3L: Language, Linguistics, Literature® The Southeast Asian Journal of English Language Studies
Vol 29(2), June 2023 http://doi.org/10.17576/3L-2023-2902-08

What a Spoken Learner Corpus Tells Us: Construction and Application of a Pronunciation Programme for English-Language Teachers

CHEN, HSUEH CHU
Department of Linguistics and Modern Language Studies
The Education University of Hong Kong, Hong Kong
hsuehchu@eduhk.hk

ABSTRACT

The linguistic backgrounds of English learners in Hong Kong are highly complex and diverse. There are local Hong Kong learners who speak Cantonese as their native language, new immigrants and students from different dialectal regions in mainland China and non-Chinese learners from the Philippines and Pakistan, among others. To optimise the effectiveness of English pronunciation teaching with such a diverse student body, it is of paramount importance to provide Hong Kong teachers with a spoken learner corpus. This paper introduces a self-developed spoken learner corpus and shows how it can be used in the design of pedagogic applications. The establishment of a learner corpus can help teachers understand the recurrent pronunciation difficulties encountered by learners of particular language backgrounds so that they can customise their teaching materials. To promote pronunciation teaching using this corpus, a teacher training programme was designed and conducted. Pre-service and in-service teachers participated in three face-to-face and real-time streaming workshop sessions, two online sessions, and a lesson plan design competition. Most of the participants appreciated the authenticity of the learner speech data, the diversity of speaker language backgrounds and the effectiveness of using the corpus as a teaching tool. Nonetheless, some suggested that the diversity of reading materials for eliciting speech data could be improved and that more dialect subgroups could be included in the corpus.

Keywords: spoken corpus; pronunciation teaching; phonological features; learner diversity; teacher training

INTRODUCTION

The value of corpus use in language teaching has been generally recognised ever since its emergence in the 1960s (Flowerdew, 2009). Since the 1990s, learner corpora, defined as electronic collections of spoken or written texts created by foreign or second language learners, have garnered significant attention as a vibrant area of study (Granger, 2004). Although there are far more written learner corpora than spoken ones, the number of spoken learner corpora is gradually increasing in Chinese contexts. For instance, the Bilingual Corpus of Chinese English Learners (BICCEL) contains both written and spoken data from Chinese students in National Oral English Tests from 2001 to 2005 (Kolesnikova & González-González, 2016); College Learners' Spoken English Corpus (COLSEC) includes transcribed speech data derived from National Spoken English Test by non-English major students (Yang & Wei, 2005); A Spoken Corpus of the English of Hong Kong and Mainland Chinese learners was established to discover mainland Chinese learners' and local Hong Kong learners' segmental and suprasegmental difficulties in acquiring English pronunciation (Chen & Wang, 2016).

A spoken learner corpus is of great pedagogical value since it enables learners to compare native and non-native speech data, and it could be more effective than native corpora in language classrooms because learners are provided with chances to discover typical and common errors made by learners from specific language backgrounds (Gut, 2005). Due to the diversity of language backgrounds of English learners in Hong Kong, catering for learner diversity should be started by understanding the differences regarding their pronunciation difficulties. Therefore, this paper aims to introduce a self-constructed spoken English corpus of Chinese and non-Chinese learners in Hong Kong and to suggest the pedagogical application of this corpus in English pronunciation teaching. This paper will help teachers, learners, and researchers have a better understanding of the major problems in learning English pronunciation by Hong Kong Cantonese learners, Mainland Chinese learners and Southeast Asian learners in Hong Kong.

LITERATURE REVIEW

Römer (2011) points to three key topics that deserve serious consideration in helping to realise the full pedagogical potential of a corpus: 'focusing on learner and teacher needs, fostering indirect uses of corpora in L2 teaching, and fostering direct uses of corpora in L2 teaching' (p. 214). In the following sections, learners' needs, teachers' needs and forms of corpus use will be reviewed.

LERNERS' NEEDS: OVERCOMING ENGLISH PRONUNCIATION DIFFICULTIES

Learners' needs are at the heart of the teaching and learning process. It is worth bearing in mind that a learners' corpus for educational purposes is designed in such a way that learners derive the greatest benefit from them. The language background of English classrooms in Hong Kong is becoming increasingly diverse. A census by the Hong Kong Government (2018) indicates that ethnic minorities comprise 8% of the Hong Kong population; 6% of the population speak Putonghua and different Chinese dialects as their first language, and over 37% speak Putonghua and Chinese dialects for communication. English learners from different language dialect backgrounds make different pronunciation errors for the same sounds, which can cause teaching difficulties.

According to previous studies, three major pronunciation difficulties are found for vowels. There is a lack of contrast between long vowels and short vowels for learners from the Chinese dialectal areas of Southwestern Mandarin, Central Plains Mandarin, Hakka, Min and Hong Kong (Hung, 2000; Lian, 2013; Qin & Wei, 2008). Diphthongs usually become monophthongs for English speakers from the Wu, Central Plains Mandarin, Lanyin Mandarin and Jiaoliao Mandarin areas (Wen & Zhou, 2014; Zhang & Chen, 2019). Erroneous insertion of a schwa can be observed in different phonological environments: for speakers of Central Plains Mandarin, insertion of a schwa is found after voiceless consonants, and for speakers of Jianghuai Mandarin, schwa insertion occurs after final consonants (Xu, 2016). Consonants such as /θ/, /ð/, /n/, /l/, /f/ and /r/ are problematic for Chinese learners in different ways. English learners from Hong Kong usually pronounce the dental fricative /θ/ as the labiodental fricative /f/ (Deterding et al., 2008; Hung, 2000). Learners from the Central Plains Mandarin and Southwestern Mandarin areas use /s/ and /z/ to replace /θ/ and /ð/, respectively (Xiao, 2014). Learners from the Hsiang, Hakka, Kan and Southwestern Mandarin areas tend to confuse /n/ and /l/ sounds, while learners who speak Jianghuai Mandarin usually replace /l/ with an /n/ sound (Xu, 2016). Moreover, learners from the Jin and Jianghuai Mandarin regions usually pronounce the word final/n/ as back-nasal /ŋ/, whereas Kan speakers pronounce the word final/ŋ/ as /n/ (Xu, 2016). Kan speakers use /f/ to replace the /h/ sound, whereas Min speakers tend to substitute /h/ for an /f/ sound (Lian, 2013). The retroflex consonant /r/ is pronounced in various ways: learners whose mother tongue is Kan, Min or Jianghuai Mandarin tend to pronounce /r/ as /l/ (Lian, 2013; Xu, 2016); Jin speakers usually use

/ʒ/ to replace /r/; and in the Southwestern Mandarin region, learners pronounce the /r/ sound like /z/ (Xiao, 2014). Chen and Wang (2016) examine the relative importance of various prosodic features on language attitudes that native and non-native English listeners hold towards Chinese-accented speech. The results revealed that Chinese speakers have a relatively slow speech rate, produce more stressed words and lack vowel reduction in their English speech compared with native English speakers. When listeners heard long and inappropriate silent pauses in the speech, the integrity rating of the speakers decreased.

Non-Chinese learners of English from the Philippines, India, Malaysia, Pakistan and Nepal also present a variety of pronunciation difficulties. For vowels, Filipino and Malaysian learners lack the contrasts between long and short vowels, while Indian and Pakistani learners usually shift diphthongs to long monophthongs (Jenkins, 2009; Mesthrie, 2008; Tayao, 2008). Filipino learners have trouble pronouncing /æ/, /ʌ/ and /eɪ/, whereas it is difficult for Nepalis to pronounce /ɒ/ and /ə/ accurately (Lesho, 2018; Tayao, 2008). As for consonants, Filipino learners from basilect and mesolect groups and Indian learners tend to produce /p^h/, /t^h/ and /k^h/ unaspiratedly (Tayao, 2008). For /θ/ and /ð/ sounds, Malaysian learners usually add alveolar stops in front of the sounds, as /tθ/ and /dð/ (Jenkins, 2009); learners from Pakistan pronounce /θ/ and /ð/ as dental /t̪/ and /d̪/, while /t/ and /d/ sounds are pronounced as retroflex /ʈ/ and /ɖ/ (Mesthrie, 2008). Li and Chen (2019) further examined the intercultural communication between Hong Kong people and Filipinos in Hong Kong, as Filipinos are the largest non-local ethnic group there. The results show that the major types of features of the Filipino English accent are consonant substitutions, deletion of consonants and consonant clusters, no distinction between long and short vowels, replacement of vowels, and shifting of word stress. These features are also found to be factors which contribute to problems in intelligibility and comprehensibility. To cater to learners from diverse language/dialect backgrounds, there is a need to develop an online, corpus-based English pronunciation platform for both learners and teachers.

