouns and their frequency of co-occurrence with high

High			
level(s)	37	standard	13
speed	22	quality	12
interest	21	degree(s)	7
cost(s)	19	rate	5
price(s)	15	percentage	5

5.2 Grammar rules and conflicting evidence

Teachers are often troubled by the fact the grammar rules that they have been taught as students do not accord with the authentic linguistic examples that they encounter (see also Tognini-Bonelli, 2001). Indeed, in the Orkut posts, one type of most frequently asked questions has to do with teachers who try to apply some usage rules but are confronted with conflicting evidence. The most frequently asked questions pertain to subject-verb agreement and the use of definite articles.

Subject-verb agreement

The following are some of the messages posted by teachers

Teacher A

Hello! Which one is correct?

There is a man and a woman outside.

Or

There are a man and a woman outside.

Please give some comments, any one.

Teacher B

Hi,

What should we use in the following sentences? is or are?

1. There _____ an apple and some oranges on the table.

2. There _____ some oranges and an apple on the table.

Thanks

It seems to me that 'are' is okay in both. Is there any rule here?

Teacher C

To me, I will use "There are a man and a woman..." It is because we are talking about two persons.

I want to make sure that if this is grammatically correct.

We can see from the response given by Teacher C that she was trying to apply the rule of subject-verb agreement to the example provided by Teacher A. Teacher B intuitively felt that *are* can be used in both sentences but she was looking for some rules.

We could respond to the teachers' questions by pointing out that usually the singular form of *be* is used when the first noun that follows is singular and the plural form of *be* is used when the noun group after it is plural (see also the *Collings Cobuild English Grammar*, p. 416). However, a search through the corpus does show an instance of the following:

According to PACE, suspects can only be detained at designated police stations where **there are a custody and a reviewing officer.**

In other words, while what is stated in the Cobuild English Grammar is correct, teachers may benefit from knowing that occasionally the plural form of *be* can be used even when the noun following is singular. Therefore, the question is not about possibility but about probability of usage.

What is interesting is that the investigation of *there's* is in a corpus of academic spoken English reveals that it often appears before a plural noun. Although Quirk et al. (1985) have also made this point, it is much more convincing to provide teachers with corpus evidence. A search on the Michigan Corpus of Academic Spoken English (MICASE) showed that there are 3,785 instances of *there's*. The following concordance lines could be provided to teachers

... and what you're saying is, if **there's two parallel lines**, well they're both being affected ...

... I'll be walking around, if **there's any questions**. Ready set go ...

... if you stand in a river what do you see I mean **there's banks, and braided channels** I mean ...

... LSD has a bigger effect cuz **there's more receptors** for it to act on ...

... what's happening in the wintering grounds **there's some areas** where there's tremendous loss, of habitat ...

By contrast there are only 30 instances of *there're* followed by a plural noun

around the room and discuss. um, sometimes **there're errors** that occur in meiosis particularly meiosis one okay? an ck actually. even even at Michigan i'm sure **there're departments** that we wouldn't wanna join as faculty members instrument as well, these little bells. now **there're ways** to indicate the sounds that these instruments make, by sel s no doubt i fully agree with you mhm mhm **there're times** that you just essentially have to set it aside and say i'm j just, it's j- absolutely full of them yeah and **there're places** like um, so so you would say that i- before the transcri now this may be just be because there there **there're delays** and timelags in publications but there's sort of some stu ut sort of the tragic dimension of life, that **there're things** you can't, control, there're things that lie beyond your d they become, generally, i mean of course **there're exceptions** and Germany in the nineteen twenties is an excepti

The discussion on subject-verb agreement led to further questions of a similar nature from other teachers regarding whether the singular or the plural verb should be used in the context of *one of the* + plural noun.

Teacher D

Should you say:

1. Peter is one of the richest boys that have/has ever studied in our school.??
2. He is one of the writers who were/was honoured yesterday.??
3. One of the boys was punished yesterday. CORRECT
4. One of the writers was honoured yesterday. CORRECT

Teacher E (responded as follows):

I've asked my panel and she said,

1. Peter is one of the riches boys who has ever studied in our school.
2. He is one of the writers who was honoured yesterday.

A search on the FROWN corpus showed that there are 71 instances of "*one of the ... (be) ...*" and both the plural and singular *be* forms are used. There is a higher frequency of occurrence of the plural form but the singular form is used often enough to be regarded as an acceptable alternative.

