The _FARHYAN_ (‘Dialect’, ‘Toptolect’), usually attributed Yang Xiong (53 BC–18 CE), a famous赋-poet and philosopher, is a collection of dialectal and regional expressions compiled during the end of the Western Han period (206 BC–9 CE, Norman 1988:185). It is the first attested study on linguistic geography and dialectology in China, possibly even worldwide (Wang 1980[2006]:17, Li 2004:1). Published under the baroque title “Dialectal Expressions from Foreign States and Glosses on Words from Extinct Eras Collected by the 'Light Chart Officials’” (Youxuan shizhe juedai yu shi bieguo farhyan 輶軒使者絕代語釋別國方言, for details regarding the title see Behr 2005: 23 and n. 36), the work is a remarkable witness of early linguistic diversity in China, and it is likely that the modern Chinese term _fARHYAN_ “dialect” goes back to the title of the work.

_Time of Creation and Authorship_

According to his own account, Yang Xiong collected the data during a period of 27 years from soldiers and candidates for imperial examinations who entered Chang'an 長安, the capital at that time, from all parts of the country (Wang 1980[2006]:17, Li 2004:1). Given Yang Xiong's biographical data and the long time that he invested into the collection of the data, the compilation of the _FARHYAN_ is usually assumed to date back to the early 1st century CE (Wang 1980[2006]:17). Some scholars, however, have raised doubts whether Yang Xiong was the real author of the book after all, since he is not mentioned as such in the earliest historiographical records (Liu et al. 1992:9–16, Huá 2007a:60-61). Despite these uncertainties, however, it seems safe to assume that the book was compiled towards the end of the Western Han period, some time before 24 CE.
In its basic structure the *Fāngyán* somewhat resembles the *Ěryǎ*, a collection of semantic glosses compiled before the Qín dynasty (221–206 BC, Malmqvist 1995:224-225). In contrast to the *Ěryǎ*, the characters do not belong to the same but to different lexical strata, including dialectal (Wáng 1980[2006]:1) and occasionally also diachronic strata of older varieties of Chinese (Lǐ 2004:19-20). Based on semantic principles which are generally rather loose (for a detailed account on the principles see Huá and Xú 2013:150), the semantic glosses of the *Fāngyán* were originally collated into 15 chapters, of which 13 survive in the extant editions (Lǐ 2004:18, Huá and Xú 2013:150, Yōng 2010: 31). Different dialect words (1666 in total, Wáng 2011b:37) are usually first listed and then explained with help of a more common gloss word (367 in total, Wáng 2011b:37), following the scheme:

甲，乙，丙，某 也
jiǎ, yǐ, bǐng, mǒu yě
‘A, B, and C are [all] X’.

In a second step, the dialect words are assigned to dialect locations or regions, following the template:

甲地，乙地 （之間） 謂 之 某
jiǎdì, yǐdì (zhījiān) wèi zhī mǒu
‘in (the area of) place A and place B one says X’,

as illustrated in (1). This principle, however, is not being strictly followed in all of the glosses. Sometimes words are only glossed without giving specific geographic information; sometimes
the assignment of words to dialect locations precedes the actual glossing, and sometimes, the dialect words are not specifically listed, as shown in (2). Apart from the basic structure of the glosses in the Fāngyán, the examples in (1) and (2) also illustrate how difficult it is to interpret the work, given that the Chinese writing system provides only limited, since indirect, use in reflecting the actual pronunciation of the dialect words. As a result, one needs to rely upon reconstructions of older stages of Chinese in order to obtain an approximate pronunciation, like Middle Chinese (here rendered in Baxter's 1992 transcription system) or Old Chinese (here cited in the reconstruction of Baxter and Sagart 2014).

