OLD CHINESE DIALECTS ACCORDING TO FANGYAN

Takashi MATSUE
Hokkaido University, Japan
matsue@let.hokudai.ac.jp

Abstract
This paper discusses Fāngyán, the oldest dialectal dictionary in China, considered to have been authored by Yāng Xiōng (53 B.C.-A.D. 18). It describes how Fāngyán was compiled as well as its basic structure and its linguistic value, and then introduces actual examples of studies of old Chinese dialects that have used Fāngyán as a linguistic resource in documenting lexical histories and in determining linguistic distance between dialectal regions.

Keywords
Old Chinese dialects, linguistic geography, Fangyan, Yang Xiong

LOS DIALECTOS CHINOS ANTIGUOS SEGÚN EL FANGYAN

Resumen
Este artículo examina el Fāngyán, el diccionario dialectal más antiguo de China, atribuido a Yang Xiong (53 a.C.-18 d.C.). Se describe cómo fue compilado el Fāngyán, así como su estructura básica y su valor lingüístico, introduciendo posteriormente ejemplos reales de los estudios de antiguos dialectos chinos que han usado el Fāngyán como un recurso lingüístico en la documentación de historias léxicas y en la determinación de la distancia lingüística entre regiones dialectales.

Palabras clave
dialectos chinos antiguos, geografía lingüística, Fāngyán, Yang Xiong
1. Introduction

The history of dictionaries in ancient China begins with Ėryǎ 爾雅, which is considered to have been created sometime between the end of the second and the beginning of the first century B.C. From the last half of the Eastern Zhou period (770-256 B.C.) through the Former Han period, there was a custom of giving semantic explanations for abstruse words written in scriptures. This custom, known as xùngū 訓詁, produced two specific forms: Interpretation with phrases or simple sentences, and explanation with synonymous words. Ėryǎ is a dictionary that includes a comprehensive collection of these xùngū, grouped into lexical items. Fāngyán 方言, which will be introduced in this paper, is a dictionary of the same general nature as Ėryǎ. However, it is unique in the following two aspects:

1. It specifies the areas of distribution of the words included;
2. Its entries reflect contemporaneous slang (colloquial expressions) as well as minority languages.

This paper aims to demonstrate the value of Fāngyán as a linguistic resource on the old Chinese dialects, and to illustrate with actual studies conducted on Fāngyán.

2. About Fāngyán

2.1 General Description of Fāngyán

The formal name of Fāngyán is Yóuxuānshǐzhějuédài yǔshìbiéguófāngyán 輻軼使者絕代語釵別国方言. It is widely believed that the dictionary was compiled by Yáng

---

1 In this paper, the Chinese names of books, names of people, and names of areas (ancient names of areas), as well as the Chinese word forms, will be spelled in Pinyin, which is widely used in China.
2 There are various theories as to the completion of Ėryǎ. The theory that it was already completed before the second century B.C. is also widely supported. This issue is not to be discussed in this paper.
3 In addition to dictionaries like Ėryǎ that organized entries according to lexical items, other ancient Chinese dictionaries catalogued their contents according to the sounds or structures of characters. The former type of dictionaries are said to have been created by around the third century, but the earliest examples are lost today. However, they are lost as of today. The oldest in existence is Qièyùn 切韻 (AD601), authored by Lù Fāyán 陸法言. The oldest among the latter type of dictionaries is Shuōwénjízì 說文解字 (ca. AD100) authored by Xū Shèn.
Xióng 揚雄 (53 B.C.-A.D. 18), who lived in the late Former Han period. The compilation process of Fāngyán can be roughly summarized as follows.

From the Zhou period (ca. 1046-256 B.C.) through the Qin period (221-207 B.C.), the imperial court sent emissaries to various regions every August to collect ancient words and regional words. When the Qin Dynasty was overthrown, these words were discarded and disappeared from sight. However, in Shǔ, which was Yáng Xióng’s hometown, there was a person who stored over 1,000 such words. Further, there was a person who knew how to sort those collected words. Yáng Xióng took an interest in these words and decided to investigate regional vocabulary himself. Whenever there were gatherings of government clerks and soldiers, he went there to search for regional words and sorted them. His investigation went on for twenty-seven years, and he collected some 9,000 characters in total.