Some further examples of questions of similar nature are whether a plural or singular verb should be used after the structure “*none of the ...*” and “*more than one...*”. For example:

Teacher F

Should we use a singular or plural verb after the structure “none of the _____”?

Teacher G

Hi, would somebody be kind enough to tell me why we should use a singular verb after ‘more than one player’?

It is clear from the teachers’ questions that they were puzzled by the lack of agreement between the subject and the verb. It could be pointed out to them that technically, *none* means literally *not one*, and thus it seems to be more logical to use a singular verb. However, because *none* of functions as a quantifier, it is often followed by a plural noun, and therefore a plural verb is used. A search on the FROWN corpus revealed both singular and plural verbs being used. For example:

istian heroes - including the gospel writer! **None of the personalities** whose thoughts **are** described is particularly colant as she pleases, outline the walkway to the patio. **None of the beds contains** permanent plantings, she says, be gs unworthy of publication - and, in fact, **none of the poems** from this period **was** ever pulled from his journals and tion, due to lack of familiarity with ejectives. Moreover, **none of the sources contains** any word of the form sak or s rcentage of Americans volunteering that **none of the three governments spends** their tax dollars wisely jumped fro far as I know. But what’s it about? Amy asked. **None of the kids** who are in it ever **talk** about it. Well, it’s zing individual sports over team activities. **None of the kids** we’re targeting **is** going to grow up to be a team player, ign language with the leader of the band. **None of the Comanches’ rifles were** aimed at Free ... for the moment. I fo n the Fitzgibbon land, work that provides **none of the fellowship** that **prevails** among the lumber handlers. He work

The above discussions sparked off a series of questions from teachers posing related questions asking whether one should say *there are no students in the room* or *there is no student in the room*; *I have no friends* or *I have no friend*. The use of corpus evidence is particularly helpful because the question is not about possibility but probability. Moreover, teachers came up with so many variations of the subject-verb agreement structure that it would not be possible to provide some kind of comprehensive

guiding principle. The best solution, I should think, would be to invite them to look for corpus evidence themselves.

Definite articles

The presence or omission of the definite article is another problematic area for teachers. They have difficulties finding some kind of consistency in the rules for using the definite article. For example, they have been told that the definite article should be used if there is only one of a kind being referred to, such as the sun, the moon, and the earth, the name of a country, and before a position, such as the Chairman and the Secretary. However they have also come across cases where the definite article was missing. For example, *He was elected Chairman of the Association, She was appointed secretary of the committee.*

Sometimes, the questions asked by the teachers can be very specific and it would not be possible to answer them without consulting a corpus. Take for example the following message posted by Teacher D:

Hello! Although I have been an English teacher for about 4 years, I still sometimes have difficulty in using articles. I would be very grateful if someone can help me in the following problem:

They watched television.

They listened to music.

They listened to the radio.

So, should I say:

“When the teacher was teaching, they listened to the CD player.”

Or “When the teacher was teaching, they listened to CD player.”

In relation to teacher’s question it would be possible to point out that the definite article *the* is used when referring to systems of communication or mass media, such as *the radio, the telephone* and *the mail*. It could be observed that the use of *the* is a bit variable with television since a search through the FROWN corpus shows instances of television, with and without *the*. For example:

Try to view your work as dispassionately as you would any other program you might watch on **the television**.

... whatever her talents are, one of them lies in getting people to watch her on **the television**.

One of the things you will notice when watching **the television** is that close-ups are ...

He watched **the television** at night, all night, making remarks like

After a week's exposure to American television's sports coverage it was a relief to **watch television** at home

There was usually literally nothing for clients to do but sit, **watch television**, or walk about.

First thing, on average how many hours a day do you **watch television** nowadays?

The question about "CD player" is a bit more problematic. There are only eight instances of "CD player" in the FROWN corpus. My intuition led me to conclude that since a "CD player" is in fact a name which has been generalized to refer to any normal size or small portable player for CDs rather than a system of communication, the tendency would be to use it as an ordinary countable noun, and therefore the indefinite article "a" or possessive pronoun would be used (the two examples in the corpus containing an article confirmed my intuition). I then conducted a further search on the BNC-online (<http://sara.natcorp.ox.ac.uk/lookup.html>) and found 56 occurrences of CD player which further confirmed my introspection (although there were two hits where "the" was used preceding "CD player"). Here are a few examples

The standard equipment list extends to electric everything, anti-lock brakes and **a CD player**

Used with **a CD player** and/or tuner and tape deck a very high quality system can be assembled at reasonably low

The package comprises **a CD player**, our anthology and the software to run it

... and a Pioneer N-92T mini system comprising **a CD player**, radio, cassette deck, 60 W amplifier and speakers

Sheila reckons police should soon be hot on the trail of the burglar, who stole **a CD player** and sentimental items.