(1) 黨、曉、哲，知也。楚謂之黨，或曰
dāng, xiǎo, zhé, zhī yě. Chǔ wèizhī dāng huò yuē
曉，齊宋之間謂之哲。
xiǎo, Qí Sòng zhījiān wèizhī zhé.
‘The words dāng (MC *tangX < OC *tˤəŋʔ?), xiǎo (MC *xewX < OC *qʰˤəwʔ?), and zhé (MC *trjet < OC *tr[e]t) all mean “know, understand”. In Chǔ they speak of dāng, or alternatively call it xiǎo. In the regions of Qí and Sòng they speak of zhé.’ (Fāngyán: 1.1)

(2) 貔，陳楚江淮之間謂之□□，北燕朝鮮
pí, Chénchū Jiāng Zhūn zhījiān wèizhī lái, Běiyān Cháoxiān
之間謂之□□，關西謂之狸。
zhījiān wèizhī péi, Guānxī wèizhī lái.
‘For “wild cat” (pí 貔), one says lái □□ (< MC *lai < OC *rˤəj) in Chén, Chū, Jiāng, and Zhūn, in the area of Běiyān and Cháoxiān one says péi □□ (< MC *pij < OC *braj), and in Guānxī one says lái 狸 (< MC *lai < OC *p.rə).’ (Fāngyán: 8.2)
Editions

The first known commentary on the Fāngyán was written by the famous Eastern Jin scholar Guō Pú 郭璞 (276–324) and the oldest complete editions which are still preserved today are from the Sòng 宋 dynasty (960–1279). For a long time, Zhōu Zūmó’s 周祖謨 (1914–1995) Fāngyán Jiàojiān 方言校箋 [Collation and Annotation to the Fāngyán] of 1956 served as an authoritative edition for contemporary investigations. Zhōu closely followed the edition of the Sòng scholar Lǐ Mèngchuán 李孟傳 (1136–1219) compiled in 1200 (Wáng & Huá 2006:163), supplemented by a great deal of the rich commentary literature on the Fāngyán which was produced during the Qīng dynasty (1636–1911). During the last two decades, several shortcomings of Zhōu’s edition were noted and criticized, including an insufficient treatment of the diverse commentaries and an improper treatment of additional sources (Huá 2007b). In two recent editions of the Fāngyán, scholars have been trying to overcome these problems. The edition by Satō (1998) compares Lǐ Mèngchuán’s edition with three additional ones from the Sòng dynasty, two of which were not included in Zhōu’s edition from 1956. A more recent edition by Huá (2006) additionally includes Wáng Niànsūn’s 王念孫 (1744 – 1832) commented version of an old Fāngyán edition from the Míng 明 dynasty which was long thought to be lost.

Terminology

The specific terminology which is used in the Fāngyán is still not fully understood and has instigated multiple debates in the world of Chinese linguistics. Apart from geographic and administrative terms which are used in order to assign words to dialect areas and regions, many discussions center around a couple of terms in the Fāngyán which further specify and
characterize the usage of dialect words. The term zhùnyǔ 轉語 (literally something like 'twisted expressions'), for example, occurs six times in the Fāngyán and is usually thought to refer to "reading variants", that is, reading differences resulting from diachronic and diatopic variation (Wáng 1989:24). This chimes well with the opinion of many scholars, that at least part of the word collections in the Fāngyán are in fact sets of etymologically related words (Wáng 1980[2006]:24-25, Wáng 1989:24-25, Serruys 1962:321-322). The term lived on in the tradition of Fāngyán commentaries and inspired later Chinese scholars, like for example, Qián Yì 錢繹 (1770-1885) in his Commentaries to the Fāngyán (Fāngyán jiānshū 方言箋疏), to carry out detailed etymological studies of the word entries in the Fāngyán (Behr 2007). The usage of the term in the Fāngyán itself, however, does not necessarily justify such an interpretation: In entry 10.6 in the Fāngyán, for example, huǐ 灰 (MC *xwoj < OC *[qʰʷəjʔ]) is described as a zhùnyǔ of huǒ 火 'fire' (MC *xwa < OC *[qʰʷəjʔ]) in the region of Chǔ 楚. In entry 10.44, however, xiè 續 (MC *sjet < OC *sat), mò 末 (MC *muwk < OC *məat), and jì 紀 (MC *ki < OC *k(r)əʔ) are glossed as xù 緒 'thread' (MC *zjo < OC *s-m-taʔ) and also labelled as zhùnyǔ in the Chǔ region. While an etymological relation between the two words in the former case (10.6) can be safely assumed, even solely based on Middle Chinese readings, this is clearly not the case for the latter example (10.44), where neither Middle Chinese nor Old Chinese readings point to etymological relations between any of the words. Even when disregarding the question whether people in Hán times had a sufficiently clear idea of the historical dimensions of linguistic diversity, it seems thus rather unlikely that the term was used to denote etymological relations, and this seems also to hold for the use of the term in the commentaries of Guǒ Pú (judging from the data reported in Wáng 1986). Wáng (2011a:103) proposes that the term zhùnyǔ was occasionally used to denote semantic similarity between words, but this does not seem very likely either, given that semantic similarity should hold for all dialect words occurring in the same gloss in the
Another term that is frequently used in the *Fāngyán* is *tōngyǔ* 通語 ‘common language’ which occurs 34 times and is used as a label for specific dialect words, following the glossing template

