

# ESL Teachers' Questions and Corpus Evidence\*

Amy B. M. Tsui

The University of Hong Kong

In the last ten years, more and more attention has been paid to the importance of raising the language awareness of language teachers. This is an area in which corpus linguistics has a unique contribution to make. With the help of a concordancer, linguistic features that may be overlooked can be made salient and intertextual information that is implicit in a single text can be made explicit. This paper reports on a study of how corpus evidence was used to address questions sent by English language teachers in Hong Kong to a dedicated website. More than one thousand grammar questions sent to the website over a period of eight years were examined. Three types of most frequently asked questions were identified. The paper discusses how corpus evidence was used to help teachers to notice features and patterns which have escaped their attention and to question long-standing assumptions and misconceptions. It shows how subsequent interrogation of corpus data stimulated by teachers' question often led to new insights into linguistic patterns and language use.

**Keywords:** teachers' language awareness, synonyms, collocation, semantic prosody

## 1. Corpus studies in applied linguistics

In the last ten years or so, interest in corpus linguistics has grown exponentially in the field of applied linguistics. One major application of corpus studies is the use of corpus evidence as a basis for making decisions about the goals and the curriculum of ESL and EFL teaching. A number of comparative studies have been conducted on the frequency of occurrence and the semantic scope of lexical and grammatical items in textbooks and in corpora. These studies show

that many linguistic items which have a high frequency of occurrence in corpora are often given little pedagogical attention in ESL and EFL curricula and textbooks (see for example Biber, Conrad & Reppen 1994; Grabowski & Mindt 1995; Holmes 1988; Kennedy 1998; Ljung 1991). They suggest that by focusing on the usual rather than the exceptional, ESL and EFL teachers can help learners acquire the target language more efficiently, especially at elementary and intermediate levels.

Another major application is the use of corpora in classroom learning. Corpus data have been integrated into curriculum materials (see for example Carter & McCarthy 1997), and corpus frequency data have been used to automatically generate cloze tests (see for example Coniam 1997). More and more corpora of various sizes are now available for access by learners on the web. To help learners see the relevance of corpus data to their learning, grammar exercises and tests are often included on these sites so that learners can consult corpus data while completing them (see for example Tim Johns' *Data-driving Learning Page* at <http://web.bham.ac.uk/johnstf/> for a list of websites of a similar nature.) As Johns (1991) points out, this kind of data-driven learning approach encourages learners to discover language for themselves and to formulate rules that can account for the patterns that they have observed (see also Tognini-Bonelli 2001).

One area that has not received much attention is the relevance of corpus linguistics to teacher education, particularly in the area of teachers' language awareness (see also Allan 1999; Berry 1994; Hunston 1995). In the last ten years, more and more attention has been paid to the importance of raising teachers' language awareness (see for example the collection of papers in *Language Awareness* 12 (2) (2003); Bygate, Tonkyn, & Williams 1994 and in James & Garrett 1991; Hawkins 1999). This is an area in which corpus linguistics has a unique contribution to make. With the help of a concordancer, linguistic features which have been overlooked can be made salient and intertextual information which is implicit in a single text can be made explicit (see Hunston 1995). Corpus-based investigations can help teachers to reflect on their knowledge of the language and to make their tacit knowledge explicit (see also Barlow 1996). Very often the impetus for teachers to reflect on the language comes from students' questions, particularly when the teachers do not have ready answers to these questions.

This paper reports on a study of lexico-grammatical questions sent by ESL teachers to a website, *TeleNex*. Over one thousand questions sent over a period of eight years were examined. The questions were categorized according

to lexico-grammatical areas. Three areas about which questions were mostly frequently asked were selected for detailed examination. The paper shows how corpus evidence was used by language specialists supporting the website to help teachers to answer students' questions that are often not addressed in reference grammars, to question long-standing assumptions, to gain a better understanding of neglected aspects of language structures and use, and to make explicit their tacit knowledge about the target language. It also shows how further interrogation of corpus data stimulated by teachers' questions led to new insights into linguistic patterns and language use.

To contextualize the discussion in the rest of this paper, I shall provide some brief background information about the *TeleNex* website.

## 2. Teachers of English Language Education Nexus (*TeleNex*)

*TeleNex* is a website set up in 1993 to provide professional support to English language teachers in Hong Kong schools. It is supported by a team of language specialists at the Teachers of English Language Education Centre (TELEC) of the Faculty of Education, The University of Hong Kong. (For a detailed discussion of the design and theoretical motivation of this website, see Tsui 1996; Tsui & Ki 2002). The website includes a conference area in which a number of discussion corners have been set up. It is open for access by the general public. However, in order to provide an anxiety-free environment in which teachers feel free to ask questions in relation to their teaching and exchange ideas, the conference area is open only to school teachers in Hong Kong who have registered as users. One of the corners is the Language Corner (formerly called the Grammar Corner) in which questions about English language are discussed. Teachers send questions to this corner to seek help and advice. The questions are responded to by both school teachers and language specialists in TELEC, some of whom are full-time staff specifically recruited to support the website and some are academic staff in the Faculty of Education. When answering teachers' questions, TELEC staff will refer to corpus data for evidence of language structure and use. Initially, the corpus used for reference consisted of a five-million word native speaker collection, referred to as *Modern English Corpus* (MEC) in this paper, which consists of one million words of spoken texts from radio phone-ins, panel discussions, casual conversations and lecture and two million words of literary and academic texts and two million words from feature articles in the *South China Morning Post*,<sup>1</sup> a major English news-

paper in Hong Kong. These feature articles were written or edited by native speakers of English in Hong Kong. Though the native speaker awareness of the writers or editors may be tempered slightly by local words and wordings, the writers were definitely “expert users” of English (Carter & McCarthy 2001). On-line corpora, such as *CoBuild Direct* and *British National Corpus* (BNC), were also used.

