CONSTRUCTION OF SPEECH CORPUS OF AESOP-SD

Yuan JIA[1], Meng WANG[2], Honghua ZHAI[2], Aijun LI[1]

[1]Institute of Linguistics, Chinese Academy of Social Sciences, China
[2] Shandong University of Science and Technology, China

ABSTRACT

The present study systematically states the construction of the corpus on the English learners, i.e., AESOP-SD. The content mainly consists of three parts: i) the background and significance of the corpus construction, which introduces the research background of the English learning in Asia and states the significance of the construction of the present corpus; ii) materials in the corpus, which contain a large amount of data, ranging from English words, English sentences, and English paragraphs to dialectal words, sentences and paragraphs; iii) recording procedure and data labeling, which states the recording environment and software in the data collection. Through the introduction of the construction of the corpus, the paper further states the theoretical and applicable value of the corpus, which can be adopted to conduct research in many research areas: e.g., second language acquisition, phonetic and phonological study, computer-aid system.

Index Terms— AESOP-SD, Corpus, English data, Dialectal data

1. INTRODUCTION

As the most widely used language in the world, English has been the most important tool of communication in the research, the cultural exchange and the international cooperation in the field of economy and trade. Therefore, English teaching has become an indispensible part of the language teaching and study in many countries. Gimson [1], the famous English phonetician, once pointed out, “To speak any language, a person must know nearly 100% of its phonetics, while only 50%-90% of its grammar and 1% of the vocabulary maybe sufficient.” Thus, the importance of the pronunciation learning is self-evident in the process of the English learning. This problem has attracted a lot of the attention of the scholars and the educators in the English teaching and research of many countries.

To effectively enhance and promote the expression and communication ability of the English learners in Asia, a number of research organizations have committed to this issue. Among these organizations, the AESOP (Asian English Speech cOrpus Project) organized by professor Sagisaka of the Waseda University intended to investigate the English pronunciation. Its main purpose is to establish the speech database of the English learners in Asian countries. On this basis, they can examine the phonetic characteristics of the English learners in different countries in order to discover their universal and different characteristics of the pronunciation from the English learners in different countries and also the impact from their mother tongues. In this way, they intend to propose the ways and means to improve the English pronunciation of the learners in different countries.

Different language has its special phonological system and pronunciation rules. There are so many dialects in Chinese, e.g. Official dialect, Wu dialect, Hui dialect, Gan dialect, Xiang dialect, etc (Li[2]), and the difference between the dialectal system and the English system is of great extent. Due to the ignorance about the differences of the phonological system, the students in the dialect area will unconsciously set the phonological rules of the dialect (mother language) to the phonological system of the English under the influence of the dialects. Therefore, it leads to many pronunciation errors and these errors will accompany with them even for their whole life. To be the worst, these problems are not only reflected on the students, but also on the teachers. The ‘dialectal pronunciation’ from teachers goes around from generation to generation through the process of teaching. Being unable to find effective solutions, this situation has become the “transparency” of the English teaching in China.

Given the above analysis, to analyze the types of errors of the English learners in the dialect area has the great significance in improving the speaking and listening level of the Chinese English levels, addressing the issue of “dialectal pronunciation” and enhancing the communication skills. Therefore, this research intends to establish a speech corpus in dialectal region following the tenet of AESOP. Specifically, the database in China was firstly constructed in Shandong Province (Hereinafter, as SD), i.e., AESOP-SD. SD Province is suited on the East China seaboard where the Yellow River empties into the sea. As for the Shandong dialect, although it deserves the feature of official language, they also present their unique features (Qian[3]).

The AESOP-SD corpus contained a mass of speech data, English materials and Chinese materials. Within the English data, it covers English words, phonetic balanced

English sentences, various kinds of English sentences with different functions, and English paragraphs and dialogues, etc. The Chinese materials contained not only dialect data, e.g., different numbers of tonal sequences, sentences with different design of foci, and paragraphs, it also covered the Dialect-Mandarin data. Consequently, the speech dada comprised eight hours recording for each participant. Based on the speech data, many phonetic and phonological research topics can be concerned. On basis of contrasting the dialect and the English sound system, we may systematically summarize the types of the errors of the pronunciation of the English learners in the aspect of segmental (consonant and vowel) and supra-segmental (stress types, boundary tone and the rhythm, etc.). Based on the result, from the perspective of phonetics, we put forward concrete ways and means for the learners in the dialect districts to improve the pronunciation problems of the learners in consonants, vowels, stress and the intonation.