\[
\text{甲、乙，通語 也}
\]

\[
\text{jiā，yǐ tōngyǔ yě}
\]

‘A and B are *tōngyǔ*’.

Many scholars assume that *tōngyǔ* refers to a “standard language” of Hàn times which served as the basic language to explain the dialectal variants (Zhōu 1956:IX, Lǐ 1990:60). Some scholars even take the term itself as evidence that a standard language existed during Hàn times (Wáng 1980[2006]:20f). However, the term is only used sporadically, often even in contexts which are not in concordance with the idea of a common standard language (Liǔ 2007:155f), and it seems at least equally likely that the term was simply used to denote “popular wordings” (*tōngxíng de shuōfǎ* 通行的說法, Liǔ 2007:155f) within the various dialect regions mentioned in the *Fāngyán*.

**Dialect Geography**

Mapping the dialect data of the *Fāngyán* to dialect areas faces certain obstacles. Not only is the terminology for place names far from coherent, including names of states and fiefdoms from the Zhōu dynasty, administrative terms of Hàn times, terms of physical geography (rivers and mountains), and terms comprising larger areas (Serruys 1959:78-85, Wáng 1980[2006]:18-19, Liū *et al.* 1992:107f). It is also not very precise, since it mixes terminology from different times to denote a complex formation of geographic regions that may intersect, overlap, even include each other. Given that the dialect words in the *Fāngyán* themselves
also may reflect speech traditions from different times (Wáng 1980[2006]:19f), it is difficult to use the Fāngyán directly to draw dialect maps as they are common in modern dialectology.

Starting with the studies of Lín Yǔtáng 林語堂 (1895–1976), e.g. Lín 1933 (a Chinese translation of a chapter from his now lost 1923 Leipzig dissertation entitled Altchinesische Lautlehre), scholars have repeatedly tried to identify the dialect areas which are mentioned in the Fāngyán. Since multiple dialect regions are often summarized in one entry, following the schema

甲，乙，丙之間 jiā，yǐ，bǐng zhī jiān

‘in the area of A, B, and C’, it is possible to count how often specific regions overlap, and use these frequencies to cluster the distinct places into larger dialect areas. Unfortunately, the dialect areas which have been proposed so far, differ greatly from each other, both in their structure and their size. While Lín (1933) postulates fourteen dialect areas, Lúo and Zhōu (1958) postulate seven, Serruys (1959) six, and Liū et al. (1992) and Lǐ (2004) twelve (see the summary in Liū et al. 1992: 104-106). Behr (2010: 571), further groups the 12 areas of Liū et al. (1992) into three major zones, depending the origin of the respective dialect vocabularies (predominantly Sinitic, mixed, or predominantly non-Sinitic). One major reason for these differences is probably that all analyses are based on intuitive data inspection, rather than strict and formal clustering principles. A formally stricter approach is presented in Matsue (2013:191-194) where dialect maps of the Fāngyán gloss words are used to calculate the degree of linguistic continuity and discontinuity between the major dialect areas in the Fāngyán. It seems, however, that this approach needs to be further enhanced by using more sophisticated approaches to modeling and visualization.
Character Use

