

# **A Local, Grammatically-Annotated Learner Corpus as A Pedagogical Tool**

CHRISTINE ANITA XAVIER  
*National Institute of Education  
Nanyang Technological University, Singapore*

## **ABSTRACT**

This article argues, through a discussion of the different views of 19 teachers, for the usefulness of a local, grammatically-annotated learner corpus as a pedagogical tool for the teaching of grammar in Singapore schools. Several practicing school teachers attended a professional development grammar course which incorporated grammar exercises developed from some data from this corpus. Based on feedback gathered from these teachers via feedback forms, this article suggests that the authentic, contextualized and accessible nature of the learner corpora in question allows teachers to make pedagogical decisions on what grammar components to teach, when to teach them, and how to teach these in the Singapore school context.

**KEYWORDS:** grammar teaching, grammar learning, learner corpus, pedagogical decisions, pedagogical tools

## Introduction

The usefulness of corpora to language teaching and learning has been amply demonstrated through a wide variety of research spanning a number of areas, including language curriculum planning (e.g. Tsui, 2004; Romer, 2011; Zante & Persiani, 2008), teacher education (e.g. O’Keefe & Farr, 2003; Wong, 2010), and in English as a Second Language (ESL) and English as a Foreign Language (EFL) classes (Lin & Lee, 2015; Quinn, 2015; Zante & Persiani, 2008) to teach language areas like collocations, vocabulary, writing, grammar, English for Academic Purposes (EAP) and English for Specific Purposes (ESP). The research has shown that the use of corpora has contributed to deeper understandings of lexico-grammatical units, collocations and language patterns in general. In much of this research, the corpora used is that of naturally-occurring, native speaker corpora, like the British National Corpus (BNC), which are “large collections of texts (books, newspapers, journals, transcribed speech, etc.), produced by native speakers of English, which are stored electronically and can be accessed using search software” (Gilmore, 2009, p. 365).

This study, however, examines the usefulness of learner corpora, which are systematic computerized collections of texts produced by language learners (Nesselhauf, 2004), as a pedagogical tool for the teaching and learning of grammar. Learner corpora has traditionally been used in Second Language Acquisition (SLA) studies and when used for language teaching and learning, usually compared with comparable samples of native-speaker corpora to show typical errors (e.g. Bernardini, 2004; O’Keefe & Farr 2003). Based on feedback gathered from 19 practicing school teachers who attended a professional development grammar course which incorporated grammar exercises developed from a local, grammatically-annotated learner corpus, this article suggests that the authentic, contextualised and accessible nature of the learner corpus in question allows teachers to make pedagogical decisions on what grammar components to teach, when to teach them and how to teach these in the Singapore school context.

### *Uses of corpora*

Corpora have been used pedagogically in indirect and direct ways (Romer, 2011). Indirect applications help with decisions on *what to teach* and *when to teach it*, having implications on the designing of the teaching syllabus and teaching materials. Direct applications, on the other hand, affect *how something is taught and learned* and actively involve the teachers and learners working directly with the corpora.

### *Indirect applications*

Native-speaker corpora have typically been incorporated in language teaching curricula, providing evidence of the frequent items learners would encounter in actual general or specialised communicative situations as used by native speakers. Such evidence is able to help teachers determine which language items to include in the teaching curricula and syllabi, and even in which sequence the language items should be taught. For example, Wong (2010) discusses how native-speaker corpus data from the BNC have been incorporated into Hong Kong’s New Senior Secondary English language curriculum to teach the use of formulaic expressions and workplace English. Zante and Persiani (2008) discuss how using frequency information from native-speaker corpora like Longman Grammar of Spoken and Written English has been useful in determining the sequence of instruction in ESL teaching.