TEACHERS' NEEDS: OBTAINING CORPUS-RELATED TRAINING

As computer technology has advanced, an increasing number of studies have focused on pronunciation education to investigate the use of corpora data for pronunciation teaching materials and pronunciation teaching effectiveness (e.g., Gut, 2005). Foreign language syllabi now incorporate computer-assisted language learning. Nonetheless, few teachers avail of a corpus, and some even regard corpus use as something restricted to higher education for research purposes. The lack of use of corpora by language teachers implies, to some extent, a still widespread lack of awareness (Pérez-Paredes, 2019). In other words, teachers may not be fully familiar with corpus use without sufficient support. Therefore, to bridge the gap between theory and practice, teacher training is essential.

It has been suggested that understanding and using a corpus-aided approach can "contribute to the personal and professional growth of language teachers and student teachers" (Ebrahimi & Faghih, 2016, p. 121). Rather than using artificial examples, teachers can improve learners' language by using authentic written and spoken data from corpora, developing teaching materials based on corpora data, and encouraging data-driven learning (DDL) in their classrooms (Hewings, 2012). Ebrahimi and Faghih (2016) conducted a qualitative study in which 32 pre-service teachers were involved in two online courses on the use of a corpus in language teaching. The results showed that the pre-service teachers benefited from the application of a linguistic corpus in their teaching. In addition to training related to the use of a corpus, ready-made corpus-based exercises designed by taking learners' needs into account are also needed (Römer, 2011).

FORMS OF CORPUS USE: DIRECT AND INDIRECT

According to Römer (2011), a corpus can be used in two ways, namely, directly or indirectly. Corpora can be indirectly applied by researchers and materials designers to inform the teaching syllabus, reference works and teaching materials. The Collins COBUILD English Course (Willis & Willis, 1989), considered the most pioneering development, would be a prime example. The course was designed by drawing on 'the commonest words and phrases in English and their meaning' and provided by the COBUILD project (Sinclair, 1987). Corpora can also be directly accessed by learners and teachers as learning and teaching tools. It is worth re-emphasising that data-driven learning is itself, in effect, the direct application of corpora. In recent decades, a number of studies have shown the effectiveness of direct corpus applications. For instance, Karras (2016) studied the influence of DDL on vocabulary acquisition for secondary students and concluded that DDL had a positive effect on vocabulary learning; students who received online-dictionary training and DDL instruction performed better than those who did not receive DDL training. In addition, Wong and Lee (2016) conducted a study on using a parallel corpus of Cantonese and Mandarin Chinese to teach Mandarin-speaking undergraduate Cantonese, and the results showed that students made great progress in Cantonese vocabulary acquisition. A learner corpus is of great pedagogical value since it enables learners to compare native and non-native data, and it could be more effective than native corpora in language classrooms because learners are provided with a chance to discover typical and common errors made by learners from specific language backgrounds (Gut, 2005). Chen and Han (2020) integrated a self-developed spoken learner corpus into a learning platform developed with a variety of resources for Mandarin learners to practice both segmental and suprasegmental aspects of their pronunciation, to discover possible causes and apply remedies for the pronunciation errors found in the corpus. They developed a set of acoustic-based Mandarin tone training materials for Hong Kong speakers using speech data from the corpus and successfully improved the pronunciation accuracy of Mandarin citation tones.

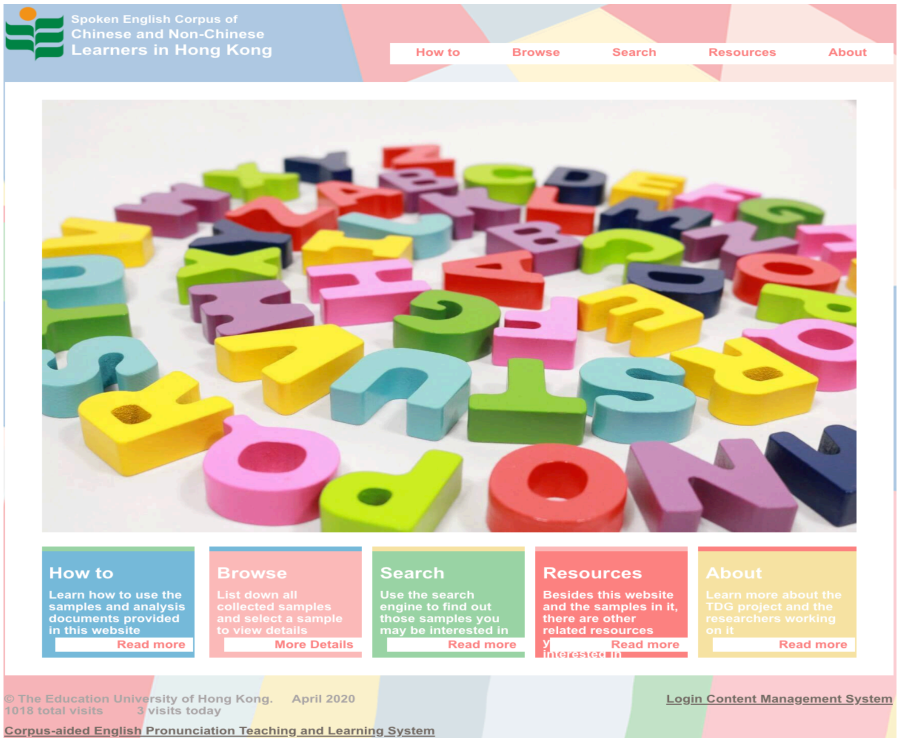
The literature reviewed above sheds light on several issues. First, corpus-aided language teaching can be of great pedagogical value to both teachers and learners, by means of which teachers can develop teaching materials that are more effective, and learners are able to access authentic language data and discern language patterns. Second, spoken learner corpora provide learners with opportunities to compare native and non-native data and thereby discover common pronunciation errors by fellow learners and have cause for reflection on their own learning. Third, English learners with different language/dialect backgrounds exhibit different pronunciation difficulties regarding the same sound. Not enough research has been conducted on the practice of spoken learner corpora in pedagogical applications; however, particularly given the complexity of the language situation in Hong Kong, the English pronunciation difficulties of English learners of different dialects or language backgrounds have not been fully investigated. Therefore, this paper aims to address the issues raised above. This paper will thus be structured by the following four domains:

1) Construction of a spoken English corpus
2) establishment of corpus-aided teaching framework and materials
3) development of the teacher training programme
4) feedback and reflection on the teacher training programme

CONSTRUCTION OF A SPOKEN ENGLISH CORPUS OF CHINESE AND NON-CHINEE LEARNERS IN HONG KONG

The spoken learner corpus introduced in this paper – Spoken English Corpus of Chinese and Non-Chinese Learners in Hong Kong – is an expansion of the previously constructed 'Spoken Corpus of the English of Hong Kong and Mainland Chinese Learners' embedded in a corpus-aided English pronunciation learning platform as introduced in an earlier study (Chen & Wang, 2016). Implementation of the previous corpus has been shown to be effective in improving learners' English pronunciation acquisition in various English and phonetic-related courses in the local university. Nonetheless, in the context of Hong Kong, much of the population come from different provinces in mainland China or are non-Chinese immigrants from countries such as the Philippines, Pakistan and India. Consequently, the 20 sets of annotated spoken data of mainland China speakers in the previous corpus are not sufficient enough to reflect the typical English pronunciation difficulties of different dialect groups in mainland China. Moreover, the pronunciation difficulties of non-Chinese learners in Hong Kong have not yet been investigated. The current corpus was established to fill the gaps and further facilitate the implementation of corpora as teaching tools for language teaching in Hong Kong.

The current corpus is an integral part of the corpus-aided English pronunciation learning platform (https://corpus.eduhk.hk/english_pronunciation/), revamped from the version introduced in Chen and Wang (2016). The cover homepage of the website is shown in Figure 1. New components were added to the learning platform to promote the use of the corpus as a teaching tool: the corpus itself, ready-made corpus-aided pronunciation teaching lesson plans and materials of the corpus-aided teacher training programme. The lesson plans and teacher training programme will be introduced in the discussion section.


Spoken English Corpus of Chinese and Non-Chinese Learners in Hong Kong
How to
Browse
Search
Resources
About

How to
Learn how to use the samples and analysis documents provided in this website

Browse
List down all collected samples and select a sample to view details

Search
Use the search engine to find out those samples you may be interested in

Resources
Besides this website and the samples in it, there are other related resources y interested in

About
Learn more about the TDG project and the researchers working on it

© The Education University of Hong Kong. April 2020
1018 total visits 3 visits today
Corpus-aided English Pronunciation Teaching and Learning System

FIGURE 1. Homepage of the corpus

The corpus contains 136 sets of high-quality recordings, providing annotated speech data collected from 20 Hong Kong learners, 96 mainland China learners (8 speakers for each dialect group [Yueh, Hakka, Hsiang, Kan, Jin, Northern Min, Southern Min and Wu] and four speakers for each Mandarin subgroup [Beijing Mandarin, Central Plains Mandarin, Jianghuai Mandarin, Jiaoliao Mandarin, Jilu Mandarin, Lanyin Mandarin, Southwestern Mandarin and Northeastern Mandarin]) and 20 non-Chinese learners from India, Indonesia, Malaysia, Pakistan and the Philippines. The speakers were university students in local universities aged from 18 to 30. Annotations in the broad transcription of both segmental and suprasegmental errors of the speech data are provided for the users. The main purpose of the corpus is threefold: a) to identify pronunciation problem patterns of English learners in Hong Kong who are local residents, mainland Chinese speakers of different dialects and non-Chinese speakers; b) to present speech data and phonological annotations; and c) to be used as a teaching tool in English as a second/ foreign language (ESL/EFL) classrooms. Figure 2 shows the language and dialect regions of the speakers in the corpus.

