Editorial 02
Advancing Corpus Research for Education 03
Visualizing Teacher Talk 04
Building a Corpus of English Writing 06
More than Meets the Eye or Ear 07
Looking at Language Learning with a Bilingual Lens 08
An English Language Corpus for Asia 09
Annotated Bibliography 10
Research Highlights 11
RESEARCH IN the field of language and language education has traditionally relied on a fine-grained analysis of large amounts of data. Each utterance in the sample had to be recorded, transcribed and annotated, before it could be analysed for meaningful patterns. It was a laborious process.

It is still a demanding endeavour because of the complex nature of language, but the process of discourse analysis has been made easier with the use of modern technology. Searching large amounts of data in a corpus is now more efficient and precise with the application of computers.

This volume focuses on the use of language corpora in education. Traditionally used for linguistics and language-related research, corpus-based techniques and discourse analysis are now being used to analyse and improve classroom teaching and learning.

At NIE, corpora-based research has helped researchers to investigate research questions in different ways. The foremost achievement is the Singapore Corpus of Research in Education (SCoRE)—the largest collection of classroom interactions in the world. Led by Dr Hong Huaiqing, this computerized corpus was designed and built by NIE researchers, and contains data from over 1,000 classroom lessons in Singapore schools.

The rich multimodal data in SCoRE lends itself to multi-level and multi-layered analyses of classroom discourse and pedagogy in Singapore. Dr Paul Doyle’s exploratory project demonstrates its potential for teacher professional development in schools by making the data on teachers’ talk in the classroom visual.

The SCoRE technology has also been applied to other projects. Dr Guo Libo and Dr Rita Silver, both language specialists, have collaborated with the corpus team to better understand students’ writing and classroom learning, in order to improve on the ways we teach the English language. The project by Dr Viniti Vaish takes this one step further, by building a bilingual database of classroom observations in an attempt to enhance the language learning of children whose first language is not English.

My own work in corpus research actually dates back to 1999, when I began work on the NIE Corpus of Spoken Singapore English. It continues to inform teacher education at NIE today. Our research is now going global. Collaborating with the Hong Kong Institute of Education, the project on “English as a Lingua Franca in Asia” will begin to establish if there is an Asian variant of English, and how that compares with English in Europe and elsewhere. It is cutting-edge as far as a corpus collection is concerned.

Development of the work in this area is truly exciting and ground-breaking. We hope this collection of articles on corpus-based research at NIE will excite you too.
NIE BOASTS the largest collection of classroom discourse data in the world. Under the Core Research Programme by the Centre for Research in Pedagogy and Practice (CRPP), researchers went to schools and recorded more than 1,000 lessons. That achievement is certainly something to be proud of. But the question was: What were they to do with so much data?

What Is a Corpus? CRPP’s answer was to turn the data into a corpus—a large collection of data (usually text) used for research purposes. Corpora have been in use for a long time. Dictionary-makers use them to collect examples of how people use a language in daily life.

Corpora are especially useful for research as they allow annotation. This means adding information about the data by tagging it. For example, the dictionary-makers may wish to categorize the content words in their corpus by tagging them as nouns, verbs or adjectives.

Designing a Corpus Turning data from more than 1,000 lessons into a corpus is no mean feat. For that, CRPP turned to Dr Hong Huaqing, a corpus expert, for help.

He and a team of researchers tirelessly transcribed, coded, annotated and compiled all the lessons into a single database. Their efforts resulted in the Singapore Corpus of Research in Education (SCoRE).

While most corpora are lexical, which means they only have texts, SCoRE is multimodal. Besides video- and audio-recorded lessons, it includes teaching materials and students’ assignments gathered from the classes they observed.

Another unique feature is the way the information is linked together via annotations that Huaqing describes as “multi-layered”. “We link the data to look at them not only from one perspective, but from multiple dimensions,” he explains.

Other corpora are usually annotated only for linguistic features, but the SCoRE data are tagged for “linguistic, discourse and pedagogical features”.

Evidence for Best Practice SCoRE is searchable anytime, anywhere. This makes it a data goldmine from which education researchers can glean many insights based on solid evidence.

“If you write a report and have no evidence to give, it’s just an impression, an intuition,” says Huaqing. With SCoRE, a researcher who’s looking at say, students’ engagement in the classroom, can search for and identify patterns of syntax structures and interactions that result in higher engagement.