Learner corpora have also been recognized to provide insights on needs of specific groups of learners (Romer, 2011). However, learner corpora are largely not used in their own right but

in comparison with native-speaker corpora. So while learner corpora have typically been used to study the language development of ESL and EFL learners, this is done by comparing the L2 language output with a similar native-speaker corpus (Granger, 2002). Nesselhauf (2004) suggests that since learner corpora are systematic, and compiled based on a number criteria, any area of learner language may be investigated by comparison to native-speaker corpora, to uncover “over- and underuse (i.e. which features learners use uncommonly often and uncommonly rarely compared to native speakers) ... in addition to mistakes and correct forms” (p. 131). Native-speaker corpora have thus been typically used to investigate whether there are disjunctures between actual language use (as used by non-native speakers of English) and the models of target language use (as used by native speakers of English) propagated in teaching materials.

### ***Direct applications***

In relation to classroom teaching practices, Gilmore (2009) discusses how native-speaker corpora from the BNC and COBUILD concordance and Collocations Sampler have helped to develop students’ writing skills in university EAP classes by getting the students to query these online corpora on how to make their writing more natural. Quinn (2015) also discusses the successful use of native-speaker corpora from Collins Wordbanks Online to support the L2 writing process at the discourse level in an EFL writing course by getting students to refer to corpora in the process of self-correcting teacher-coded errors. The successful use of native-speaker corpora from the BNC in the creation of data-driven learning (DDL) materials and DDL-centred activities to teach grammar to EFL college students is also reported by Lin and Lee (2015).

O’Keefe and Farr (2003) also discuss the use of language corpora in a teacher education programme through the use of corpus-based tasks for increasing students’ understanding of word classes, register-related grammatical choices and socioculturally conditioned grammatical choices. In their study, they discuss the use of both native-speaker corpora like the Cambridge International Corpus and learner corpora that they have built from their own English language teaching classroom corpus. They use their own English language teaching classroom corpus as they feel that

“many local contextualization cues are lost in their reproduction and extraction for third-party analysis operating in far-removed realities. In other words, nonpresent third parties in different educational or cultural surrounds cannot easily capture in their entirety the sociocultural and environmental factors that create and cast the lesson” (p. 392).

They state that the bulk of their teacher education materials available commercially in their Irish context are either British or U.S. produced and this does not match the experiences of their students. However, it must be noted that their students are still encouraged to compare data across corpora, which includes native-speaker corpora. Therefore, as mentioned earlier, there is a tendency not to use learner corpora in their own right.

The studies discussed so far show some ambivalence in the use of learner corpora in pedagogical applications in their own right. When learner corpora are used, there is a tendency to compare learner corpora with samples of native-speaker corpora to show typical errors. The general belief is that it is only in comparisons with the target language of ‘native speakers’ that we are able to discover and address learner difficulties (Nesselhauf, 2004).

This study, however, discusses the use of a local, grammatically-annotated learner corpus as a pedagogical tool for the teaching and learning of grammar, in its own right.

### *The learner corpus in question*

A 271,300 word electronic (longitudinal) corpus, developed from Primary 2 to 6 essays written by the same 233 Singapore primary school children over five years of their primary school years, was comprehensively electronically grammatically-annotated for errors in a research project at the National Institute of Education (NIE) in Singapore (some findings from this research project were reported and discussed in Alsagoff, 2016). This 271,300 word corpus is a subset of a larger corpus originally collected as part of the Singapore Ministry of Education's (MOE) longitudinal Early Literacy Research Study, conducted between 2007 and 2012. The Early Literacy Research Study, in general, allowed for the systematic study of the English language development of Singaporean school children in their primary school years.

The value of this learner corpus lies in the fact that it is composed of writings of Singapore learners of the English language, which is the official language in Singapore. Singapore, which is a multilingual postcolonial society, has a linguistic landscape that is heterogeneous with English as the 'first language' (as opposed to the 'mother tongue languages of Mandarin, Malay and Tamil which are generally referred to as 'second languages') and the primary medium of instruction of all schools. In addition, it plays the utilitarian role of a language for knowledge and technology transfer, for accessing science, technology and global markets, and as a 'neutral' language for the country's various ethnic groups (Alsagoff, 2010). English is also the predominant language at home and at work for most speakers and learners in Singapore (Alsagoff, 2016). There are two different standards of English used in Singapore – the Singapore Colloquial English (SCE) variety which is used in a wide variety of informal domains in Singapore and is heavily influenced by the mother tongue languages in Singapore, and the Standard Singapore English (SSE) variety which is used in institutionalized formal domains in Singapore.