Subsequently, in addition to using commercially available corpora, TELEC built its own corpora, referred to as *TeleCorpora*, to include a much bigger corpus of feature articles from the *South China Morning Post* (20 million words, referred to as *SCMP Corpus*) and a learner corpus. The learner corpus is still being constructed and is around 2.2 million words at present. It consists of a sub-corpus of primary school students’ writing and a sub-corpus of secondary school students’ written and spoken English. *TeleCorpora* is now available for on-line access by registered users of the *TeleNex* website (<http://www.telenex.hku.hk>).

### 3. Teachers’ questions and corpus evidence

The questions sent by teachers to *TeleNex* are either students’ questions that teachers have difficulty answering or teachers’ own queries about lexicogrammatical rules when they come across conflicting linguistic evidence in textbooks or in other published materials. Over one thousand questions sent by teachers to the website during a period of eight years were examined and categorized according to lexico-grammatical areas. Altogether 16 categories were identified (see Table 1).

As shown in Table 1, the most frequently asked questions (225 in total) are questions about lexical items that teachers take to be synonyms or near-synonyms and have problems explaining to students their difference in meaning or usage. The second type relates to linguistic evidence which seems to contradict the grammar rules that teachers were taught as learners. Questions relating to countable and uncountable nouns are among the most frequently asked questions in this type (129 in total). The third type pertains to prescriptive stylistic rules which seem to have been passed on from generation to generation. These questions relate mostly to sentence structures and the use of connectives (147 in total). (For a discussion of other types of questions, see Tsui 2004a<sup>2</sup>). In the ensuing discussion, I shall cite sample messages that teachers sent and discuss how corpus data was used to address teachers’ questions. In

**Table 1.** Categorization of teachers' questions

	Lexico-grammar	No. of Questions
1	Commonly confused expressions / words which are semantically close / meanings of words	225
2	Sentence structure / connectives	147
3	Countable and uncountable nouns	129
4	Prepositions	116
5	Agreement (singular and plural)	99
6	Adverbs	81
7	Adjectives	77
8	Tenses	70
9	Active / passive voice	70
10	Determiners	65
11	Collocation	42
12	Phrasal Verbs	42
13	Statements and Questions	38
14	Modals	35
15	Pronouns	29
16	Conditionals	29
	Total	1294

order to preserve the flavour of the messages, the messages are cited verbatim with no editing. The teachers and TELEC staff cited shall remain anonymous.<sup>3</sup>

### 3.1 Type I: Synonyms

The question of how synonymous synonyms are has plagued many language teachers. They are often asked by students if there is any difference in the meanings and the usage of lexical items which are considered synonymous. Partington (1998) observes that new vocabulary items are often introduced to learners by building on items that they have already learnt. A strategy commonly used by teachers to explain a new word is to provide students with a synonym for the word. In fact, "definition through synonym" is one of the distinctive features of dictionaries (Partington 1998: 29). However, items that are completely synonymous in all their meanings and contexts of occurrence, which Lyons (1981: 149) refers to as "absolute synonyms", are very rare. This is because it would be very inefficient for a language to have words that have identical meaning and can be used interchangeably in all contexts (see also Cruse 1986; Partington 1998). Yet synonyms are often presented in dictionaries and thesauruses as though they are absolute synonyms and important contextual information is often not

included. This is one of the major sources of confusion not only for second language learners but also for teachers.

The following are two questions teachers asked about the synonyms “big” versus “large”, and “finally” versus “lastly”.

### 3.1.1 “Big” versus “large”

Teacher A sent a message asking if there is any rule governing the use of “big” and “large”.

Teacher A:

Hi there,

A student asked my colleague when to use “big” and when to use “large”. I couldn’t give her a definite answer. Is there some kind of rule that we should follow?

In response, Teacher B wrote as follows.

Teacher B:

*Longman Language Activator* writes: ‘big - of greater than average size (slightly more informal than large)’ and ‘large - of greater than average size (slightly more formal than big)’.

The stylistic differentiation in the usage and the mutually defining explanations of these two words provided by the *Longman Language Activator* were probably not very helpful. Subsequently, Teacher C asked Teacher B for further explanations. Instead of providing an explanation, Teacher B told her to look it up in Michael Swan’s *Practical English Usage* (1995) and said that she might be able to find some helpful examples.

The differences between “big” and “large” have been discussed by Biber et al. (1998) on the basis of corpus evidence. Using the *Longman-Lancaster Corpus*, they found that “big” has a much higher frequency of occurrence in fiction (1,235 counts) than in academic texts (84 counts) whereas “large” occurs much more frequently in academic texts (2,342 counts) than in fiction (701 counts). In both fiction and academic prose, “big” is most commonly used to describe physical size. “Large”, on the other hand, is most commonly used in academic prose to describe quantity or amount. In fiction, however, “large” is commonly used to describe physical size. Since there is a much higher frequency of occurrence of “large” in academic texts in which it is used to refer to quantity or amount, one could say that the evidence provided by Biber et al. seems to suggest that the major difference between “big” and “large” is that

the former is used most frequently to describe physical size whereas the latter is most frequently used to describe amount and quantity.

To respond to Teacher A's question, TELEC staff searched the MEC and found that there are several ways in which "large" and "big" are different, some of which accord with Biber et al.'s observation.

Firstly, as Biber et al. observed, there is a strikingly high number of instances of "large" collocating with measurement words. For example,

this on a large	<b>scale,</b>	the the material requirements are so stringent
were large	<b>in size</b>	and low in cost in relation to their size?
to a very large	<b>extent,</b>	<P 2> a subjective thing, people who can be
quite a large	<b>degree,</b>	the time when they prepare their new lecture
sing for large	<b>increase</b>	in the public sector (all talking)... now he

As pointed out by one of the TELEC staff in his response to the teachers, neither the dictionary nor the book by Swan mentions that there is a high frequency of "measurement" words collocating with "large" but not with "big".