2. MATERIALS IN THE CORPUS

When designing the text, we need to maximally investigate the characteristics of the ‘dialectal English pronunciation’ of the English learners from the SD dialect region. The design of the text has been completed, and the tentative recording has been conducted in Shandong Province. This text mainly includes two aspects, namely the English and the Chinese corpus, and the concrete content are as follows:

2.1 English data

English data is the most significant part in this research. All speakers in SD dialect areas read the same English data in order to make a comparative study of the segmental and supra-segmental features. The data mainly includes three parts: (i) English words, which include various word stress patterns, abundant vowel and consonant combinations. In this part, 90% are real words while in line with phonetic features 10% are artificially designed to include limited clusters of the vowels and the consonants to investigate the ‘negative transfer’ expressed by the learners under the influence of the dialect in the articulation of consonants, vowels and word stress; (ii) English sentences, which include various sentential forms and types: a) phonetic balanced sentences, mainly meet different speech combinations; b) interrogative sentences, including general questions, interrogative questions, alternative questions and echo questions to investigate the articulation features of the speakers and interrogative sentences in different dialect areas and the differences with the native speakers, correction methods and standards; c) imperative sentences, which mainly include the imperative ones and the corresponding declaratives, with the aim to investigate the acoustic features of the speakers when expressing imperatives; d) exclamatory sentences, which mainly contain exclamatory ones and corresponding declaratives, aiming to investigate the error types of the speakers in different dialect areas; e) focus sentences, which contain the foci distributing in different positions of sentences. It aims to investigate the difference of stress distributing position and realization type between SD learners and Standard English speakers. Further, the sentences also combine focuses with the interrogatives with the aim to investigate the existing problems of the learners when expressing questions and focuses; f) sentences with different structures, which mainly contain subjective clause, predicative clause, and more structures, aiming to investigate the coherence between the sentence structures and the pronunciation; (iii) English discourses, e.g., ‘The North Wind and the Sun’ and ‘Cinderella’, as well as the natural dialogues, which are about air-ticket booking, and dialogue-making according to the pictures. The discourses and dialogues data can be adopted to examine the prosodic features above the sentence level.

2.2 Chinese data

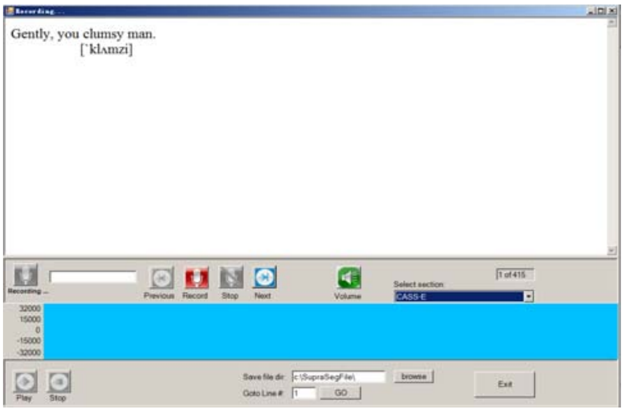
It mainly includes two aspects: (i) Dialectal-Mandarin data with regional accents. It aims to observe and examine the prosodic features of consonants, vowels and tones reflected in the process of speaking, and thus revealing the interference that the Mandarin and the dialects exert on the English pronunciation. This part mainly includes mono-syllabic and disyllabic combinations. The former mainly includes the combinations of all the Mandarin initials and the vowels of [a, i, u, ü], each of which includes four tones: the level tone (Tone1), the rising tone (Tone2), the falling-rising tone (Tone3) and the falling tone (Tone4). The latter mainly includes the common used Mandarin words whose sequences include all the sixteen tonal combination types. Its purpose is to investigate the features of the Mandarin tones influencing the production of the English intonation; (ii) Dialect data, which aims to examine the prosodic features of consonants, vowels and tones of SD dialect, through which the ‘negative transfer’ exerted from SD dialect on the English pronunciation. The dialect data was selected according to the most representative constituents in the dialect. It includes mono-syllabic, di-syllabic, tri-syllabic and phrasal constituents. Mono-syllabic items include all the possible combinations of initials and finals, and each syllable includes all of the tones in SD dialect. And, the data consists of all the combinations of tones in the two-word phrases and three-word phrases. Phrase data collected the distinctive phrases used in SD dialects.

Shandong dialects, which belong to northern mandarin area, are mainly divided into four districts: east district of Lai, east district of Wei, west district of Lu and west district of Qi (Qian [3]). Shandong dialects were selected in the way of generation in that although there were differences within the four districts, the differences are comparatively subtle. The four districts adopted the same dialect data which includes all the possible mono-syllabic, di-syllabic, tri-syllabic and phrasal constituents in the four districts. Dialect data also includes the focus sentences with the purpose to investigate the features of dialectal pronunciation of the speakers inspected and the effect of ‘negative transference’ that exerts on the production of English focus sentences. This kind of sentences mainly includes the focus components of two-word phrases of all the tonal combinations. The dialectal data also contains paragraph like ‘The North Wind’. It was collected for the purpose of examining the expression features of ‘negative transference’ that the dialect exerts on the pronunciation of the English discourse.