Not only the aphonetic nature of the Chinese writing system, but also the way Chinese characters are used to denote dialect pronunciations make it difficult to interpret the data in the Fāngyán. Apart from regularly used characters which were taken to reflect the pronunciation of specific words (jiājièzi 假借字 'loan characters' according to the classical liūshū 六書 classification of Chinese characters), the Fāngyán also shows a considerable amount of extremely rare and even unique character variants (qízì 奇字, i.e. ‘rare’ or ‘hapax characters’, Huá 2000a:45). While scholars assumed for a long time that most of the hapax characters were specifically created by Yáng Xióng in order to represent dialect readings (e.g. Luó 1956:II), research on excavated sources has shown that many of the rare characters are variants of paleographically attested ancient characters (Huá 2000a:46f). A closer inspection of the relation between phonetic components and actual readings (usually derived from fǎnqiè 反切 spellings in Guǒ Pú’s commentary) seems to provide further evidence that the majority of the hapax characters were created before the Fāngyán was compiled (Huá 2000b:47), reducing the number of unresolved hapax characters from originally around 300 to about 100 (Huá 2000b:48, Wáng 2006:42-54). Whether these characters were actively created for the purpose of dialect documentation remains an open question.

A further challenge for the investigation of character use in the Fāngyán is the phonological interpretation of words denoted by two character-entries (also called binoms), since it is not always clear whether they were used to denote monosyllabic words with consonant clusters, or pure multisyllabic words (Serruys 1959: 103-120). The majority of these cases is still not well understood, and in some cases there is even no agreement whether certain two-character-sequences correspond to one word or two. As an example, consider the case of the dialect word qián-xuān 虔儇, OC *gran-qwhen (Fāngyán 1.2),
glossed as *huì 慧 ‘wise’, which Bái (1990:18) analyses as one word, while Zhōu (1956) analyses it as two separate words. A first explicit attempt to resolve two-character-words in the Fāngyǎn was made by Serruys (1959:102-140), who proposed to distinguish “dimidiation binoms”, binoms resulting from clitisation, and pure compounds of two independent words or morphemes. Dimidiation is hereby understood as a process by which bisyllabic words evolve from the simplification of complex syllable onsets, such as the Fāngyǎn word for “locust” in the Nánchǔ 南楚 area, zhèmǎng 蟾蟒 < MC *tsyaeH-mangX, which Serruys (1959: 108) reconstructs as a reflex of an OC form *ts-maŋ (transliteration of Serruys was adjusted according to Baxter and Sagart 2014). Clitisation is described as a process by which syllabic elements are added before or after monosyllables, such as cānggē 鶬鴚 < MC *tshang-ka, a Fāngyǎn word for “goose” in the Nánchǔ area, which Serruys (1959:110) analyses as an enclitic expansion of an older word *tsʰaŋŋk. Unfortunately, these analyses were not pursued further by later scholars, and a full critical account of Serruys at times very idiosyncratic proposal is still pending. Wáng (2011a:90-94) investigates to which degree monomorphemic two-character words already occur in older texts. The study provides interesting statistical accounts, but does, unfortunately, not include a phonological investigation of the data.