“At any time, you can show them the examples and say: ‘This is the way and the best practice for how teachers can engage students effectively.”’

An Educational Legacy Huaqing believes that corpora-based research can contribute much to education. He is looking at repackaging the data to train in-service teachers in NIE.

“We have very good evidence to show what can be considered good lessons, not only in terms of language but also pedagogical strategies, like questioning strategies in the classroom.”

SCoRE will be invaluable for those who wish to trace the development of education in Singapore.

“When we’re reading research articles from 10 or 20 years ago about Singapore classrooms, you can only read and imagine what happened in the classroom. You don’t have data to compare with.” That will not be a problem anymore.

SCoRE will also be a reminder of the effort that CRPP researchers have put in to document and improve teaching and learning in Singapore.

“20 years later, even when academics may have forgotten what CRPP has done for Singapore education, you can still find the data in the corpus,” Huaqing points out. “It’s a kind of asset, a heritage.”

Huaqing, a Research Scientist with CRPP, is an expert in designing and using corpora.
A PICTURE paints a thousand words. For Paul Doyle, the opposite of this old adage holds the key to his research—using words to create a picture.

**Investigating Pedagogy** Paul is interested in the ways language is used in the classroom. Much of his research at NIE is based on the Singapore Corpus of Research in Education (SCoRE), a database of classroom lessons from Singapore schools.

But having a database of millions of words isn’t of much use unless it can be understood by the key contributors to this collection—teachers. The next step was to make all that research data useful to them.

An expert in corpus linguistics, Paul wanted to find a way to apply the corpus data to teacher professional development. For him, it boiled down to this question: How do we make what teachers say in the classroom visible?

“The essential aim of this project was to use visualization techniques to prompt teacher reflection on what they do.”

**Making Practice Visible** Paul and his team experimented with various ways to represent “teacher talk”—the things that teachers say in the classroom—in a visual form. “The issue was finding a visualization which would be relatively easy for a teacher to understand at first sight.”

Why visualization? Why not just show the video or the transcript of the lesson?

Traditional approaches to analysing classroom talk require specialised knowledge, such as discourse analysis, in order to make sense of the data. Most teachers are not well versed in this.

It can also be overwhelming to view an hour-long video of teacher teaching. What should one look for? “We needed a way of analysing what was done and representing it in a way that was concise and meaningful,” Paul explains.

**Prompting Teacher Reflection** For Paul, it was essential that the visual would induce teachers to reflect on their practice, to ask why they did what they did and how their teaching could be improved.

He also wanted it to have analytical value for researchers. “I was trying to capture both the analytical aspect of a concordance and the more visual elements semantic networks or webs.”

So they looked at “wordles” (or word clouds), where the font size is proportional to the frequency of occurrence of words or phrases in a text. They explored interactive data visualizations that allow users to change the focus of the analysis by clicking on words.

They also experimented with a kind of scatter plot, which was quite powerful when comparing the types of questioning techniques used, in this case the frequency of open and closed questions in a Math lesson.
By visualizing teacher talk, teachers can begin to explore the pedagogy they use. For example, a quick glance at Figure 1 will tell us that the most common phrase in this sample of classrooms is “I want you to”.

**Improving Pedagogy** They finally arrived at the use of bar charts to represent the use of open and closed questions (Figure 2). Bar charts are familiar to teachers, so they find them easier to ‘read’ than denser visualizations.

Moreover, they are interactive—teachers can zoom in on particular points in time in the lesson and look at what was said. “We want to make it interactive because we want the teacher to be able to manipulate it.”

So after video-taping the lesson, the research team transcribed, annotated and uploaded the transcript to the corpus, and generated a visual representation of it. Paul then met the teacher to discuss and reflect on the lesson.

The team found that seeing is indeed believing. The teacher was able to engage with and critically reflect on her pedagogy from her lesson. “It’s a quick way of focusing on an issue in pedagogy,” notes Paul.

**Breaking New Ground** While still exploratory, Paul believes this research is breaking new ground. “I don’t think people have really looked at this in terms of visualizing text or classroom talk. It’s certainly done in science, but here, I’m not dealing with the use of visualization to model concepts in class.”