3. RECORDING AND DATA LABELING

3.1 Recording procedures

The recording was conducted in a quiet indoor environment, e.g., the classroom, the office or the laboratory in Shandong University of Science and Technology. The total recording duration for a speaker is nearly ten hours. The equipment of the recording are the laptop and the head-wear with microphone and its type is Sennheiser PC166, with the built-in type sound card. The recording software is developed by Chinese University of Hongkong. In accordance with the purpose of the recording experiment, the recording software of the project AESOP-SD, made a preliminary improvement by phonetic lab of Chinese Academy of Social Sciences. In this recording software, the recording interface can show the sentence, the chapter, the picture and so on. The sampling frequency is 16000 Hz, and the sampling rate is 16-digit and the sound track is mono. A word or a sentence is saved for a separate ‘wav’ file, in order to be convenient for the later speech data processing and the analysis. In the recording, the speakers wear the earphone, and sit in front of the computer screen, and the recording was manipulated by the operator. The corpus has already collected the speech from nearly ninety speakers, and the data from one hundred and ten speakers will be collected in the future. The following pictures are the software interface of recording and the type of the headphones:



Gently, you clumsy man.
[ˈklʌmzi]

Select section: CASS-E

Save file dir: c:\SupraSegFile\

Goto Line #: 1

Figure 1 Window of recording software

3.2 Data labeling

2.4.1 Segment labeling

Speech database annotation is divided into two parts, specifically, basic annotation and extensive annotation. The former is to annotate the regular pronunciation, such as phonemic boundary, syllable boundaries, word boundaries and sentence boundaries; the latter includes the annotation of the actual pronunciation and the prosodic labeling. The automatic segmentation software is used to label the boundaries of phones, words, sentences and discourses. The annotative symbols adopt the ARPABET Symbol Set (see the table below). As for the segmentation of the Chinese data, the automatic segmentation software is adopted. The boundaries of consonants, vowels and word boundary can be automatically labeled; then the manual proofreading is proceeded. The labeling of dialectal feature mainly relies on the handwork, using the SAMAP-C Symbol Set. The task is accomplished by the graduate students of English majors or the professional person engaging in the study of the phonetics. The annotation system is conducted by the adaptation of the Praat annotation software. Table 1 exhibits the symbol inventory of ARPABET.

Table 1 symbol inventory of ARPABET

No.	ARPA BET	IPA	Examples	No.	ARPA BET	IPA	Examples	No.	ARP ABE T	IPA	Examples
1	i:	i:	beat		r2	r	red		v	v	very
2	I	i	bit		j	j	yes		f	f	fly
3	E	e	bet		w	w	wash		D	D	they
4	ae	æ	bat		m	m	must		T	T	thief
5	V	ʌ	cut		N	ŋ	sing		_h		弱化
6	u:	u:	boot				Washingt on		_u		清化
7	U	u	put		tS	ʧ	chair		_?		紧喉或者喉化
8	aU	au	bout		dZ	ʤ	just		dr2	dr	drive
9	s@r	ə:	curt		b	b	boy		tr2	tr	tree
10	eI	ei	bait		p	p	play		ts	ts	cats
11	aI	ai	bite		d	d	do		dz	dz	goods
12	OI	ɔi	boy		dt	ɾ	butter		_r		r 音色
135	O:	ɔ:	caught		t	t	to		sp	停顿	Word 内
14	oU	au	boat		g	g	go		sil	停顿	Word 之间
15	l	l	led		k	k	kick		silv		
16	-l	-l	little		z	z	zoo		spv		
					s	s	sit		_^		鼻化
					Z	ʒ	measure		A:	§:	长 a
					S	ʃ	shoe				

2.4.2 Prosody labeling

Prosodic feature is labeling by the combination of IViE and ToBI labeling system. The label tiers include:

(i) Orthographic Tier: transcriptions of the spoken words

(ii) Prominence Tier: location of prominent syllables (stressed and accented)

(iii) Break Index Tier: transcription of intermediate and intonational.

(iv)Target Tier: Phonetic transcriptions, syllable-based, allowing transcribers to draw up a first set of hypotheses about accent alignment in phrase boundary;

(v) Phonological Tier: formal linguistic representations of speakers' intonational choices

(vi) Comment Tier: alternative transcriptions and notes

The following figure 2 is the labeling sample of the corpus contained both the dialectal and prosodic features transcription.