Given then the unique position and role of English in Singapore, adopting a comparative approach against any particular target native-speaker corpora is not straightforward (Alsagoff, 2016) as in Singapore, English is spoken not with one voice but many (Alsagoff, 2010). It is thus debatable whether speakers of English in Singapore should be classified as ESL learners, non-native speakers, native speakers or otherwise. As such, Alsagoff (2016), who examined the error patterns of the same set of grammatically-annotated learner corpus as used in this study, stresses the unique nature and role of English as used in the Singapore context and argues that for any meaningful interpretation of such a local learner corpus, researchers (and here, it is suggested, teachers) need to understand the specific sociolinguistic context of Singapore. In analysing the said learner corpus, Alsagoff (2016) even used a comparative approach "mediated by local norms" (p.116) by referring to a locally produced grammar text that is in line with the MOE's English Language Syllabus 2010 in Singapore.

In order to grammatically-annotate the learner corpus for the NIE research project, a coding scheme was specially developed, to align with the way grammar is described in the MOE's English Language Syllabus 2010 in Singapore, and how it is generally taught to student teachers at the NIE and to students in Singapore schools. In this way, the coding scheme developed was pedagogically oriented. It was also sociolinguistically grounded as it was developed to investigate specific contextual issues with regard to the English used in Singapore - for example, examining whether errors bear similarity to SCE features/structures. The hierarchical coding scheme encompassed 11 broad grammatical categories that covered the major grammatical classes such as noun, verb, preposition and adjective and these categories were further sub-divided into main types of grammatical areas featured in the

English Language Syllabus. The learner corpus was then grammatically-annotated electronically using an online coding platform developed for this project. (For a more detailed discussion of the coding scheme and how the coding was done, see Alsagoff, 2016.)

This study then argues through a discussion of the different views of 19 practicing Singapore teachers, for the usefulness of such a local, grammatically-annotated learner corpus, which provides critical insights into the language development and needs of actual language learners in complex, multilingual Singapore, as a pedagogical tool for the teaching and learning of grammar in Singapore schools.

### **Background to the study**

Some data from this local, grammatically-annotated learner corpus were extracted and developed into in-class grammar analysis exercises that focused on specific patterns of error in a teacher professional development course at the NIE, which is the sole teacher education institute in Singapore. The course, 'Exploring Grammar' was one of the courses run within the Advanced Diploma in Primary English Language Education and Certificate in Primary English Language Education programme, for in-service teachers wishing to upgrade themselves. The 19 programme participants were all local primary school teachers.

One of the researchers in this study was the tutor of the 'Exploring Grammar' course. The 36-hour course, which ran over 12 3-hour sessions, explores the foundations of English grammar, and is intended for teachers who would like to extend their awareness of how the English language functions. It aims to deepen participants' understanding of the basic structure of the English language and develop some basic skills in language analysis for language teaching purposes. References to common errors made by local students are also made to raise participants' awareness of these common errors. Ultimately, the course aims to equip the in-service teachers with greater confidence in teaching grammar - by using appropriate metalanguage for teaching and learning of grammatical items and structures, and analysing and explaining students' grammatical errors.

Grammatical categories and concepts like that of noun phrases, verb phrases, and clause and sentence structures were taught and discussed in-depth in classes supported by a variety of assessment exercises that tested their understanding. Then to ensure that the participants understood and could apply their grammar knowledge, they worked on in-class exercises developed from some data extracted from the grammatically-annotated learner corpus. Sentences from the corpus that displayed specific patterns of errors produced by primary school students, stemming from particular grammatical categories and concepts taught, were extracted and listed as samples of authentic data to analyse and categorise. The participants were asked firstly, to analyse the errors in relation to the content learnt and then secondly, to practise explaining the errors using the content learnt. For example, if a particular session was on Noun Phrase structures, the participants were asked to identify and analyse errors that had to do with Noun Phrase structures and explain the errors according to the grammatical knowledge taught on Noun Phrases during that session. The aim of such exercises was thus to provide opportunities for the participants to demonstrate their applied knowledge of the grammatical categories and concepts learnt in their NIE grammar classes.