The above shows that most of the words included in Fāngyán are regional words actually used in the late Former Han period — words that Yáng Xióng himself collected through field surveys — and a portion of those are regional words dating from before the Han period.

What motivated Yáng Xióng to perform such surveys was his eagerness to make accurate interpretations of the abstruse words used in Confucian scriptures. In other words, Yáng Xióng believed that old words that had already been lost in the common
language were still preserved in the dialects, and that they were useful in correctly understanding the Confucian scriptures.

2.2 Basic Structures of the Contents

The entries in Fāngyán come in several different patterns. Following is an example of one simple type:

・黨、曉、哲，知也。楚謂之黨，或曰曉，齊宋之間謂之哲。(Volume1 of Entry1)

Here is a translation of this entry:

The word forms dǎng 黨, xiǎo 曉, and zhé 哲 all belong to the lexical item zhī 知 (= they all mean the same). In the area called Chǔ 楚, the word dǎng 黨 is used. Some use xiǎo 曉 (in this area) as well. In the areas such as Qí 齊 and Sòng 宋, zhé 哲 is used.

As this example shows, we can obtain detailed information on a certain lexical item, including its standard word form, its regional word forms, and their areas of distribution. This is only one of the several patterns, but nearly all of them provide the same general information.6 In Fāngyán, there are 263 total lexical entries that specify the areas of distribution of regional words, as Entry 1 of Volume 1 does. The dictionary includes a wide range of word types, from abstract ones such as adjectives and verbs, to concrete names of weapons, clothing, housewares, animals, and insects.

2.3 Value as a Linguistic Resource

As described in § 1.2, the entry provides evidence that dǎng 黨 and xiǎo 曉 are used in Chǔ as the regional forms of the standard word zhī 知, and zhé 哲 is used in the area straddling Qí and Sòng. Fāngyán can be used as a linguistic resource on old Chinese dialects if the following three things are possible from its records:

(1) Determination of the meaning of the entry in question,

---

6 However, for most of the words included in Volume 12 and Volume 13, the areas of distribution are not specified. This indirectly indicates that Fāngyán is an unfinished work. The contents of the correspondence exchanged between Yáng Xióng and Liú Xin also implies that Fāngyán is unfinished.
(2) Presumption of the actual geographical locations of the areas of distribution described by the names in the entry,

(3) Reconstruction of the sounds of each word form.

With regard to (1), annotations were made by Guō Pú 郭璞 (A.D. 276-324) at an early stage, and many scholars made subsequent additions. The most comprehensive and systematical of these annotations is the work called Fāngyánshūzhèng 方言疏訶 by Dài Zhèn 戴震 (A.D. 1723-1777), a philologist from the Qing period. In it, Dài Zhèn empirically examined the meanings of all the entries included in Fāngyán. It retains its value as a research material even today; for example, Dài Zhèn points out that the “知” here represents the word form that is usually written with the character zhì “智”. It is my own belief that Dài Zhèn’s theory is basically correct. Based on this, and based on how regional words such as dǎng 黨, xiǎo 曉, and zhé 哲 are used in ancient Chinese texts, I would like to propose that the meaning of the words included in this entry is “to have a thorough knowledge (of the nature of things).”

With regard to (2), the names of the areas described in Fāngyán include ones actually used in the late Han period, but many of them are from the Zhou period, which poses a problem. Geographical ranges of the names used in Zhou times vary significantly at different periods. Therefore, the question here is which geographical range the names used in Fāngyán are referring to, and in which period. Philological studies have made striking progress in resolving this question. Thanks to studies such as Dīng (1991), Liú (1992), and Huà (2006), many questions are being clarified.