He is keen to see how far visualization can be applied to the classroom. “The next step for me is going to the student level. Because what is missing here is, what are the students doing and saying?”

“This brings us back to where I started, the critical issue—what is the best visualization? What will make sense to teachers? I’m still interested in that because I fundamentally believe if researchers are to communicate with teachers, it’s got to be through visual means.”

He is already looking ahead at the possibilities. “I think there are immense potential in visualizing corpus data,” he says of the rapidly developing technology for corpus-based research.

“In the future, for teacher training and teacher professional learning, this is the way to go.”
Building a Corpus of English Writing

**PROJECT TEAM**

**Principal Investigator** Guo Libo, National Institute of Education, Singapore

**Co-Principal Investigator** Hong Huaqing, National Institute of Education, Singapore

**THERE IS** more to students’ written work than just the grade. Dr Guo Libo is building up a corpus of English writing by students—the first in Singapore. Such corpora are usually used only for linguistic research, but Libo wants his to be more than that.

“This corpus has a pedagogical intention,” explains Libo, an Assistant Professor of English Language. Teachers, students, researchers, and even the public can benefit from it.

**Linguistic Profiles** Libo’s team is collecting and analysing writings of students in Primary 6, Secondary 4 or 5, and Junior College Year 2. These levels are considered key stages in the writing development of learners.

“At different stages, we can see what they are capable of doing,” he says. They want to get a better idea of how students progress through the stages.

Such research on students’ writing has been done before in Australia and the UK, but rarely with such a large sample. Four pieces of writing will be collected from about 3,000 students in this project.

“Previously, 30 or 40 student essays were good enough to analyse to generate some patterns. But whether this sample size was representative or not, we couldn’t tell.”

As all the works are graded, the team will explore what an “A” piece of writing looks like and the linguistic features it contains. The “linguistic profile” of learners generated from this analysis will be of great use to language teachers.

**A Corpus for Pedagogy** Using computerized corpus technology, students’ writing can be annotated in multiple ways. “We can analyse the same piece of text using different frameworks and theories, so this will help us from a research point of view.”

Both teachers and students will gain from this. Libo notes, “When a teacher looks at particular pieces of writing by their students, sometimes they base their judgement on their intuition.”

With evidence from the corpus to back up their intuition, teachers can help students to learn better. The findings can help teachers to design learning tasks and to give more specific feedback to students.

“The basis for a teacher’s judgment becomes more visible and tangible. They can then talk about it and share it with the students,” he explains. “This is what we mean by a corpus for pedagogy.”

**Improving Language Learning** Libo is certain that corpus-based research can contribute to better research in language teaching and learning in Singapore.

He wants to compare the texts collected by the team against the English Language syllabus, to see if our students are meeting the curriculum requirements. In this way, policymakers and teachers can set more suitable benchmarks for students.

Libo also sees potential benefits if this corpus can be made accessible to the public. For example, we can compare the writing development of Singapore students with those in other countries with similar corpora, such as Hong Kong or Australia.

“Different researchers can access it for different purposes. Once the infrastructure is created, we can reap the benefits.”

*Libo (centre) wants his corpus to benefit all.*
DR RITA SILVER is a consummate eavesdropper of classroom conversations. To put it more academically, she is a researcher in classroom discourse analysis.

Classroom Discourse Analysis Rita is interested in classroom discourse—that is, how students and teachers talk in the classroom. Specifically, she wants to know what happens when students work together in peer groups.

“Most of what I do is look at the way that teachers and students talk in class, and how that talk either demonstrates their learning or helps them with their learning,” she explains.

“When I say discourse analysis, I don’t mean it in a very technical sense,” clarifies Rita. “What I mean is basically the way that talk is used to accomplish things in the classroom, or the way that the talk and the writing are used.”

They got teachers to carry out peer work activities in the English Language classroom to see how students would take to them. At the same time, they looked at what the teachers thought of the activities.

“We looked at what happened between the teacher and the students in the classroom and how the discourse—or the talk—was used in the classroom to get things done,” she explains.

Corpus-based Technology Classroom discourse analysis is not new. In the days before computers, scholars would record their classroom sessions, write out the transcripts, get out their coloured pens and manually code the data. But those days are over.