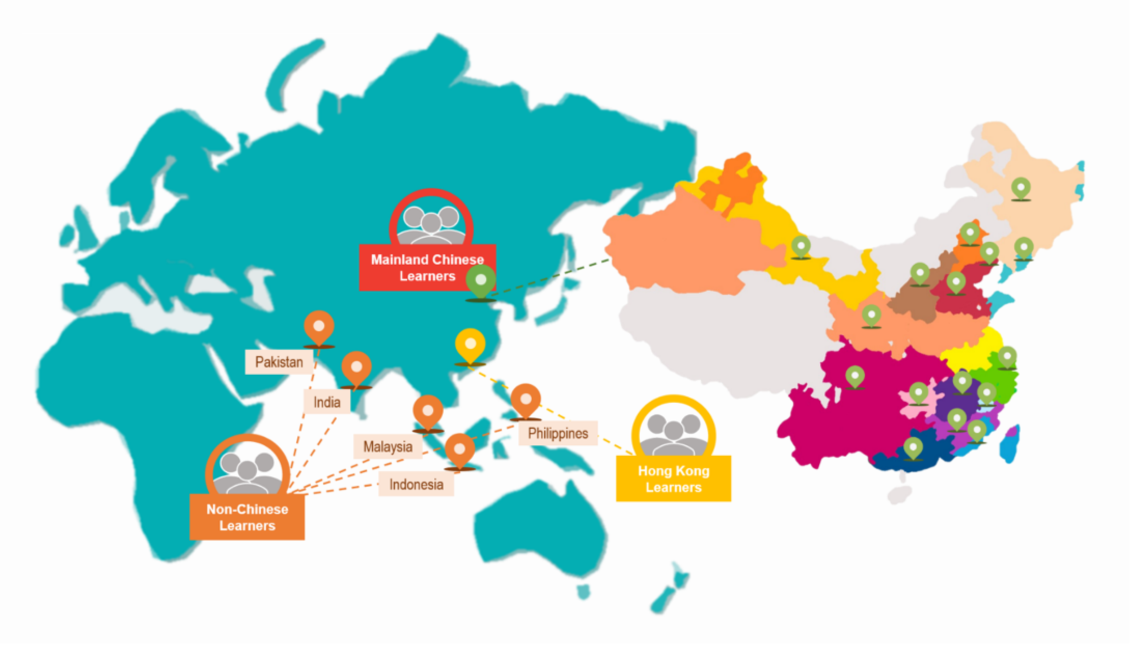
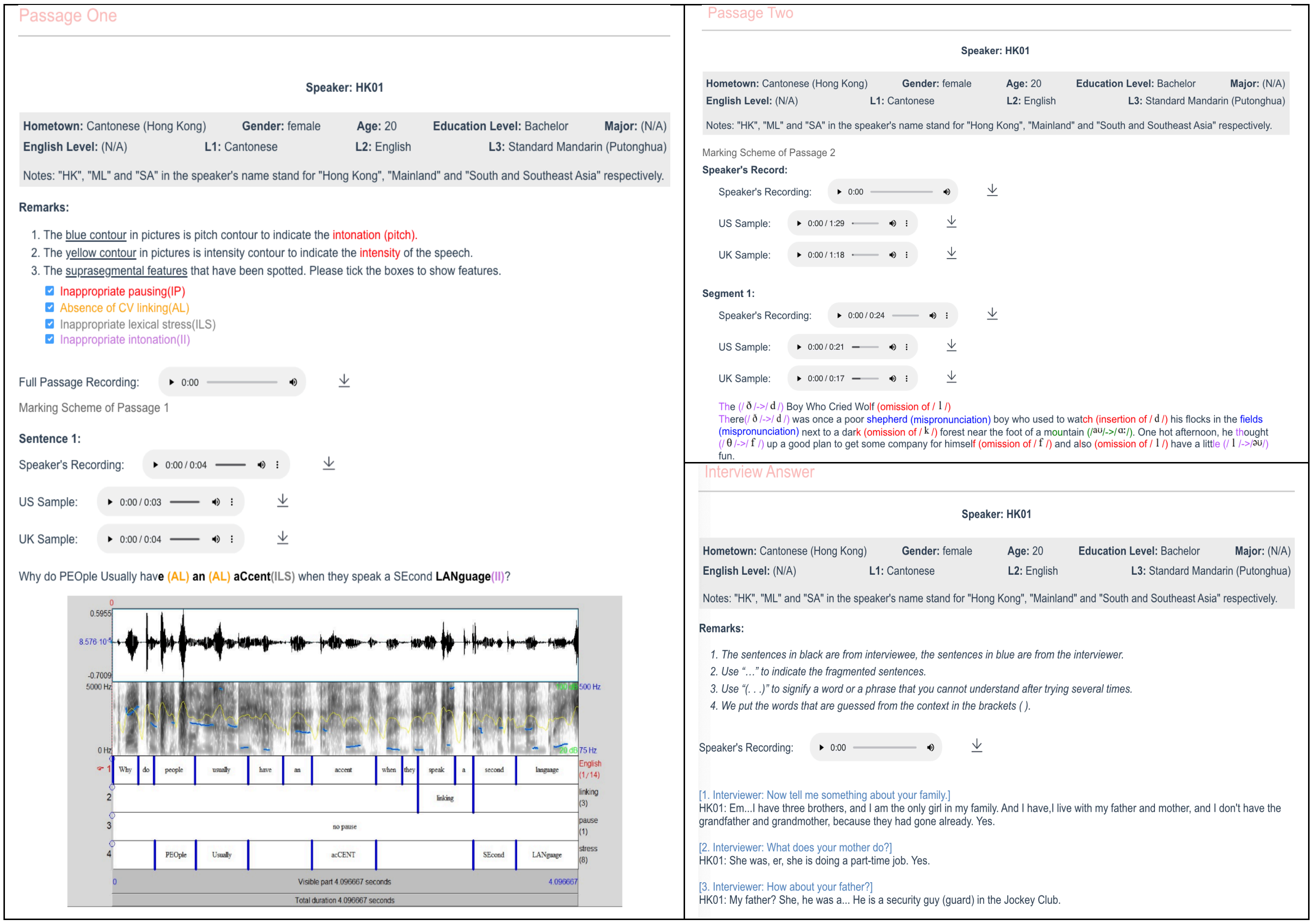


FIGURE 2. Language and dialect regions of speakers in the corpus

The 25 hours of speech data in the corpus were derived from the speakers' completion of three reading/speaking tasks – Passage One: reading of sentences; Passage Two: reading of the story 'The Boy Who Cried Wolf'; and an interview. Three evaluators with high English proficiency and knowledge of phonetics and acoustics were recruited to annotate the speech data. Two of the evaluators processed the annotation of speech data simultaneously, and the third evaluator compared the consistency of their annotations and then finalised the annotations. Annotations for Passage One focused on the suprasegmental properties of the speech data performed using Praat software for five features: the presence of pausing, appropriateness of pausing, presence of consonant and vowel (CV) linking, appropriateness of lexical stress and appropriateness of intonation. Annotations for Passage Two focused solely on segmental properties and the syllable structure changes in speech data, that is, the accuracy of vowels and consonants and the insertion or omission of sounds. The Interview section was transcribed orthographically without phonetic annotation. Figure 3 shows screenshots of the three tasks.



Passage One

Speaker: HK01

Hometown: Cantonese (Hong Kong) **Gender:** female **Age:** 20 **Education Level:** Bachelor **Major:** (N/A)
English Level: (N/A) **L1:** Cantonese **L2:** English **L3:** Standard Mandarin (Putonghua)

Notes: "HK", "ML" and "SA" in the speaker's name stand for "Hong Kong", "Mainland" and "South and Southeast Asia" respectively.

Remarks:

1. The blue contour in pictures is pitch contour to indicate the intonation (pitch).
2. The yellow contour in pictures is intensity contour to indicate the intensity of the speech.
3. The suprasegmental features that have been spotted. Please tick the boxes to show features.

☑ Inappropriate pausing(IP)
☑ Absence of CV linking(AL)
☑ Inappropriate lexical stress(ILS)
☑ Inappropriate intonation(II)

Full Passage Recording: 0:00

Marking Scheme of Passage 1

Sentence 1:

Speaker's Recording: 0:00 / 0:04

US Sample: 0:00 / 0:03

UK Sample: 0:00 / 0:04

Why do PEOple Usually have (AL) an (AL) aCcent(ILS) when they speak a SEcond LANguage(II)?