With regard to (3), which is the issue of reconstructing the sound system of proto Chinese languages, Berhard Karlgren (1889-1978) laid the groundwork for this project when he succeeded in reconstructing Ancient Chinese used around 600 A.D. and Archaic Chinese used around 800 B.C.; and the results came to fruition in Karlgren (1957). Thanks to the efforts of many researchers since then, the accuracy of the reconstruction is much improved today. However, many of the regional words included in Fāngyán are ones used in the late Former Han period, belonging neither to

---

7 Satō (2000: 21) translated this lexical item as “知恵のあるさま (state of having wisdom)”. This paper uses this interpretation as a reference as well.
Ancient nor Archaic Chinese, but rather falling somewhere between them. The accurate reconstruction of the phonology of the Han period is a task to be addressed in the future. In this paper, I will use the presumed sounds of Old Chinese (Chinese used between the eleventh century B.C. and third century B.C.), which was reconstructed by Schuessler (2009). Then, the words in Entry 1 of Volume 1 shown above will be reconstructed as .zhí 知 (智) [*treh], .dtdǎng 黨 [*tâŋʔ], .d.xiǎo xiǎo [＊hiâuʔ], and zhé 哲 [*trat].

As described above, 方言 is a valuable resource on old Chinese dialects. However, it has certain limitations as a dialect resource:

(i) The number of places investigated is small, and the regional vocabulary data are neither systematic nor comprehensive.

(ii) Because it is an ancient text, there are issues such as how the meanings of the words included in the entries should be interpreted, or what geographical ranges the area names actually refer to.

Despite these flaws, many have conducted productive studies of old Chinese dialects using 方言 as a linguistic resource. I will briefly describe one of these linguistic studies in the following section.

3. Studies of the History of Words

As described above, 方言 offers data regarding regional words used in the Han period and their areas of distribution. This alone makes it an important resource for studies of the history of words. However, we will be able to obtain more concrete information if a map of the Chinese dialects used in Han times can be created based on 方言. In this section, I will explain in specific terms what kind of information 方言 can provide for studies of the history of Chinese words.

---

8 As stated in § 2.1 of this paper, 方言 actually includes regional words used before the Han period. Therefore, it would be technically preferable to say “the Han period and some parts of certain periods before that,” but for the sake of convenience, I will merely say “the Han period” here.
3.1 Comparison with the Map of Modern Chinese Dialects

If the lexical items listed in Fāngyán are mapped by the studies of modern Chinese dialects, comparing the two maps can provide us with important information. Here, I would like to use as an example the character from Fāngyán’s (Entry 1 of Volume 1), zhì 知, which means “to have a thorough knowledge (of the nature of things).” Map 1 was created based on the data obtained from Entry 1 of Volume 1 and shows the Chinese regional words used in the Han period.

Map 1. Map showing the regional words for the lexical item zhì 知 [*treh] ‘to have a thorough knowledge (of the nature of things)’ in the Han period

To my knowledge, no map exists for the same lexical item in modern Chinese dialects. However, there is a map showing regional words for the lexical item zhīdào 知道 (‘to know’) (Iwata 2012: 140-141), which is semantically very similar to zhì 知. In that map, a regional word form that uses the morpheme xiǎo 曉 is also distributed in the area corresponding to the Chǔ dialect area of the Han period. Although there is a semantic difference between Map 1 and Iwata’s (2012) dialect map created by Iwata, it is extremely likely that xiǎo 曉, seen in both of them, is the same morpheme. It can be therefore presumed that xiǎo 曉, which means ‘to have a thorough knowledge (of
the nature of things)’ or ‘to know,’ has existed in this Chǔ dialect area for 2,000 years — since the Han period.

We should be able to obtain more information on other lexical items in the future, as we make further progress mapping modern Chinese dialects.

3.2 Use of Geolinguistical Method

In § 3.1, I used the method of comparing maps of Chinese dialects from the Han period and of modern Chinese dialects. Now, I would like to examine whether it is possible to analyze the map of Chinese dialects from the Han period using a geolinguistical method.9

For example, let’s take a look at Map 2, which shows distributions of the lexical item ài 愛 (Entry 17, Volume 1), which means ‘to care for.’ In this map, the word form jí 於 is distributed in the western areas, Qín and Jín, and in the eastern areas, Dōngqí and Hǎidài. The word forms wǔ 五 and yán 言 are distributed in the areas between them. At first glance, it seems that this exhibits what Japanese dialect geographers call “ABA distribution” (i.e., jí 於 is A, and wǔ 五 and yán 言 are B; see Iwata (2010: 20-21; 31-33). If this is true, it can be presumed that jí 於 is relatively older than wǔ 五 and yán 言.