Now, with computerized corpus-based tools, analysing large amounts of data is faster and more efficient. By collaborating with the Singapore Corpus of Research in Education (SCoRE) team, Rita used customised annotation schemes to analyse the data.

The lesson-transcripts are uploaded into a shared system and different researchers can analyse the data using different annotation schemes to meet specific research needs. Because the system is able to help search through the data easily, it can pick up things that might have been missed by traditional methods.

Rita is interested in what students and teachers are talking about in the classroom.

And the technology isn’t complicated. “The different corpus tools can help you think about ways to organize your data, to search your data, and to use your data to find out things you want to find out—and it can be about anything.”

Multiplying the Gains According to Rita, a corpus should be planned with a clear purpose in mind. “A good corpus should be a representative sample of data that is connected through some primary purpose. It’s not just going through and coding for a particular pattern, like we used to do with felt markers.”

When you have a good corpus, the technology helps take you that extra mile. “You can keep adding data to it and you can change the analytical framework, but you have to be very thoughtful as a lot of work goes into starting it up.”

Best of all, the data can be widely shared. For Rita’s project on peer work and peer talk, the technical expertise of the SCoRE team allowed the research team to apply a multi-level analysis, in every sense of the word.

Multiple researchers can look at the same transcript at the same time for different things. So while one researcher is looking at how students code-switched from one language to another, another could focus on how certain words were used.

“That allows for different ways of looking at your data, and that is when the corpus becomes really powerful,” says Rita. “Multiple people can add to it. And so you get more access, more data, more opportunities.”
Looking at Language Learning with a Bilingual Lens

PROJECT TEAM

Principal Investigator Viniti Vaish, National Institute of Education, Singapore
Co-Principal Investigator Goh Yeng Seng, National Institute of Education, Singapore

OUR CLASSROOMS today are made up of students from multicultural backgrounds. They bring an assortment of language skills to the classroom.

In an education system where much emphasis is placed on English as the first language, many struggle to keep up. How can we develop their language skills so that they don’t lag too far behind? More importantly, where do we start?

Easing Linguistic Pains Dr Viniti Vaish believes that developing language skills and addressing weak spots at an early stage of learning is important. “Early literacy determines the entire schooling outcome of children,” she says. “Early interventions for children with weak literacy skills in English can greatly improve educational achievement. That is why my focus is on very young learners.”

Focusing on pupils in the Learning Support Programme (LSP), Viniti is studying the extent teachers use the pupil’s first language (L1) to teach the second language (L2).

In Singapore, it is often assumed that English is the first language. In the case of the LSP pupils, their L1 could actually be their mother tongue and English their L2.

Viniti’s research looks at the use of Chinese to teach English, or in the case of Chinese mother tongue classes, the use of English to teach Chinese.

Harnessing Bilingual Resources Pupils in the LSP are often deemed to be academically weak. They are in the programme because of their weak literacy skills, in either L1 or L2.

But Viniti sees these pupils as gifted. “They bring the gift of bilingualism into a very globalized and diverse classroom,” she says, “and I think it is important for teachers to harness this gift.”

She believes that these pupils can be helped by using the language that they know better to help them learn the one they are struggling with. The question is, how?

By looking at classroom examples, Viniti hopes to analyse how teachers use one language to teach another. She also wants to see how pupils’ engagement and motivation levels change when a familiar language is used to enhance their learning.

“If we harness their linguistic resources in the classroom to teach the L2, I think we will achieve optimal biliteracy, which is the goal of our bilingual education programme.”

Teaching Diversely “The focus on Chinese and English is quite exciting because these are languages of enormous cultural and linguistic capital,” explains Viniti.

“In Singapore, we have a very rich site where these two languages are taught to bilingual learners, and I think the international academic community would like to know how Singapore does it.”

To explore how the L1 is used to teach L2, the research team developed a bilingual coding instrument called the Mandarin-English Coding Scheme (ME-CODES). This will be used to code video recordings of the LSP lessons and Chinese mother tongue classes.

“When do they use L1 and how do they use it? The coding system will help us pinpoint the specific activity of speech where we can insert the mother tongue,” explains Viniti.

She hopes the examples they amass can help teachers harness their pupils’ bilingual resources.