5.3 Rationalization of collocations

The third type of frequently asked question has to do with rationalization of collocations. Teachers often try to look for rules governing which words can go with certain words and why. For example, the following is a message from a teacher asking whether one can say *well-experienced*.

Teacher H

Hello everyone!!!

If we say someone is experienced, we mean this person has certain knowledge or expertise, right? Do we have 'well-experienced' as well? If so, does it mean there is an even higher level of expertise?

A search conducted on the FROWN corpus on the adjective *experienced* revealed only fourteen instances. Of these fourteen occurrences there were only two where *experienced* was modified by the superlative *most*. It became obvious that a corpus of 1 million words would not suffice if we were to give a sound explanation for this teacher. I then used the BNC-online search (100 million words) to investigate what types of intensifiers (e.g. *vastly*, *highly*, *very*) and comparatives (e.g. *more*, *less*) could modify the adjective *experienced*. A search for *well experienced* yielded the following eight instances:

the intention being (and Dudek was very **well experienced** in this sort of work) for him to take
 As a rule, the examining doctor will be **well experienced** in dealing with sexual symptoms and problems
 Firms which are **well experienced** in overseas employee transfers often have international personnel departments
 guard of a godling despot must be assumed to be numerous and very well armed - though how **well experienced**?
 A life science graduate already **well experienced** as a CRA, you should ideally have worked with antibacterial...
 However, encouragingly, our Export Department is **well experienced** and appears to be well placed to take full ...
 relating to head chefs who have only ever worked in pubs;
 CVs of head chefs **well experienced** in take-aways
 I mean as you as you say, Nick's a long serving and **well experienced** reliable

To investigate whether there is any difference in the behavior between *experienced* and other adjectives that take *well* as the modifier, a search was carried out on *well* in the BNC and it yielded the following compound adjective: *well qualified*, *well educated*, *well organized*, *well equipped* and *well-known*. To see whether the rare occurrence of *experienced* being pre-modified by the adverb *well* has to do with the semantics of *experienced*, a search was conducted on the modifying adverbs, *highly*, *very*, *poorly* and *badly*. The following are the results of the search on the BNC-online (see Table 2).

Table 2. Adjectives and their most typical adverb modifiers

	well	highly	very	poorly	badly	very well
experienced	8	33	78	0	0	1
qualified	89	86	1	1	0	8
educated	76	39	3	8	2	7
organized	57	39	3	7	6	10
equipped	151	1	0	20	0	11
known	1756	0	0	21	3	52

The figures in Table 2 show that there are several ways in which *experienced* behaves differently from the other five adjectives. First, despite the fact that *well experienced* is found in the BNC, its occurrence is much more restrict than the rest. Second, while there is a large number of instances of *experienced* taking the intensifier *very*, there are very few or no instances of the other five adjectives co-occurring with *very*. Third, these five adjectives, however, take the intensifier *very* when they combine with *well* to form compound adjectives. Fourth, while *educated*, *organized*, *equipped* and *known* can be modified by *poorly* and/or *badly*, *experienced* cannot. These four characteristics suggest that it is likely that *experienced* denotes a positive quality which renders the modification by *well* superfluous and the contradictory modification by *poorly* and *badly* unacceptable. By contrast, except for *qualified*, the other adjectives can be modified by adverbs denoting negative qualities, suggesting that they can be used neutrally, though they commonly denote positive qualities.

6. Implications for language teacher education

In the above discussion, we have seen that teachers often look for generalizations about grammar rules so that they can provide some guidelines to their students. This is perfectly legitimate especially in second/foreign language learning situations where learners do not have the same amount of exposure to the language as in first language learning situations. The problem is whether the rules and generalizations indeed capture how language is actually used rather than how language is perceived to be used, and whether they reflect the dominant patterns of use. The easy accessibility of corpora allows teachers to check prescribed rules and generalizations against linguistic data, it encourages them to be sensitive to patterns that emerge from the data and to make their own interpretations and generalizations of these patterns (see also Hunston, 1995).

Indeed, the constant use of corpus evidence in addressing such questions may help teachers to reflect on their knowledge of the language as well as critically examine grammar rules and patterns that they have always taken for granted. They may begin to look at corpus evidence for answers, instead of just relying on dictionaries and reference grammars. For example, messages like the following have emerged in the post quite frequently:

Hi!