Unfortunately, however, it is difficult to determine that it is an ABA distribution based on Map 2 because the data in Fāngyán are not all-encompassing. In fact, Map 2 does not contain data for the areas north of the Yellow River. If jí 於 were distributed in the areas north of the Yellow River, and if the distributions of jí 於 in the west and in the east were connected, an ABA distribution would not apply to these word forms. As just described, due to imperfection of Fāngyán as a linguistic resource, even if we apply a geolinguistical method to dialect maps created based on Fāngyán, its effects are limited. Nonetheless, even if we cannot definitively determine the actual distributions, we can still use geolinguistical evidence to identify and rule out

9 Serruys (1959) has also conducted a geolinguistical study of Fāngyán.
possibilities, and to this extent, geolinguistical research has an important supplementary role to play.

Map 2. Dialect map of the lexical item $\dot{a}i$ 爱 ('to care for') in the Han period

3.3 Comparison with Words of Minority Languages

Some of the words included in $Fängyán$ do not have a Chinese character of their own, which means that these words appear in no or very few ancient Chinese texts except for $Fängyán$. These words are Chinese slang words or the words of minority languages. The fact that these words are recorded in $Fängyán$, which was completed approximately 2,000 years ago, is extremely important in considering the history of slang words and the words of minority languages.

In $Fängyán$, Yáng Xióng used two methods to describe word forms that did not have a Chinese character of their own. One involved applying a character used to write another word form as a phonogram; the other entailed creating a new character. Let me introduce an example of the former method.

$Fängyán$’s (Entry 1 of Volume 8) gives us the information that, for the lexical item $hū$ 虎 ‘tiger’, the character $lìfù$ 李父 [$*rəʔpaʔ$] (or [$*rəʔbaʔ$]) is used in areas such as Chén, Wèi, Sòng, and Chǔ, and $lìěr$ 李耳 [$*rəʔnəʔ$] is used in areas such as Jiānghuái
and Nánchǔ. The Chinese characters such as “李”, “父”, “耳” used in this entry do not represent the words that they normally do. Here, they are used as phonograms to describe morphemes having the same sounds (or similar sounds). According to Zhāng (1988), “李” means ‘tiger,’ “父” means ‘male tiger,’ and “耳” means ‘female tiger’ in the Tujia language. This implies that “李父” and “李耳” seen in Volume 8 of Fāngyán represent some words of ancient minority languages which are etymologically related to the “父” and “耳” of today’s Tujia.

As shown above, studies on Fāngyán can offer insights into the historical content of Chinese minority languages and popular slang from 2000 years ago — items not typically found in texts — as well the geographic areas across which those words were distributed. Fāngyán includes quite a number of these rarely encountered words.

4. Studies of Linguistic Distances between Areas

4.1 Discussion of Demarcation

Next, I will introduce studies that have statistically analyzed the distribution of lexical items in ancient China and examined the linguistic distances between demarcated areas.

Among these studies, the most widely discussed subject may be demarcation based on Fāngyán. The earliest of these is a study by Lin (1927), who proposes a demarcation into fourteen dialect areas. Following is a summary of the method that he used to come to that conclusion:

1. Each area name appearing in Fāngyán will be regarded as one unit. When the same regional words are frequently distributed in Area A and Area B, it will be regarded as a phenomenon that indicates that A and B are linguistically proximate.

2. When an area has many regional words that are distributed only in that area,

---

10 See Map 3 for the actual geographical ranges of the areas.
11 Although the difference between “父” and “耳” used in today’s Tujia language is a male-female distinction, the difference between “李父” and “李耳” is regarded as a regional one. How to interpret this discrepancy remains an issue.
it will be regarded as a phenomenon that indicates that the area is linguistically distant from adjacent areas.

Since then, many researchers, including Serruys (1959) and Liú (1992), have proposed their own demarcation theories. Their studies adhere fundamentally to the method used by Lin (1927). However, some have further discussed cultural and historical elements.