Passage Two

Speaker: HK01

Hometown: Cantonese (Hong Kong) **Gender:** female **Age:** 20 **Education Level:** Bachelor **Major:** (N/A)
English Level: (N/A) **L1:** Cantonese **L2:** English **L3:** Standard Mandarin (Putonghua)

Notes: "HK", "ML" and "SA" in the speaker's name stand for "Hong Kong", "Mainland" and "South and Southeast Asia" respectively.

Marking Scheme of Passage 2

Speaker's Record:

Speaker's Recording: 0:00

US Sample: 0:00 / 1:29

UK Sample: 0:00 / 1:18

Segment 1:

Speaker's Recording: 0:00 / 0:24

US Sample: 0:00 / 0:21

UK Sample: 0:00 / 0:17

The (/ ð /->/ d /) Boy Who Cried Wolf (omission of / l /)
There(/ ð /->/ d /) was once a poor shepherd (mispronunciation) boy who used to watch (insertion of / d /) his flocks in the fields (mispronunciation) next to a dark (omission of / k /) forest near the foot of a mountain (/aʊ/->/ɑː/). One hot afternoon, he thought (/ θ /->/ f /) up a good plan to get some company for himself (omission of / f /) and also (omission of / l /) have a little (/ l /->/əʊ/) fun.

Interview Answer

Speaker: HK01

Hometown: Cantonese (Hong Kong) **Gender:** female **Age:** 20 **Education Level:** Bachelor **Major:** (N/A)
English Level: (N/A) **L1:** Cantonese **L2:** English **L3:** Standard Mandarin (Putonghua)

Notes: "HK", "ML" and "SA" in the speaker's name stand for "Hong Kong", "Mainland" and "South and Southeast Asia" respectively.

Remarks:

1. *The sentences in black are from interviewee, the sentences in blue are from the interviewer.*
2. *Use "…" to indicate the fragmented sentences.*
3. *Use "(. . .)" to signify a word or a phrase that you cannot understand after trying several times.*
4. *We put the words that are guessed from the context in the brackets ( ).*

Speaker's Recording: 0:00

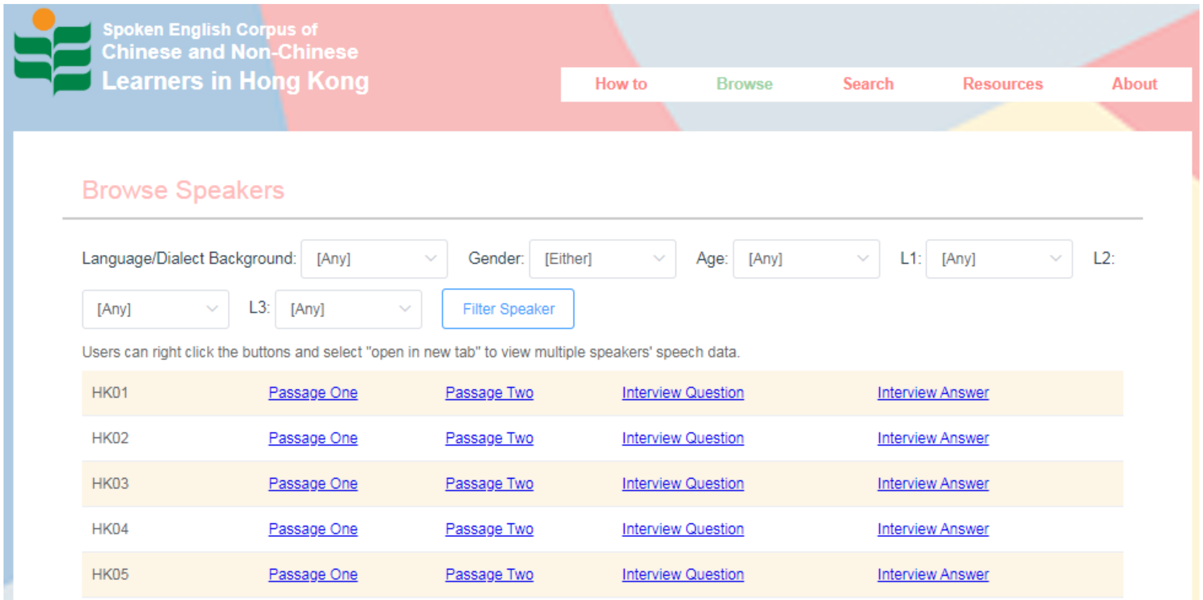
[1. Interviewer: Now tell me something about your family.]
HK01: Em...I have three brothers, and I am the only girl in my family. And I have,I live with my father and mother, and I don't have the grandfather and grandmother, because they had gone already. Yes.

[2. Interviewer: What does your mother do?]
HK01: She was, er, she is doing a part-time job. Yes.

[3. Interviewer: How about your father?]
HK01: My father? She, he was a... He is a security guy (guard) in the Jockey Club.

FIGURE 3. Screenshots of the three tasks

Users are able to access the speech data and annotations by using the 'browse' or 'search' functions. On the browse page of the corpus website, users can find specific speakers by setting filters for language/dialect background, gender, age and L1, L2 or L3, then select the reading/speaking task they would like to view in detail. For Passage One, annotations of different categories of suprasegmental errors can be viewed by clicking the boxes so that the annotations appear, marked in colour. A screenshot of the acoustic analysis for each sentence is provided. Users can click on each annotation label, and a pop-up window will appear with its corresponding analysis highlighted. For Passage Two, annotations are divided into four categories and can also be accessed by clicking the corresponding box for each category: vowels, consonants, syllable change and mispronunciation. Figure 4 shows a screenshot of the browse page.



Browse Speakers

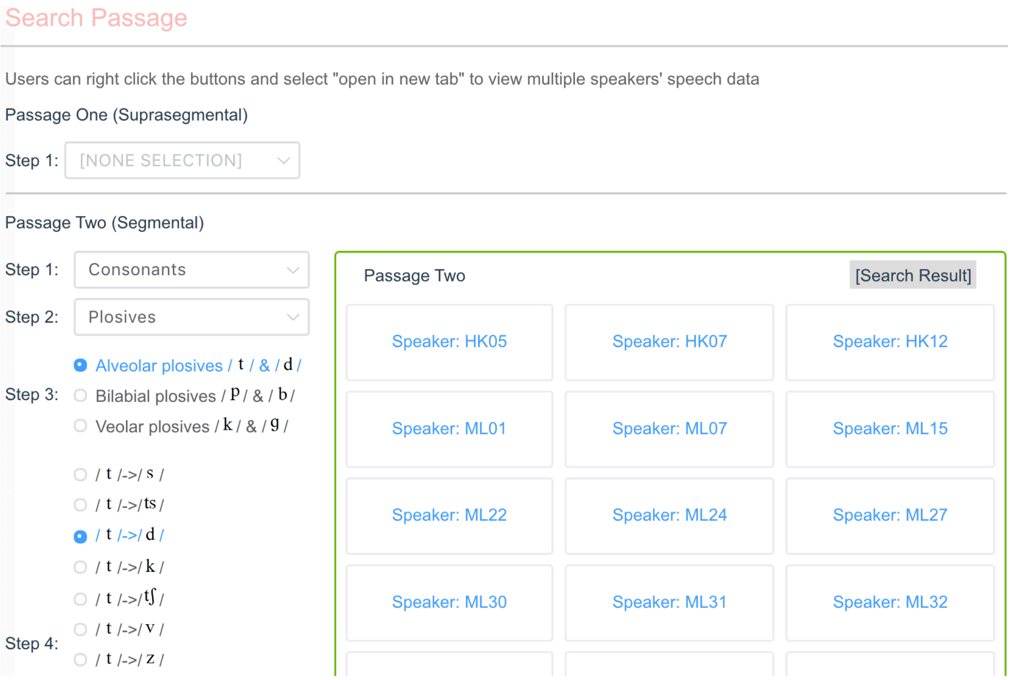
Language/Dialect Background: [Any] Gender: [Either] Age: [Any] L1: [Any] L2: [Any] L3: [Any] Filter Speaker

Users can right click the buttons and select "open in new tab" to view multiple speakers' speech data.

HK01	Passage One	Passage Two	Interview Question	Interview Answer
HK02	Passage One	Passage Two	Interview Question	Interview Answer
HK03	Passage One	Passage Two	Interview Question	Interview Answer
HK04	Passage One	Passage Two	Interview Question	Interview Answer
HK05	Passage One	Passage Two	Interview Question	Interview Answer

FIGURE 4. Screenshot of the browse page

If users are interested in a specific sound or phonological error, they can use the search function. Suprasegmental errors can be searched by specifying pausing, linking, lexical stress or intonation in one step. Segmental errors are specified by four steps due to the complexity of classification. Take the error '/t/ pronounced as /d/', for example. Users can select 'consonants' for step 1, 'plosives' for step 2, 'alveolar plosive /t/ & /d/' for step 3, and '/t/->/d/' for step 4. Figure 5 shows an example of the search function (/t/ pronounced as /d/).