Secondly, "large" can be followed by a wide range of expressions of quantity whereas "big" cannot.

to pay large	<b>sums</b>	of money to jump the queue in congested
sometimes large	<b>number</b>	of successive steps. For example
a very large	<b>amount</b>	of time for many complicated elementary
burning large	<b>quantities</b>	of fossil fuels. The build up of carbon
such a large	<b>volume</b>	of replies at such short notice. It may well
to my large	<b>collection</b>	of ties and braces; and will pay for them
cover a large	<b>range</b>	of things like the patients age, the patient

A subsequent search on a much bigger corpus, BNC, confirmed this finding. There is only one instance of "big" collocating with "range" as an expression of quantity in "a big range of sites both in size and location", and one instance of "big" in "a big collection of jazz records."

Thirdly, there seems to be a high frequency of the occurrence of "large" in the context of discussing part-whole relationships. For example,

to donate such a	large	<b>part</b>	of the collection to the academy
represent a	large	<b>percentage</b>	of industrial production."
salaries, housing, a	large	<b>portion</b>	of research and development,
{para}	A large	<b>proportion</b>	of the teaching profession would
building	large	<b>sections</b>	of sea wall in Kowloon and Hong
using a	large	<b>share</b>	of the environmental budget for
, since a	large	<b>slice</b>	of its healthy surplus came from
, of	large	<b>tracts</b>	of forest, even as much as 1 mil

**Table 2.** Frequency of occurrence of *large* and *big* as descriptions of part-whole relationship

Part-whole	large	big
Part in / of	22	3
Percentage of	6	0
Portion(s) of	6	0
Proportion(s) of	14	0
Section(s) of	8	0
Slice of	4	1
Shares of	2	0
Total	62	4

From the concordance lines given immediately above, we can see that “large” is used to indicate the size of a part of the whole. The context in which each of the above lines occurs will make this clear. For example, the following is the context in which the first line occurs.

The decision to donate such a **large part** of the collection to the academy was taken because the band will amalgamate with the Scottish Highland and Lowland bands, which already have much of the music.

The whole is the “collection” and “large” is used to describe the size of the part of the collection donated to the academy.

By contrast, “big” is used much less frequently to describe the size of a part of the whole. Table 2 shows the frequency of occurrence of “big” and “large” in the context of part-whole relationship in the MEC.

As we can see from Table 2, there are 62 instances of “large” compared to four instances of “big” used to indicate part-whole relationships.

The above three differences between “big” and “large”, supported by concordance lines given above, were presented to teachers. Subsequently, stimulated by teachers’ questions, further interrogation was conducted on the *SCMP Corpus* and *BNC*. The corpus evidence suggests that there are further differences between “large” and “big” which throw into question whether the major difference between “big” and “large” is indeed one of physical size versus amount or quantity.

An examination of the contexts in which both “big” and “large” can be used interchangeably shows that both items can be used when they co-occur with concrete noun used literally, as opposed to figuratively. For example,



have the manpower to patrol the	<b>big area,</b>	he said. {para} Mr Yan
If true, this suggests that the	<b>big firms</b>	will step in themselves to
erm pool, you know, not a very	<b>big pool,</b>	I'd like one of those -
two, and erm, I'd make it into a	<b>big room</b>	...Charan stop laughing .
of developing tourism on quite a	<b>big scale,</b>	as they were beginning to
Israeli troops were flattening a	<b>large area</b>	of ground less than four
the lack of a limit would mean	<b>large firms</b>	with more directors
so easily, Mr Fung said. The	<b>large pool,</b>	often dried out and
counselling and interview rooms, a	<b>large room</b>	serving as a drop-in

However, there seems to be a high frequency of abstract nouns collocating with “big” but not with “large”. For example,

effect as 240 volts D.C. The	<b>big advantage</b>	of a transformer is when
another US\$500,000. It's a	<b>big adventure,</b>	isn't it? she exclaims.
ceremony is a very, very	<b>big affair.</b>	The actual ceremony is
all-dancing miner whose	<b>big ambition</b>	is to become a musical
Now this year, there's a	<b>big argument</b>	going on, because in one
which could just be the	<b>big attraction.</b>	{para} This is where the
his legacy, have become a	<b>big bone of contention</b>	in the leadership.
on the land - {B} I have a	<b>big complaint</b>	about English law. it
since Colombia began its	<b>big crackdown</b>	on the cocaine barons six
I believe are the two really	<b>big decisions.</b>	AC In politics we have
walk each day can make a	<b>big difference</b>	to health. <p> From now
And this is one of the	<b>big difficulties.</b>	But I so much agree wit
Bull's absence would be a	<b>big disappointment</b>	for Wolves as his fell
Policy spokeswoman said. The	<b>big discount</b>	is only an illusion. If
hich, of course, was a very	<b>big hazard</b>	with the number of (er)
striking, they want to get a	<b>big impact</b>	from them for every sort
still had a lot of other	<b>big issues</b>	to deal with together.
nuclear war are really the	<b>big worry.</b>	That we really cannot be

In addition, when “big” is used in a figurative sense, it is not interchangeable with “large”. For example,

let us in because there was this	<b>big queue</b>	growing up behind us. He let
dignified reserve and gave me a	<b>big hug.</b>	<p> It had been a remarkable
cinema lately (unless you're a	<b>big fan</b>	of Michael Jackson), with
{para} Section 78 could be a	<b>big headache</b>	because guilty people

A “big hand” is a loud applause, a “big hug” refers to an effusive hug, a “big fan” of somebody refers to having a strong liking and admiration for him or her, and a “big headache” means a severe headache. In other words, the word “big” is used figuratively and its meaning changes when it occurs in a different environment. It is not possible to replace “big” with “large” when the former is used figuratively.