There are some studies that have attempted to clarify the differences between dialects’ sound systems based on such demarcation theories. For example, Fāngyán’s (Entry 12 of Volume 11) talks about the lexical item yíng 翼 ‘fly,’ and it says that yáng 羊 is used in Dōngqí as a word form that corresponds to the standard form yíng 翼 used in the areas such as Chén, Chū, Qín, and Jìn. The character “羊” is evidently used as a phonogram. Yíng 翼 [*ləŋ] and yáng 羊 [*jaŋ] were probably recognized as different word forms because the sound systems were quite different between the dialects, even though they were originally the same word. Based on this, Dīng (1991: 87) believes that a syllable with the final [*-aŋ] used in the standard language sometimes changed to [*-ioŋ] in the dialect area that included Dōngqí (i.e., the Hāidái dialect area). Dīng’s (1991) study is an extremely meaningful one. However, “every word has its own history”, as is well known. Therefore, there is some possibility that the above is a phenomenon only seen with the lexical item yíng 翼, and that it has nothing to do with the differences among dialects’ sound systems between the dialects.

4.2 Matsue’s Calculation of Distances between Dialects

As stated above, many researchers have proposed their own theories on the demarcation of dialects. However, the second element of Lin’s (1927) method,
explained above in § 4.1, is problematic. When some regional words are distributed only in a certain area, as explained in (2), there are two theoretically possible cases: a case in which an isogloss exists between the area and the adjacent areas, and a case in which no isogloss exists. Only the former should be used as an index of linguistic remoteness. To distinguish between the former and the latter, it is necessary to create a map for each lexical item.

Matsue (1999) presented a collection of maps for each entry. These are the maps that show the distributions of regional words for basically every lexical item although there are some incompletions because semantic interpretations of the entries are not included and the sounds of the regional words are not reconstructed. These have enabled us to estimate the locations of the isoglosses for each lexical item in the Han-period dialects and calculate the linguistic distances between them by quantitatively analyzing the isoglosses between the areas.

In Matsue (2006), the linguistic relationships of the areas described in Fāngyán with adjacent areas were measured using the following two indexes, with reference to the maps created in Matsue (1999).

(1) Index $\alpha$: Frequency of the border of two adjacent areas overlapping with the isogloss of regional words (i.e., linguistic discontinuity).

(2) Index $\beta$: Frequency of the same regional words being distributed over the border of two adjacent regions; that is frequency of being able to determine that the border of the two areas does not overlap with the isogloss of the regional words (i.e., linguistic continuity).

The results are shown below in Map 3, where curved lines separating two areas indicate that the value of $\alpha$ between the two areas is high (10 points or higher), which means that linguistic discontinuity is strong. Arrows connecting two areas indicate that the value of $\beta$ is high (10 points or higher), which means that linguistic continuity is strong. The thick line indicates that linguistic discontinuity is extremely strong, whereas the thick arrow indicates that linguistic continuity is extremely strong.\(^\text{16}\)

\(^\text{15}\) Technically speaking, some of the entries include two or more lexical items, or vice versa. However, this will not be discussed in this paper.

\(^\text{16}\) See Matsue (2006) for specific values of $\alpha$ and $\beta$. 

©Universitat de Barcelona
Map 3 can be interpreted from various perspectives. In this paper, I would like to point out the following three notable characteristics:

(i) There is a noticeable south-north contrast in modern Chinese dialects. However, it is evident from Map 3 that the east-west contrast is more noticeable in old Chinese dialects. For example, in areas between the Yellow River and the Huai River, there are east-west contrasts in several separate blocks, as seen between (1) Qín and Jin, (2) Hán and Zhèng, (3) Wèi and Sòng, and (4) Dōngqí and Hǎidài.¹⁷

(ii) Modern Chinese dialects have a large-scale bundle of isoglosses along the Yangtze River and the Huai River (see Iwata 2010b, etc.), whereas no such bundle is

¹⁷ Ōnishi (2003) points out that the old Chinese dialects also have an east-west contrast in terms of grammar.
clearly shown in Map 3. However, there is a certain level of linguistic discontinuity between Hāidài and Jiānghuái, and there is also linguistic discontinuity between Chū and Nánchū, indicating the possibility that these may be the beginnings of the isoglosses later known as the Huai River line and Yangtze River line, respectively.

(iii) Uniqueness of the Chū dialect area (Chū and Chén) was emphasized with demarcations discussed in the previous studies. This is certainly correct because there is indeed extremely strong linguistic discontinuity between Chū and Qín. On the other hand, however, there is also a certain level of linguistic continuity from Chū to Chén, Sòng, Lù, and Qí. Apparently, Chū had linguistic connections with the areas to its north.

