

A comprehensive corpus of three varieties of Jin Dialects, Chinese (Yu Dialect, Pingding Dialect, Jiaoqu Dialect)

This corpus provides the data that reflects the phonological system of three dialectal varieties spoken in the City of Yangquan, Shanxi, China. It provides the data of my dissertation (Jia, 2021).

Contents of the corpus

This corpus includes data of the following six sections of each dialectal variety:

Table 1. The Contents of the Corpus.

Sections	Number of files	Description of the data
(1) Initial system	YD: 113 PD: 114 JD: 114	The pronunciations of monosyllabic characters cover all initials in each dialectal variety.
(2) Final system	YD: 128 PD: 110 JD: 109	The pronunciations of monosyllabic characters cover all finals in each dialectal variety.
(3) Tonal system	YD: 228 PD: 255 JD: 108	The pronunciations of monosyllabic characters cover all tone categories in each dialectal variety.
(4) Syllabic system	YD: 2025 PD: 952 JD: 823	The pronunciations of monosyllabic characters that cover all syllables which are allowed by the constraints in the phonological system and are really attested in each dialectal variety
(5) Disyllabic tone sandhi	YD: 490 PD: 738 JD: 310	The attested pronunciations of the disyllabic segments in combinations of all tonal categories in each dialectal variety.
(6) Other materials	YD: 36 PD: 129 JD: 8	In the language investigation, trisyllabic tone sandhi and sentence intonations were also investigated. But due to the time limit, they have not been fully sorted out. Therefore, only the materials used in my dissertation are provided for the time being. The systematic material on trisyllabic tone sandhi and sentence intonation will be provided later.

(YD: Yu Dialect; PD: Pingding Dialect; JD: Jiaoqu Dialect)

1. Language Investigation

This corpus is built based on the result of a language investigation conducted from March to April 2019.

i. Investigator: Pingping Jia

ii. Informants:

Table 2. The information of Informants.

	Location	Age	Gender	Occupation	Education
Informant1	Jiao-qu – Village of Yuxiagou	64	Female	Farmer	Primary School
Informant2	Pingding (Near to Jiao-qu) – Village of Luomianzui	64	Male	Wireman	Primary School
Informant3	Pingding – Village of Libi	68	Male	Farmer	Primary School
Informant4	Pingding – Village of Libi	63	Female	Farmer	Primary School
Informant5	Yuxian – Xiushui Town	74	Female	Farmer	Primary School
Informant6	Yuxian – Xiushui Town	55	Male	Businessman	Middle School

iii. Investigation sites:

- a. Yu Dialect: the Town of Xiushui, the County of Yu, Yangquan City, Shanxi Province, China
- b. Pingding Dialect: the Village of Libi, the County of Pingding, Yangquan City, Shanxi Province, China
- c. Jiaoqu Dialect: the village of Yuxiagou, Jiaoqu District, Yangquan City, Shanxi Province, China
- d. Jiaoqu Dialect: the village of Luomianzui, the County of Pingding, Yangquan City, Shanxi Province, China

Detailed locations are shown on the map below.



审图号：晋S(2018)021号

2. Retrieval methods

This corpus provides data lists for each of the three dialectal varieties. The data lists of each variety contain all six sections described in Table 1. In these lists, every monosyllabic, bisyllabic, or polysyllabic segment is provided with a Data Number. This number is followed by the basic information of this segment.

Section 1 provides the IPA of the initials and the syllables, the Chinese characters, and the English glosses.

Section 2 provides the IPA of the finals and the syllables, the Chinese characters, and the English glosses.

Section 3 provides the IPA of the syllables, the tone category of the characters or syllables in Medieval Chinese and the current dialect, the Chinese characters, and the English glosses.

Section 4 provides the IPA of the initials, of the finals, and of the syllables, the tones, the Chinese characters, and the English glosses.

Section 5 provides the citation tones and sandhi tones of each disyllabic segment, the IPA, the Chinese characters, and the English glosses.

Section 6 provides the Chinese characters, English glosses, IPA, and the notations of Pinyin.

Each file is stored in the format of *wav*. and named with a data No.. A *wav*. file can be retrieved through the Data No. on the data lists, or the basic information can be checked on the data lists according to the name of the *wav*. file.