## The Study

The course, where the focus is on enhancing grammar content knowledge in current teachers to aid in the teaching of grammar, was thus an appropriate site for the researchers in this study to gather feedback on the ways practicing teachers would find such a local, grammatically-annotated learner corpus pedagogically useful to the teaching and learning of grammar in Singapore schools.

### *The participants*

The 19 programme participants were all local primary school teachers, holding qualifications ranging from Masters to Degrees to Diplomas to Certificates in Education, with teaching experience ranging widely from 2.5 – 33 years, teaching grade levels Primary 1-6.

### *Data collection*

After the final set of in-class grammar analysis exercises was done by the 19 participants, they were asked by the course tutor (who was also one of the researchers in this study) to evaluate the effectiveness of using such a learner corpus as a pedagogical tool for the teaching of grammatical knowledge in a professional development class and, also as a potential pedagogical tool for the teaching and learning of grammar to students. It was explained to the participants that the evaluation was not part of the coursework requirements but for the course tutor's research purposes. The participants were given participant information sheets that provided information on the piece of research being conducted and consent forms to fill. They were also assured that their identities and feedback would be treated confidential with access restricted to the researchers. All 19 participants gave their consent to participate in the evaluation task, which required them to complete a feedback form (a sample of the feedback form is provided in Appendix A).

The main questions asked of the participants in the feedback form dealt with the following:

1. The usefulness/usability of the corpus to the course in helping teachers *learn* about grammar
2. The usefulness/usability of the corpus to the course in helping teachers *teach* grammar
3. Suggestions on how the corpus could be used by practicing teachers as a pedagogical tool in class

This article, however, only discusses the feedback gathered for questions 2 and 3 on the feedback form as the focus of this article is on ascertaining the potential usefulness of a local, grammatically-annotated learner corpora as a pedagogical tool for the teaching of grammar in Singapore schools.

### *Data analysis*

In analysing the data, which is the written feedback on feedback forms, a grounded theory-inspired approach (Dörnyei, 2007; Mills, Birks & Hoare, 2014; Silverman 2014) was employed. This approach to data analysis does “not begin with a prior hypothesis but ... hypotheses [are induced] from close analysis” (Silverman, 2014, p. 119) from the data. The approach generally starts with a close inspection of the data, where segments of the data, be it at word, phrase, sentence or paragraph level, are coded/labelled, based on what the data is ‘telling’ the researcher. This kind of coding allows the researcher to “attach labels to bits of data to distil it” (Charmaz, 2003, p. 3, as cited in Silverman, 2014: 119) to allow for the comparison of data. Based on further analysis and re-analysis of the data with the codes/labels, “[i]nitial coding progresses into intermediate coding as categories start to take shape”

(Mills, Birks & Hoare, 2014, p. 114). Hence, content categories that are linked to the data emerge and this is a re-iterative process (Dörnyei, 2007). This “sequential coding system” and analysis then “produces some theory as an outcome of the investigation” (Dörnyei, 2007, p. 258). Dörnyei (2007) argues that ‘theory’ here does not refer to ‘grand theories’ but rather that the “main point about ‘theory’ is that researchers should go beyond merely describing or categorizing the target phenomenon and should offer some basic understanding of the principles, relationships, causes and/or motives underlying it” (p. 260).