4.3 Presumption of Dynamics: Presumed Possibility of Mutual Influences between Areas

In § 4.2, I discussed the synchronic situation of the Chinese dialects of the Han period. However, in some cases, diachronic changes can be presumed from the dialect maps created based on Fāngyán. I will introduce concrete examples in this section.

When looking at the three areas Jiānghuái, Chū, and Nánchū in Map 3, we can see that there is strong linguistic discontinuity between Chū and Nánchū, but there is strong linguistic continuity between Jiānghuái and Chū and also between Jiānghuái and Nánchū. Superficially, these seem to be logically contradictory.

However, when paying attention to the data from each lexical item, we can tell that the above phenomena are not logically contradictory. To explain in concrete terms:

(1) When the same word is distributed in Jiānghuái and Nánchū, either there is no data regarding Chū, or another word is distributed there;

(2) When the same word is distributed in Jiānghuái and Chū, either there is no data regarding Nánchū, or another word is distributed there.

I suggest that (1) and (2) reflect different language layers, and, assuming this is true, the above phenomena can be understood with no logical contradictions. From the historical fact that Chū expanded its influence to the north during the Eastern Zhou period, and from the descriptions in ancient Chinese texts, it can be speculated that Chū and Nánchū had a close relationship from the beginning. Based on such
observation, I suggest that the relationship between Jiānghuái, Nánchǔ, and Chǔ went through the following process.

(i) First Phase: Jiānghuái, Chǔ, and Nánchǔ were all linguistically proximate.

(ii) Second Phase: Chǔ expanded its influence to the northeast, generating dialectal contacts with the areas in the northeast, resulting in the absorption of a large number of new words from the northern areas. As a result, its linguistic distance with Jiānghuái and Nánchǔ became remote.

(iii) Third Phase: Chǔ also expanded its influence to the east and gave new words to Jiānghuái. As a result, its linguistic distance with Jiānghuái became proximate again.

I therefore propose that the regional words common to Jiānghuái and Nánchǔ belong primarily to the old layer (first phase), and the regional words common to Jiānghuái and Chǔ belong to the new layer (third phase). The above is just one hypothesis, and further scrutiny into these questions is necessary. However, questions like these are important ones, and as these kinds of studies accumulate in the future, we may be able to obtain a great wealth of information from Fāngyán.

6. Conclusion

In this paper, I introduced Fāngyán and reviewed studies of proto Chinese languages conducted based on it. It is expected that there will be major advancements in studies of lexical history as well as of linguistic distances between areas. These will depend, of course, on further developments and modifications of the original maps created in Matsue (1999), including phonological reconstructions and semantic interpretations of the entries.
References

DÀI, Zhèn 戴震 (ca.1777) Fângyánshûzhèng 方言疏证 [Interpretations of Fângyán and Examples], in Dài Zhènquánji Vol. 5, Beijing: Qinghuadaxuechubanshe, 2304-2487.

DÎNG, Qîzhèn 丁啓震 (1991) Qin'hàn fângyán 秦漢方言 [Chinese Dialects of the Qin-Han Period], Beijing: Dogfangchubanshe.


LÎN, Yûtâng 林語堂 (1927) “Qiánhànfângyínqûyûkâo” 前漢方音区域考 [Studies of Chinese Dialect Areas in the Former Han Period] , in Yûyánxuélûncóng 語言學論叢 (1933), Shanghai: Kaimingshudian, 16-44.

LIÛ, Jûnhui 劉君惠 (eds.) (1992) Yáng Xióngfângyánynjî 揚雄方言研究 [Studies of Fângyán], Chengdu: bashushushe.

MATSUE, Takashi 松 江 崇 (2006) “Hàndàifǎngyánzhōng de tóngyánxianshù-yētángēnjùFāngyán de fāngyánqūhuàlùn” 漢代方言中的同言線束——也談根據方言的方言區劃論 [Isogloss in the Han era—with discussion of the dialect demarcation according to Fāngyán], in Hhuáxuéchéng (eds.), 1509-1533.