Phonological Investigations

Apart from Serruys (1959) no further exhaustive phonological investigations of the Fāngyǎn have been carried out so far. In a couple of studies, however, scholars have tried to identify which linguistic varieties are reflected in the Fāngyǎn, and whether they can be compared with the contemporary Chinese dialects. The studies are very diverse both regarding scope and methodology. The reconstruction systems for older character readings vary greatly, with most studies being restricted to Middle Chinese readings (e.g. Yáng 2008 and Lǐ 1987) and
some studies even disregarding the phonological aspects completely (e.g. Cài 2003 and Dù and Gào 2012). Given the large time span between the compilation of the Fàngyán and the Chinese dialects spoken today, all studies which restrict themselves to sporadic matches between modern dialect words and the dialect readings which occur in the Fàngyán have to be taken with a considerable amount of care, since with a limited number of comparanda it is difficult to control for the possibility of chance resemblances. As an example, consider the proposal by Lǐ (1987:67) who identifies the Fàngyán expressions lǐfù 李父 (MC *liX-pjuX < OC *C.rə?-p(r)aʔ) and lǐěr 李耳 (MC *liX-nyiX < OC *C.rə?-nəʔ) with the words [li³¹pu³¹] ‘tiger’ and [li³¹ŋi³⁵] ‘tigress’ in Tǔjīā 土家, an unclassified Sino-Tibetan language spoken in South-Central China. Although the similarity between the words seems to be striking, both regarding their phonetic and semantic similarity (at least in Middle Chinese), this word comparison can, unless more matchings are found, by no means be taken as proof that the Fàngyán really documents an ancestor of the Tǔjīā language (Wang 1996:242-243, Zhào and Huáng 1998). The situation is similar for the proposal by Zhào and Huáng (1998:107), who trace Fàngyán entries for “wild cat” (Fàngyán 8.2), as illustrated in (2), back to Hmong origin, proposing that lateral reflexes, like lǎi 拉, and labial reflexes, like pěi 皮, reflect former complex consonant clusters of [p] and [l] which are still preserved in some Hmong varieties, such as Chuanqian Miao, where “cat” is pronounced as [pli]: As long as these examples remain single instances, not backed by larger numbers, it is hard whether they are just chance resemblances or reflect real language history. As Baxter and Sagart (2014:112-116) illustrate for dialect expressions in the Hán glossary of paronomastic glosses Shìmíng 釋名 (Bodman 1954, Xū 1989, Wú 2010), a thorough comparison of dialect data from Hán times with proposed reconstructions for Hán time readings can yield valuable insights into the dialect diversity of China's past. Regarding the Fàngyán, this work has yet to be done.
Primary Sources

Fāngyán jiàojīn方言校笺 [A collatied and annotated edition of the Fāngyán], ed. by Zhōu Zūmó 周祖謨, Shànghǎi 上海: Kēxué Chūbānshè 科学出版社, 1956, I–V.


References


Bái Zhàolín 白兆麟, “Fāngyán shuāngyīncí tànxī《方言》双音词探析” [Investigation of bisyllabic words in the Fāngyán], Gǔjì Zhēnglǐ Yánjiū Xuékān 古籍整理研究学刊 2, 1990, 18–22.

Bodman, N. C., A Linguistic Study of the Shih Ming: Initials and Consonant Clusters.

Cài Xiǎo 蔡晓, “Yóu Yánghónggōng Fāngyán kàn Mìyánghuà zhōng gǔyǔ de yíliú 由揚雄《方言》
看泌阳话中古语的遗留” [Investigating old strata in the Miyáng dialect from the
perspective of Yáng Xióng’s Fāngyán], Tiānzhōng Xuékàn 天中学刊, 18(3), 2003, 98–
100.

Dù Jùnpíng 杜俊平 and Gāo Liàng 高亮, “Lún Yǎnshīhuà zhōng bāocún de Fāngyán gǔ yǔcí
论偃师话中保存的《方言》古语词 [On the preservation of ancient expressions from the
Fāngyán in the Yǎnshī dialect], Héngshuǐ Xuéyuàn Xuébào 衡水学院学报, 15(5), 2013,
80–82.

Huá Xuéchéng 华学诚, “Yáng Xióng Fāngyán 'qízi' kǎo (shàng) 揚雄《方言》“奇字”考(上)”
[On the strange characters in Yáng Xióng’s Fāngyán (first part)], Qīnzhōu Shìfān

Huá Xuéchéng 华学诚, “Yáng Xióng Fāngyán 'qízi' kǎo (xià) 揚雄《方言》“奇字”考(下)” [On
the strange characters in Yáng Xióng’s Fāngyán (second part)], Qīnzhōu ShìfānGāodǎi

Huá Xuéchéng 华学诚, ed., Yáng Xióng Fāngyán jiàoshì huìzhèng 揚雄方言校释汇证 [A
critical compilation of glosses to Yáng Xióng's Fāngyán], Běijīng 北京: Zhōnghuá 中华书
局, 2006.