“I hope to find aspects of best practices so that we can show teachers on the ground how to use the linguistic resources,” shares Viniti. “This is the change I wish to initiate in terms of teacher education through this project.”
The computer-based corpus allows them to analyse the recordings of local speech, and highlight the areas that need the most attention in teaching pronunciation. They can then address these common problems that affect Singaporean speakers.

An Asian Lingua Franca  The next project Ee Ling is undertaking extends beyond our shores. She is working with Andy Kirkpatrick of the Hong Kong Institute of Education to develop an Asian Corpus of English (ACE).

ACE will include English speech recordings from Hong Kong, Singapore, Malaysia, Vietnam, the Philippines, Brunei, Japan, Australia, and China. They aim to collect at least one million words of informal interactions from Asia.

By analysing English as it is spoken in Asia, Ee Ling says they hope to identify an Asian variety of English which could enhance communication in the region. “We ask what are the features that should be emphasized for the purpose of English as a lingua franca in Asia."

The project also has a European component, the Vienna-Oxford International Corpus of English (VOICE). Work on this has only just begun, but promises to bridge cultures across the East and West.

“It is definitely an international corpus,” says Ee Ling, “so it is getting quite exciting.”

**Corpus-based Teaching**  A key concern for Ee Ling as she continues to understand new varieties of English, such as Singapore English, is the applications of the corpus for teachers.

“Being a teacher educator and an educator at heart, it is to attempt to come up with pedagogical applications based on our empirical research.”

Ee Ling and her colleagues have developed new ways of teaching English pronunciation to NIE student teachers based on the NIECSSE.

**PROJECT TEAM**

**Principal Investigator** Low Ee Ling, National Institute of Education, Singapore
This book is a basic how-to for anyone interested in undertaking a corpus-based project. Paul Baker provides clear explanations of what corpus linguistics is and the advantages of linking it to discourse analysis. His introductory chapter provide useful background and definitions. The book provides information on how to put a corpus together (“corpus building”), which is useful for those who want to construct their own corpus. There is also a section on using a pre-existing and publicly available corpus (see “Using a reference corpus”, pp. 43–44).

Drawing on systemic functional linguistics and its application and development in language education in Australia, this book describes the trajectory of language development in schooling from age 6 to 18. The author shows how successful students expand their meaning-making resources across the years as they work on the curricular tasks of science, history, and language arts. The author also suggests practical ways of incorporating the findings from the study into the classroom activities.

All this chapters of this edited volume based on data from the NIE Corpus of Spoken Singapore English. Included in the volume is a good collection of papers on research into various aspects of Singapore English pronunciation by scholars from Singapore, Japan, Taiwan, Australia, Germany, the UK and USA. These aspects of Singapore English are described in detail under five categories, namely consonants, vowels, suprasegmentals, conversation analysis, and intelligibility.

This book provides a more detailed look at the how-to’s of corpus analysis. The first chapter covers some basic but important questions (e.g., “What is a corpus?”) and provides some historical background on corpus-based studies. There is a full chapter on using existing corpora and another on “DIY corpora”.

After a brief introduction to basic concepts in corpus linguistics, this book moves on to provide specific guidance for language teachers and teacher educators on how to “successfully bring corpora and corpus resources into their language classrooms” (p. xii). Topics discussed include: using corpus studies to inform language teaching, using corpus Internet resources in the classroom, and creating corpora for class use.
Research Highlights

CONGRATULATIONS TO our NIE colleagues whose research projects were approved for funding in the 7th Request for Proposals by the Office of Education Research.

The full list of approved projects is available on the NIE website (www.nie.edu.sg) under Research@NIE.