Search Passage

Users can right click the buttons and select "open in new tab" to view multiple speakers' speech data

Passage One (Suprasegmental)

Step 1: [NONE SELECTION]

Passage Two (Segmental)

Step 1: Consonants

Step 2: Plosives

Step 3:
● Alveolar plosives / t / & / d /
○ Bilabial plosives / p / & / b /
○ Veolar plosives / k / & / g /

Step 4:
○ / t /->/ s /
○ / t /->/ts /
● / t /->/ d /
○ / t /->/ k /
○ / t /->/tʃ /
○ / t /->/ v /
○ / t /->/ z /

Passage Two [Search Result]

Speaker: HK05	Speaker: HK07	Speaker: HK12
Speaker: ML01	Speaker: ML07	Speaker: ML15
Speaker: ML22	Speaker: ML24	Speaker: ML27
Speaker: ML30	Speaker: ML31	Speaker: ML32

FIGURE 5. An example of a "search" function (/t/ pronounced as /d/)

On the basis of the annotations, common suprasegmental and segmental errors of Hong Kong, mainland Chinese and non-Chinese speakers are summarised. Most of the language/dialect groups exhibit similar suprasegmental errors, namely, ‘absence of linking’ and ‘inappropriate intonation’. Common segmental errors show strong inter-group differences. All of the language/dialect group present drastically different segmental error patterns from one another (visit https://corpus.eduhk.hk/english_pronunciation/index.php/for-teachers/ for a detailed pronunciation error list).

ESTABLISHMENT OF COPRUS-AIDED TEACHING FRAMEWORK AND MATERIALS

Speech data from the current corpus can be integrated into the corpus-aided pronunciation teaching framework proposed in Figure 6. Common pronunciation errors found in the corpus of learners with a specific language background inform both the learning and instruction of English. Learners have the chance to perform a phonological analysis using speech data from the corpus. They can observe the common pronunciation errors and discover patterns shared by the speakers in the corpus and themselves in the learning process. Teachers can provide explicit articulatory and perceptual explanations for the target sounds and guide learners to perform comparisons between native and learner speech data. To further facilitate learners’ acquisition of the target sounds, in the development of teaching materials, speech data can be flexibly integrated into either meaning-focused activities or form-focused exercises. Meaning-focused activities include information-gap tasks, picture-based storytelling, and role-play. Form-focused exercises are minimal-pairs, read-aloud and listen-and-repeat exercises. With the help of the pronunciation feature list compiled by the author, teachers can convert it into an assessment checklist and provide learners with quality feedback for their pronunciation learning.

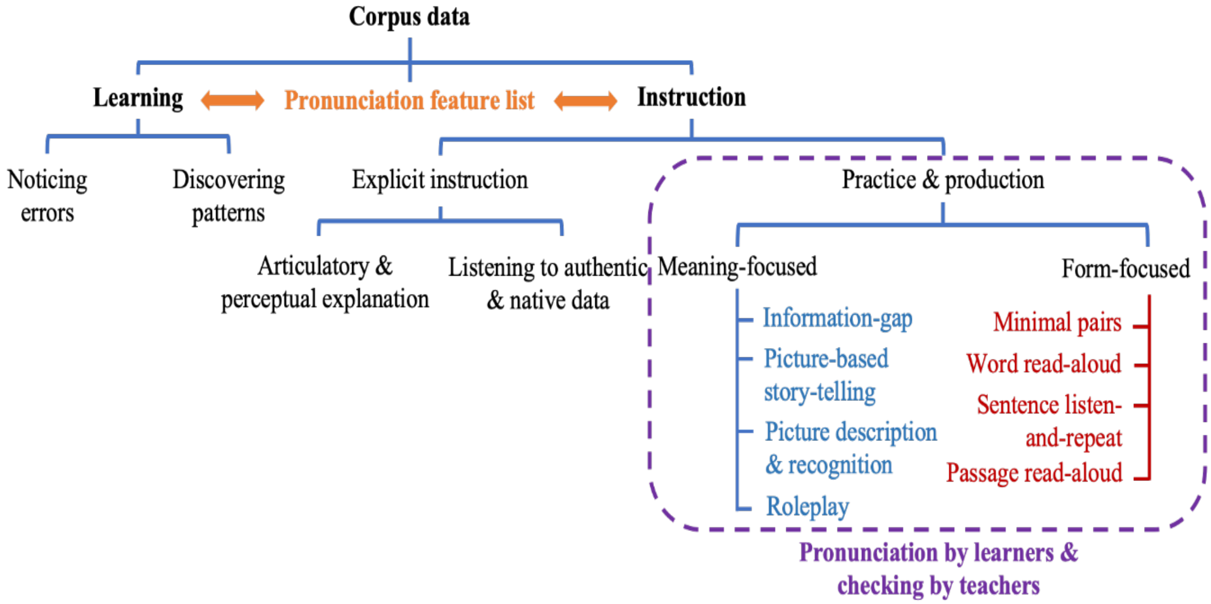
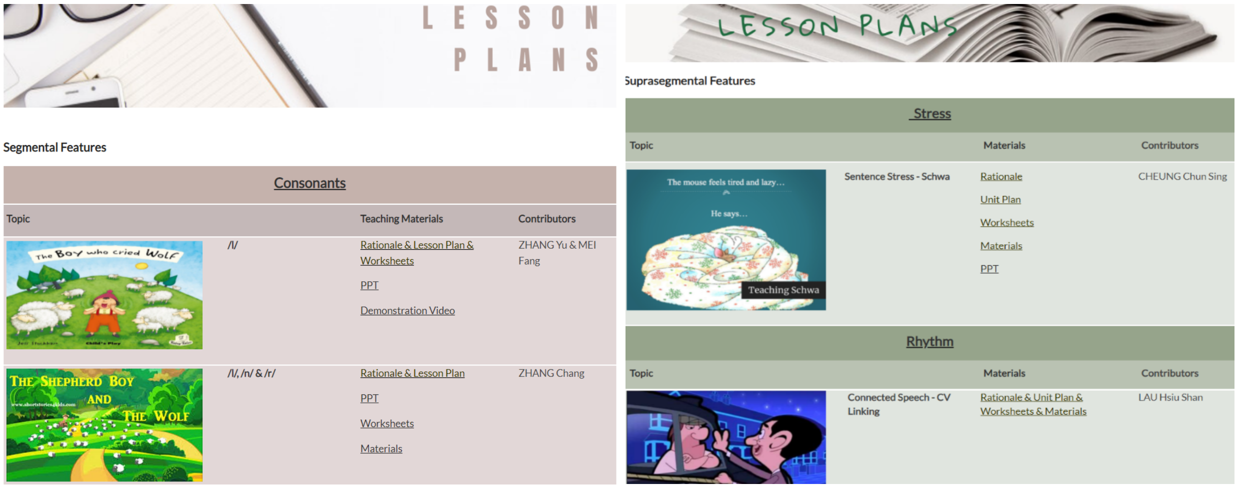

Corpus data
Learning
Pronunciation feature list
Instruction
Noticing errors
Discovering patterns
Explicit instruction
Practice & production
Articulatory & perceptual explanation
Listening to authentic & native data
Meaning-focused
Form-focused
Information-gap
Picture-based story-telling
Picture description & recognition
Roleplay
Minimal pairs
Word read-aloud
Sentence listen-and-repeat
Passage read-aloud
Pronunciation by learners & checking by teachers

FIGURE 6. Corpus-aided pronunciation teaching & learning framework

SAMPLE CORPUS-AIDED PRONUNCIATION LESSON PLANS

More than ten sets of ready-made lesson plans are designed to integrate the current corpus into corpus-aided pronunciation teaching based on the abovementioned framework. Two of the lesson plans are selected as representatives in this paper to demonstrate corpus data integration into teaching (for more corpus-aided lesson plans, visit https://corpus.eduhk.hk/english_pronunciation/index.php/for-teachers/). Figure 7 shows a screenshot of the ready-made lesson plans.



Segmental Features

Consonants			
Topic		Teaching Materials	Contributors
	/l/	Rationale & Lesson Plan & Worksheets PPT Demonstration Video	ZHANG Yu & MEI Fang
	/l/, /n/ & /r/	Rationale & Lesson Plan PPT Worksheets Materials	ZHANG Chang

Suprasegmental Features

Stress			
Topic		Materials	Contributors
	Sentence Stress - Schwa	Rationale Unit Plan Worksheets Materials PPT	CHEUNG Chun Sing

Rhythm			
Topic		Materials	Contributors
	Connected Speech - CV Linking	Rationale & Unit Plan & Worksheets & Materials	LAU Hsiu Shan

FIGURE 7. Screenshot of the ready-made lesson plans

FOR CHINESE LEARNERS

The first lesson plan primarily focuses on the consonant sound /θ/. Hong Kong learners of English are more likely to use the voiceless labiodental fricative /f/ to substitute the voiceless dental consonant /θ/, which is one of the most salient phonological errors made by Hong Kong students (Deterding et al., 2008); the data in the corpus support this. For the 20 Hong Kong speakers in this corpus, the mean occurrence of the pronunciation error /θ/ pronounced as /f/ is 1.65 times out of 3 tokens containing /θ/ in the reading passage.