Following such an approach, once the data was collected from the participants, the data was closely read and re-read by the researchers, before the researchers highlighted words, phrases and sentences that held some form of lexical or grammatical relations between them and repeatedly emerged in the data. For example, in the feedback on the corpus as a pedagogical tool, words and phrases like “authentic setting”, “authentic errors”, “true reflections” “realistic representations”, “real situations/data”, “realistic platform” and “authentic data” were repeatedly used by the participants. The words “authentic”, “true” and “realistic”, for example, hold synonymous relations. These words and phrases were coded with the labels “authentic”, “true” and “real” and then later categorised within the category of “authenticity of data”. This process was done for all of the data, using a variety of codes and categories. These codes and categories were discussed and mutually agreed upon by both researchers working on this study. The categories that were derived then led to an understanding of the usefulness of such a local, grammatically-coded learner corpus as a pedagogical tool to the participants in this study.

## Findings and discussion

All 19 course participants provided their feedback on the feedback forms. The participant feedback addressing questions 2 and 3 on the usefulness of the corpus in question as a pedagogical tool seemed to cluster around indirect (*what and when to teach*) and direct (*how to teach*) applications. In analysing the data at the level at which the data was categorised, it was found that the usefulness of the local, grammatically-annotated learner corpora as a pedagogical tool centred, in its own right, around the authentic, contextualised and accessible nature of the corpus.

### *Authentic and contextualised corpus*

#### *Indirect applications*

In order for teachers to make pedagogically-sound decisions in designing curriculum and materials, it is imperative that there is learner awareness. The participant feedback seems to suggest that the authentic and contextualised nature of the corpus in question contributes to this learner awareness. The following are some of the feedback from participants emphasizing how the corpus allows them, as teachers, to know the common errors made by pupils and thus make pedagogical decisions based on these.

- S2: It provides an authentic setting for us to identify the common errors made by the pupils in their writing.
- S6: It helps in identifying the common mistakes made by the pupils. It also enables me to decide if I need to reteach a certain grammar item.
- S7: It informs us of the extent to which the ‘errors’ are prevalent so that we can decide on how much time we should spend on teaching these.
- S17: Real situations/data allow us to be more aware of our pupils’ weaknesses
- S19: The authentic data is very real to us teachers and we can identify with many of the errors presented.

The repeated use of words like “authentic” and “real” to describe the corpus suggests that teachers view the corpus as authentic data that are similar to the kinds of output their own students in the local primary schools they teach in, would produce. The teachers also seemed to stress how the “common” and “prevalent” errors in the corpus would allow them to make appropriate pedagogical decisions like what to teach and how much time to allocate to teaching particular grammar items. This suggests that the teachers identify the errors in the corpus as those similar to what their own students produce. The teachers thus seem to be of the view that the authentic and contextualised nature of the corpus would allow them to better understand their own learners and their language development, and so then make pedagogically-sound decisions in designing curriculum and teaching materials. This finding seems to echo Mendikoetxea, Bielsa and Rollinson’s (2010) hypothesis in their study of the use of a small, locally-produced learner corpus focusing on errors, that “by having a good understanding of learners’ difficulties, teachers and teaching materials can help students ... by fostering language awareness...” (p.181) as with a local learner corpora, we are “able to design pedagogical materials which are more ‘locally’ oriented for learners...” (p. 182). They add that in these ways, “a learner corpus can be used as a powerful pedagogical tool” (p. 181).

### ***Direct applications***

The feedback also suggests that such authentic and contextualised corpus would allow for “learner-aware teaching” (Granger, 2002: 14). Some of the comments are as follows:

S3: Teachers will be more aware of the mistakes made and we can address these areas in our teachings.

S12: I can use the data to highlight the common mistakes to my students in future when I am teaching grammar and also more aware of the mistakes made.

S15: They usually surface the common mistakes that pupils make so it helps us to teach better.

S12: I find them useful. I can use the data to analyse students’ mistakes and I also can use them to highlight the common mistakes to my students in future when I am teaching. The data provide very authentic examples.

Feedback also suggests that not only do such corpora aid teachers in being aware of their learners but also heighten students’ awareness of their own language development and errors. When presented with the corpora, students would feel supported that they are not alone in making similar errors and be more engaged in learning from the mistakes. The following are some of the feedback from the teachers highlighting this:

S1: The pupils would also relate well to the similar mistakes.