Huá Xuéchéng 华学诚, “Jìn 15 nián lái de Yáng Xióng Fāngyán zhěnglǐ yǔ wǒmen duì
Fāngyán de zhěnglǐ 近 15 年来的揚雄《方言》研究与我门对《方言》的整理” [An
overview on research on Yáng Xióng’s Fāngyán during the last 15 years and our
collation of the text of the Fāngyán], Nánkǎi Yǔyánxuékǎn 南开语言学刊, 9(1), 2007a,
59–69.

Huá Xuéchéng 华学诚, “Èrshí shìjì yǐlái de Fāngyán zhěnglǐ 二十世纪以来的《方言》整理”
[Collation of the text of the Fāngyán in and after the 20th century], Zhōngwén Zìxué

Huá Xuéchéng 华学诚 and Xú Yányàn 徐妍雁, “Yáng Xióng Fāngyán jí qí yánjiū shùpíng 揚雄《方言》及其研究述评” [Review of research on Yáng Xióng's Fāngyán], Sūzhōu Dàxué Xuébào 苏州大学学报, 1, 2013, 149–157.

Liú Jūnhuì 刘君惠, Lǐ Shùháo 李恕豪, Yáng Gàng 杨钢, and Huá Xuéchéng 华学诚, Yáng Xióng Fāngyán yánjiū 揚雄方言研究 [Investigations into Yáng Xióng's Fāngyán], Chéngdū 成都: Bā-Shū 巴蜀书社, 1992.


Lín Yǔtáng 林語堂, “Qián-Hàn fāngyīn qūyù kǎo 前漢方音區域考” [Investigation of the geographical distribution of dialect pronunciations during Qín and Hán times], in: Yǔyánxué lùncóng 語言學論叢, ed. by Lín Yǔtáng 林語堂, Shànghǎi 上海: Kāiming 開明書店, 1933, 16–44.

Lǐ Jìngzhōng 李敬忠, “Fāngyán zhōng de shāoshū mínzú yǔcí shīxī《方言》中的少数民族词语试析” [Preliminary analysis of words from minority languages in the Fāngyán], Minzú Yǔwén 民族语文 3, 1987, 64–68.

Lǐ Shùháo 李恕豪, “Lùn Yáng Xióng Fāngyán zhōng de jīgè wèntí 论扬雄《方言》中的几个问题, 古汉语研究” [Some questions regarding Yáng Xióng's Fāngyán], Gǔhàn yǔ Yǎnjiǔ 古

李恕豪 杨雄《方言》与方言地理学研究 [Yáng Xióng’s Fangyán and research on dialect geography], Chéngdū 成都: Bāshǔ 巴蜀书社, 2004.


Wáng Zhìqún 王智群, Fāngyán yǔ Yáng Xióng cíhuìxué 《方言》与扬雄词汇学 [On Fāngyán and Yáng Xióng’s Lexicology], Bēijīng 北京: Gāodēng Jiàoyù 高等教育, 2011b.

Wú Zhùi 吴锤, Shimíng shēngxùn yánjū 習名声训研究 [Research on the sound explanations in the Shimíng], Bēijīng 北京: Mínzú 民族出版社, 2010.

Xū Fāngmīn 徐芳敏, Shimíng yánjū 習名研究 [Research on the Shimíng], Táiběi 台北: Guólí 國立 Taiwan 大學, 1989.

Yáng Xiǎohóng 杨晓红, “Fāngyán yǔ Lǔ nán fāngyán cíyì de gǔjīn chuánchéng 《《方言》与鲁南方言词义的古今传承” [Fāngyán and the transmission of word meanings across times into the Lǔnán dialect], Jīnān Zhiyè Xuéhuán Xuébào 济南职业学院学报 68(3), 2008, 92–94.

Yōng Péng 雍鹏, Fāngyán wénxiàn zōngshù 《方言》文献综述 [Summary on documentation of the Fāngyán], Yūwén Xuékān 语言学刊, 5, 2011, 31–32.