The following instances of “big” are similar to the preceding examples in that it is used non-literally.

that if he did not follow	<b>big brother</b>	he had to pay \$10,800, the
home-grown boy can take on the	<b>big boys.</b>	{para} South African Basil
fuelled by an obsession for	<b>big bucks,</b>	some Chinese are charging
made you someone in the town, a	<b>big fish</b>	in a small pool. Arnold
catching on in Hongkong in a very	<b>big way.</b>	Not only is the Asian

In the above concordance lines, the words that collocate with “big” seem to form an extended unit of meaning (see Sinclair 1996 cited in Tognini-Bonelli 2001: 19) which cannot be decomposed. For example, “big brother” in the context of the concordance line above refers to the ring-leader in a triad society. It is also used in a wider context to refer to an authoritarian person who is in power. The term “big brother” therefore carries a single meaning and cannot be decomposed as “big” and “brother”. Similarly, “big boys”, usually preceded by the definite article “the”, carries a single meaning to refer to a group of people who are at a higher level of skill, violence and usually carries a negative prosody implying menace (Sinclair 1991). “Big fish” cannot be decomposed as “big” and “fish” because a “fish” in a small pool (or pond) carries a fairly different meaning from a “big fish in a small pond”.

The above examples show that “big” has an idiomatic usage (in the sense of collocational restrictions) which “large” does not seem to have. It seems that the “big” is much more “malleable” in the sense that it can be used in a much wider range of contexts than “large”, apart from contexts in which physical sizes are referred to.

The subsequent investigation reveals further insights into the usage of these two lexical items and the pragmatic functions they serve. It has helped us to make explicit our tacit knowledge of the differences between “big” and “large”.

3.1.2 “Finally” versus “Lastly”

Another example of a frequently asked question is whether “finally” and “lastly”, which are often taken to be synonymous by both students and teachers, can be used interchangeably. The following is an extract from the exchanges between teachers.

Teacher D:  
A student asks a teacher, “What is the difference between lastly and finally, can they replace each other?” How should the teacher answer the student?

Teacher E felt that there was no difference and that they could be used interchangeably. She wrote,

Teacher E: I have been asked such a question too. My first impression is that they are more or less the same, so students can use either one when they want to mention the last element in writing or oral. However, I wonder whether “at last” is a common expression when talking about the last element.

Teacher E's impression is correct as far as the commonality between “finally” and “lastly” is concerned. However, it does not cover the difference in usage between the two. Moreover, her question reveals that she was thinking about the semantic scope of listing and that, like the students, she confused “at last” with “lastly”.

Teacher F, who is a native speaker of English, responded to Teacher E by citing the following from Michael Swan's *Practical English Usage*.

- (1) FINALLY and LASTLY have the same meaning when they introduce the last element in a series i.e. Firstly ... Secondly ... Finally/Lastly
- (2) FINALLY (but not LASTLY) can suggest that someone has been waiting a long time for something e.g. He finally passed his driving test (after taking it several times).

Teacher F pointed out that the problem students had was to use “at last” when they meant “finally” or “lastly” in the sense of (1) above. He observed that “at last” meant more or less the same as “finally” in (2) above except that it could occur at the beginning or the end of a sentence. For example, “At last, he passed his driving test”, or “He passed his driving test at last”.

In response to Teacher F's reply, Teacher G sent a message asking for further clarification. She wrote:

If “finally” can mean something has been waited for a long time and finally got done, what about “lastly”? Are there other ways to use “lastly” other than introducing the last element?

In response to Teacher G's question, Teacher F pointed out that “lastly” could only be used to indicate the last item in a series; it could never be used in the same way as “finally” in the context of “He finally returned the money to me.” (i.e., after a long period of time) and “She finally agreed to marry me.” (i.e., after I'd asked her a lot of times.)

Teacher F's native speaker intuition, supported by citation from Swan, is correct, and his explanation of the difference between “finally” and “lastly” is helpful. There are, however, aspects of the linguistic context in which these two

words are found which have not been noted in textbooks or reference grammars and which could not have been detected without going through a corpus. TELEC staff searched the MEC and found that there were far fewer instances of “lastly” (15 instances) than “finally” (441 instances). “Lastly”, and its variant “And lastly”, which is used to introduce the last item in a series, were found to occur only at the beginning of a sentence. For example,

until office instead. {Q}	<b>Lastly</b> , what was the name of that
non-renewable sources. And	<b>lastly</b> , what do we mean by
child’s school perhaps. {I}	<b>Lastly</b> , where do you think
At this late stage, yes.	<b>Lastly</b> , I would like to ask you
rules of intestacy. <p>	<b>Lastly</b> , we may notice that even a
with a moving observer. And	<b>lastly</b> , er, I will be talking
flavour of the honey. {I} And	<b>lastly</b> , is this year going to
spending money in the store.	<b>Lastly</b> , I hope in the future that
followers not G.Best. {O} And	<b>lastly</b> , Geoffrey, is it fair to
police for their attention.	<b>Lastly</b> , I should be grateful if Mr

“Finally” is found more often in the middle of a sentence than at the beginning and the meaning changes depending on its position. When it is found in the middle, it carries a very different meaning (a long duration) from when it is found at the beginning (the last in the series). For example,

be conquered and the war ended	<b>finally</b> in Germany’s favour.
lessons” to the point where he	<b>finally</b> complained to M. Aumont.
less of the people inside, who	<b>finally</b> became quite alarmed”.
the allied operation Overlord	<b>finally</b> commenced, there was a
of cells, then the wrist, and	<b>finally</b> the hand with its digits
leave the hospital, and he	<b>finally</b> asked, and he realized
structures in the wrist, and	<b>finally</b> the digits. Overall, it
best known and most accepted.	<b>Finally</b> , there have been numerous
.....udly replied ‘Four’ ==	<b>Finally</b> , Father McKenna said he had
and Georgie Fame. {para} And,	<b>finally</b> , we hear an amusing story

It was also found that most instances of “lastly” found in MEC occurred in spoken texts whereas most instances of “finally” occurred in written texts.