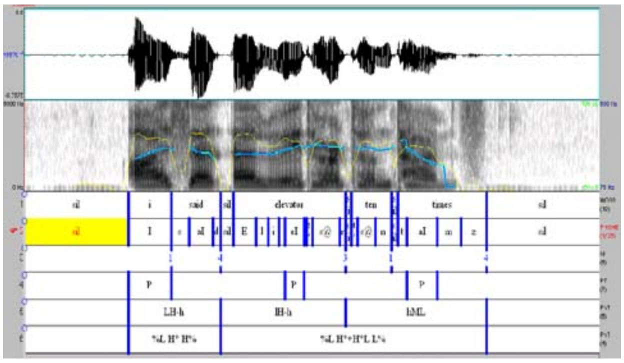


Figure2 Labeling sample

4. CONCLUSIONS AND DISCUSSIONS

The study introduces the construction of the corpus of English learners from dialect region, i.e., AESOP-SD CORPUS. The study states in detail the related information of the database, specifically, it mainly contains the following aspects: i) the background and significance of the corpus construction, which introduces the research background of the English learning in Asia and states the significance of the construction of the present corpus; ii) materials in the corpus, which contain a large amount of data, ranging from English words, English sentences, and English paragraphs to dialectal words, sentences and paragraphs; iii) recoding procedure and data labeling, which states the recording environment and software in the data collection. The labeling part mainly describes the segmental and prosodic labeling criteria.

The significance and application value of the corpus mainly includes:

(i) This speech database is a comprehensive large-scale corpus, including the Standard English materials, the dialect materials and the Dialect-Mandarin data. It provides an important source of data for large-scale study of the dialectal English, SD dialect and the Dialect-Mandarin, especially the comparative study between their phonological systems.

(ii) The establishment of the phonetic annotation system of the English learners in the Chinese dialect, i.e., SD districts can reflect the phonetic errors of the English learners in the dialect area more directly and provide a lot of speech resources for the second language speech research. Previously, in the research about the pronunciation problems, most researchers summarized the dialect pronunciation features of the speaker only based on perceptual data. Due to the limitation of the objects and the means of investigation, the coverage of the result tends to be narrow and not systematic enough. The research based on the large-scale dialect speech database can effectively improve the depth and width of the research of English learning, and it provides an important reference value for the teaching of the English phonetics.

(iii) Based on the speech analysis of the large-scale corpus, we can provide practical correction method for the pronunciation of the English learners in SD dialectal regions. According to the collected corpus, we can systematically compare the pronunciation of the consonants and the vowels of the speakers in dialect region with the acoustic characteristics of the standardized English speaker. Through applying the acoustic parameters in the phonetic research, we can classify the types of the error and make personalized study program.

(iv) It is more significant in the theoretical and practical aspects for the integrated study of the pronunciation of the English learners in the dialect area. This research synthesizes the second language acquisition, the corpus and the phonetic analysis into the study of English learning in dialect region. Comparing with the previous dispersion research, the study of the accent problem of the English learners can improve the level of the English learners more practically and effectively by combining many research advantages in many research fields in English learning.

(v) It provides the evidence from the phonetic perspective, and it may enrich the second language acquisition theory. Referential examples for the previous second language acquisition researches mostly come from the classroom teaching directly, and the data is in limited amount; while based on this study, the corpus established of the English learners in the dialect areas which provides extensive speech data and data sources for quantitative analysis for the second language acquisition research. Therefore, it provides more evidence to explore the ‘positive transfer’ and the ‘negative transfer’ features of the second language speech acquisition, as well as the development of the second language acquisition theory.

(vi) The ways and means proposed on the basis of the phonetic research to improve the spoken English of the people in the dialect district are more targeted and more in line with the status of the Chinese English learners. When the previous English learners in the dialect districts were faced with the pronunciation problem, they can only refer to their English teachers for help, so the teaching effect was different according to the English level of the different teachers. With the method of improving the English pronunciation through the phonetic and phonological analysis, the English learners from dialect districts can find the ways to deal with those problems, such as to correct the consonants in the place of articulation, the vowels in the opening degree, the distribution and the realization types of the stress, etc. Therefore, the application of the results, as to the English teachers of the primary and secondary schools, can help to improve the quality of the English teaching; as to the majority of the English learners, can be targeted to improve their pronunciation level.

5. ACKNOWLEDGEMENTS

This search is financially supported by Innovation Program of Chinese Academy of Social Sciences (AESOP—Corpus and Phonetic Study of English Learners from Chinese Dialect Regions).

6. REFERENCES

[1] Gimson, A.C. “An Introduction to the Pronunciation of English”, 2nd Edition. Edward Arnold Press, 1976.

[2] Li, Rong. “Hanyu Fangyan de Fenqu (The classification of Chinese dialect regions)”, Dialect, 4, 241-259, 1989.

[3] Qian, Zingyi. “Shandong Fangyan Yanjiu(Study of Shandong dialect)”, Qilu Press, 2001.

[This paper was published in O-COCOSDA 2011]