Speech data from the passage 'The Boy Who Cried Wolf' were drawn from the corpus and integrated into the lesson as material for listening practice. This lesson plan aims to teach Primary 3 students the articulation of the voiceless consonant /θ/. It is also suitable for learners at other proficiency levels who need to improve the articulatory and phonological aspects of the voiceless dental fricative /θ/. Centring upon the story, the teacher first delivers the gist of the story with the aid of puppets and plays the recording of the story read by a native speaker. The relevant story scenes are shown by clear and lucid PowerPoint slides to help students comprehend the story. The teacher then asks the students to notice and find the consonant sound /θ/ while listening to the recording for a second time. To raise students' awareness of the phonological errors in context, the teacher plays a recording read by a Hong Kong learner and asks students to highlight the mispronounced words. After this, the students perform a listening-and-speaking activity with their partners. In this lesson plan, the presentation, practice and production approach are employed to teach students the target consonant sound and provide them with specific support, phonological practice and the opportunity to relate the pronunciation difficulties learned in the lesson to their real life.

FOR NON-CHINESE LEARNERS

The second sample lesson plan focuses on segmental errors made by non-Chinese learners of English in Hong Kong. About 8% of the Hong Kong population consists of ethnic minorities, of which 50% are from South and Southeast Asia (Hong Kong Government, 2018). It is, therefore, worth helping English learners in Hong Kong to notice the diversity of English phonological errors occurring in Hong Kong so that South and Southeast Asian learners may increase awareness of their pronunciation difficulties. By making use of a corpus that provides authentic data produced by non-Chinese learners in Hong Kong, errors with regard to confusion of two voiceless-voiced consonant pairs (i.e. /t/vs/d/ and /p/vs/b/) are frequently found for Filipino learners. For the 14 Filipino speakers in the corpus, the mean occurrence of the error 'non-word-final /t/ pronounced as /d/' is 3.85 times out of 19 tokens, and the mean occurrence of the error 'non-word-final /p/ pronounced as /b/' is 1.28 times out of 5 tokens. The pedagogical sequence for pronunciation stated by Ranta and Lyster (2007) is fully applied in this lesson, which includes seven stages (i.e. lead-in, gist listening, error discovery, listening for the specific information, focus on forms instruction, reading aloud task and spontaneous speech task). Speech data from the corpus are integrated into the listening tasks.

DEVELOPMENT OF THE TEACHER TRAINING PROGRAMME

According to Granger's observation (2004), learner corpus research has fallen short of achieving its full potential, leading to a scarcity of concrete pedagogical applications.. Indeed, the use of learner corpora and corpus tools is also new to English learners and teachers in Hong Kong. In order to maximise the effectiveness of the corpus from this study, it is of paramount importance to provide training for in- and pre-service language teachers to develop teaching materials and facilitate students' pronunciation acquisition using spoken corpora and corpus tools.

To help teachers understand how to use a corpus for their daily classroom teaching, a training programme was developed by means of workshop sessions and online sessions. The face-to-face workshop sessions were conducted according to the structure proposed by Aston (2001), who suggests three ways in which corpora can be implemented in teaching: teaching about corpora, exploiting corpora to teach languages and teaching to exploit corpora. The two online lessons served as supplements for the training programme and supported participants in obtaining a better and deeper understanding of corpus-aided language teaching and learning and in enhancing the learning engagement of students beyond the physical learning environment. Three types of activities were included: a) resource-based activities, including a full range of learning resources such as narrated lecture slides or mini Massive Open Online Course (MOOC) videos; b) response-based activities in which the participants were required to submit their responses, including online quizzes and short answer questions; and c) collaborative activities such as fora in which the participants could interact and collaborate in an asynchronous fashion.

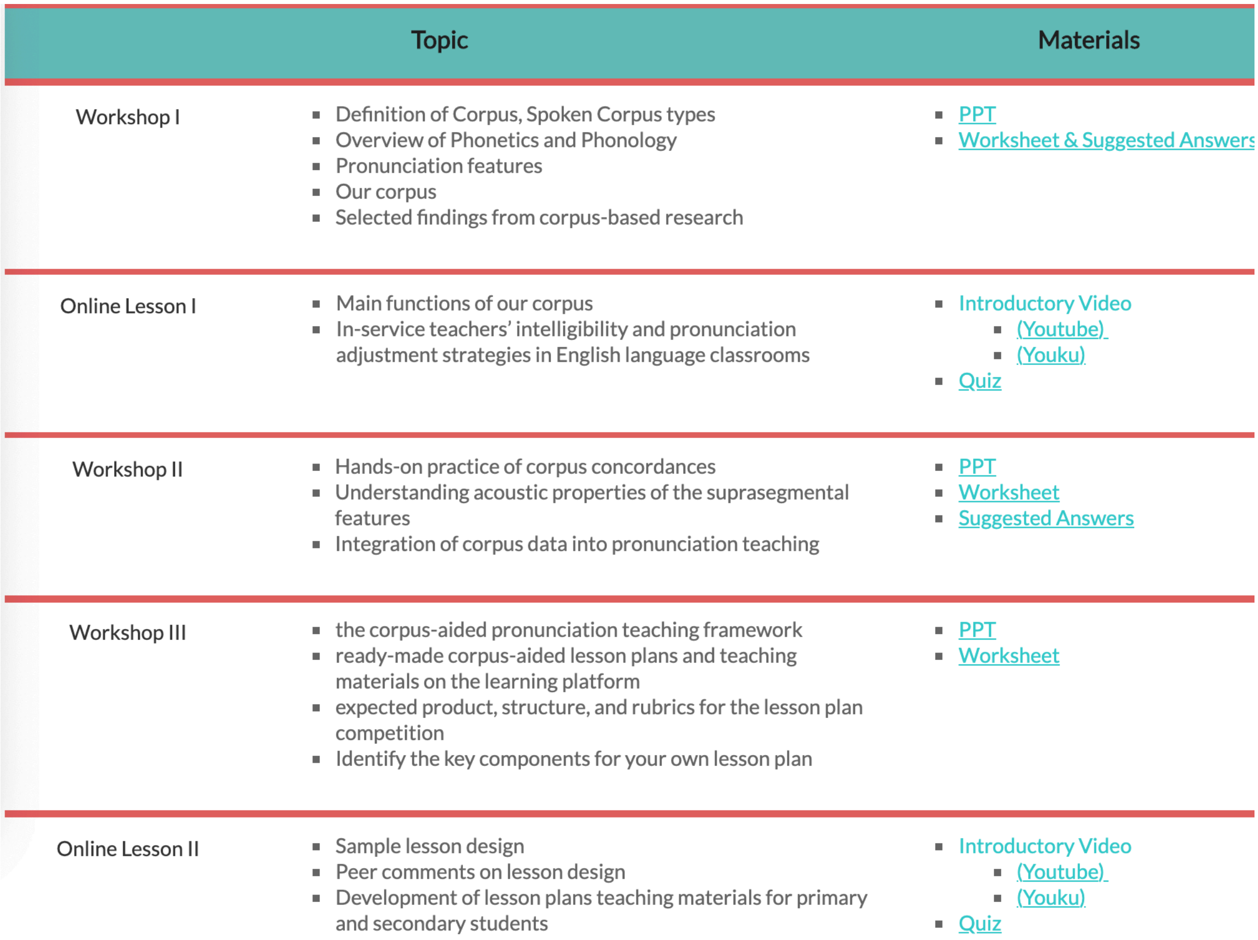
The current corpus comprised the core of this teacher training programme. Upon completion of the programme, participants were expected to a) use a corpus-aided approach to support their English pronunciation learning with a variety of pronunciation assessment tools in workshop sessions and online lessons; b) discover common pronunciation errors produced by non-Chinese, mainland Chinese and Cantonese speakers of English and reflect on their own pronunciation; c) identify recurrent segmental and suprasegmental difficulties in learners' English pronunciation with different language backgrounds; and d) understand possible remedies that may reduce or eliminate English pronunciation difficulties and raise awareness of the issues involved in achieving a native-like pronunciation or a comfortable, intelligible accent.

This training programme consisted of three two-hour workshops and two one-hour online lessons. The first two workshops and the first online lesson were intended to provide participants with basic knowledge concerning English phonetics and phonology and practical skills in using the corpus; the last two workshops and the second online lesson provided sample lesson plans and introduced the teaching framework to participants. Each session was systematically and meticulously integrated into the entire programme. Meanwhile, the three activity types were included in two online lessons in the training programme. Specifically, each online lesson, uploaded to the easily accessed platform Schoology, included a mini video and online quizzes. Response-based activities and collaborative activities were included in the competition session, in which participants were required to work in pairs and design a corpus-aided English pronunciation lesson and submit comments on other groups' lesson designs. Details of the programme are shown in Table 1.