<table>
<thead>
<tr>
<th>Project No.</th>
<th>Project Title</th>
<th>Principal Investigator</th>
<th>Programme</th>
</tr>
</thead>
<tbody>
<tr>
<td>OER 1/12 LYJ</td>
<td>Hands-on and Minds-on Learning of Science Using a Microbial Fuel Cell</td>
<td>Lee Yew Jin</td>
<td>4</td>
</tr>
<tr>
<td>OER 2/12 TTW</td>
<td>Curriculum Evaluation and Change for a Re-envisioned NIE Chemistry Program</td>
<td>Teo Tang Wee</td>
<td>1</td>
</tr>
<tr>
<td>OER 3/12 KKH</td>
<td>Understanding Singaporean Teachers’ Conceptions of Assessment</td>
<td>Koh Kim Hong</td>
<td>1</td>
</tr>
<tr>
<td>OER 4/12 TCS</td>
<td>Exploring the Dialogic Space in Teaching: A Study of Pre-university Classroom Talk in Singapore</td>
<td>Peter Teo</td>
<td>4</td>
</tr>
<tr>
<td>OER 5/12 TKS</td>
<td>Harnessing Popular Media for Science Learning and Critical Literacy</td>
<td>Tang Kok Sing</td>
<td>4</td>
</tr>
<tr>
<td>OER 6/12 TSS</td>
<td>An Exploratory Study on the Learning Needs and Home Background of Socio-economically Disadvantaged “New Singaporeans”</td>
<td>Teng Siao See</td>
<td>5</td>
</tr>
<tr>
<td>OER 7/12 CYH</td>
<td>Teacher and Student Perceptions of Peer Assessment for English Writing</td>
<td>Cho Young Hoan</td>
<td>4</td>
</tr>
<tr>
<td>OER 8/12 NYY</td>
<td>Norm-referenced, Criterion-referenced, and Self-referenced Feedback: Their Effects on Motivation and Achievement</td>
<td>Nie Youyan</td>
<td>5</td>
</tr>
<tr>
<td>OER 9/12 LYH</td>
<td>Mathematical Progress and Value for Everyone</td>
<td>Leong Yew Hoong</td>
<td>4</td>
</tr>
<tr>
<td>OER 10/12 WJS</td>
<td>Design Study Approach to Teacher Professional Development to Support the Implementation of the Revised 2013 Lower Secondary Science Curriculum</td>
<td>Darren Wong</td>
<td>1</td>
</tr>
<tr>
<td>OER 11/12 HHQ</td>
<td>Integrating Classroom Discourse Corpus for Reflective Practice and Professional Development</td>
<td>Hong Huaqing</td>
<td>4</td>
</tr>
<tr>
<td>OER 12/12 MK</td>
<td>Implicit Social Cognition as a Predictor of Academic Performance</td>
<td>Manu Kapur</td>
<td>5</td>
</tr>
<tr>
<td>OER 13/12 LWS</td>
<td>Assessing Students’ Academic Motivation, Emotion and Strategies</td>
<td>Luo Wenshu</td>
<td>5</td>
</tr>
<tr>
<td>OER 14/12 HS</td>
<td>The Impact of Community-Based Teacher Learning on Student Learning Outcomes</td>
<td>Hairon Salleh</td>
<td>1</td>
</tr>
<tr>
<td>OER 15/12 LJJ</td>
<td>Transition into Inclusive Society for Youths with Special Needs in Singapore: An Enquiry into Transition Outcomes</td>
<td>Li Jen-Yi</td>
<td>4</td>
</tr>
<tr>
<td>OER 16/12 RB</td>
<td>Examining Kindergarten Approximation Skills as a Predictor of Children Requiring Learning Support for Mathematics</td>
<td>Rebecca Bull</td>
<td>5</td>
</tr>
<tr>
<td>OER 17/12 ED</td>
<td>Learning through Popular Music, Lessons for the General Music Programme Syllabus in Singapore</td>
<td>Eugene Dairianathan</td>
<td>4</td>
</tr>
<tr>
<td>OER 18/12 LWO</td>
<td>A Value-mediated Approach to Profiling Professional Learning Communities in Singapore Schools</td>
<td>Lee Wing On</td>
<td>1</td>
</tr>
<tr>
<td>OER 19/12 LJJ</td>
<td>The Development of Confidence</td>
<td>Lee Jihyun</td>
<td>5</td>
</tr>
<tr>
<td>OER 20/12 CYS</td>
<td>Developing Cognitive Diagnostic Assessment for Primary Mathematical Teachers to Enhance Pupils’ Learning</td>
<td>Cheng Yuan Shan</td>
<td>4</td>
</tr>
</tbody>
</table>

Legend

Programme 1: Teacher Learning
Programme 2: Technology in Formal and Informal Learning Environments
Programme 3: Leadership and Organizational Change
Programme 4: Pedagogy and Classroom Practice
Programme 5: Student Motivation, Cognition and Learning Outcomes