Subsequently, a search was conducted on a 54 million word corpus, *Corpus Universal Examiner* (CUE), a corpus compiled at the Tuscan Word Centre (TWC), a corpus linguistics research centre directed by John Sinclair, which consists of a 33-million-word corpus of written texts and a 21-million-word corpus of spoken English. The results are provided in Table 3.

The above findings converged with the search on the MEC. They show that “lastly” was used much less frequently than “finally” and that “lastly” was typically sentence initial whereas “finally” was more commonly non-sentence initial.

**Table 3.** Frequency of occurrence of *finally* and *lastly* in CUE

	Written Corpora (33m)	Spoken Corpora (21m)
<i>Lastly</i>		
– Sentence initial	194	21
– Non-sentence initial	3	1
Total	197	22
<i>Finally</i>		
– Sentence initial	1031	54
– Non-sentence initial	2331	606
Total	3362	660

In answering the teachers' questions given above, TELEC staff provided the above sample concordance lines so that they could see the patterns for themselves. Teachers' attention was also drawn to the features identified in corpora analysis mentioned above, namely that the frequency of occurrence, the position of the two words in a sentence, and whether they were commonly found in written or spoken texts.

To see whether the usage of "finally" and "lastly" is a common problem for students, a search was conducted on the 2.2-million-word Learner Corpus in *TeleCorpora*, and it was found that there were 155 instances of "lastly". This is a much higher frequency than that found in the MEC (15 instances). A more detailed examination of these instances showed that students used "lastly" when they should have used "finally", and that they tended to use "lastly" in narratives when they were recounting a sequence of events. For example:

That day, when I was walking on Mei La Road. Since I wanted to go home. I saw a man who carried a big bag. He went out from the Bank fastly. After that, I heard the police bell was ringing. And someone cried 'Help'. I started to think if the man is a robber. **Lastly**, I thought he is a robber, so I write this letter to you.

One possible reason for this misuse is that the distinction between "finally" and "lastly" is not found in their first language. These subtle differences are often a source of difficulty for learners. Another possible reason is that while explanations such as those provided by Swan have highlighted the essential differences between "finally" and "lastly" and are pedagogically defensible, they may not be able to draw teachers' or learners' attention to more subtle differences which are nevertheless important. The use of "lastly" in students' writing above to indicate the last event in a series of events is a case in point.

### 3.2 Type II: Grammar rules and conflicting evidence – countable and uncountable nouns

Teachers often come across linguistic evidence which seem to contradict the grammar rules they have been following. An example is given below of how a teacher tried to apply the rules governing the use of the quantifiers “less” for uncountable nouns and “few” for countable nouns to “no less than” and “no fewer than”.

#### 3.2.1 “Less than” versus “fewer than”

The following is a question from Teacher H on “not less than”.

Teacher H:

In reading instructions for compositions, it is quite often that we will read: Write not less than 300 words. As ‘word’ is a countable noun, why not ‘fewer than’ but ‘less than’?

According to Quirk, Greenbaum, Leech and Svartvik (1985), “few” only co-occurs with plural count nouns. They point out that there is a tendency to use “less” (instead of “fewer”) with count nouns. For example, “You’ve made less mistakes than last time.” This usage however, as Quirk et al. (1985) point out, is often “condemned”. One of the TELEC staff responded to Teacher H as follows

... the distinction between [*less* and *fewer*] is becoming increasingly blurred ... we are exhorted to Use Less Plastic Bags and we also hear people talking about less cars ... a large amount of people etc.!

The distinction between “less” and “fewer” becomes even more blurred when negative forms, “no/not less than”, were interrogated. In the following concordance lines, the use of “no/not fewer than” would seem to be more logical than “no/not less than”.

on the word ‘bank’ to give it	<b>no less than</b> three meanings;
hot meal with two vegetables..	<b>no less than</b> two, sometimes
memorandum requirement that	<b>no less than</b> \$25 billion should be
legislators, it should contain	<b>no less than</b> 400 members from the
passports and equipment for	<b>no less than</b> 5,000 others. {para} A
gains of last year he’d made	<b>no less than</b> 78% altogether. Only
in banking jargon _ is worth	<b>no less than</b> £400m annually. For
forests, in areas receiving	<b>not less than</b> 100 mm of precepitation
retire at 50 or above after	<b>not less than</b> 15 years’ service.
grant extra land to build	<b>not less than</b> 20,000 units for the
must be left with reserves of	<b>not less than</b> \$25 billion while the
a K factor for 1990-91 of	<b>not less than</b> 4.5 per cent, while
in service for a period of	<b>not less than</b> one year. An agency
for a continuous period of	<b>not less than</b> seven years and have

*Collins Cobuild English Grammar* (1990: 126) points out that “less than” is also used “to indicate that a number is a maximum figure and that the actual figure is or may be smaller”. Indeed, the above concordance lines suggest that “no/not less than” is used in contexts where the specified quantity is taken as a total amount, in this case, a minimum quantity.

In response to the teacher’s question, TELEC staff presented the above concordance lines and pointed out that the distinction between “less” and “fewer” is increasingly blurred, particularly in the negative forms “no/not less than”. However, they remarked that “fewer than” is still the typical quantifier for countable nouns.