S3: The patterns would be useful for me to show pupils that they are not alone – thus increasing their self-confidence. Pupils can then learn from others’ mistakes.

S13: As they are real and the students will be able to relate to those mistakes. They will have ownership to those mistakes and will be more engaged in learning.

According to Romer (2011), such a data-driven approach can empower learners to “to find out things for themselves, ... rais[ing] the learners’ language awareness...” (p.213). Seidlhofer (2002) also highlights that “current thinking advocates learning by discovery, which encourages learners to bring their inductive abilities to bear on real data” (p. 214)

The findings thus reveal that the teachers believe that an authentic and contextualized corpus is a beneficial pedagogical tool to make pedagogically-sound decisions in curriculum and material design, and to complement the teaching of grammar, which includes the corpus as a



self-directed learning tool for students. According to Mishan (2004), authenticity as an attribute is relevant and essential to language teaching. Authentic texts are rich “in terms of their cultural and linguistic content”, such texts provide the opportunities “to select materials that are relevant, appropriate, and interesting to particular groups of learners” and there are “the motivational aspects of learning from authentic rather than didactic material, and so on” (Mishan, 2004, p.219). Wong (2010) too stresses that “[a]uthenticity is vital to language teaching as it is always regarded as a motivating force for learners.... [as opposed to] contrived, isolated texts”(p.3).

Tan (2005) also argues that corpora research applied to language teaching should take into account how “the social and cultural practices of learners’ local contexts could influence [learners’] use of English” (127) as specific patterns of errors identified in the learner language may then in fact be attributable to users’ social and cultural identities and not classified as inauthentic target language behavior. In the same vein, Alsagoff (2016) argued that for any meaningful interpretation of such local learner corpus, researchers need to understand the specific sociolinguistic context of Singapore. It is also argued here that for any corpora to be pedagogically useful for teachers, the corpora needs to be authentic and contextualised within the sociolinguistic context of Singapore and so, relevant to the teachers, who have also been trained to become English language teachers within such a sociolinguistic context and are also familiar with the local educational system. Such a learner corpus, reflecting the language use of students in complex, multilingual Singapore, is certainly valued as a pedagogical tool in its own right based on the feedback from the teachers.

### ***Accessibility of corpus***

#### ***Direct applications***

Participants also provided feedback that the ease of exploring an electronic grammatically-annotated learner corpus would facilitate practicing teachers in investigating patterns of errors in student writing. This is because they would have ready access to identified common errors at various grade levels, and so be able to achieve greater focus in their remediation by designing curriculum and teaching materials that would cover relevant grammatical knowledge needed for their students’ language development. Some of the feedback included the following:

S3: They have been categorized for easy usage and identification

S2: More like a corpus where we can search for errors based on ‘phrases’ and their frequency of use

S6: It helps in identifying the common mistakes made by the pupils.

S5: Data can be uploaded onto a Data Bank e.g. common errors made by P4, P5 and P6

S14: The errors can be consolidated into genres. As such, the teachers can refer to the common errors of each genre when teaching the grammar to students.

#### ***Indirect applications***

The feedback also suggests that the accessibility of the corpus in terms of the ease of searching for common errors and frequencies of use would help with classroom teaching. Some comments are as follows:

S14: It help to classify all the errors into groups to help teach the students.

S16: These errors can be compiled for pupils to be made aware of and the ‘correct’ concepts can be taught to them

Phrases used by the participants like “categorized for easy usage and identification”, “can search for errors”, and “helps in identifying the common mistakes” stress the importance of the ease of searching for specific grammatical categories and common groups of errors. This would allow teachers to easily search for relevant data instead of manually searching for them, and in turn facilitating them in making sound pedagogical decisions in curriculum and material design, and classroom teaching.

The findings thus reveal that the authentic, contextualized and accessible nature of the local, grammatically-annotated learner corpus is certainly a valuable pedagogical tool in its own right that allows teachers to make pedagogical decisions on what grammar components to teach, when to teach them and how to teach these in the Singapore school context.