When I was at school, one of my English teachers taught me patterns like

I help him with his English / (to) do something.

I assist him in doing something.

I wonder if it's alright to say, for example, "I helped him doing sth"?

The process of using corpus evidence to investigate questions like the above may be beneficial for the teachers because in this process they will probably notice linguistic patterns and pragmatic loads carried by linguistic items that they might not be aware of, some of which may not be found in typical reference materials. The following is just one example, among many, of how a question from a teacher can lead to interesting discoveries of linguistic facts.

Imply and Infer

These two words cause some confusion because some people use the word *infer* to mean *imply*, as observed by the *Collins Cobuild English Dictionary* (p. 862, entry *infer*). For example,

The police inferred, though they didn't exactly say it, that they found her behaviour rather suspicious.

A teacher posted a question regarding these two words in the *English as a Second Language* list:

Can anyone please provide some examples for me to explain the use of the words *imply* & *infer*?

Thank you very much!

Another teacher responded by saying that *to imply* means "to suggest something indirectly" whereas *to infer* means "to guess something is the case or to conclude".

The explanation provided by the above teacher is very much in agreement with that provided by the *Collins Cobuild English Dictionary* for *infer* and *imply*, which is given below.

If you infer something is the case, you decide that it is true on the basis of information that you already have. *I infer from what she said that you have not been well.*

1. If you imply that something is the case, you say something which indicates that it is the case in an indirect way. *'Are you implying that I have something to do with those attacks?' She asked coldly.*

2. If an event or situation implies that something is the case, it makes you think it likely that it is the case. *Exports in June rose 1.5%, implying that the economy was stronger than many investors had realized.*

By carefully looking at the semantics of *infer* and *imply* in the FROWN corpus I would like to suggest that if the teachers in question had access to corpus data they could not only clarify their doubts in relation to how these items are used (by relying on several examples authentically contextualized), but also note that there are a number of instances where *imply* is modalized, and where the writer makes it clear that what has been said does not imply what the reader has inferred. In fact, out of a total of 61 instances of *imply* found in the FROWN corpus, there are 27 instances of *imply* being modified by either modal verbs, *may*, *might*, *would*, *could* or lexicalized modality, such as *seem to*, *appear to*, *tend to*. For example:

Vaillanc, head of the press office, **seemed to imply** that the next session would be the last and and personal pensions scheme. This **would imply** a lower rate of return on pension investment from a location somewhere in our galaxy **would imply** more than the existence of one other time. The reason is that sometimes the law **will imply** the grant of, amongst other things, a right [f] speaks of vast stretches of time. It **may imply** eternity, but it does not state it in so many thing to say, but a layered structure **could imply** the presence of say sedimentary rocks of some shifts of register suggestive, for they **tend to imply** that we are listening to this or that person's alization of linguistic knowledge, as Long **appears to imply**, or whether it does not also develop preoccupation with his commercial affairs, which **might imply** a neglect of his wife's conjugal and material

There are 18 instances of negative forms of *imply*, such as *does not imply*, *should not be taken to imply*, *need not to imply* and so on. For example,

components onto a Service Contract **does not imply** that

it is installed satisfactorily or to the
and inclusion on the list **does not imply** any formal approval or recommendation by the
Mr Hurd said the draft resolution **would not imply** immediate military action once any deadline
departure of its chief executive **did not imply** any change in its policy, its international
In or Getting Out. [p] Getting Out **may not imply** blowing up the Channel tunnel. But it is
out, however, that such a discovery **need not imply** that life has evolved independently elsewhere
to simplify the analysis and **should not be taken to imply** that the aggregate level of unemployment
This is **not meant to imply** that the details of the interaction are the same in both
ed in terms of a united Europe, **does not necessarily imply** a negative connotation.
It would be **misleading to imply** from the experience of Mazda

A search on the FROWN corpus for instances of *infer* showed that there is a similar tendency for *infer* to be modalized. There are 12 instances of *infer*, out of which nine co-occur with the modals *can*, *would*, *may* and lexicalized modality, such as *tempting* and *tend to* (two are in negative forms).