Subsequently, further interrogation was conducted on “no/not fewer than”. The evidence suggests that in fact “no/not fewer than” also give a sense of total amount rather than a countable quantity, again indicating a minimum quantity.

these same squads each had	<b>no fewer than</b> 10 northern players
for Christmas, a parcel of	<b>no fewer than</b> 12 CDs by The Carpenters,
allocated to each board to	<b>no fewer than</b> 15 and no more than 30.
Prado Museum has contributed	<b>no fewer than</b> 16 paintings, making
system is fully installed.	<b>No fewer than</b> 20 legislative acts will
government, and accompanied by	<b>no fewer than</b> 25 Swedish journalists,
statistics. Jason Nicolle wor	<b>no fewer than</b> 28 rallies in an opening
at York, has examined	<b>no fewer than</b> 990 horror films released
seems to have evolved	<b>no fewer than</b> eleven times <i>
came under severe attack from	<b>no fewer than</b> four different directions.
and June 1941 he broadcast	<b>no fewer than</b> nine major speeches,
is known to have visited Rome	<b>no fewer than</b> seventy-two times. This
work” will be reduced from	<b>not fewer than</b> 24 to not fewer than 16
from not fewer than 24 to	<b>not fewer than</b> 16 hours a week from April
must be in writing, signed by	<b>not fewer than</b> the required number of

What is interesting is that in BNC, “no less than” is twice as frequent (6.03 instances per million words) as opposed to “no fewer than” (2.97 instances per million words) and there are only three instances of “not fewer than” in the entire corpus. This suggests that “less” may be becoming the default quantifier in such contexts.<sup>4</sup>

### 3.3 Type III: Prescriptive stylistic rules

Teachers often wonder whether some of the rules that they were taught as a student are right or whether they are too rigid. One question that was repeatedly asked by teachers in the Language Corner is whether or not conjunctions, such as “because”, “and” and “but”, can be used to start a sentence.

### 3.3.1 *Sentence initial conjunctions*

The following is a question sent by Teacher I.

I have been wondering, every now and then, whether I can use 'because' to start a sentence. I was taught that it's a must to start with 'It is / was because ...', but then the concept was shattered.

This question generated a series of messages in which teachers shared their own learning experiences. Some said that they were taught the same rule by their primary or secondary teachers. Some were taught that starting a sentence with conjunctions like "Because" and "And" was poor style. Others insisted that their students start a sentence with "As" or "Since" when answering questions requiring a reason and avoid "Because" or "It is because". Still others said that they were taught by their university tutors not to start a sentence with "It is because" but to use "This is because" instead.

The following are some of the exchanges between the teachers.

Teacher J wrote:

I never start a sentence with Because. This is a remark given by my teacher when I was in the primary school. She said to start a sentence with Because is poor style. Is this really the reason?

Teacher K responded as follows:

With regard to your question whether we can start a sentence with "because", I have learned this issue from one of my lecturers. He said that sentences start with "because" or "it is/was because" are not good style: "it is/was" only refers to the noun in the previous sentence. Therefore, he suggested that we should use "this is because" in order to give the full picture of the explanation as "this" refers to the whole sentence.

Teacher L disagreed with these prescriptive rules. However, she could not articulate her intuition apart from citing a remark from one of her students that starting a sentence with "because" sounded "boring" and "old".

I was taught not to start sentence with 'Because' when I was in primary school too. But now I don't find it necessary to do so. One of my bright students even said to me that the sentence would become so boring and 'old' with that structure. I kind of agree.

To address questions relating to prescriptive statements about style, corpus evidence is very powerful because it shows how language is actually used rather than how it is perceived to be used (see also Kennedy 1998). The following



concordance lines from the written and spoken texts of MEC were presented to teachers. They show that sentence initial “because” is in fact very common in both written and spoken English.

(Written texts)

**Because** of too many patients we cannot achieve the internationally e  
**Because** the HKPA doesn't have any polo ponies yet or even players for  
**Because** we are interns we have to be responsible for patients admitted  
**Because** it was produced entirely with donated services, all proceeds  
**Because** of Mr Patten's changing travel plan, the programme originally  
**Because** now they will definitely think twice, balancing the pros and

(Spoken texts)

enthusiasm for these sanctions.	<b>Because</b> I'd always like to see that
gonna be too happy. CF Mmm, no. RB	<b>Because</b> they'll have to convert all
Er yes. RB be sufficient. CF Mm. RB	<b>Because</b> , what they've actually got,
sons for that purpose. {JM} No.	<b>Because</b> the majority of people who
numbers. {BR} Cheap... {MTS}	<b>Because</b> even smaller numbers of
t's the word - what do I mean? {I}	<b>Because</b> they get tenure under the
beat because he can do it himself?	<b>Because</b> he can play a guitar, or, or

TELEC staff also pointed out that it is important to analyse the contexts in which “Because” is sentence initial. For example, the first five lines of the concordance data in the written texts actually show that the sentence initial “Because” is used to indicate a cause-effect relationship rather than to provide a reason. In the spoken texts, the sentence-initial “Because” is used in the context of answering a question which asks for reasons.

*TeleCorpora* which, like most other corpora, allow the user to view the linguistic context in which a given word or expression is found, is very useful for teachers who have questions like the following.

Teacher M:

When we are answering questions, can we start with Because?

a: Why didn't you tell me?

b: Because I didn't see you at that time.

Can some native speakers tell me if the answer acceptable?

Or you would probably say, 'I didn't see you at that time.'? Or 'It's because I didn't see you at that time.'?

In response to Teacher M, TELEC staff pointed out that the use of “Because” seems to be much more natural than “It is because” and “This is because” which tend to be used for emphasis. Further corpus data was provided to show teachers that when “It is because” does occur, it is often followed by a that-clause. For example,

**It is because** of this sort of Continental earnestness **that** 'there is in  
**it is because** it is a three-sided plane figure **that** it has them. Proper  
**it is because** something has the form it has, **that** it is the kind of  
**it is because** of that **that** Russia hasn't invaded Europe, is a valid  
**It is because** the State has recognized this fact **that** the law has  
**It is because** Britain will resist that idea in Malaysia **that** Whitehall

### 3.4 Reframing the question

From the above discussion, we can see that the questions posed by teachers are often in the form of "Is it grammatical?" or "Is it acceptable?". In addressing teachers' questions, TELEC staff helped teachers to reframe the question as "Is it appropriate?", "Do people say that?" or "Is it commonly used?" by providing corpus evidence. This kind of reframing reflects a paradigm shift in the approach to linguistic description that has been taking place in the last few decades. This shift is very much facilitated by the accessibility of technology that provides huge bodies of naturally occurring texts. The robustness of linguistic descriptions based on introspection can be easily put to the test by analysing corpus data. The above discussion shows that this kind of linguistic enquiry should not just be the preoccupation of linguistic researchers but should also become an integral part of the work of language teachers. The interrogation of corpus data will raise teachers' awareness of how the language works and enable them to explore the language collaboratively with students.