### **Challenges and the way forward**

It must be noted though that while the teachers recognize the value of such a local, grammatically-annotated learner corpus as a pedagogical tool based on the authentic, contextualised and accessible nature of the corpus, the teachers also added in the feedback form that they would need more institutional support and guidance to use such a corpus tool and data. They recommended having workshops and resources for teachers to guide them in the use of such corpora, and the provision of corrected answers to the errors in the grammatically-annotated corpus. This suggests that implementing any use of such an electronic, grammatically-coded corpus as a pedagogical tool in schools with teachers and students may not be without its challenges and here, two challenges are highlighted. Firstly, the corpus is an electronic tool and teachers need to be trained in how to use such an electronic tool, in for example, how to retrieve certain required data following certain error patterns. Secondly, using such grammatically-annotated data as a pedagogical tool assumes that teachers are competent and confident in their knowledge about grammar. If the teachers are not confident and competent in such knowledge, they would certainly have to enhance their grammar content knowledge by attending content upgrading courses. So while such a local, grammatically-annotated learner corpus is indeed a valuable pedagogical tool in its own right as attested by the teachers’ feedback, a realistic implementation of the use of such a corpus in the classroom would entail that teachers be provided with institutional support in using such corpora pedagogically.

### **Acknowledgement**

The author would like to express gratitude to Dr Lubna Alsagoff for her invaluable advice on previous drafts, and to Dr Peter Teo for his helpful suggestions in the revision of this paper.

## References

- Alsagoff, L. (2010). English in Singapore: Culture, capital and identity in linguistic variation. *World Englishes*, 29(3), 336-348.
- Alsagoff, L. (2016). Interpreting error patterns in a longitudinal primary school corpus of writing. *The Asian Journal of Applied Linguistics*, 3(1), 114-124.
- Bernardini, S. (2004). In the classroom: Corpora in the classroom. In Sinclair, J.M. (Ed.), *How to Use Corpora in Language Teaching* (pp. 15-36). Philadelphia: John Benjamins.
- Dörnyei, Zoltán. (2007). *Research Methods in Applied Linguistics*. Oxford: Oxford University Press.
- Gilmore, A. (2009). Using online corpora to develop students' writing skills. *ELT Journal*, 63(4), 363-372.
- Granger, S. (2002). A bird's-eye view of learner corpus research. In S. Granger, J. Hung, & S. Petch-Tyson (Eds.), *Computer Learner Corpora, Second Language Acquisition and Foreign Language Teaching* (pp. 3-33). Philadelphia/Amsterdam: John Benjamins.
- Lin, M. H., & Lee, J-Y. (2015). Data-driven learning: Changing the teaching of grammar in EFL classes. *ELT Journal*, 69(3) 264-274.
- Mendikoetxea, A., Bielsa, S. M., & Rollinson, P. (2010). Focus on errors: Learner corpora as pedagogical tools. In M. C Campoy-Cubillo, B. Belles-Fortunato, & M. L. Gea-Valor (Eds.), *Corpus-Based Approaches to English Language Teaching* (180-194). New York: Continuum.
- Mills, J., Birks, M., & Hoare, K. (2014). Grounded theory. In J. Mills, & M. Birks (Eds.), *Qualitative Methodology: A Practical Guide* (pp. 107-121). London: SAGE.
- Mishan, F. (2004). Authenticating corpora for language learning: a problem and its resolution. *ELT Journal*, 58(3), 219-227.
- Nesselhauf, N. (2004). Learner corpora and their potential for language teaching. In Sinclair, J.M. (Ed.), *How to Use Corpora in Language Teaching* (pp. 125-152). Philadelphia: John Benjamins.
- O'Keefe, A., & Farr, F. (2003). Using language corpora in initial teacher education: Pedagogic Issues and practical applications. *TESOL Quarterly*, 37(3), 389-418.
- Quinn, C. (2015). Training L2 writers to reference corpora as a self-correction tool. *ELT Journal*, 69(2), 165-177.
- Romer, U. (2011). Corpus research applications in second language teaching. *Annual Review of Applied Linguistics*, 31, 205-225.
- Seidlhofer, B. (2002). Pedagogy and local learner corpora: working with learner-driven data. In S. Granger, J. Hung, & S. Petch-Tyson (Eds.), *Computer Learner Corpora, Second Language Acquisition and Foreign Language Teaching* (pp. 213-234). Philadelphia/Amsterdam: John Benjamins.
- Silverman, D. (2014). *Interpreting Qualitative Data*. London: SAGE.
- Tan, M. (2005). Authentic language or learner errors? Lessons from a learner corpus. *ELT Journal*, 59(2), 126-134.
- Tsui, A. B. M. (2004). What teachers have always wanted to know and how corpora can help. In Sinclair, J.M. (Ed.), *How to Use Corpora in Language Teaching* (pp. 39-66). Philadelphia: John Benjamins.
- Wong, M. L. Y. (2010). Hong Kong's new senior secondary (NSS) English Language Curriculum: Perspectives from corpus linguistics. *Hong Kong Journal of Applied Linguistics*, 12(2), 1-24.
- Zante, J. V., & Persiani, R. (2008). Using corpus-linguistics findings in the classroom: A rationale with practical applications. *The CATESOL Journal*, 20(1), 95-109.