speed of processing, it does seem **tempting to infer** that the brighter one is, the more efficient
with laboratory rats, Petit **tends to infer** that it does. When animals are given an
the former. Furthermore, **it is not easy to infer** from Bell's writings why he believes that the
1987) makes clear, this **should not lead us to infer** the absence of an ethnic dimension in social
to pull a fast one and been rumbled: **one can infer** that from the fact that the protests forced
s favourite causes. [p] **It would be easy to infer** that the two women rely on each other for
the author's choice of details **may lead us to infer** his or her attitude, but also choice of
a slightly inhuman appearance, **but we may infer** that the spectators accepted such conventions
you were just saying about how **you can** you know **infer** weather patterns from looking say around the

The common characteristics shared by *imply* and *infer* is that both pertain to what is not explicitly stated. Therefore, people tend to hedge statements about implications and inferences with modals or lexicalized modality. The much higher proportion of the negative form of *imply* as compared to that of *infer* suggests that there is a difference in the semantics of the two lexical items. The negative form of *imply* serves to pre-empt possible misinterpretations of what is not directly said. This kind of evidence may help teachers to better understand the difference in meaning between these two items spelled out in the *Collins Cobuild English Dictionary*. These characteristics would not have been easily detected without the help of the corpus and the concordancer showing the environments in which these two words occur.

Studies of applications of corpus linguistics to second/foreign language teaching and learning have emphasized the importance of adopting a data-driven approach to language learning so that learners go through a process of self-discovery (see for example Johns, 1991). The discussion in this paper attempts to show that it might be equally important for teachers to go through this process of self-discovery and to experience formulating generalizations about linguistic patterns that they have observed so that they try to grasp the grammar as much as linguistic researchers do.

References

- Berry, Roger (1994). Using concordance printouts for language awareness training. In C. S. Li, D. Mahoney, & J. Richards (Eds.), *Exploring Second Language Teacher Development* (pp. 195-208). Hong Kong: City University Press.
- Biber, Douglas (1988). *Variation across Speech and Writing*. Cambridge: Cambridge University Press.
- Biber, Douglas;
Conrad, Susan;
Reppen, Rendi. (1994). Corpus-based approaches to issues in applied linguistics. *Applied Linguistics*, 15(2), 169-189.
- Bygate, Martin;
Tonkyn, Alan;
Williams, Eddie
(Eds.). (1994). *Grammar and the Language Teacher*. New York: Prentice Hall.

- Carter, Ronald;
McCarthy, Michael (1997). *Exploring Spoken English*. Cambridge: Cambridge University Press.
- Carter, Ronald;
McCarthy, Michael (2001). Size isn't everything: Spoken English corpus, and the classroom. *TESOL Quarterly*, 35(2), 337-340.
- Collins Cobuild English Grammar* (1990). London: Harper Collins.
- Hawkins, Eric (1999). Foreign language study and language awareness. *Language Awareness*, 8(3/4), 124-142.
- Holmes, Janet (1988). Doubt and certainty in ESL textbooks. *Applied Linguistics*, 9, 21-44.
- Hunston, Susan (1995). Grammar in teacher education: The role of a corpus. *Language Awareness*, 4(1), 15-31.
- Johns, Tim (1991). Should you be persuaded: Two examples of data-driven learning. In T. Johns & P. King (Eds.), *Classroom Concordancing*. *ELR Journal* 4 (pp. 1-16). Birmingham: CELS University of Birmingham.
- Kennedy, Graeme (1998). *An Introduction to Corpus Linguistics*. London: Longman.
- Kjellmer, Goran (1994). *A Dictionary of English Collocations Based on the Brown Corpus*, 3 Vols. Oxford: Clarendon Press.
- Ljung, Magnus (1991). Swedish TEFL meets reality. In S. Johansson & A.-B. Stenstrom (Eds.), *English Computer Corpora* (pp. 245-256). Berlin: Mouton de Gruyter.
- Mindt, Dieter (2000). *An Empirical Grammar of the English Verb System*. Berlin: Cornelsen.
- Partington, Alan (1998). *Patterns and Meanings. Using Corpora for English Language Research and Teaching* [Studies in Corpus Linguistics 2]. Amsterdam: John Benjamins.
- Quirk, Randolph;
Greenbaum, Sydney;
Leech, Geoffrey; Svartvik, Jan (1985). *A Comprehensive Grammar of the English Language*. London: Longman.
- Sinclair, John (1991). *Corpus, Concordance, Collocation*. Oxford, UK: Oxford University Press.
- Tognini-Bonelli, Elena (2001). *Corpus Linguistics at Work* [Studies in Corpus Linguistics 6]. Amsterdam: John Benjamins.