TABLE1. The Topics Covered in the Program

Category	Systematisation	Topic
Workshop I	Introduction	• Definition of Corpus, Spoken Corpus types • Overview of Phonetics and Phonology • Pronunciation errors • Our corpus • Selected findings from corpus-aided research
	Exploitation	• Hands-on practice: Phonological analyses through our corpus.
Online lesson I	Introduction	• Two main functions of our corpus • In-service teachers' intelligibility and pronunciation adjustment strategies in English language classrooms
	Exploitation	• Frequent segmental errors (e.g. vowels and consonants) of a given word in our corpus
Workshop II	Exploitation	• Suprasegmental system of English in alignment with the suprasegmental errors in the corpus (pausing, intonation, lexical stress, CV linking) • Hands-on practice of corpus concordances • Understanding acoustic properties of the suprasegmental errors • Integration of corpus data into pronunciation teaching
Workshop III	Introduction	• The corpus-aided pronunciation teaching framework • Ready-made corpus-aided lesson plans and teaching materials on the learning platform
	Exploitation	• Expected product, structure, and rubrics for the lesson plan competition
	Transformation	• Identify the key components for your own lesson plan
Online lesson II	Transformation	• Sample lesson design • Peer comments on lesson design
Competition	Transformation	• Development of lesson plans and teaching materials for primary and secondary students

The teacher training programme attracted great attention from participants worldwide. Registration was received from 121 participants across six countries: China, the UK, the US, Australia, Germany and Singapore. The Chinese participants came from different provinces and regions, including Hong Kong, Macao, Beijing, Hebei, Liaoning, Inner Mongolia, Shandong, Shanghai, Sichuan, Hubei and Guangdong provinces. Participants were mainly pre-service teachers and in-service teachers. Pre-service teachers, including undergraduates and postgraduates, accounted for 77%; in-service teachers, comprising primary school teachers, secondary school teachers and university professors, accounted for 23%. Three workshops were held on campus, broadcast live on Facebook and Tencent synchronously. There were 40 participants on average for each workshop. For the online lessons, 44 participants created Schoology accounts, watched the online videos and completed the online tasks. Materials used for the training programme can be accessed via https://corpus.eduhk.hk/english_pronunciation/index.php/teacher-training-programme/. Figure 8 shows a screenshot of the training programme materials.



	Topic	Materials
Workshop I	▪ Definition of Corpus, Spoken Corpus types ▪ Overview of Phonetics and Phonology ▪ Pronunciation features ▪ Our corpus ▪ Selected findings from corpus-based research	▪ PPT ▪ Worksheet & Suggested Answers
Online Lesson I	▪ Main functions of our corpus ▪ In-service teachers' intelligibility and pronunciation adjustment strategies in English language classrooms	▪ Introductory Video ▪ (Youtube) ▪ (Youku) ▪ Quiz
Workshop II	▪ Hands-on practice of corpus concordances ▪ Understanding acoustic properties of the suprasegmental features ▪ Integration of corpus data into pronunciation teaching	▪ PPT ▪ Worksheet ▪ Suggested Answers
Workshop III	▪ the corpus-aided pronunciation teaching framework ▪ ready-made corpus-aided lesson plans and teaching materials on the learning platform ▪ expected product, structure, and rubrics for the lesson plan competition ▪ Identify the key components for your own lesson plan	▪ PPT ▪ Worksheet
Online Lesson II	▪ Sample lesson design ▪ Peer comments on lesson design ▪ Development of lesson plans teaching materials for primary and secondary students	▪ Introductory Video ▪ (Youtube) ▪ (Youku) ▪ Quiz

FIGURE 8. Screenshot of training program materials

FEEDBACK AND REFLECTION ON THE TEACHER TRAINING PROGRAMME

An evaluation form on the effectiveness of each workshop, online lesson, corpus and the entire training programme was sent to the participants. Feedback from the workshop series and online lessons was very positive. With regard to the workshop series, 96% of the participants considered the workshops worth attending. In addition, 97% of the participants, after each workshop, stated that they had gained knowledge and skills for using corpus tools in English pronunciation teaching and learning and that they had a better understanding of the pronunciation difficulties of their students. With regard to online lessons, over 95% of the participants thought that they would recommend the new corpus to their colleagues and students.

In general, upon completion of the programme, all participants stated that they had successfully grasped the two core functions (i.e. the browse and search functions) of the corpus and that they were satisfied with the lesson plans and other relevant resources provided. On top of that, participants saw great value in using the newly launched corpus and corpus-aided approach to teach and learn English pronunciation. For example, two participants said that "'the corpus provides genuine data to study local students' EFL speech" and "I have gained new insights from this workshop, especially from the comparison between Hong Kong and mainland students. Very helpful." All participants viewed the newly launched corpus as a useful data source for their English pronunciation teaching and said they would design corpus-aided language learning tasks for their students and integrate the corpus into their English pronunciation teaching. Moreover, it was fully recognised by all participants that it was necessary for English teachers to receive training in corpus-aided English pronunciation teaching.

Among the participants in the lesson plan competition, ten pre-service English teachers were invited to conduct an in-depth interview on a voluntary basis. These participants' views on the design of the training programme were categorised according to three key components incorporated into this training programme, namely, the learner spoken corpus, the teaching framework and the sample lesson plans, and the online lessons.

LEARNERS SPOKEN CORPUS

In comparison with the natively spoken corpus, all the participants acknowledged that the learner-spoken corpus was more useful because the students would be able to identify particular difficulties they shared with speakers in the corpus. For instance, two participants said the following:

> P1: Learner corpora make it easier to discover errors. It is hard for learners to spot the differences between their own pronunciations and the correct ones. When learners find that 80% of all learners make the same error, they may pay attention to it when pronouncing the sound.
>
> P2: I read an article about effective methods of learning English. One method is to observe your classmates' errors. Their errors are the same errors you are likely to make. Learner corpora allow learners to hear errors that speakers with the same L1 make, and so to discover their own errors.

However, the natively spoken corpus was normally regarded as a model for students to simply imitate and follow, as stated below:

> P3: When using native speakers' corpora, learners can only imitate the native speakers' speech. They do not know whether they are making errors. It is just easier for them to compare.

In addition, making use of the corpus in English lessons may hold students' attention, as one participant stated:

> P4: Using learner corpora motivates students to learn pronunciation. When listening to the recordings with similar errors to those of the learners, they may find that the lesson is more interesting and attractive.

The majority of the participants said they would use a spoken corpus in their future English pronunciation teaching. One participant said this:

> P8: I will definitely use the spoken corpus in future teaching. I think it is useful because it highlights all the different mistakes made by speakers from different regions, and we can also select which specific mistakes we want to highlight. It's a lot better than the teacher trying to make mistakes to show the students. It's better to have authentic readings for each of them. We can use different readings depending on which mistake we want to highlight for the students, so they can also spot the mistake. We can also use the native speaker function (recording) so that they can compare how the words or passages should be properly read.

When asked about the limitations of using spoken corpus data for their language learning and teaching, participants gave feedback mainly centred on the quantity of data and the diversity of passage readings. Two participants explained this in their comments:

> P5: More data could be collected. Speakers from different sub-dialect groups could be added.

> P6: Sometimes, I cannot find the data I need. More data could be collected. More speakers from each dialect group in mainland China should be collected. It would be easier for teachers to find the common errors.

This learners' spoken corpus currently contains 96 sets of data collected from nine dialect groups. Of these dialect groups, the data for Mandarin were collected more systematically and included eight subgroups. Each dialect group of the Chinese language can be categorised into three or more detailed subgroups, but this corpus covers less than two sub-dialect groups for each dialect group at present, a fact that will be taken into account for future supplementation and renovation of the corpus. In terms of passage readings, most participants said that more passages could be incorporated since the more passages that are included, the more pronunciation errors of relevance would be shown. One participant further commented thus:

> P5: When we want to highlight specific errors for the students, we can only choose from those three passages (including the interviews), and sometimes some of those errors may only be shown once or twice in the particular passage.

In addition, they also suggested that the difficulty of the passages should be taken into consideration:

> P6: More passages could be added. Students with different levels could read different passages. Teachers who teach different levels could find different resources to use. The interviews in the corpus could be annotated because errors in reading-aloud tasks and speech delivery are different. Teachers should also know the errors in speech delivery.

Learners’ needs can be viewed as a top priority in teaching, regardless of the subject. However, this corpus was designed to provide all learners, teachers and researchers with a full range of recordings and phonological annotations and to help them identify Hong Kong, mainland Chinese and non-Chinese learners’ recurrent difficulties in English pronunciation learning. The corpus could be expanded to a more comprehensive one on a larger scale in the near future.

TEACHING FRAMEWORK AND SAMPLE LESSON PLANS

On the basis of the teaching framework, teachers are able to identify learners' pronunciation errors through the spoken corpora first and then identify learners' recurrent difficulties in using segmental errors and suprasegmental errors in English, designing lessons to reduce such difficulties. According to the participants' feedback, most groups in the lesson plan competition applied the framework in their lesson design. They would appear to have paid full attention to the use of the corpus when designing lessons. One participant who won first place in the competition tactfully used the teaching framework and considered it quite useful.