### Author information

**Dr. Christine Xavier** is a lecturer at the National Institute of Education, Singapore. She teaches both the pre-service and in-service courses, specialising in pedagogical grammar courses. Her research focuses on the intersections between theory and practice in the areas of pedagogical grammar and English language variation and use.

### Appendix A

Please provide information where you are comfortable in doing so, and as much or as little as you wish.

Thank you!

=====

Before we begin, please tell us something about yourself:

- a. Grade levels that you have taught:
  - i. Primary \_\_\_\_\_
  - ii. Secondary \_\_\_\_\_
  - iii. Junior college \_\_\_\_\_
  - iv. Others (please specify) \_\_\_\_\_
- b. Number of years that you have taught at the primary or secondary school level: \_\_\_\_\_
- c. What is your highest qualification:  
\_\_\_\_\_
- d. Subjects you currently teach:  
\_\_\_\_\_
- e. How comfortable are you in teaching grammar? \_\_\_\_\_  
(1 = not comfortable at all → 4 = very comfortable)
- 1. Did you find the use of actual data from primary school children’s writings as a beneficial pedagogical tool to complement and inform the grammatical concepts taught during the workshop? If yes, why? If no, please tell us why? Please include any other feedback or comments.

\_\_\_\_\_

\_\_\_\_\_

\_\_\_\_\_

\_\_\_\_\_

\_\_\_\_\_

\_\_\_\_\_

\_\_\_\_\_

\_\_\_\_\_

\_\_\_\_\_

\_\_\_\_\_

\_\_\_\_\_

\_\_\_\_\_

\_\_\_\_\_

\_\_\_\_\_

\_\_\_\_\_

\_\_\_\_\_

\_\_\_\_\_

\_\_\_\_\_

\_\_\_\_\_

2. As practising school teachers, would you find actual data from school children's writings as a beneficial pedagogical tool to complement the teaching of grammar and to investigate patterns of 'errors' in student writing?  
If yes, why? If no, why? Please include any other feedback or comments.

---

---

---

---

---

---

---

---

---

---

---

---

---

---

---

---

3. Can you provide suggestions as to how such data could be presented and used by practicing teachers as a beneficial pedagogical tool to complement the teaching of grammar and to investigate patterns of 'errors' in student writing?

---

---

---

---

---

---

---

---

---

---

---

---

---

---

---

---

Please tick the box provided if you allow the researchers to use your feedback as data in conference presentations or research papers. All the data collected will be kept under lock and key by the researchers, and all data will be treated as confidential with access restricted to these researchers. Your identity will be kept confidential and not be identifiable in any presentation or publication. No information which collected will be used other than for the purposes and aims of presentations or publications.

If you agree to have your feedback used as data by the researchers, please also provide an email address, should the researchers need to contact you. Thank you.

---