### 3.5 Linguistic forms and meaning

In the past few decades, there has been an unhealthy dichotomization of form-focused instruction and meaning-focused instruction. Corpus studies have shown that linguistic forms, contexts and meanings are inextricably linked. The corpus evidence presented in the previous discussion demonstrates that the co-occurrence of lexical items in different contexts is crucial to the meanings that they take on and the pragmatic functions that they perform. The engagement of teachers in corpus enquiry will help them to gain a better understanding of the relationship between form and meaning, which can in turn redress the balance between form and meaning in the language curriculum.

### 3.6 Discovering language

Finally, the most exciting dimension of corpus enquiry is the constant discovery of new linguistic facts that have escaped the attention of dictionary and reference grammar writers. Indeed, one of the most rewarding dimensions of

engaging in discussions with teachers on grammar questions in *TeleNex* is precisely this process of discovery. To illustrate this point, I shall cite just one of the many examples. A teacher asked a question about the difference between “based on” and “in view of”.

I know there is a difference between “In view of” and “Based on” ... but are there any examples which show its explicit differences?

It was not difficult to address the teacher's question because “in view of” and “based on” have fairly different meanings. The former means that you take into consideration facts that have just been mentioned or are about to be mentioned, and the latter means that the first thing develops from the second (see *Collins Cobuild English Language Dictionary*). What is interesting is that after going through MEC of *TeleCopora*, it was found that there is a large number of instances of “in view of” where the collocates are unfavorable events, situations or conditions. For example,

for political development in view of **China's disapproval**, another 36.7 line transmissions in view of **growing evidence** they may **cause cancer** In view of how often **overdoses** result from **relationship problems** was not sufficient in view of **Hui Kem-ming's plea of guilty** and Margaret Hospital said in view of Mrs Ng's **persistent abdominal pain** Zealand's **less than dynamic start** to their tour, is one of the **dangers** trafficker. {para} In view of the **quantity of drugs** involved, 2 be as bright as it seems in view of the **current unemployment rate** of over China's response in view of the **airport row** is a factor in as demand dropped. {para} In view of the **shortages in beds** for the proposed amendments in view of the **Government's sweeping powers** please. {JW} John Watson. In view of the **undignified scramble** for {RS} Richard Simpson. In view of the **outburst** in Liverpool Parish Thank you Teresa Lee. In view of the **relentless increase in violence** {BF} Barbara Fearon. In view of the **recent sacking** of the Editor

The above concordance lines show that “in view of” collocates with unpleasant events such as “China's disapproval”, “relationship problems”, “less than dynamic start”, “airport row”, “undignified scramble”, “relentless increase in violence”, “recent sacking”, and so forth. To confirm that the high proportion of negative collocations is not the result of the fact that two-fifths of the MEC consists of newspaper articles which have a propensity for negative news reporting, a further search was conducted on BNC. The findings show a similar tendency for “in view of” to collocate with unfavorable events, such as “stringent financial requirements”, “economic crisis”, “depressed state”, “financial risk”, “atrocious transport situation”. The discovery of the negative semantic prosody of “in view of” would not have been possible without the help of the corpus.

#### 4. Conclusion

In this paper, I have reported on a study of about one thousand grammar questions sent by teachers to the *TeleNex* website. I have focused on three major types of grammar questions which were identified amongst these questions and have demonstrated how corpus evidence was used by language specialists managing the website to address teachers' questions. The discussion in this paper shows that presenting teachers with corpus data is a much more effective way of helping teachers understand the meanings and usages of the linguistic items(s) in question than simply providing them with dictionary definitions or statements about usage. Furthermore, in the course of going through corpus data, existing conceptions are often challenged and new insights are often gained. Sharing such evidence-based insights with teachers encourages them to formulate hypotheses about language and to look for answers by interrogating corpus data. This is a very important aspect of language awareness-raising. Since the website was set up some ten years ago, there have been an increasing number of teachers asking for access to *TeleCorpora* as well as other corpora such as the *Bank of English* (the Collins Cobuild Corpus). As teachers have been exposed to more and more corpus data, they have begun to refer to the corpora as the source of authority rather than TELEC staff. This is evidenced by the fact that increasingly teachers attached remarks like "What does the corpus say?" after putting their questions on the website. In response to requests from teachers to access *TeleCorpora*, a user-friendly concordancer called *PatternFinder* was designed and made available on the *TeleNex* website for teachers. In order to help novices make use of the corpora, a tutorial was built in to provide instructions and demonstrations on how to generate KWIC concordance lines, and how to make use of them to find answers to their questions. A glossary of key terms such as corpus, concordance, KWIC and citation is also provided. (Tsui 2004b discusses the design of *PatternFinder* in greater detail.)

Studies of the relevance of corpus linguistics to language teaching and learning emphasize the importance of actively involving students in formulating their own hypotheses, interrogating linguistic data and generating explicit grammatical statements about the patterns and regularities that they have observed. Johns (1991) proposed that in this kind of data-driven learning, the learner's role is to discover the target language and the teacher's role is to provide the context in which learners can develop strategies for discovery. The discussion in this paper shows that in fact teachers can benefit a great deal from being actively engaged in corpus enquiry, perhaps even more so than their students.