> P8. It is useful. It teaches me how to design logical and systematic lessons. It helps to design an effective pronunciation lesson. Students can learn better by following the sequence of the framework step by step. I learned a useful teaching approach and different task types to integrate corpus data into my lessons. It provides practical suggestions.

According to the participants’ feedback, almost all the participants tended to pay more attention to and positively commented on the sample lesson plans. The following extracts illustrate positive views on the sample lesson plans provided in the corpus-aided training programme.

> P1: I am aware of how to organise the sequence of my lessons and how to choose the target sounds that I will teach.

> P9: I learned the teaching approach to designing a lesson. I may use the approaches used in the sample lesson, such as stories and self-corrections. I learned different approaches for different students (primary and secondary school students).

ONLINE LESSONS AND TASKS

Online lessons and tasks were used to expose participants to an interactive forum in which the participants could interact and collaborate in an asynchronous fashion. Over 60% of the participants agreed that online lessons and tasks could enhance and consolidate their learning, as exemplified in the comments below:

> P7: The online session can be considered a supplement to the workshop. I think this combination is the best.

> P2: I prefer online tasks because they have fewer limitations. I can do them whenever I want.

Furthermore, all participants were actively engaged in the response-based activity and shared their views with others. All of them considered comments from others to be useful and valuable. For example, a participant claimed that:

> *P2: I can receive comments from others. I can communicate with others through online tasks.*

CONCLUSION

In this paper, a self-developed corpus was introduced that included speech data from English learners from Hong Kong, nine major dialectal regions from mainland China and non-Chinese speakers who were living in Hong Kong. English learners in Hong Kong are not only local Hong Kongers but also from different ethnic groups worldwide and various provinces in mainland China. Given the complexity of their language backgrounds and the fact that second language acquisition is influenced by L1, the spoken learner corpus can serve as an effective tool for both teachers and learners to facilitate English pronunciation teaching and learning. English phonological errors made by learners of different language backgrounds were identified in the corpus. Accordingly, sample corpus-aided lesson plans and materials were provided for English teachers to increase their awareness of the corpus-aided teaching approach. A corpus-aided teacher training programme was developed and implemented through a workshop and online sessions to help pre-service, and in-service teachers understand the corpus-aided teaching approach and the use of the spoken learner corpus as a tool to optimise teaching effectiveness. Participants in the training programme expressed positive views on the corpus, lesson plans and the programme itself.

For future improvement of this corpus-aided English pronunciation learning and teaching website, as suggested, the speech data size from different Chinese dialectal backgrounds can be expanded. To help learners obtain a better understanding of the distinction between the native speakers and their own reading, similar to learners' speech data, the same acoustic analysis and IPA transcriptions of native speakers' reading can be added in more detail. Finally, more pre-and in-service teachers from primary to tertiary levels can be invited to participate in corpus-based learning and teaching activities. Systematic evaluations and feedback from them can be collected in order to improve the quality of the corpus and the website. Both teachers' and learners' needs can be further identified and accommodated.

ACKNOWLEDGEMENTS

This work was supported by the Teaching Development Fund provided by the Education University of Hong Kong.

REFERENCES

Aston, G. (2001). Learning with corpora: An overview. In G. Aston (Ed.), *Learning with corpora* (pp. 6–45). Houston, TX: Athelstan.

Chen, H. C., & Han, Q. W. (2020). Designing and implementing a corpus-based online pronunciation learning platform for Cantonese learners of Mandarin. *Interactive Learning Environments*, *28(1)*, 18–31.

Chen, H. C., & Wang, Q. (2016). Development and Application of a Corpus-based Online Pronunciation Learning System for Chinese Learners of English. *English Teaching and Learning*, *40*(2), 83-114.

Deterding, D., & Wong, J., & Kirkpatrick, A. (2008). The pronunciation of Hong Kong English. *English Worldwide, 29*, 148-175. 10.1075/eww.29.2.03det.

Ebrahimi, A., & Faghih, E. (2016). Integrating corpus linguistics into online language teacher education programs. *ReCALL, 29*(1), 120-135.

Flowerdew, L. (2009). Applying corpus linguistics to pedagogy: A critical evaluation. *International Journal of Corpus Linguistics, 14*, 393–417. doi: 10.1075/ijcl.14.3.05flo

Granger, S. (2004). Computer Learner Corpus Research: Current Status and Future Prospects. In Connor, U., Upton, T. (Eds*.) Applied corpus linguistics: a multidimensional perspective*, (pp. 123–145). Amsterdam and Atlanta: Rodopi.

Gut, U. (2005). Corpus-based pronunciation training. *Proceedings of the Phonetics Teaching and Learning Conference*, London.

Hewings, M. (2012). Using corpora in research, teaching, and materials design for ESP: An evaluation. *Taiwan International ESP Journal, 4*(1), 1–24.

Hong Kong Government. (2018). *2016 Population by Census.* Retrieved from https://www.bycensus2016.gov.hk/tc/bc-mt.html

Hung, T. T. N. (2000). Towards a phonology of Hong Kong English. *World Englishes, 19*, 337–356.

Jenkins, J. (2009). *World Englishes: A resource book for students* (2nd ed., Routledge English language introductions series). London; New York: Routledge.

Karras, J. N. (2016). The effects of data-driven learning upon vocabulary acquisition for secondary international school students in Vietnam. *ReCALL (Cambridge, England), 28*(2), 166-186.

Kolesnikova, O., & González-González, O. A. (2016). Spoken English Learner Corpora. *Research in Computing Science*, *130*, 111-132. *learning, 60*(3): 534–572.

Lesho, M. (2018). Philippine English (Metro Manila acrolect*). Journal of the International Phonetic Association: Illustrations of the IPA*.

Li, H. P., & Chen, H. C. (2019). Intelligibility and comprehensibility of the Filipino English accent to Hong Kong English speakers. *3L: Language, Linguistics, Literature, 25*(1), 23-42

Lian, X. L. (2013). *Minanyu dui chuzhongsheng yingyu yuyin fuqianyi de diaocha yanjiu – yi Zhangzhou shiyan zhongxue chusan xuesheng weili* [An investigation of negative transfer from Southern Min to English pronunciation learning – a case study of Secondary 3 students in Zhangzhou Experimental School]. Master Thesis.

Mesthrie, R. (2008). English is circling the globe*. English Today, 24*(1), 288–32.

Pérez-Paredes, P. (2019). A systematic review of the uses and spread of corpora and data-driven learning in CALL research during 2011–2015. *Computer Assisted Language Learning*, 1-26.

Qin, Y. L., & Wei, X. M. (2008). Analysis of the Negative Transfer of Dialects to English Phonemes in Southeast of Guangxi. *Journal of Yulin Normal University*.

Ranta, L., & Lyster, R. (2007). A cognitive approach to improving immersion students' oral language abilities: The awareness-practice-feedback sequence. In DeKeyser, R. (Ed.), *practice in a second language: Perspectives from applied linguistics and cognitive psychology* (pp. 141–160). New York: Cambridge University Press

Römer, U. (2011). Corpus research application in second language teaching. *Annual Review of Applied Linguistics, pp. 31*, 205–225.

Sinclair, J. M. (Ed.). (1987)., *Looking Up: An Account of the COBUILD Project in Lexical Computing.* London: Collins ELT.

Tayao, M. L. G. (2008). A Lectal Description of the Phonological Features of Philippine English. In M. Lourdes, S. Bautista & Kingsley Bolton (Eds.), *Philippine English: Linguistic and literary perspectives*. Hong Kong: Hong Kong University Press.

Wen, L., & Zhou, Y. Z. (2014). Yingxiang Ningxia shanqu xuesheng yingyu yuyin shuiping de xianzhi yinzi yu duice fenxi [Factors influencing English pronunciation level of students in Ningxia mountain area and analysis of solutions]. *Journal of Inner Mongolia Normal University (Educational Science).*

Willis, D., & Willis, J. (1989). *Collins COBUILD English Course*. London: HarperCollins.

Wong, T. S., & Lee, J. S. Y. (2016). Corpus-based learning of Cantonese for Mandarin speakers. *ReCALL (Cambridge, England), 28*(2), 187-206.

Xiao, C. C. (2014). Qiantan Sichuan is fangyan yuyin tedian dui yingyu fayin de yingxiang [A discussion of the influence of Sichuan dialect on English pronunciation]. *Yingyu jiaoshi*.

Xu, H. X. (2016). Youshisheng yinyu fuyin fayin zhiliang diaocha – yi jiangsu lianyungang fangyanqu weili [An investigation of English consonant pronunciation by pre-service early childhood teachers – a case study in Lianyungang, Jiangsu province]. *Neimenggu Jiaoyu*.

Yang, H., & Wei, N. (2005). *Construction and data analysis of a Chinese learner spoken English corpus*. Shanghai: Shanghai Foreign Language Education Press.

Zhang, H. X., & Chen, R. (2019). Dalian fangyan dui Dalian Yingyu xuexizhe de yingyu fayin de fuqianyi yanjiu [A study on negative transfer from Dalian dialect to the English pronunciation of local learners]. *Xiandai jingji xinxi*.