## Notes

\* I wish to thank John Sinclair for going over this paper in great detail and for his comments, which are always encouraging and insightful. I also wish to thank Suzi Nicholson for proof-reading this paper and her editorial advice. Any fault with this paper is of course my own.

1. Features articles are downloaded to TELEC two or three times each week by the *South China Morning Post* (SCMP). I wish to thank the SCMP for their generosity and assistance in providing the texts.

2. The other questions reported in Tsui (2004a) include (a) questions about lexical items that teachers are aware of the difference in usage but have problems explaining the difference to students, for example, "tall" and "high", (b) questions about words and phrases which appear to be synonymous but are not "absolute synonyms" (Lyons 1981; Partington 1998), for example, "day by day" and "day after day"; and (c) collocations.

3. I wish to thank the teachers and the TELEC staff for allowing to cite the messages. I wish to thank especially Quentin Allan who has made a major contribution in compiling *TeleCorpora* and overseeing the construction of the *PatternFinder* in *TeleNex*.

4. I am grateful to my colleague Leo Hoyer for sharing his observation with me and providing the figures for the BNC.

## References

- Allan, Q. (1999). Enhancing the Language Awareness of Hong Kong Teachers through the Corpus Data: The *TeleNex* experience. *Journal of Technology and Teacher Education*, 7 (1), 57–74.
- Barlow, M. (1996). Copora for theory and practice. *International Journal of Corpus Linguistics*, 1 (1), 1–37.
- Berry, R. (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, D., Conrad, S., & Reppen, R. (1994). Corpus-based approaches to issues in applied linguistics. *Applied Linguistics*, 15 (2), 169–189.
- Biber, D., Conrad, S., & Reppen, R. (1998). *Corpus Linguistics*. Cambridge: Cambridge University Press.
- Bygate, M., Tonkyn, A., & Williams, E. (1994). (Eds.) *Grammar and The Language Teacher*. New York: Prentice Hall.
- Carter, R., & McCarthy, M. (2001). Size Isn't Everything: Spoken English Corpus, and the Classroom. *TESOL Quarterly*, 35 (2), 337–340.
- Carter, R., & McCarthy, M. (1997). *Exploring Spoken English*. Cambridge: Cambridge University Press.
- Collins Cobuild English Grammar*. (1990). Sinclair, J. et al. London: Collins.
- Coniam, D. (1997). A preliminary inquiry into using corpus frequency data in the automatic generation of English language cloze tests. *CALICO Journal*, 14 (2–4), 15–33.
- Cruse, D. (1986). *Lexical Semantics*. Cambridge: Cambridge University Press.

- Grabowski, E., & Mindt, D. (1995). A corpus-based learning list of irregular verbs in English. *ICAME Journal*, 19, 5–22.
- Hawkins, E. W. (1999). Foreign Language Study and Language Awareness. *Language Awareness*, 8 (3 & 4), 124–142.
- Holmes, J. (1988). Doubt and certainty in ESL textbooks. *Applied Linguistics*, 9, 21–44.
- Hunston, S. (1995). Grammar in teacher education: The role of a corpus. *Language Awareness*, 4 (1), 15–31.
- James, C., & Garrett, P. (Eds.). (1991). *Language Awareness in the Classroom*. London: Longman.
- Johns, T. (1991). Should you be Persuaded: Two Examples of Data-driven Learning. In T. Johns & P. King (Eds.), *Classroom Concordancing, ELR Journal*, 4, 1–16. CELS University of Birmingham.
- Kennedy, G. (1998). *An Introduction to Corpus Linguistics*. London & New York: Longman.
- Ljung, M. (1991). Swedish TEFL meets reality. In S. Johansson & A. B. Stenström (Eds.), *English Computer Corpora* (pp. 245–256). Berlin: Mouton de Gruyter.
- Lyons, J. (1981). *Language and Linguistics*. Cambridge: Cambridge University Press.
- Partington, A. (1998). *Patterns and Meanings. Using Corpora for English Language Research and Teaching*. SCL 2. Amsterdam and Philadelphia: John Benjamins.
- Quirk, R., Greenbaum, S., Leech, G., & Svartvik, J. (1985). *A Comprehensive Grammar of The English Language*. London: Longman.
- Sinclair, J. (1991). *Corpus Concordance and Collocation*. Oxford: Oxford University Press.
- Sinclair, J. (1996). The Search for Units of Meaning. *Textus*, IX, 75–106.
- Swan, M. (1995). *Practical English Usage*. Oxford: Oxford University Press.
- Tognini-Bonelli, E. (2001). *Corpus Linguistics at Work*. SCL 5. Amsterdam: John Benjamins.
- Tsui, A. B. M. (1996). The participant structures of *TeleNex* – a computer network for ESL teachers. *International Journal of Educational Telecommunications*, 2 (2/3), 171–197.
- Tsui, A. B. M. (2004a). What Teachers Have Always Wanted to Know – and how corpora can help. In John Sinclair (Ed.), *How to Use Corpora in Language Teaching* (pp. 39–61). Amsterdam: John Benjamins.
- Tsui, A. B. M. (2004b). *PatternFinder*: Corpus Evidence and Teachers' Language Awareness. Plenary paper presented at The First National Symposium on Corpus Linguistics and EFL Education, 11–13 October, South China Normal University, Quanzhou, People's Republic of China.
- Tsui, A. B. M. & Ki, W. W. (2002). Socio-psychological dimensions of teacher participation in computer conferencing. *Journal of Information Technology for Teacher Education*, 11 (1), 23–44.

### *Author's address*

Amy B. Tsui  
 Chair Professor  
 Faculty of Education  
 The University of Hong Kong  
 Pokfulam Road, Hong Kong SAR  
 E-mail: bmtsui@hkucc.hku.hk